SPEECH SOUND CODING AND TRAINING-INDUCED PLASTICITY IN PRIMARY AUDITORY CORTEX

by

Crystal Tasha Engineer

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Michael P. Kilgard, Chair

Dr. Peter F. Assmann

Dr. Marco Atzori

Dr. Lucien T. Thompson

SPEECH SOUND CODING AND TRAINING-INDUCED PLASTICITY IN PRIMARY AUDITORY CORTEX

by

CRYSTAL TASHA ENGINEER, B.S., M.S.

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN COGNITION AND NEUROSCIENCE

THE UNIVERSITY OF TEXAS AT DALLAS

December, 2008

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper or papers submitted for publication. The dissertation must still conform to all other requirements explained in the "Guide for the Preparation of Master's Theses and Doctoral Dissertations at The University of Texas at Dallas." It must include a comprehensive abstract, a full introduction and literature review and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author(s). The signatures of the Supervising Committee which precede all other material in the dissertation attest to the accuracy of this statement.

ACKNOWLEDGEMENTS

I would first like to thank the massive team of people who have helped me to complete these experiments over the past 5 years. In particular, the authors on the following manuscripts have been invaluable, and I could not have finished without their help. Claudia Perez has been with me for the entire duration of these experiments, and has helped tremendously with experimental design, behavioral training, mapping the overnight shifts, and has always been willing to let me bounce crazy analysis ideas off of her. Additionally, she was in charge of the vowel training experiment in Chapter 4, and designed that experiment with Matt Perry. Helen Chen helped design the speech stimulus set and mapping procedure, and we went through all the pain together of setting up brand new experiments. Ryan Carraway was an amazing behavioral training manager, and made sure that each rat was trained correctly, which is an incredibly difficult task! Ryan also assisted with the vast majority of the mapping experiments. Amanda Puckett was the best behavioral training partner I could have ever asked for. She was always willing to drop everything and help me, whether it was helping with analysis, letting me record her saying hundreds of words until she went hoarse, or flying to another country to be my maid of honor. Vikram Jakkamsetti performed the mapping surgeries for many rats, and was always willing to help, no matter how busy he was with his own set of experiments. Kevin Chang helped with many mapping experiments. Jai Shetake recorded A1 responses to speech sounds in awake rats. Jarod Roland helped with both behavioral training and mapping experiments.

Dr. Kilgard has been an amazing mentor. His lab is a very good environment to work in, where all the graduate students are closely knit, good friends, and always willing to help each other out. In his lab, I have slowly learned how to troubleshoot when everything breaks down all at the same time, yet Dr. Kilgard is always there to show me what I missed when I can't figure out how to fix something. He has let me follow my experimental interests, and over the years, I've grown to truly enjoy writing code and performing analysis. I am extremely grateful that I have had the opportunity to learn how to write scientific papers as a graduate student.

My dissertation committee members have given me plenty of useful input over the years. Dr. Assmann spent a tremendous amount of time going over experimental design with me when I was first beginning these experiments, and he taught me everything I know about speech! I can count on Dr. Atzori to always be able to think up new ways of doing the analysis, and new questions to ask. Dr. Thompson was there to help when I took Grant Writing, and first began trying to write down these experiments.

I want to thank the lab's behavioral training team that did an amazing job training so many rats: Jonathan Riley, Chris Heydrick, Matt Perry, Pei-Lan Kan, Joanna Gibbons, Alyssa McMenamy, Anushka Meepe, Vismay Badhiwala, Jonathan Seale, Maria Luisa de la Vega, Robert Frazier, Cathy Haupstueck, Jeans Lee Choi, Chikara Dablain, Ashley Hudson, Cherri Whang, Fern Yoon, Nick Hatate, and Ryan Cheung. Roshini Jain and Dwayne Listhrop helped with mapping experiments. Pritesh Pandya, Cherie Percaccio, and Raluca Moucha were wonderful mentors, and taught me so much when I was first starting out in the lab.

Finally, none of this would have been possible without the support of my parents and my brother, David. Kyra and Sixy were always there to listen when I was practicing a presentation, or talking through something out loud. My husband Navzer has been incredibly supportive, and this journey would never have been possible without him.

September, 2008

SPEECH SOUND CODING AND TRAINING-INDUCED PLASTICITY IN PRIMARY AUDITORY CORTEX

Publication No._________

Crystal Tasha Engineer, Ph.D.
The University of Texas at Dallas, 2008

Supervising Professor: Michael P. Kilgard, Ph.D.

It is important to understand how the human brain processes speech sounds in order to lead to perception. fMRI and EEG studies have shown that certain cortical regions are activated after hearing speech, but these techniques lack the temporal and spatial precision necessary to document the unique pattern of activation evoked by each speech sound. Previous studies have shown that the primary auditory cortex (A1) pattern of activity evoked in response to speech sounds is altered by the temporal features of the sound. In this study, we show that rats' (Rattus norvegicus) behavioral performance on consonant discrimination tasks is similar in key respects to human performance, and can be predicted from the spatiotemporal pattern of rat A1 activity when temporal information is maintained. When temporal information is ignored and the mean firing rate is used, behavioral performance cannot be predicted as accurately. We also document the ability of rats to categorize speech sounds by voicing or gender on the first day of training. This categorization ability is also predicted by the spatiotemporal pattern of A1 activity. Finally, we show that training on multiple speech discrimination tasks increases the proportion of neurons responding to low frequency tones, the threshold of A1 neurons, the response strength to tones, the receptive field size, and response latencies. Passive exposure to speech sounds increases the proportion of neurons responding to high frequency tones, and decreases the threshold of A1 neurons and response latencies. These results indicate that training on multiple speech tasks does not result in stimulus specific response enhancement in primary auditory cortex, but instead, results in generalized enhancement of untrained sounds following speech training. This result suggests that non-primary or higher cortical areas, as opposed to A1, may exhibit stimulus specificity after speech sound training.

TABLE OF CONTENTS

Vita

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Processing speech sounds is something that we humans do on a daily basis. Throughout the day, we are able to perceive the words that someone has spoken, process what was said, and then respond. When the ability to perceive speech sounds is lost, as a result of stroke or traumatic brain injury, the road to recovery can be long to impossible. Understanding the neural mechanisms behind cortical speech sound processing could potentially lead to better treatment options to facilitate recovery in patients with speech processing disorders.

fMRI and EEG studies show that certain areas of the brain are activated by speech sounds (Callan et al. 2003). However, there are limitations on the spatial and temporal resolution of these techniques (Rosen et al. 1998). People with speech processing disorders need treatment options, and this study may facilitate the development of new therapeutic paradigms. Multi-unit recordings offer better temporal precision than fMRI and better spatial resolution than EEG recordings. As a result, it is possible to document which neurons are firing down to the single unit level, and precisely when they are firing on a millisecond timescale. Using this technique, we are able to show that the precise spectrotemporal activation pattern of primary auditory cortex neurons strongly predicts rat behavioral performance on a variety of speech discrimination tasks (Chapter 2).

Previous studies have documented the neural responses to speech sounds in primary auditory cortex (A1). Monkey A1 neurons respond distinctly to sounds that differ in voice onset time (VOT), which is the amount of time between the onset of the consonant burst release and the onset of vocal fold vibrations (Kent and Read 2002). Sounds with short VOTs, corresponding to voiced stop consonants in English, like 'dad', evoke a single peak of activity in response to the onset of the consonant /d/, while sounds with long VOTs, like 'tad', evoke two peaks of activity in response to the consonant onset and the vowel onset (Steinschneider et al. 2003). This finding has been obtained in cats, awake and anesthetized monkeys, as well as humans (Steinschneider et al. 2003; Steinschneider et al. 2005; Steinschneider et al. 1999; Wong and Schreiner 2003). Sounds that differ in place of articulation also evoke distinct responses in A1. Low frequency neurons have a stronger response to the consonant /b/, while high frequency neurons have a stronger response to the consonant /d/ (Steinschneider et al. 1995). These studies have shown that primary auditory cortex neurons generate different spatial and temporal patterns of activity to speech depending upon the acoustic features of the sound.

Numerous studies have documented the ability of animals to discriminate speech sounds. Chinchillas can respond to consonants with either short or long VOTs in order to avoid foot shock (Kuhl and Miller 1975). When the sounds are synthesized to lie on a VOT continuum between the voiced and voiceless endpoints, the perceptual boundary for chinchillas is nearly identical to the identification function found in humans. Japanese quail are able to accurately peck for words starting with /d/ and refrain from pecking for words that start with either /b/ or /g/, in multiple vowel contexts (Kluender et al. 1987). Rats have also been shown to discriminate speech sounds, and (for example) are able to accurately discriminate 'shad' from 'chad' (Reed et al. 2003).

In Chapter 2, we document the ability of rats to successfully discriminate nine out of eleven different consonant pairs. This behavioral performance is strongly predicted by A1 neural responses when temporal information is maintained, but is not predicted when using the mean firing rate. This study supports the view that both spatial and temporal information are necessary to accurately predict rat behavior.

The stimuli used for the experiments in Chapter 2 are spoken by a single speaker. However, in the real world, humans are able to categorize speech sounds regardless of the gender of the speaker, age of the speaker, or distance from the speaker. The word 'dad' could be spoken by a young child or a deep voiced adult male, but it still needs to be put into the 'dad' category regardless of the speaker. This ability to normalize across the variability is important not only for human speech sounds, but also for many visual stimuli and other auditory stimuli, including animal vocalizations.

Both animals and humans are able to accurately categorize sounds; however the neural mechanisms involved in sound categorization are not yet known. When Vervet monkeys spot an eagle, snake, or leopard, they produce an alarm call specific to the type of animal sighted (Seyfarth et al. 1980). Surrounding monkeys who hear the alarm call will produce an animal-specific reaction: after hearing an eagle call, monkeys will look up at the sky, while a snake call causes monkeys to look down at the ground. Monkeys run up a tree in response to hearing a leopard alarm call. In this case, being able to correctly categorize the sound can mean the difference between life and death for the monkey. Monkeys are also able to categorize vocalizations based on the quality of the food that another monkey has found (Gifford et al. 2003). These studies show that animals are able to categorize complex acoustic stimuli, including vocalizations.

In Chapter 3, we show that rats are able to accurately categorize a single set of sounds on the first day of training either by voicing (while ignoring gender information) or by gender (while ignoring voicing information). Voicing categorization is also possible when the sounds are temporally compressed. Behavioral performance on each of these categorizations tasks can be predicted from A1 responses. This chapter extends the experiments in the previous chapter by showing that animals can categorize multiple tokens despite large acoustic variability, and that A1 responses are still able to predict behavior even with the increase in acoustic variability.

It is well known that training on visual, auditory, or somatosensory tasks changes the brain. Monkeys trained on a tone frequency discrimination task had an enlarged representation of the tone in A1 (Recanzone et al. 1993). Expansion of the representation of a trained tone has also been seen in rats trained on a frequency discrimination task (Polley et al. 2006). However, when rats were trained on an intensity discrimination task using the same stimuli, there was no increase in frequency representation; instead, there was an increase in the response to the trained intensity (increased non-monotonicity). We hypothesized that speech training would alter the receptive fields, cortical maps, and speech responses in A1.

In Chapter 4, we show that long-term speech training on multiple tasks alters the receptive field properties in A1, and shows an opposite effect compared to rats passively exposed to speech sounds. Long-term speech training also causes an increase in the response to low frequency tones, and a decrease in the response to mid frequency tones. Passive exposure results in an increased response to high frequency tones. This chapter shows that speech training alters A1 responses.

In Chapter 5, we discuss the clinical implications of these experiments, and explore how these experiments fit in with the previous literature. This dissertation consists of 5 chapters and 1 appendix that contains supplementary data and figures. Chapter 2 explores consonant discrimination, Chapter 3 explores categorization ability, and Chapter 4 documents A1 plasticity after speech training.

CHAPTER 2

CORTICAL ACTIVITY PATTERNS PREDICT SPEECH DISCRIMINATION ABILITY

Crystal T. Engineer, Claudia A. Perez, YeTing H. Chen, Ryan S. Carraway, Amanda C. Reed, Jai A. Shetake, Vikram Jakkamsetti, Kevin Q. Chang, Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

* This paper was published in Nature Neuroscience in May 2008.

ACKNOWLEDGEMENTS

Crystal Engineer, Claudia Perez, Ryan Carraway and Amanda Reed collected behavioral training data. Crystal Engineer, Claudia Perez, Helen Chen, Ryan Carraway, Vikram Jakkamsetti, and Kevin Chang recorded anesthetized cortical responses. Jai Shetake recorded awake cortical responses. Michael Kilgard and Crystal Engineer wrote the manuscript and performed data analysis. All authors discussed the paper and commented on the manuscript. The authors would like to thank Jarod Roland, Roshini Jain and Dwayne Listhrop for assistance with microelectrode mappings. We would like to thank Robert Rennaker for technical assistance and training and for providing microelectrode arrays and inserter. We would also like to thank Matt Perry, Chris Heydrick, Alyssa McMenamy, Anushka Meepe, Chikara Dablain, Jeans Lee Choi, Vismay Badhiwala, Jonathan Riley, Nick Hatate, Pei-Lan Kan, Maria Lazo de la Vega and Ashley Hudson for help with behavioral training. We would also like to thank Sheila Blumstein, Yale Cohen, Heather Read, Susan Denham, Lee Miller, S. Edelman, Valentin Dragoi, Herve Abdi, Peter Assmann, Xiaoqin Wang and Ranulfo Romo for their suggestions about earlier versions of the manuscript.

ABSTRACT

Neural activity in the cerebral cortex can explain many aspects of sensory perception. Extensive psychophysical and neurophysiological studies of visual motion and vibrotactile processing show that the firing rate of cortical neurons averaged across 50–500 ms is well correlated with discrimination ability. In this study, we tested the hypothesis that primary auditory cortex (A1) neurons use temporal precision on the order of 1–10 ms to represent speech sounds shifted into the rat hearing range. Neural discrimination was highly correlated with behavioral performance on 11 consonant-discrimination tasks when spike timing was preserved and was not correlated when spike timing was eliminated. This result suggests that spike timing contributes to the auditory cortex representation of consonant sounds.

INTRODUCTION

The debate about the importance of spike timing began with the first recordings of neural activity and remains unresolved (Parker and Newsome 1998; DeWeese, Hromadka et al. 2005). Although coding strategies based on precise timing have the potential to transmit more information than strategies based on firing rate averaged over long intervals, psychophysical studies of tactile modulation rate and visual movement indicate that rate-based descriptions of sensory events provide the best predictions of behavioral discrimination ability (Britten, Shadlen et al. 1992; Parker and Newsome 1998; Pruett, Sinclair et al. 2001; Romo and Salinas 2003; Liu and Newsome 2005). The use of precise temporal information by somatosensory cortex has been rejected because neurometric analysis predicts much better discrimination ability than is observed behaviorally (Romo and Salinas 2003).

The auditory system is sensitive to precise temporal information and is a logical place to study perceptual correlates of neural representations based on precise spike timing (DeWeese, Hromadka et al. 2005; Narayan, Grana et al. 2006; Schnupp, Hall et al. 2006; Walker, Ahmed et al. 2007). However, few behavioral studies have examined the relationship between neural activity and auditory discrimination (Ahissar, Nagarajan et al. 2001; Orduna, Mercado et al. 2005; Schnupp, Hall et al. 2006; Wang, Narayan et al. 2007). Psychophysical studies have demonstrated that newborn and adult humans, as well as rats and chinchillas, can reliably distinguish consonants based on acoustic information found within 40 ms of sound onset (Miller and Nicely 1955; Kuhl and Miller 1975; Blumstein and Stevens 1979; Bertoncini, Bijeljac-Babic et al. 1987; Jongman 1989; Fowler, Brown et al. 2003; Reed, Howell et al. 2003). Similarly, the onset response of neurons in the central auditory system recorded in awake and anesthetized subjects reliably encodes the rapid acoustic transitions that provide information about consonant identity (Steinschneider, Reser et al. 1995; Steinschneider, Fishman et al. 2003; Steinschneider, Volkov et al. 2005; Tavabi, Obleser et al. 2007; Young 2007). A1 lesions impair judgments of complex sounds, including speech (Dewson, Pribram et al. 1969; Heffner and Heffner 1989; Wetzel, Ohl et al. 1998; Rybalko, Suta et al. 2006; Cooke, Zhang et al. 2007). Here we report that the precise spatiotemporal activity pattern evoked by the onset of consonant sounds is well correlated with the ability of rats to discriminate these sounds.

METHODS

Speech stimuli.

We recorded 20 English words ending in /ad/ (as in 'sad') in a double-walled, soundproof booth. The initial consonants differed in voicing (voiced /d/ versus voiceless /t/), place of articulation (lips /b/ versus back of mouth /g/) or manner of articulation (fricative /sh/ versus nasal /n/) (Figure A2.1). The fundamental frequency and spectrum envelope of each word was shifted up in frequency by a factor of two using the STRAIGHT vocoder (Kawahara 1997) in order to better match the rat hearing range (Figure B2.9). The intensity of each sound was adjusted so that the intensity during the most intense 100 ms was 60 dB SPL.

Operant training procedure and analysis.

Eleven rats were trained using an operant go/no-go procedure to discriminate words differing in their initial consonant sound. Each rat trained for two 1-h sessions each day (5 d/week). Rats first underwent a shaping period during which they were taught to press the lever. Each time the rat was in close proximity to the lever, the rat heard the target sound and received a pellet (45-mg sugar pellet). Eventually, the rat began to press the lever without assistance. After each lever press, the rat heard the target sound and received a pellet. The shaping period lasted until the rat was able to obtain at least 100 pellets per session for two consecutive sessions. This stage lasted on average 3.5 d. After the shaping period, rats began a detection task in which they learned to press the lever each time the target sound was presented. Silent periods were randomly interleaved with the target sounds during each training session. Sounds were initially presented every 10 s, and the rat was given an 8-s window to press the lever. The sound interval was gradually decreased to 6 s, and the lever-press window was decreased to 3 s. Once rats reached the performance criteria of a *d*' ≥1.5 for ten sessions, they advanced to a consonant discrimination task. The quantity *d*' is a measure of discriminability of two sets of samples based on signal detection theory.

During each consonant discrimination task, rats learned to discriminate the target sound from the distractor sounds. Trials began every 6 s, and silent catch trials were randomly interleaved 20–33% of the time. Rats were only rewarded for lever presses to the target (conditioned) stimulus. Pressing the lever on a stimulus other than the target resulted in a time-out during which the house light was extinguished and the training program paused for a period of approximately 6 s. Training took place in a soundproof, double-walled training booth that included a house light, a video camera for monitoring, a speaker (Optimus Bullet Horn Tweeter) and a cage (8 inches length ×8 inches width ×8 inches height) that included a lever, lever light and pellet receptacle. A pellet dispenser was mounted outside the double-walled, foam-lined booth to reduce noise. Rats were food deprived to motivate behavior, but were fed on days off to maintain between 80% and 90% ad lib body weight. Rats were housed individually and maintained on a reverse 12-h light-dark cycle.

Each consonant discrimination task lasted for 20 training sessions over 2 weeks. Six rats performed each of four different consonant-discrimination tasks (/d/ versus /s/, /d/ versus /t/, /r/ versus /l/, and /d/ versus /b/ and /g/), and five rats performed each of three different consonant-discrimination tasks (/m/ versus /n/; /sh/ versus /f/, /s/ and /h/; and /sh/ versus /ch/ and /j/). Each group of rats trained on each of the tasks for 2 weeks, in the order given. We subsequently tested them for 2 d on each task to ensure that discrimination ability was not strongly influenced by task order (see Appendix B: Supplementary Data). Data shown in Figure A2.4 was collected on the seventh and eighth days (that is, four sessions) of training on each task. Over these 2 d, each rat performed 940 ±173 trials (mean ±s.d.). An example learning curve is shown in Figure B2.4.

Recording procedure.

We recorded multiunit (n = 445) and single-unit (n = 16) responses from right primary auditory cortex (A1) of anesthetized, experimentally naive, female Sprague-Dawley rats in a soundproof recording booth (n = 11 rats). Rats were anesthetized with pentobarbital (50 mg kg^{-1}) and received supplemental dilute pentobarbital (8 mg ml^{-1}) every 0.5–1 h as needed to maintain areflexia, along with a 1:1 mixture of dextrose (5%) and standard Ringer's lactate (Engineer et al., 2004) to prevent dehydration. Heart rate and body temperature were monitored throughout the experiment. Four Parylene-coated tungsten microelectrodes (1–2 MΩ, FHC) were simultaneously lowered to 600 µm below the surface of the right primary auditory cortex (layer 4/5). Electrode penetrations were marked using blood vessels as landmarks.

We recorded multiunit A1 responses (n = 40) in six awake rats using chronically implanted microwire arrays, which have been described in detail in previous publications (Rennaker et al., 2005a; Rennaker et al., 2005b). Briefly, 14-channel microwire electrodes were implanted in the right primary auditory cortex using a custom-built mechanical insertion device to rapidly insert electrodes in layers 4/5 (depth, 550 µm) (Rennaker et al., 2005b). Restraint jackets were used to minimize movement artifacts during recording sessions, conducted 1–7 d after implantation.

Twenty 60-dB speech stimuli were randomly interleaved and presented every 2,000 ms for 20 repeats per site at each recording site. Brief (25-ms) tones were presented at 81 frequencies (1–32 kHz) at 16 intensities (0–75 dB) to determine the characteristic frequency of each site. All tones were separated by 560 ms and randomly interleaved. 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). Single units refer to well isolated waveforms likely to have been evoked by a single neuron. Multiunits include action potentials from more than one nearby neuron. The University of Texas at Dallas Institutional Animal Care and Use Committee approved all protocols and recording procedures.

Statistical analysis.

Neurogram similarity was computed using euclidean distance. The euclidean distance between any two neurograms (*X*, *Y*) is the square root of the sum of the squared differences between the firing rate at each bin (*j*) for each recording site (*i*). For the analysis in Figure A2.3, part 1 and Figure A2.5a, we used activity from 40 1-ms bins from all 445 sites to compute the similarity between neurogram pairs. For the analysis in Figure A2.3, part 2 and Figure A2.5b, we used activity from a single 40-ms bin from each of 445 sites to compute the similarity between neurogram pairs

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n_{\text{sites}}} \sum_{j=1}^{n_{\text{bins}}} (x_{ij} - y_{ij})^2}$$

where n_{sites} and n_{bins} are the total numbers of sites and bins, respectively.

We used a nearest-neighbor classifier to quantify neural discrimination performance based on single-trial activity patterns (Schnupp et al., 2006; Foffani and Moxon, 2004). The classifier binned activity using 1-ms to 700-ms intervals and then compared the response of each single trial with the average activity pattern (PSTH) evoked by each of the speech stimuli presented. The given trial being considered was not included in the average activity pattern, to prevent artifact. This model assumes that the brain region reading out the information in the spike trials has previously heard each of the sounds 19 times and attempts to identify which of the possible choices was most likely to have generated the trial under consideration. It uses euclidean distance to determine how similar each response was to the average activity evoked by each of the sounds. The classifier guesses that the single-trial pattern was generated by the sound whose average pattern it most closely resembles (that is, minimum euclidean distance). The onset response to each sound is defined as the 40-ms interval beginning when neural activity exceeded the spontaneous firing rate by three s.d. Error estimates are s.e.m. Pearson's correlation coefficient was used to examine the relationship between neural and behavioral discrimination on the 11 tasks ($n = 11$).

RESULTS

Neural responses to consonant sounds

We recorded neural responses to 20 English consonants (Figure A2.1 and Figure B2.1) from single neurons and multiunit clusters of A1 neurons in awake and barbiturate-anesthetized rats. To illustrate the response to each sound, we constructed neurograms from the average onset response of 445 multiunit A1 recording sites ordered by characteristic frequency (Figure A2.2 and Figure B2.2). As expected, each consonant evoked a distinct spatiotemporal activity pattern in A1 (Supplementary Video 1 online: http://www.nature.com/neuro/journal/v11/n5/suppinfo/nn.2109_S1.html).

Consonants differing only in their place of articulation resulted in different spatial activity patterns (Blumstein and Stevens, 1979; Steinschneider et al., 1995; Tavabi et al, 2007). For example, the /s/ sound activated high frequency neurons, whereas /sh/ activated mid-frequency neurons (Figure A2.2, third column and Appendix B: Supplementary Data). Manner of articulation (for example, stop, fricative or glide) substantially altered the temporal profile of the population response (Figure A2.2, top row). As in earlier studies, stop consonants generated the sharpest onset peaks (Steinschneider et al., 1999; Steinschneider et al., 2003; Wong and Schreiner, 2003). Nasals, glides and liquids resulted in the weakest onset responses; fricatives and affricates resulted in intermediate onset responses (Appendix B: Supplementary Data). Whereas the voiced stop consonants (/b/, /d/, /g/) evoked a single burst of activity, unvoiced stop consonants (/p/, /t/, /k/) resulted in a second peak of activity at voicing onset, consistent with previous reports in cats, monkeys and humans (Figure B2.3, Appendix 2: Supplementary Data, and Supplementary Video 2 online: http://www.nature.com/neuro/journal/v11/n5/suppinfo/nn.2109_S1.html) (Steinschneider et al., 1999; Steinschneider et al., 2003; Steinschneider et al., 2005; Wong and Schreiner, 2003).

Relating onset response similarity and behavior

Although it is reasonable to expect that sounds that evoke similar cortical responses will be more difficult to discriminate than sounds that evoke distinct responses, this is the first study to test whether this relationship requires precise temporal information (that is, 1-ms bins) or whether the rate-based strategies observed in visual and somatosensory cortex (that is, 50- to 500-ms bins) predict behavioral performance.

We quantified the difference between each pair of neurograms using euclidean distance (Figures A2.2 and A2.3). When 1-ms windows were used, the spatiotemporal patterns evoked by the consonants /d/ and /b/ were much more distinct than the patterns evoked by /m/ and /n/ (Figure A2.3, part 1), leading to the prediction that /d/ versus /b/ would be one of the easiest consonant pairs to discriminate and /m/ versus /n/ would be one of the hardest. Alternatively, if information about precise timing is not used, /d/ versus /b/ was predicted to be a very difficult discrimination (Figure A2.3, part 2). To test these contrasting predictions, we evaluated the ability of rats to distinguish between these and nine other consonant pairs using an operant go/no-go procedure wherein rats were rewarded for a lever press after the presentation of a target consonant. The tasks were chosen so that each consonant pair differed by one articulatory feature (place, voicing or manner; Figure A2.1). Rats were able to reliably discriminate 9 of the 11 pairs tested (Figure A2.4 and Figure B2.4). These results extend earlier observations that monkeys, cats, birds and rodents can discriminate consonant sounds (Kuhl and Miller 1975; Kluender, Diehl et al. 1987; Dooling, Okanoya et al. 1989; Sinnott and Brown 1997; Ramus, Hauser et al. 2000; Reed, Howell et al. 2003; Wong and Schreiner 2003; Toro, Trobalon et al. 2005). The wide range of difficulty across the 11 tasks is advantageous for identifying neural correlates.

Consistent with our hypothesis that A1 representations make use of precise spike timing, /d/ versus /b/ was one of the easiest tasks (Figure A2.4), and differences in the A1 onset response patterns were highly correlated with performance on the 11 tasks when 1-ms bins were used ($R^2 = 0.75$, $P = 0.0006$, Figure A2.5a). A1 responses were not correlated with behavior when spike timing information was removed ($R^2 = 0.046$, $P = 0.5$; Figure A2.5b; Figure B2.5a and Supplementary Data).

Neural discrimination predicts behavioral discrimination

Although it is interesting that the average neural response to each consonant was related to behavior, in practice, individual speech sounds must be identified during single trials, not based on the average of many trials. Analysis using a nearest-neighbor classifier makes it possible to document neural discrimination on the basis of single trial data and allows the direct correlation between neural and behavioral discrimination in units of percentage correct. This classifier (which compares the poststimulus time histogram (PSTH) evoked by each stimulus presentation with the average PSTH evoked by each consonant and selects the most similar; see Methods) is effective in identifying tactile patterns and animal vocalizations using cortical activity (Foffani and Moxon 2004; Schnupp, Hall et al. 2006).

Behavioral performance was well predicted by classifier performance when activity was binned with 1-ms precision. For example, a single sweep of activity from one multiunit cluster was able to discriminate /d/ from /b/ 79.5 ±0.8% (mean ± s.e.m.) of the time and /m/ from /n/ 60.1 ±0.7% of the time; 50% is chance performance. Consistent with previous psychophysical evidence that the first 40 ms contain sufficient information to discriminate consonant sounds (Blumstein and Stevens 1979; Bertoncini, Bijeljac-Babic et al. 1987; Jongman 1989; Fowler, Brown et al. 2003), the correlation between the behavioral and neural discrimination was highest when the classifier was provided A1 activity patterns during the first 40 ms of the cortical response ($R^2 = 0.66$, $P = 0.002$; Figures A2.5c and A2.6a, part 1, Figure B2.6 and Supplementary Data). This correlation was equally strong in awake rats ($R^2 = 0.63$, $P = 0.004$; Figure B2.7). Neural discrimination correlated well with behavior provided that onset responses were used (5–100 ms) and temporal information was preserved (1–10 ms, Figure B2.5b).

Because of a ceiling effect caused by greatly improved neural discrimination, the correlation between the behavioral and neural discrimination was not significant ($R^2 = 0.02$, $P = 0.6$) when the classifier was given all 700 ms of activity (Figure A2.6b, part 3). Neural discrimination was greatly reduced when temporal information was eliminated (that is, mean firing rate over 700 ms) and no relationship with behavior was observed ($R^2 = 0.06$, $P = 0.5$). For example, on the easiest task (/d/ versus /s/), rats were correct on 92.5 ±0.8% of trials, whereas the classifier was correct only 55.4 ±0.6% of the time when spike timing was removed (Figure A2.6b, part 4). The correlation between classifier and behavior was also not significant when the mean onset response rate was used (40-ms bin, $R^2 = 0.14$, $P = 0.2$; Figures A2.5d and A2.6a, part 2). These results show that the distinctness of the precise temporal activity patterns evoked by consonant onsets is highly correlated with discrimination ability in rats.

Influence of population size on neural discrimination

To determine the neural population size that best correlates with behavior, we compared behavioral discrimination with neural discrimination using individual single units, 16 single units, individual multiunits and sets of 16 multiunits. Stringent spike-sorting criteria were used to increase our confidence that we were recording from individual neurons. We collected a total of 16 well isolated single units from 16 different recording sites distributed across A1. Consonant discrimination was evaluated for each of the 16 single units individually and for the set of all 16 single units. When the classifier was provided with activity from all 16 sites, each pattern was a matrix of 16 columns and a number of rows determined by the bin size used. We used the same technique to evaluate classifier performance using multiunit activity from sets of 16 recording sites randomly selected from the full set of 445 recording sites. Each population size was evaluated with or without precise temporal information using the onset response (that is, 1-ms or 40-ms bins) or the entire response (that is, 1-ms or 700-ms bins).

Neural discrimination using single units did not correlate with behavior regardless of the coding strategy used in the analysis (Figure A2.7a). The poor correlation may be related to the poor neural discrimination of single units (Figure A2.7b), which was probably due to the small number of action potentials in single-unit responses compared to multiunit responses. Although discrimination using all 16 single units was better than individual single units, neural discrimination on the 11 tasks was still not significantly correlated with behavior (Figure A2.7), perhaps because of the anatomical distance between the 16 recording sites.

Neural discrimination using 16 randomly selected multiunit sites correlated with behavior but did so only when temporally precise onset responses were used (Figure A2.7a). Although the dependence on temporally precise onset responses was similar to results based on single multiunit sites, the average neural performance using 16 multiunit sites significantly exceeded actual behavioral performance (Figure A2.7b). This excessive accuracy resulted in a ceiling effect, which probably explains why the correlation with behavior was lower when large populations were used. After exploring a large set of neural readouts using various time windows and population sizes, we found that discrimination using onset activity patterns from individual multiunit sites correlated best with behavioral discrimination.

Our observation that multiunit responses were highly correlated with behavioral performance is consistent with earlier reports that multiunit responses are superior to single-unit responses for identifying complex stimuli. For example, V1 single units provide an unreliable estimate of the local contrast in natural images, whereas multiunit responses encode this information efficiently (Weliky, Fiser et al. 2003). Similarly, multiunit clusters in the bird homolog of A1 are better than single units at discriminating song from simpler sounds, including tones, ripples and noise (Grace, Amin et al. 2003).

DISCUSSION

Although theoretical studies have suggested that precise spike timing can provide a rapid and accurate code for stimulus recognition and categorization (Buonomano and Merzenich 1999; VanRullen, Guyonneau et al. 2005), studies in visual and somatosensory cortex have indicated that firing rates averaged across 50–500 ms are best correlated with behavior (Britten, Shadlen et al. 1992; Parker and Newsome 1998; Pruett, Sinclair et al. 2001; Romo and Salinas 2003; Liu and Newsome 2005). Our results suggest that the representation of consonant sounds in A1 is based on time windows that are approximately 50 times more precise.

The greater temporal precision observed in this study could be specific to the auditory system (DeWeese, Hromadka et al. 2005; Narayan, Grana et al. 2006; Schnupp, Hall et al. 2006; Walker, Ahmed et al. 2007). However, it is also possible that spike timing is important in all modalities when transient stimuli are involved (Weliky, Fiser et al. 2003; Grace, Amin et al. 2003; Richmond, Optican et al. 1990). The latter hypothesis is supported by observations of a rate-based code for steady-state vowels (Ohl and Scheich 1997; Versnel and Shamma 1998; Qin, Chimoto et al. 2004; Young 2007) and by computational studies showing that cortical neurons can efficiently extract temporal patterns from populations of neurons in a manner that promotes accurate consonant categorization (Buonomano and Merzenich 1995). It will be important to test whether neural correlates of transient visual and tactile stimuli make use of spike timing.

Error-trial analysis in an awake behaving preparation, as well as lesion and microstimulation experiments, are needed to evaluate our hypothesis that consonant processing depends upon precise spike timing in A1. Recordings in higher cortical areas will be needed to establish whether temporal patterns or mean firing rates are better correlated with behavioral discrimination.

APPENDIX A

FIGURES

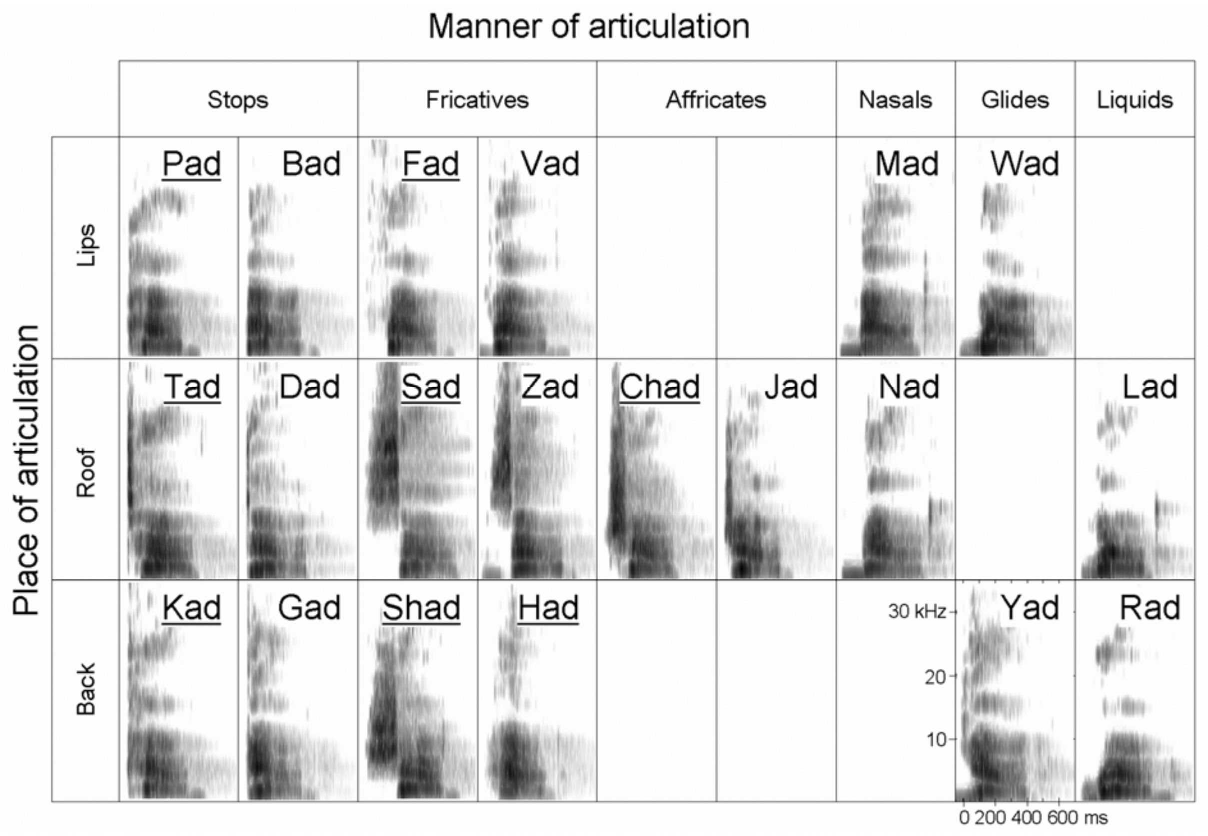


Figure A2.1. Spectrograms of each speech sound grouped by manner and place of articulation. Words with unvoiced initial consonants are underlined. Frequency is represented on the *y* axis (0–35 kHz) and time on the *x* axis (–50 to 700 ms). Speech sounds were shifted one octave higher to accommodate the rat hearing range. See Figure B2.1 for a more detailed view of the first 40 ms of each spectrogram.

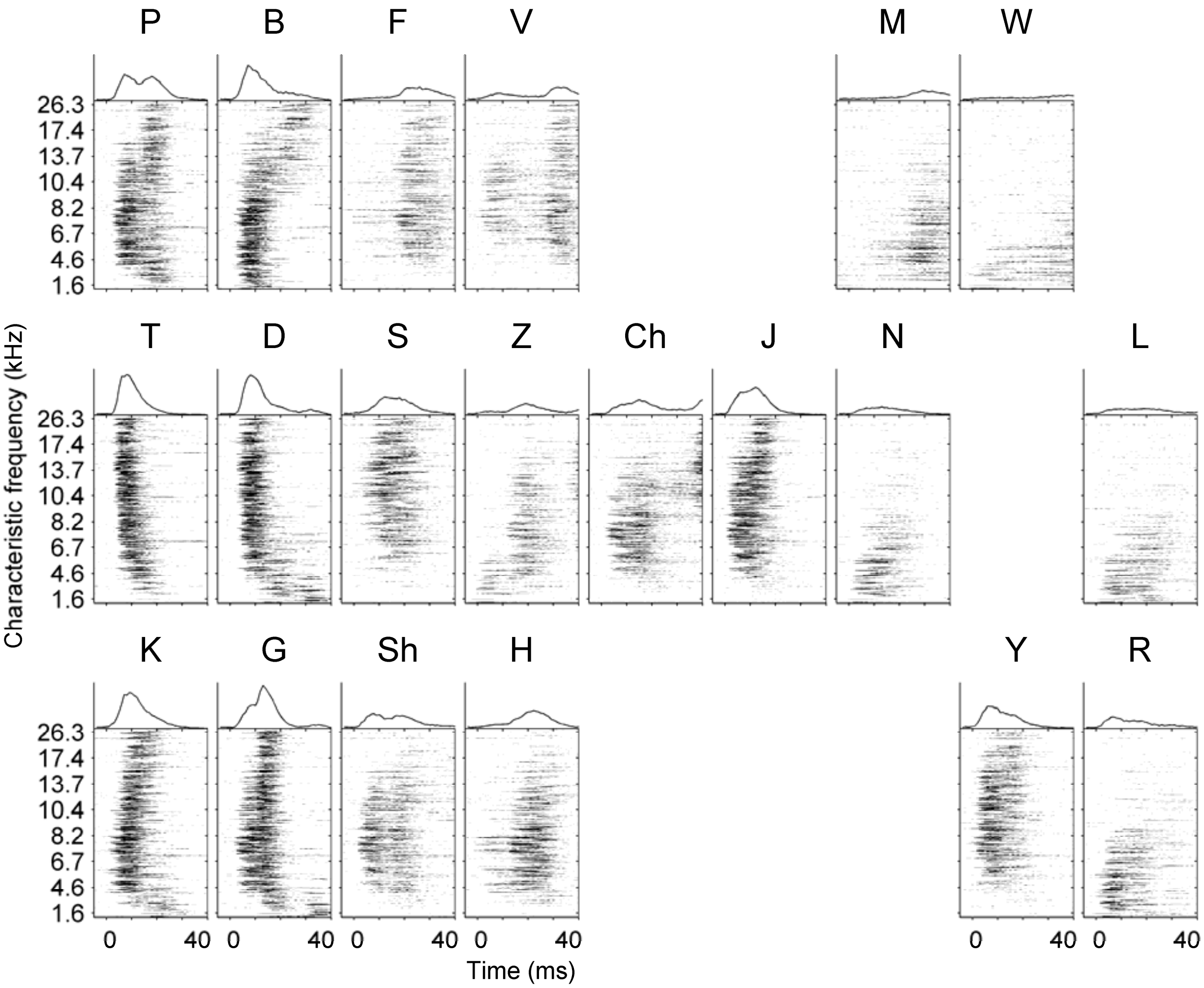


Figure A2.2. Neurograms depicting the onset response of rat A1 neurons to twenty English consonants. Multi-unit data was collected from 445 recording sites in eleven anesthetized experimentally naïve adult rats. Average post-stimulus time histograms (PSTH) derived from twenty repeats were ordered by the characteristic frequency (kHz) of each recording site (*y* axis). Time is represented on the *x* axis (–5 to 40 ms). The firing rate of each site is represented in grayscale, where black indicates 400 Hz. For comparison, the mean population PSTH evoked by each sound is plotted above the corresponding neurogram. For reference, 'gad' evoked the highest population firing rate of 288 Hz. As in Figure A2.1, rows differ in the place of articulation of each consonant, while columns differ in the manner of articulation. See Figure B2.2 for a direct comparison between stop consonant spectrograms and neurograms.

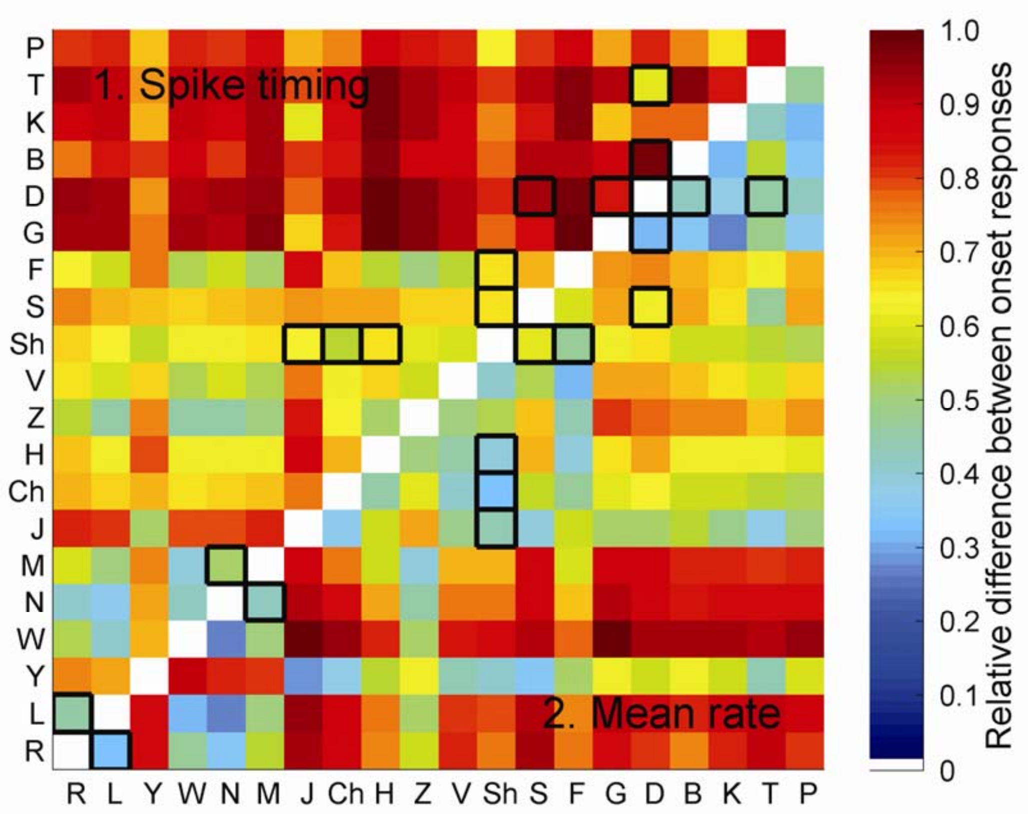


Figure A2.3. Predictions of consonant discrimination ability based on onset response similarity. The Euclidean distance between the neurograms evoked by each consonant pair (Figure A2.2) was computed using (**1**) spike timing with 1 ms bins or (**2**) average firing rate over a 40 ms bin (see Methods). Distance measures were normalized by the most distinct pair such that dark red squares indicate pairs that are highly distinct, and blue indicates pairs that are similar. Pairs that were tested behaviorally are outlined in black. (**1**) refers to the upper left half of the figure and (**2**) refers to the lower right.

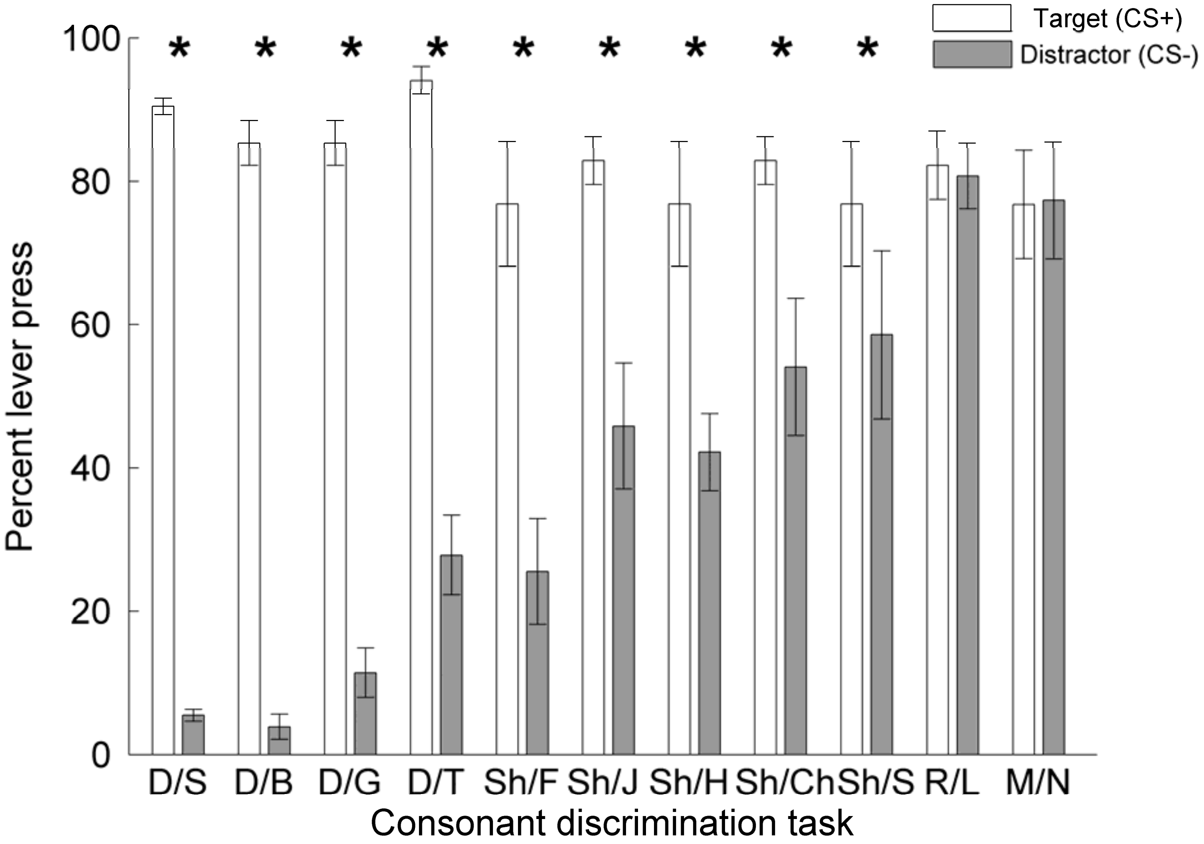


Figure A2.4. Behavioral discrimination of consonant sounds. Rats successfully discriminated nine of eleven consonant pairs evaluated. Open bars represent the target sound for each go/no-go task and filled bars represent the non-target sound. Error bars indicate s.e.m. across rats. Asterisks indicate significant discrimination (one tailed paired *t*-test, $P < 0.05$).

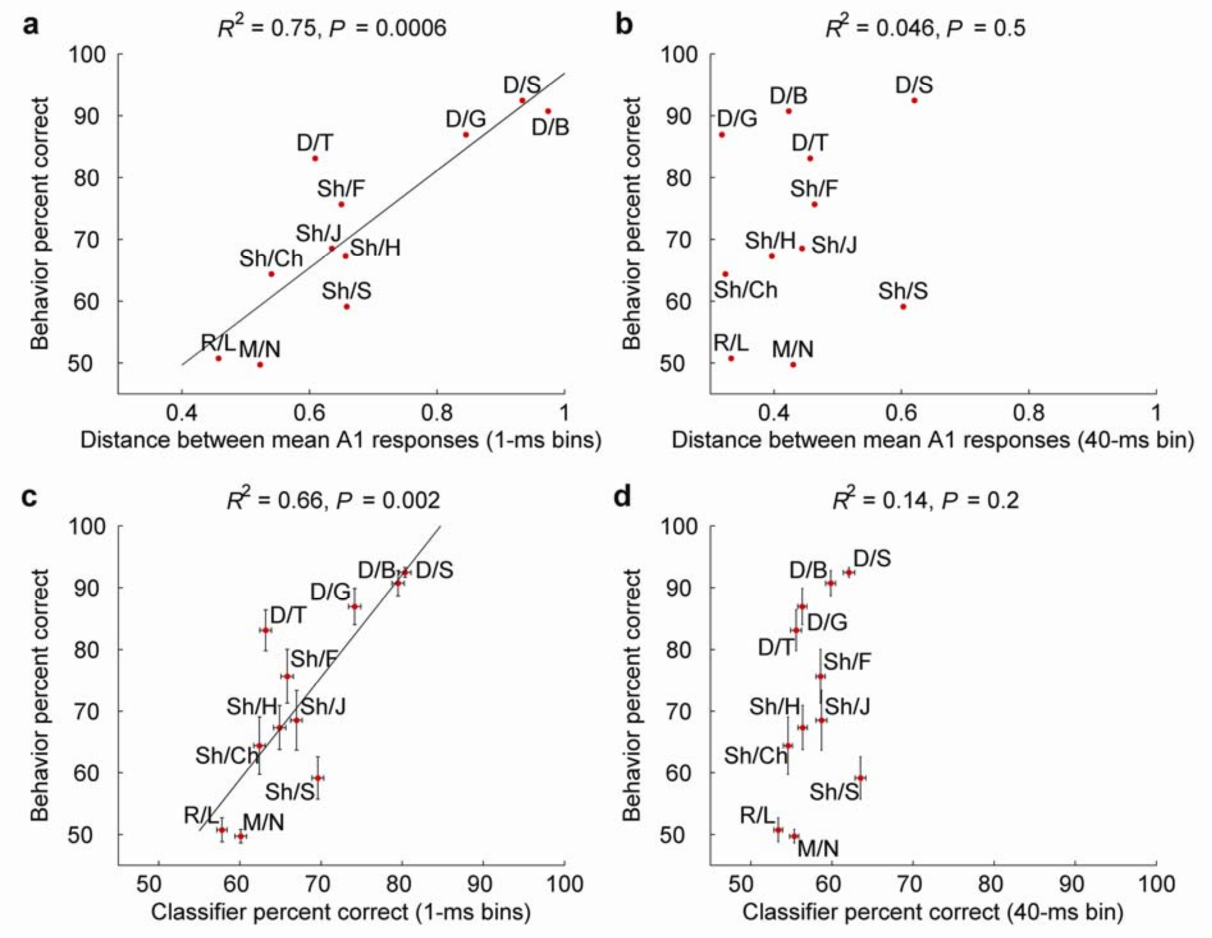


Figure A2.5. Both average A1 responses and trial-by-trial neural discrimination predicted consonant discrimination ability when temporal information was maintained. (**a**) The normalized Euclidean distance between neurogram onset patterns (Figure A2.3.1) was correlated with behavioral performance when action potential times were binned with 1 ms precision. (**b**) Distance was not correlated with behavior when a single 40 ms bin was used (Figure A2.3.2). (**c,d**) The average neural discrimination was correlated with behavioral discrimination when 1 ms bins were used (**c**), but not when a 40 ms bin was used (**d**). Neural discrimination was performed by a nearest-neighbor classifier (see methods) using a single sweep of neural activity. The eleven consonant pairs are printed next to each data point. Error bars indicate s.e.m. for behavioral and neural discrimination performance. Solid lines indicate the best linear fits when statistically significant. Analysis using other distance measures is shown in Table B2.1. Classifier performance on each task as a function of tone characteristic frequency is shown in Figure B2.5.

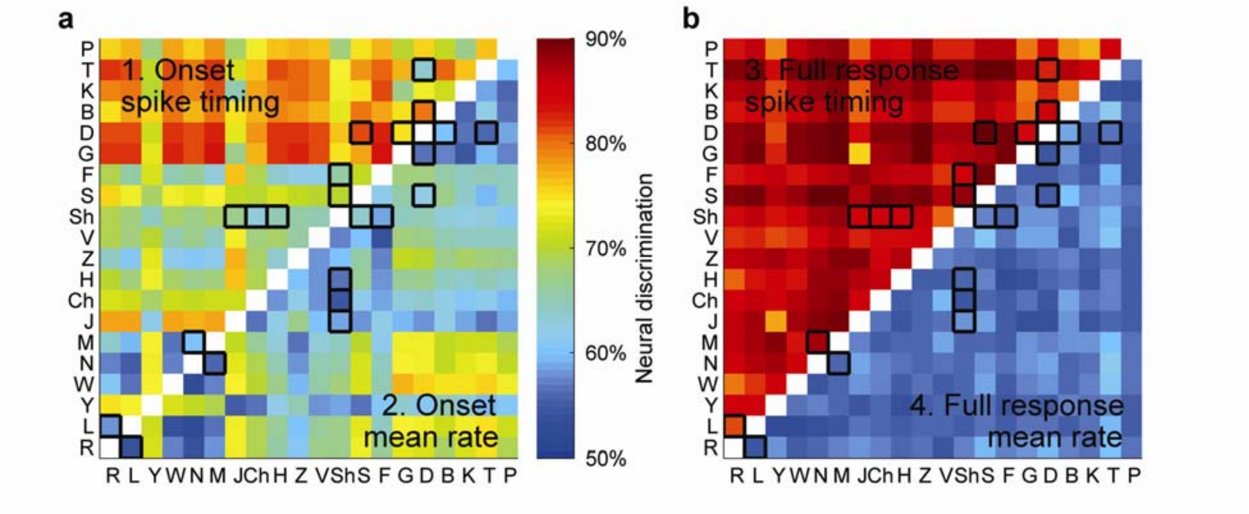


Figure A2.6. Predictions of consonant discrimination ability based on nearest-neighbor classifier. (**a.1**) Neural discrimination of each consonant pair using the 40 ms onset response with 1 ms bins. (**a.2**) Neural discrimination of each consonant pair using the average firing rate over the first 40 ms of the onset response. (**b.3**) Neural discrimination using the entire 700 ms stimulus duration with 1 ms bins. (**b.4**) Neural discrimination using the average rate over the 700 ms stimulus duration. Pairs that were tested behaviorally (Figure A2.4) are outlined in black.

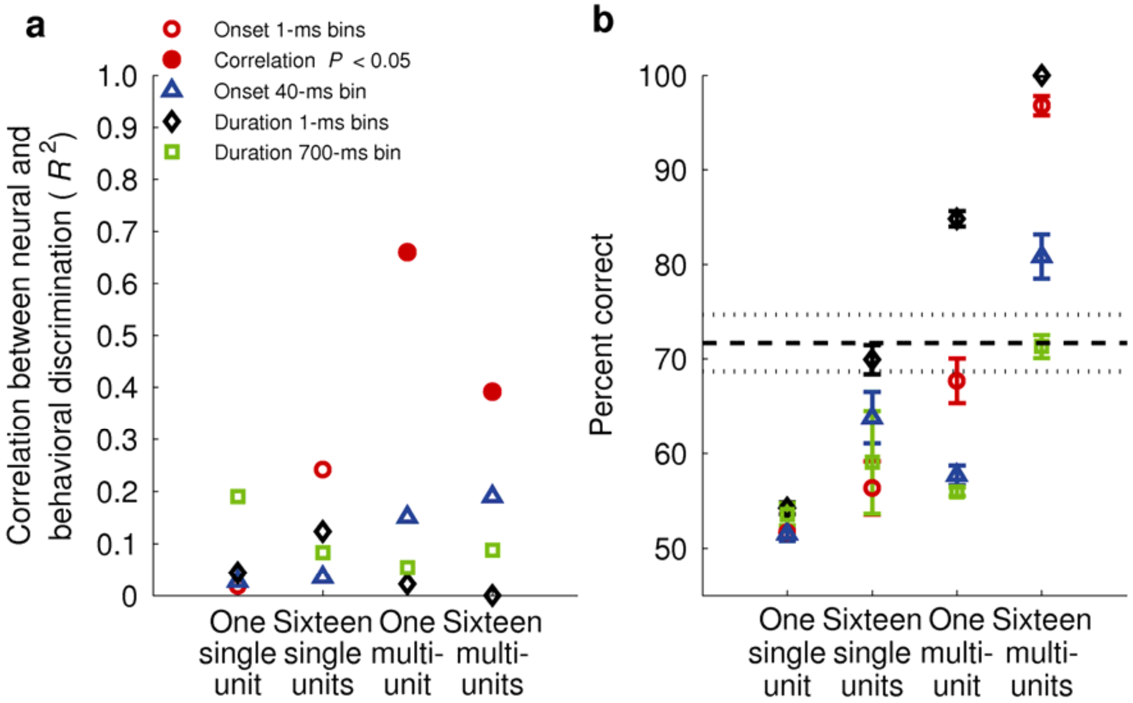


Figure A2.7. Neural discrimination using the onset activity pattern from individual multi-unit sites was best correlated with behavior. (**a**) Percent of variance across the eleven behavioral tasks that was explained using individual single units, sixteen single units, individual multi-units, and sets of sixteen multi-units. Each population size was evaluated with or without precise temporal information using the onset response or the entire response. Filled symbols indicate statistically significant correlations between neural and behavioral discrimination (Pearson's correlation coefficient, $P < 0.05$). Regardless of population size, neural discrimination based on mean firing rate (i.e. 40 or 700 ms bin) was not correlated with behavioral performance on the eleven tasks. (**b**) Mean neural discrimination on the eleven consonant discrimination tasks using individual single units, sixteen single units, individual multi-units, and sets of sixteen multi-units. The solid line indicates average behavioral discrimination. Error bars indicate s.e.m. across the eleven tasks. The complete distributions of neural discrimination by single and multi-unit activity for each task are provided in Figure B2.8.

APPENDIX B

SUPPLEMENTARY DATA

Neurogram quantification and comparison with earlier physiology studies

Consistent with earlier observations in humans and non-human animals (Steinschneider et al., 1999; Steinschneider et al., 2003; Steinschneider et al., 2005; Wong and Schreiner, 2003), more spikes were evoked during the second peak of the unvoiced stop consonant /p/ than during the second peak of the voiced stop consonant /b/ (1.05 ± 0.08 additional spikes from 70 to 170 ms after consonant onset, two tailed *t*-test, $n = 445$, $P < 0.00001$). An increase in firing rate was also seen in the unvoiced stop consonants /k/ and /t/ compared to the voiced stop consonants /g/ and /d/ (two tailed *t*-test, $n = 445$, $P < 0.00001$, Figure B2.3 and Supplementary Video 2). The stop consonants generated the strongest onset response (40 ms), the fricatives and affricates had an intermediate response, and the sonorants (nasals, glides, and liquids) generated the weakest response. The six stops had an average peak firing rate of 524 ± 13 Hz, the eight fricatives and affricates had an average peak firing rate of 284 ± 9 Hz, and the six sonorants had an average peak firing rate of 203 ± 7 Hz (ANOVA $F_{19,444} = 284$, $MSE = 43745$, $P < 0.00001$, Figure A2.2).

Consistent with earlier observations in human and monkey auditory cortex, consonants that differ in place of articulation generate different spatial patterns of A1 activity (Blumstein and Stevens, 1979; Obleser et al., 2003; Steinschneider et al., 1995). In the low frequency region of A1 (< 9 kHz), the consonant /d/ had the weakest response of the voiced stop consonants (two tailed *t*-test, $n = 445$, $P < 0.001$, 40 ms onset window), while in the high frequency region of A1

(> 9 kHz), the consonant /b/ had the weakest response (two tailed *t*-test, $n = 445$, $P < 0.00001$). The differential spatial patterns were most apparent when comparing activity produced during 1 ms windows. For example, 8 ms after sound onset, /b/ primarily activated sites with 2–9 kHz CF's, /g/ activated sites with 6–14 kHz CF's, and /d/ activated sites with 8–25 kHz CF's (Supplementary Video 1).

The fact that voicing, manner, and place of articulation alter A1 activity as in earlier studies conducted in humans and non-human animals with or without anesthesia supports the use of the rat model and indicates that frequency shifting speech sounds to match the rat hearing range does not dramatically distort the previously reported pattern of cortical responses (Steinschneider et al., 1995; Steinschneider et al., 1999; Steinschneider et al., 2003; Steinschneider et al., 2005; Wong and Schreiner, 2003; Blumstein and Stevens, 1979; Obleser et al., 2003).

Behavioral discrimination time course

Average performance during the seventh and eighth days of training was used to estimate the relative difficulty of different consonant discriminations. By the end of day eight, each rat had performed 3544 ± 757 (mean ± s.d.) trials. All rats reached asymptotic performance by the seventh day of training on each task.

Two additional days of training (i.e. days 9 and 10) were used to confirm asymptotic performance (paired *t*-test, $P > 0.05$). However, for /d/ versus /t/, only one rat was trained for ten days and the others were inadvertently trained for only nine ($n = 4$ rats) or seven ($n = 1$ rat) days. The learning curve for the /d/ versus /t/ task is shown in Figure B2.4a. Although the false alarm rate was trending down during the last three days of training, discrimination quantified by *d'* did not improve because the hit rate was also trending down (Figure B2.4b).

To ensure discrimination ability was not strongly influenced by task order, each rat underwent a review session (after completion of all of the two week consonant discrimination tasks) during which it reviewed each of the tasks it was trained on for two additional days per task. Performance was not significantly different between the seventh and eighth days of training and the two days of review training (ANOVA $F_{9,36} = 0.49$, $MSE = 0.28$, $P = 0.87$).

Onset neurograms only predict behavior when spatial and temporal information is preserved

The correlation between behavioral discrimination and the Euclidean distance between neurogram pairs was highest when the classifier was provided 40 ms of A1 onset activity in 1 ms bins ($R^2 = 0.75$, $P = 0.0006$, Figure A2.5a). When spike timing information was removed by using a single 40 ms bin, the distance between neurogram pairs was not significantly correlated with behavior ($R^2 = 0.046$, $P = 0.5$, Figure A2.5b). When the entire duration (i.e. including the vowel and final /d/) was included, the distances between neurogram pairs (Figure B2.3) were not significantly correlated with behavior whether 1 ms bins were used ($R^2 = 0.07$, $P = 0.4$) or a single 700 ms bin was used ($R^2 = 0.08$, $P = 0.4$). The critical observation is that cortical responses are only correlated with behavior when spike timing is considered.

To test whether spatial information is also required for accurate consonant discrimination, we compared the Euclidean distances between the mean population PSTH's (i.e. the PSTH above each neurogram in Figure A2.2) for each pair of sounds. Although precise temporal information was maintained, behavior was not significantly correlated with the similarity of responses averaged across all 445 A1 sites ($R^2 = 0.23$, $P = 0.1$). Similarly, no correlation was observed when both spatial and temporal information was removed (i.e. comparing the total number of spikes in each neurogram in Figure A2.2, $R^2 = 0.14$, $P = 0.2$). These results indicate that the precise spatiotemporal pattern of A1 activity is best correlated with behavior.

The Euclidean distances between the first 40 ms of the consonant spectrograms (Figure B2.1) were modestly correlated with behavior ($R^2 = 0.40$, $P = 0.038$), and no correlation was observed when the entire spectrogram was used ($R^2 = 0.03$, $P = 0.61$). This result confirms our expectation that behavioral discrimination ability is related to acoustic similarity.

Relationship between tone and consonant responses

As expected, tone response properties explain much of the variance in the ability of different recording sites to discriminate each of the eleven pairs evaluated. For example, low frequency neurons were better than high frequency neurons at distinguishing nasals and liquids (Figure B2.6 black lines), and the opposite pattern was seen for fricatives and affricates (Figure B2.6 green lines). Stops which begin with broadband noise could be reliably discriminated by neurons from any region of A1. These observations are consistent with the acoustics of the consonants studied (Figure B2.1) (Stevens, 2000). Broader tone frequency tuning, shorter response latency, lower tone thresholds, and greater tone evoked response strength were each positively correlated with neural discrimination ability ($R^2 = 0.15$, $P = 1.1 \text{ x } 10^{-17}$; $R^2 = 0.2$, $P = 7.8 \text{ x } 10^{-23}$; $R^2 = 0.1$, $P = 3.2 \text{ x } 10^{-12}$; $R^2 = 0.38$, $P = 2.1 \text{ x } 10^{-47}$). Collectively, these factors account for 47.3% of the variance in average consonant discrimination ability ($P = 0$).

Other analysis models

Although there are an infinite number of potential models that could be used to read out neural activity, we attempted to use common analysis methods that relied on as few assumptions as possible. Euclidean distance was selected to compare average and single trial responses (Figures A2.3 and A2.5–2.7) as it is the most commonly used metric of similarity. Our results are not dependent on the use of Euclidean distance. When Chebychev and city block distances were computed with 1 ms bins, neural discrimination was also well correlated with behavior

(Table B2.1). Neither method was significantly correlated with behavior when a single bin of activity was used.

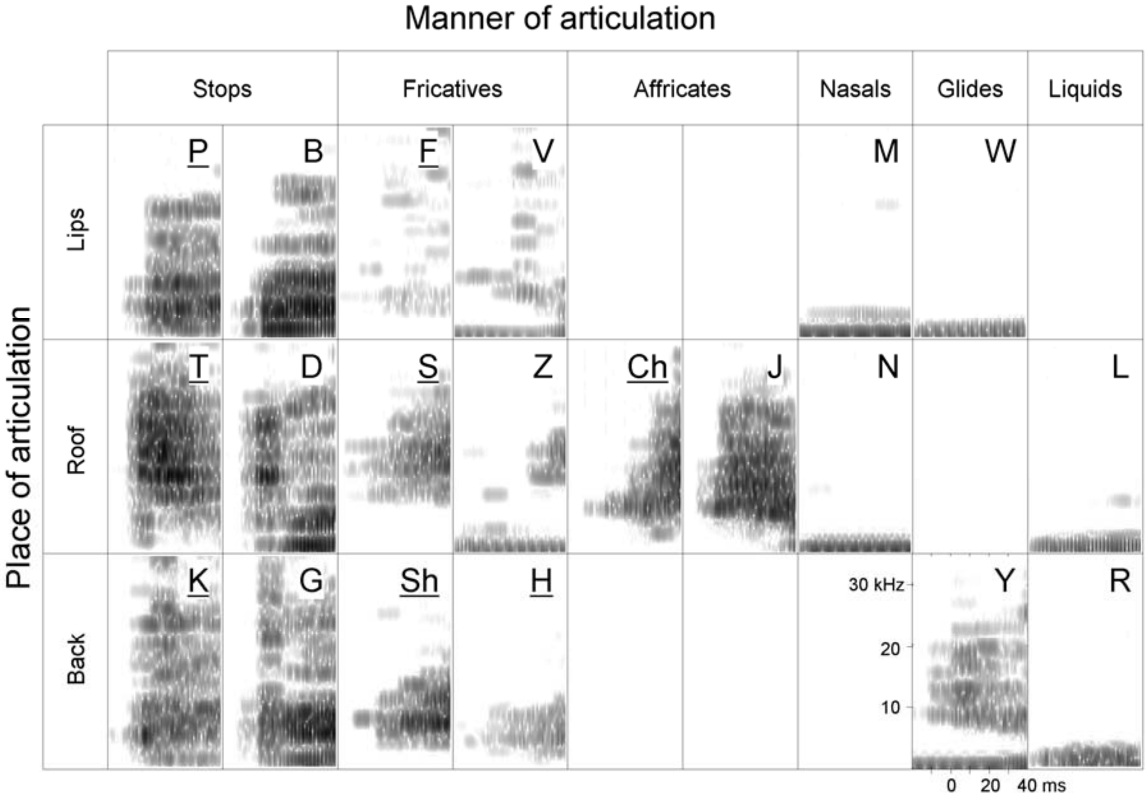


Figure B2.1. Spectrograms of each speech sound grouped by manner and place of articulation. Words with unvoiced initial consonants are underlined. All conventions are identical to Figure A2.1, except that the *x* axis has been limited to 40 ms to facilitate comparison with Figure A2.2.

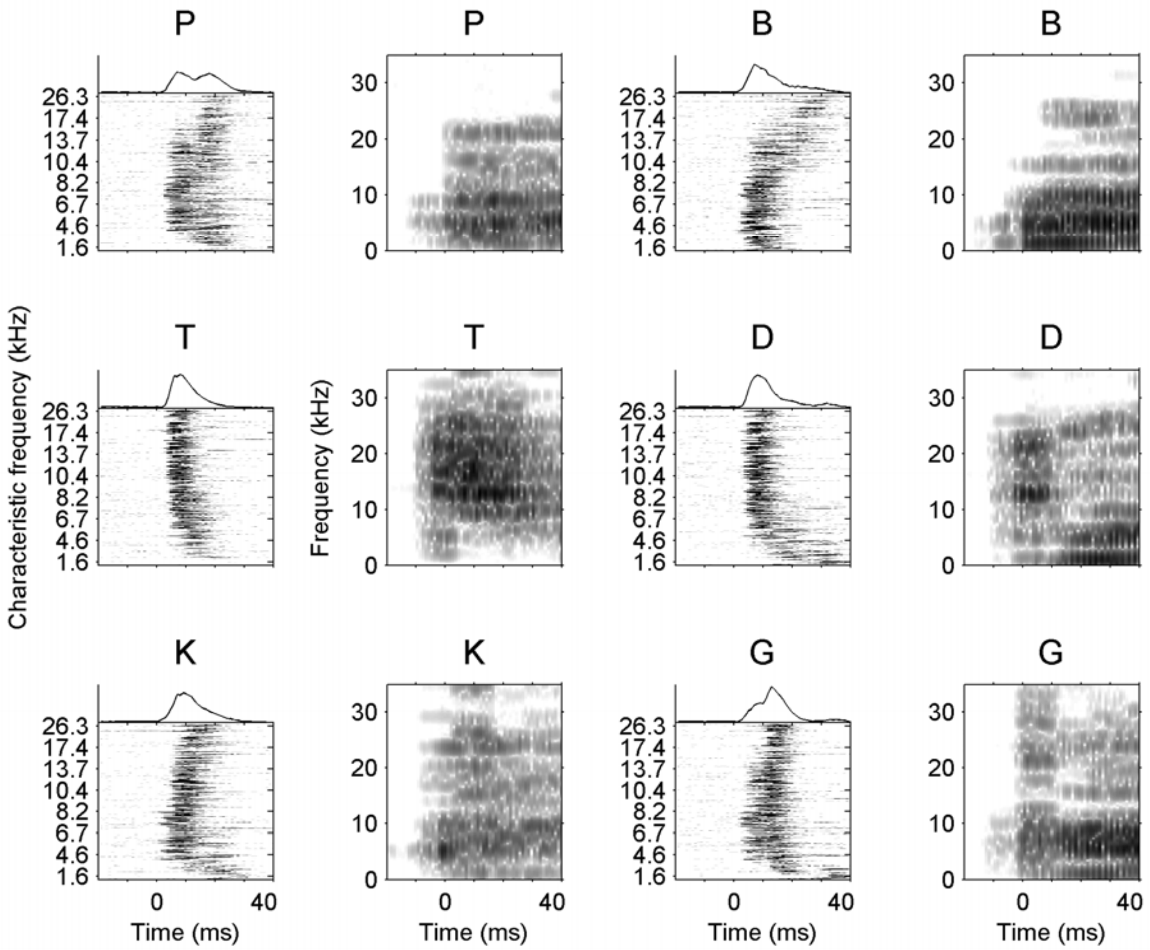


Figure B2.2. Onset neurograms and spectrograms of each stop consonant are shown to facilitate visual comparison. All conventions are identical to Figure A2.2 and Figure B2.1.

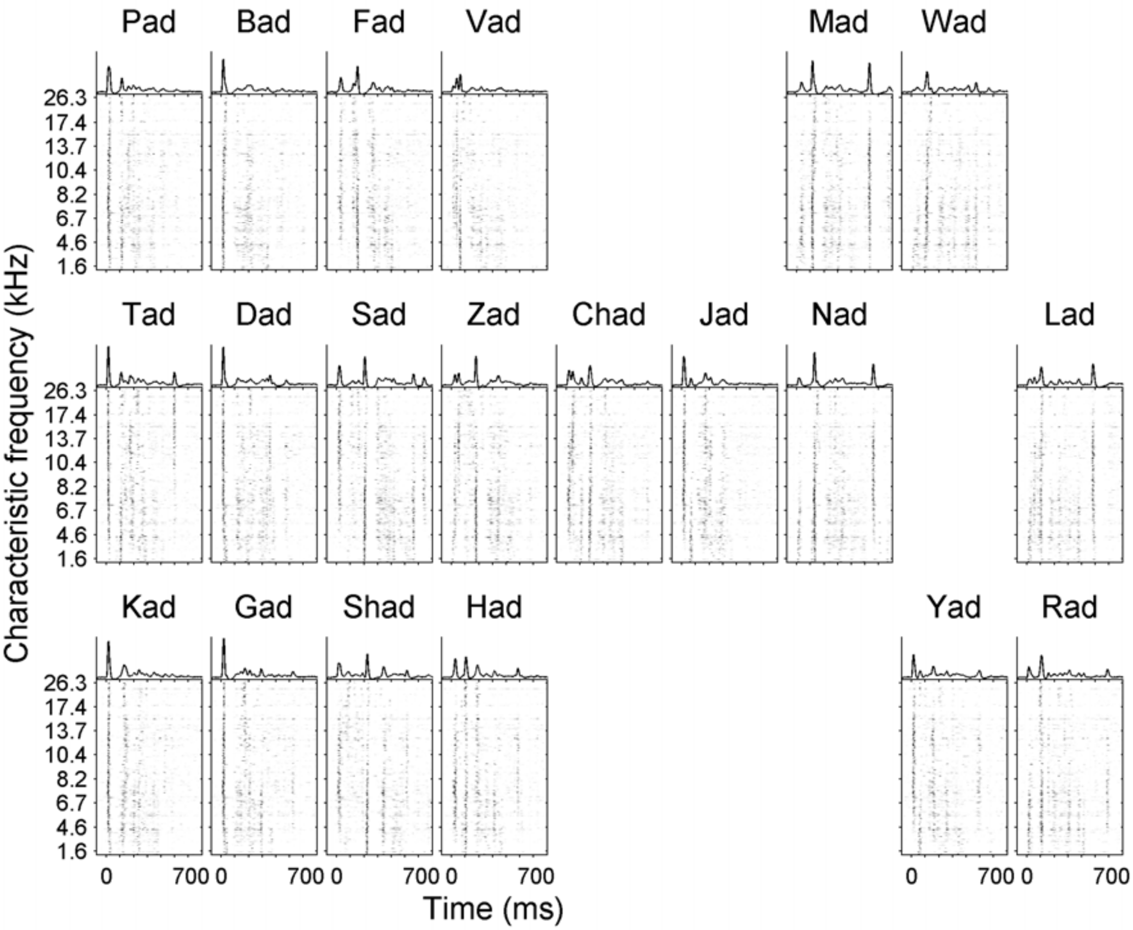


Figure B2.3. Neurograms and population PSTH's of rat A1 neurons to twenty English consonants. Post-stimulus time histograms (PSTH) were ordered by the characteristic frequency (kHz) of each recording site (*y* axis). Multi-unit data was collected from each of 445 multi-unit recording sites in eleven experimentally naïve adult rats. Conventions are identical to Figure A2.2, except for the time interval included on the *x* axis (–75 to 700 ms). Late response peaks were due to the onset of the vowel /a/ and the terminal /d/ (as in 'sad').

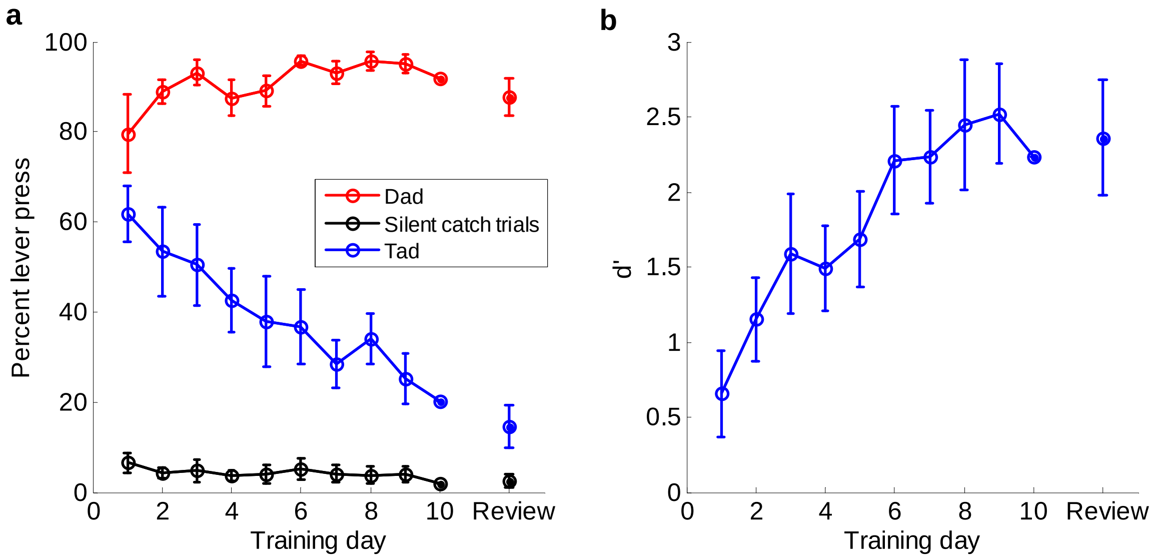


Figure B2.4. Average learning curve illustrates asymptotic performance by the seventh and eighth days of /d/ versus /t/ discrimination (*n* = 6 rats). In (**a**) the percent of trials in which rats pressed the lever is given. In (**b**), the discrimination between /d/ and /t/ is shown using *d'*. Review indicates performance during review sessions conducted after several weeks of training on other discriminations. For this task, we only had data from four rats on day 9 and one rat on day ten (see methods).

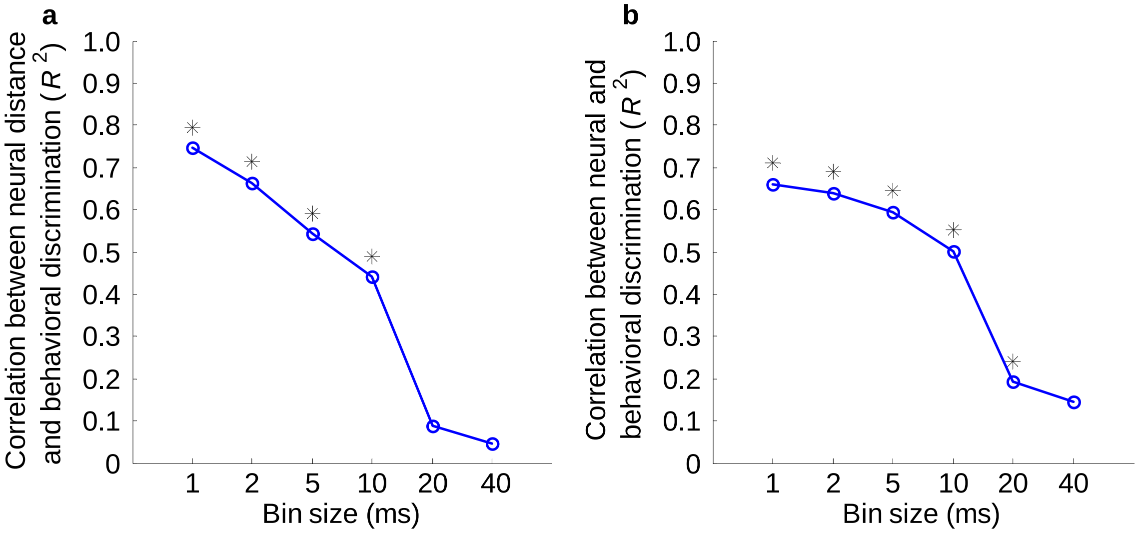


Figure B2.5. Percent of variance explained decreases as bin size increases. (**a**) Percent of variance across the 11 consonant discrimination tasks that is explained using Euclidean distance calculated from the 40 ms onset response. (**b**) Percent of variance explained using classifier percent correct. The observation that neural discrimination in the auditory system depends upon temporal information is consistent with earlier studies [9–11]. Asterisks indicate correlations that are statistically significant (Pearson's correlation coefficient, $P < 0.05$).

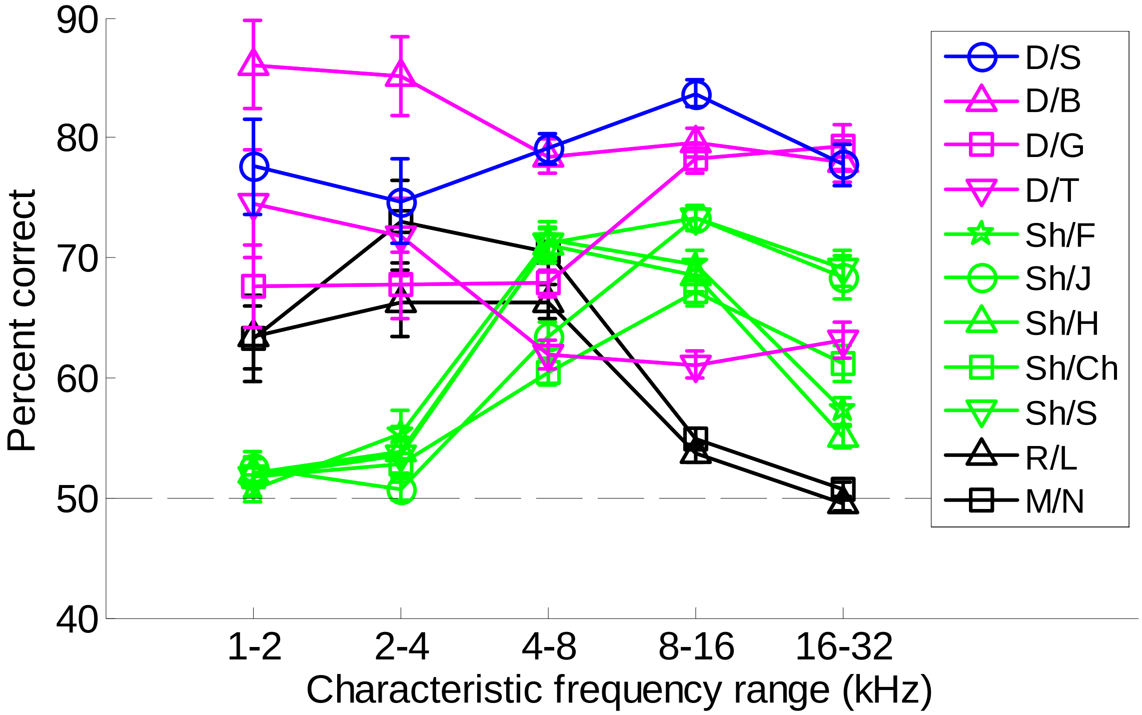


Figure B2.6. Classifier performance on each task as a function of tone characteristic frequency. High frequency neurons provide little information about nasals and liquids (black). Low frequency neurons provide little information about fricatives and affricates (green). Both high and low frequency neurons provide information about stop consonants (purple) and manner of articulation (blue). Classifier data is identical to that used in Figure A2.5c.

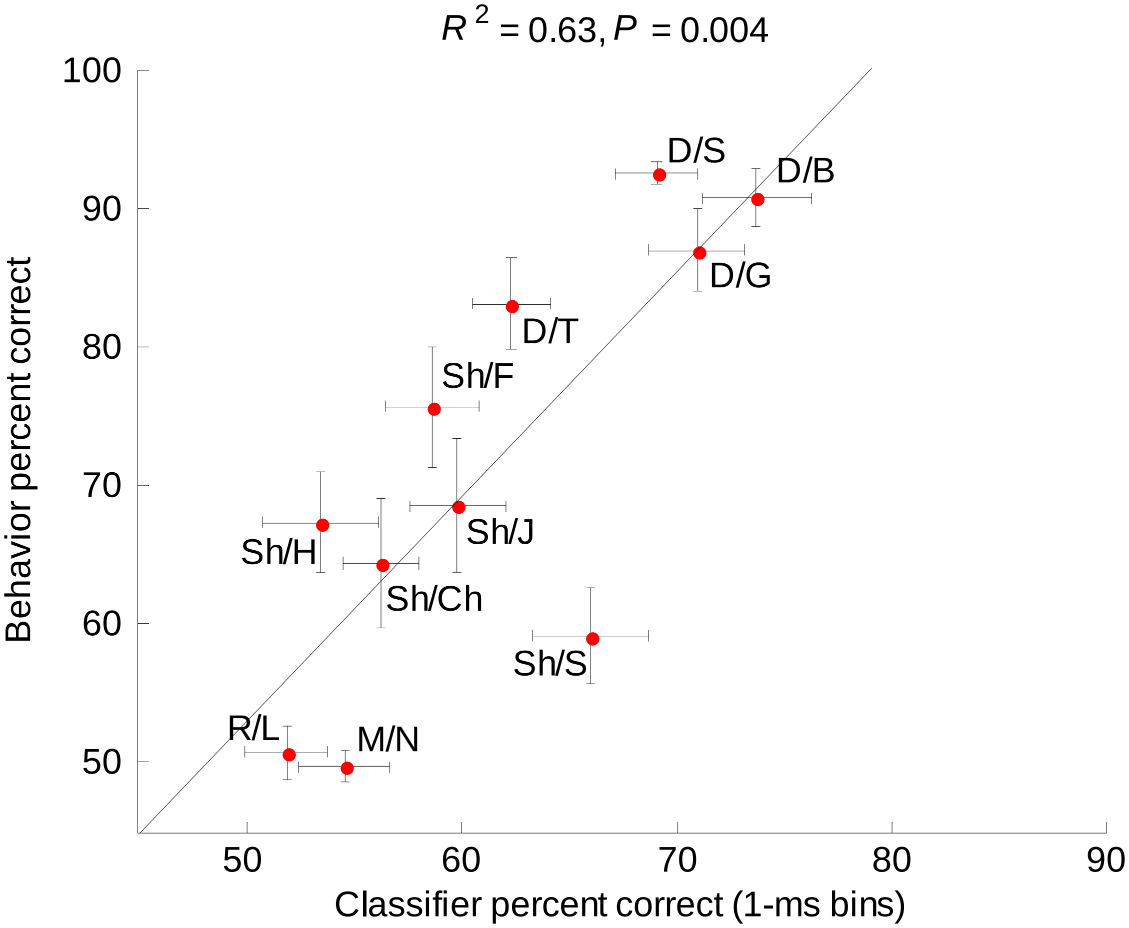


Figure B2.7. Strong correlation between behavioral performance on 11 consonant discrimination tasks and neural discrimination using data from 40 A1 multi-unit recording sites in 6 chronically implanted awake rats. Conventions are identical to Figure A2.5c, except data was recorded without anesthesia. The solid line indicates the best linear fit ($R^2 = 0.63$, $P = 0.004$ compared to $R^2 = 0.66$, $P = 0.002$ in anesthetized rats).

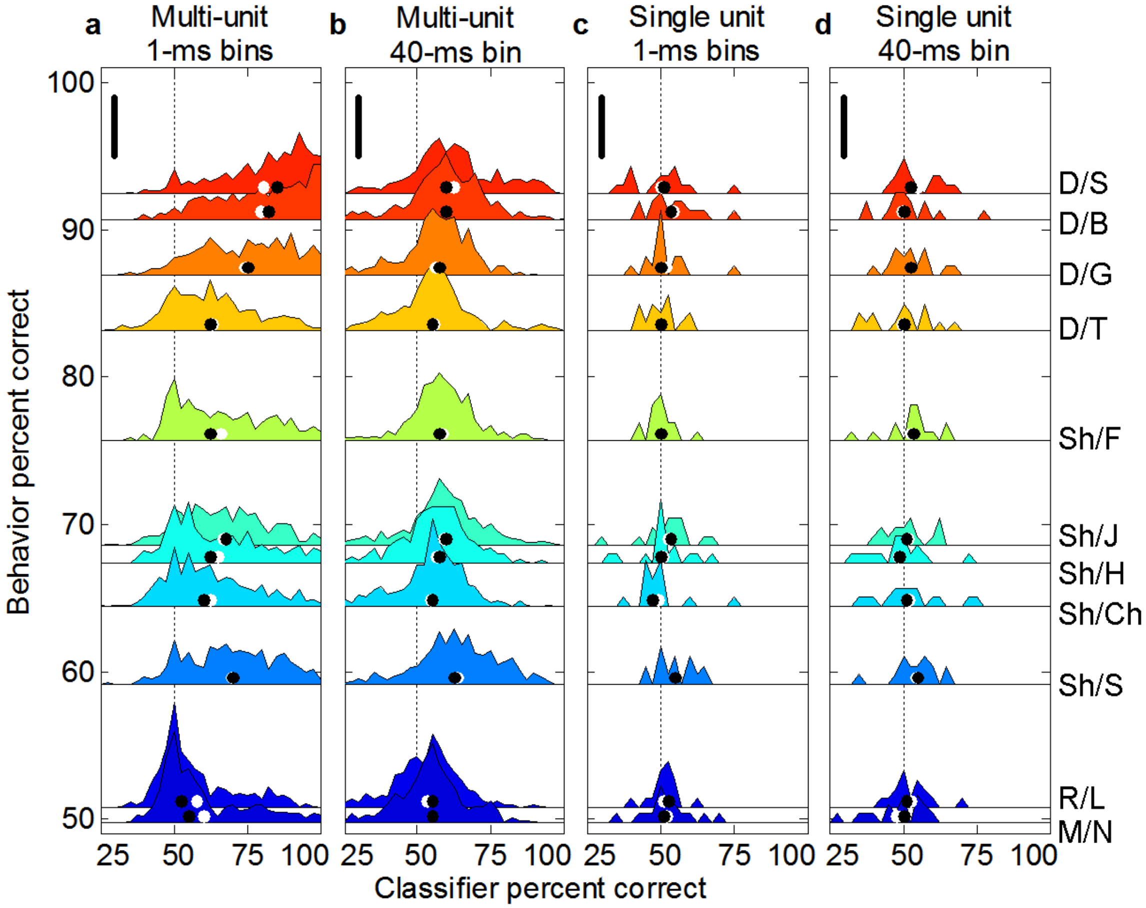


Figure B2.8. Neural discrimination using the onset response of individual multi-units with precise temporal information predicts behavioral discrimination. (**a**) The population distribution of the onset response of 445 multi-unit sites using forty 1 ms bins and (**b**) a single 40 ms bin. (**c**) The population distribution of the onset response of 16 single unit sites using forty 1 ms bins and (**d**) a single 40 ms bin. Black circles indicate the median classifier percent correct for each of the 11 consonant discrimination tasks, and white circles indicate the mean classifier percent correct. The dashed line indicates chance performance. The scale bar at the top left of each subplot represents 10% of multi-units sites (**a,b**) and 40% of single unit sites (**c,d**). Classifier percent correct is binned into 2.5% correct bins. Each task is labeled to the right of the figure.

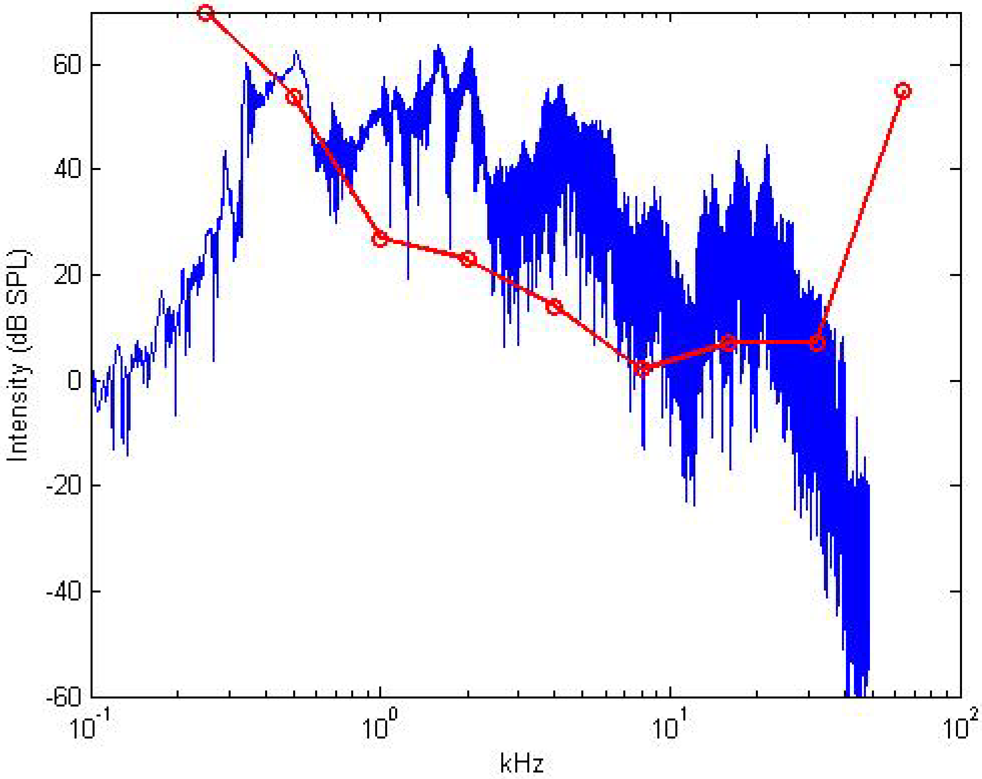
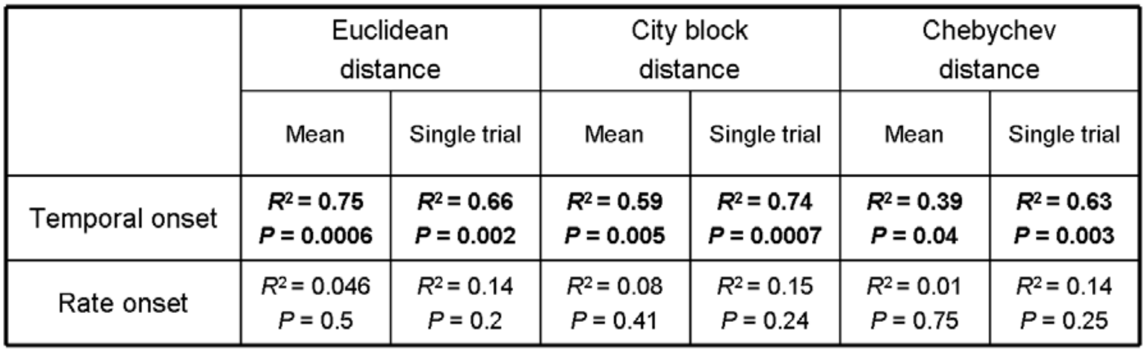


Figure B2.9. The spectrum sampled from a typical stimulus (the word ‘sad’, blue line) shifted up using STRAIGHT superimposed on the rat audiogram (Kelly and Masterton 1977, red line). The amplitude scale (in decibels) of the speech was shifted to match the spectrum of the stimulus to the audiogram.

Table B2.1. Neural activity was only correlated with behavioral performance when temporal precision was preserved. Both mean activity and single trial activity were significantly correlated with behavior when temporal information was preserved using Euclidean distance, City block distance, and Chebychev distance. There was no significant correlation between behavior and neural activity when temporal information was eliminated using mean or single trial activity for each of the three models. Significant correlations are shown in bold.



	Euclidean distance		City block distance		Chebychev distance	
	Mean	Single trial	Mean	Single trial	Mean	Single trial
Temporal onset	**R^2 = 0.75** **P = 0.0006**	**R^2 = 0.66** **P = 0.002**	**R^2 = 0.59** **P = 0.005**	**R^2 = 0.74** **P = 0.0007**	**R^2 = 0.39** **P = 0.04**	**R^2 = 0.63** **P = 0.003**
Rate onset	R^2 = 0.046 P = 0.5	R^2 = 0.14 P = 0.2	R^2 = 0.08 P = 0.41	R^2 = 0.15 P = 0.24	R^2 = 0.01 P = 0.75	R^2 = 0.14 P = 0.25

Supplementary Video 1. The video illustrates the spatiotemporal activity patterns evoked in primary auditory cortex by twenty different consonant sounds (Figure B2.1). The color of each polygon indicates the multi-unit activity at each of 63 recording sites in a single rat. Pairs of sounds that evoke similar patterns (i.e. /m/ and /n/) are difficult for rats to discriminate, while pairs that evoke distinct patterns are easy (i.e. /d/ and /b/, Figure A2.4). The map of characteristic frequency (CF) illustrates the topographic organization of tone frequency tuning. Red indicates instantaneous firing rates above 800 Hz, yellow and orange denote 550–800 Hz, light blue and green denote 100–550 Hz, and dark blue denotes rates below 100 Hz. Activity patterns reflect the average response to twenty repeats. The blue lines under each activity map indicate the average firing rate for all 63 A1 sites. The red lines mark the time at which each spatial activity pattern occurs relative to response onset. The patterns were smoothed using a 3 ms sliding window.

Supplementary Video 2. The video illustrates the spatiotemporal activity patterns evoked in primary auditory cortex by twenty different consonant sounds (Figure A2.1). The conventions are the same except that the frame rate is faster and the video continues until 250 ms after response onset. Late response peaks were due to the onset of the vowel /a/ (as in ‘sad’). The terminal /d/ of each sound was excluded to reduce video size.

REFERENCES

Ahissar, E., S. Nagarajan, et al. (2001). "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex." Proc Natl Acad Sci U S A **98**(23): 13367-72.

Bertoncini, J., R. Bijeljac-Babic, et al. (1987). "Discrimination in neonates of very short CVs." J Acoust Soc Am **82**(1): 31-37.

Blumstein, S. E. and K. N. Stevens (1979). "Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants." J Acoust Soc Am **66**(4): 1001-17.

Britten, K. H., M. N. Shadlen, et al. (1992). "The analysis of visual motion: a comparison of neuronal and psychophysical performance." J Neurosci **12**(12): 4745-65.

Buonomano, D. V. and M. Merzenich (1999). "A neural network model of temporal code generation and position-invariant pattern recognition." Neural Comput **11**(1): 103-16.

Buonomano, D. V. and M. M. Merzenich (1995). "Temporal information transformed into a spatial code by a neural network with realistic properties." Science **267**(5200): 1028-30.

Cooke, J. E., H. Zhang, et al. (2007). "Detection of sinusoidal amplitude modulated sounds: deficits after bilateral lesions of auditory cortex in the rat." Hear Res **231**(1-2): 90-9.

DeWeese, M. R., T. Hromadka, et al. (2005). "Reliability and representational bandwidth in the auditory cortex." Neuron **48**(3): 479-88.

Dewson, J. H., 3rd, K. H. Pribram, et al. (1969). "Effects of ablations of temporal cortex upon speech sound discrimination in the monkey." Exp Neurol **24**(4): 579-91.

Dooling, R. J., K. Okanoya, et al. (1989). "Speech perception by budgerigars (Melopsittacus undulatus): the voiced-voiceless distinction." Percept Psychophys **46**(1): 65-71.

Engineer, N. D., C. R. Percaccio, et al. (2004). "Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons." J Neurophysiol **92**(1): 73-82.

Foffani, G. and K. A. Moxon (2004). "PSTH-based classification of sensory stimuli using ensembles of single neurons." J Neurosci Methods **135**(1-2): 107-20.

Fowler, C. A., J. M. Brown, et al. (2003). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks " Journal of Memory and Language **49**(3): 396-413.

Grace, J. A., N. Amin, et al. (2003). "Selectivity for conspecific song in the zebra finch auditory forebrain." J Neurophysiol **89**(1): 472-87.

Heffner, H. E. and R. S. Heffner (1989). "Effect of restricted cortical lesions on absolute thresholds and aphasia-like deficits in Japanese macaques." Behav Neurosci **103**(1): 158-69.

Jongman, A. (1989). "Duration of frication noise required for identification of English fricatives." J Acoust Soc Am **85**(4): 1718-25.

Kawahara, H. (1997). "Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited." Proceedings of the ICASSP **2**: 1303-1306.

Kelly, J.B. and Masterton, R.B. (1977). “Auditory sensitivity of the albino rat.” Journal of Comparative and Physiological Psychology **91**(4): 930-936.

Kluender, K. R., R. L. Diehl, et al. (1987). "Japanese quail can learn phonetic categories." Science **237**(4819): 1195-7.

Kuhl, P. K. and J. D. Miller (1975). "Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants." Science **190**(4209): 69-72.

Liu, J. and W. T. Newsome (2005). "Correlation between speed perception and neural activity in the middle temporal visual area." J Neurosci **25**(3): 711-22.

Miller, G. A. and P. E. Nicely (1955). "An Analysis of Perceptual Confusions Among Some English Consonants." J Acoust Soc Am **27**(2): 338-352.

Narayan, R., G. Grana, et al. (2006). "Distinct time scales in cortical discrimination of natural sounds in songbirds." J Neurophysiol **96**(1): 252-8.

Obleser, J., Lahiri, A. & Eulitz, C. Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 millisecinds post syllable onset. *NeuroImage* **20**, 1839–1847 (2003).

Ohl, F. W. and H. Scheich (1997). "Orderly cortical representation of vowels based on formant interaction." Proc Natl Acad Sci U S A **94**(17): 9440-4.

Orduna, I., E. Mercado, 3rd, et al. (2005). "Cortical responses in rats predict perceptual sensitivities to complex sounds." Behav Neurosci **119**(1): 256-64.

Parker, A. J. and W. T. Newsome (1998). "Sense and the single neuron: probing the physiology of perception." Annu Rev Neurosci **21**: 227-77.

Pruett, J. R., Jr., R. J. Sinclair, et al. (2001). "Neural correlates for roughness choice in monkey second somatosensory cortex (SII)." J Neurophysiol **86**(4): 2069-80.

Qin, L., S. Chimoto, et al. (2004). "Spectral-shape preference of primary auditory cortex neurons in awake cats." Brain Res **1024**(1-2): 167-75.

Ramus, F., M. D. Hauser, et al. (2000). "Language discrimination by human newborns and by cotton-top tamarin monkeys." Science **288**(5464): 349-51.

Reed, P., P. Howell, et al. (2003). "Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds." J Exp Anal Behav **80**(2): 205-15.

Rennaker, R. L., A. M. Ruyle, et al. (2005). "An economical multi-channel cortical electrode array for extended periods of recording during behavior." J Neurosci Methods **142**(1): 97-105.

Rennaker, R. L., S. Street, et al. (2005). "A comparison of chronic multi-channel cortical implantation techniques: manual versus mechanical insertion." J Neurosci Methods **142**(2): 169-76.

Richmond, B. J., L. M. Optican, et al. (1990). "Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations." J Neurophysiol **64**(2): 351-69.

Romo, R. and E. Salinas (2003). "Flutter discrimination: neural codes, perception, memory and decision making." Nat Rev Neurosci **4**(3): 203-18.

Rybalko, N., D. Suta, et al. (2006). "Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats." Eur J Neurosci **23**(6): 1614-22.

Schnupp, J. W., T. M. Hall, et al. (2006). "Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex." J Neurosci **26**(18): 4785-95.

Sinnott, J. M. and C. H. Brown (1997). "Perception of the American English liquid /ra-la/ contrast by humans and monkeys." J Acoust Soc Am **102**(1): 588-602.

Steinschneider, M., Y. I. Fishman, et al. (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**(1): 307-21.

Steinschneider, M., D. Reser, et al. (1995). "Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey." Brain Res **674**(1): 147-52.

Steinschneider, M., I. O. Volkov, et al. (2005). "Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter." Cereb Cortex **15**(2): 170-86.

Steinschneider, M., I. O. Volkov, et al. (1999). "Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex." J Neurophysiol **82**(5): 2346-57.

Stevens, K.N. *Acoustic Phonetics* (MIT Press, Cambridge, MA, 2000).

Tavabi, K., J. Obleser, et al. (2007). "Auditory evoked fields differentially encode speech features: an MEG investigation of the P50m and N100m time courses during syllable processing." Eur J Neurosci **25**(10): 3155-62.

Toro, J. M., J. B. Trobalon, et al. (2005). "Effects of backward speech and speaker variability in language discrimination by rats." J Exp Psychol Anim Behav Process **31**(1): 95-100.

VanRullen, R., R. Guyonneau, et al. (2005). "Spike times make sense." Trends in Neurosciences **28**(1): 1-4.

Versnel, H. and S. A. Shamma (1998). "Spectral-ripple representation of steady-state vowels in primary auditory cortex." J Acoust Soc Am **103**(5 Pt 1): 2502-14.

Walker, K. M., B. Ahmed, et al. (2007). "Linking Cortical Spike Pattern Codes to Auditory Perception." J Cogn Neurosci.

Wang, L., R. Narayan, et al. (2007). "Cortical discrimination of complex natural stimuli: can single neurons match behavior?" J Neurosci **27**(3): 582-9.

Weliky, M., J. Fiser, et al. (2003). "Coding of natural scenes in primary visual cortex." Neuron **37**(4): 703-18.

Wetzel, W., F. W. Ohl, et al. (1998). "Right auditory cortex lesion in Mongolian gerbils impairs discrimination of rising and falling frequency-modulated tones." Neuroscience Letters **252**(2): 115-118.

Wong, S. W. and C. E. Schreiner (2003). "Representation of CV-sounds in cat primary auditory cortex: intensity dependence." Speech Communication **41**(1): 93-106.

Young, E. D. (2008). "Neural representation of spectral and temporal information in speech." Philos Trans R Soc Lond B Biol Sci. **363**, 923-945.

CHAPTER 3

CORTICAL ACTIVITY PATTERNS PREDICT SPEECH SOUND CATEGORIZATION BY RATS

Crystal T. Engineer, Claudia A. Perez, Ryan S. Carraway, Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

ACKNOWLEDGEMENTS

The authors would like to thank Kevin Chang, Amanda Reed, Jarod Roland, and the Kilgard lab behavioral training team for their help with these experiments.

ABSTRACT

The brain processes sights and sounds into meaningful categories. Prefrontal neurons faithfully represent category membership. Although it is assumed that similarities in earlier cortical regions contribute to stimulus categorization, few studies have tested this hypothesis with real-world stimuli. In this study, we demonstrate that rats rapidly categorize speech sounds using voicing and gender cues. Our results provide the first evidence that spatiotemporal patterns in primary auditory cortex (A1) contain sufficient information to explain behavioral categorization of speech sounds.

INTRODUCTION

Both humans and animals categorize the sights and sounds they encounter on a daily basis. Vervet monkeys emit different warning calls for each class of predator in their environment, and troop mates exhibit distinct behaviors that indicate they understand the categories each call type represents (Seyfarth et al., 1980). After a few weeks of training, macaque monkeys can categorize images of animals, plants, and objects (Seyfarth et al., 1980; Vogels, 1999; Freedman et al., 2001). Several species have been shown to categorize animal vocalizations, music and human speech sounds (Kuhl and Miller, 1975; Kluender et al., 1987; Wyttenbach et al., 1996; Wright et al., 2000; Gifford et al., 2003; Reed et al., 2003). Rodents, for example, can accurately categorize speech sounds based on differences in voice onset time (/d/ vs. /t/, (Kuhl and Miller, 1975)) and rise time (/sh/ vs. /ch/, (Reed et al., 2003)).

Prefrontal and parietal cortex neurons develop categorical responses based on observed contingencies. For example, after training many prefrontal neurons respond more robustly to images of cats compared to images of dogs, and respond similarly to visually distinct images within the same category, such as a housecat and a cheetah (Freedman et al., 2001). Prefrontal and parietal neurons can also respond categorically to drifting dot patterns, vibrotactile stimuli and vocalizations (Freedman et al., 2001; Romo and Salinas, 2003; Gifford et al., 2005; Cohen et al., 2006).

These categorical boundaries develop in a space of features (i.e. orientation, color, motion, etc.) created by earlier sensory areas. Studies using simple visual and somatosensory stimuli indicate that the transformation of sensory information to guide behavioral responses is a hierarchical process involving many brain regions (Freedman et al., 2003; Romo and Salinas, 2003). To understand how early sensory processing of complex stimuli relates to categorization behavior, we tested the ability of rats to categorize speech stimuli that differed in both voicing and gender, and compared categorization behavior with predictions derived from responses in primary auditory cortex (A1).

Neurons in the primary auditory cortex of humans, monkeys and cats show distinct responses to speech sounds that differ in voice onset time (VOT). For example, syllables with a short VOT evoke a single peak of activity in response to the onset of the consonant, while syllables with a long VOT evoke an additional peak of activity to the onset of voicing (Steinschneider et al., 1999; Steinschneider et al., 2003; Wong and Schreiner, 2003; Steinschneider et al., 2005). While cortical responses to speech stimuli that differ in voicing have been well studied, it is not known what stimulus features are highlighted in A1 that provide useful information that higher areas can extract to form behaviorally relevant categorical boundaries.

Our earlier study demonstrates that the unique spatiotemporal pattern of A1 responses to each consonant accurately predicts consonant discrimination performance (Reed et al., 2003; Toro et al., 2003, 2005; Engineer et al., 2008). This study expands on our earlier consonant discrimination study that used single exemplars by testing the ability of rats to accurately categorize speech sounds when presented with multiple exemplars. We trained rats to categorize speech stimuli that differed in both voicing and gender, and recorded A1 responses to these sounds to determine if A1 responses can accurately predict categorization performance.

METHODS

Eleven rats were trained to categorize sounds by voicing or gender. Each rat performed the categorization task for two weeks. Behavioral performance on the voicing multiple speaker, gender multiple speaker, and voicing temporal compression categorization tasks was compared to neural activity in 441 primary auditory cortex (A1) sites in 11 experimentally naïve rats in response to the ‘dad’ and ‘tad’ stimuli. Behavioral training and A1 recording procedures are identical to previous studies and the 11 trained and 11 naïve rats used in this study are the same rats involved in consonant discrimination tasks in our previous study (Engineer et al., 2008).

Speech stimuli

We recorded the words ‘dad’ and ‘tad’ spoken by 3 male and 3 female native English speakers ($n = 12$ sounds, Figure C3.1). Additionally, ‘dad’ and ‘tad’ spoken by a single female speaker were temporally compressed in 10% increments, down to 10% of the original stimulus length, resulting in 20 temporally compressed ‘dad’ and ‘tad’ stimuli, using the STRAIGHT vocoder (Kawahara, 1997). As in our earlier study, the speech sounds were shifted up an octave in order to match the rat hearing range and presented at 60 dB (Kawahara, 1997; Engineer et al., 2008).

Behavioral training

To test for the ability of rats to discriminate differences in voicing, we trained six rats to categorize by voicing when the stimuli were temporally compressed. These six rats were previously trained for two weeks to discriminate ‘dad’ from ‘tad’ when spoken by a single female speaker. Immediately following the single speaker voicing task, rats were trained to categorize the 20 temporally compressed ‘dad’ and ‘tad’ stimuli. Following two weeks of training on the temporal compression voicing task, these six rats began training on the voicing multiple speaker task.

To test for the ability of rats to discriminate gender, we trained five rats to categorize gender when presented with ‘dad’ and ‘tad’ sounds spoken by multiple speakers. The rats were previously trained on other tasks for a month. For the first two weeks, the rats were trained to discriminate pitch using the word ‘dad’ spoken by a single female speaker. For the remaining two weeks, the rats were trained to categorize gender using the 6 male and female ‘dad’ stimuli. Immediately following the ‘dad’ gender task, rats began training on the gender multiple speaker task.

Training took place in double-walled booths that each contained a speaker (Optimus Bullet Horn Tweeter), house light, and cage (8” length x 8” width x 8” height) which included a lever, lever light, and pellet dish. The pellet dispenser was mounted outside of the booth to minimize the sound. Rats received a 45 mg sugar pellet reward for pressing the lever in response to the CS+ sounds, and received a time out where the house light was extinguished for a period of approximately 6 seconds for pressing the lever in response to the CS- sounds. Rats were food deprived to encourage motivation, and were kept between 80% and 90% body weight.

Rats were first trained to press the lever in order to receive a sugar pellet reward. Once they reached the criteria of independently pressing the lever 100 times per session for two sessions, they advanced to the detection phase of training. During this phase, rats learned to press the lever after hearing the ‘dad’ speech stimulus spoken by a female speaker. Rats started with a long 8 second hit window after each sound presentation, and the hit window was decreased in 0.5 second increments every few sessions as performance increased, down to a hit window of 3 seconds. When rats reached the criteria of a d’ ≥ 1.5 for 10 sessions, they advanced to the discrimination task. From this phase on, rats performed each task for 20 sessions over 2 weeks (2 training sessions per day). Six rats trained on a ‘dad’ vs. ‘tad’ discrimination task for two weeks, followed by a ‘dad’ vs. ‘tad’ temporal compression categorization task for two weeks, followed by a ‘dad’ vs. ‘tad’ multiple speaker categorization task for two weeks (Figure C3.2). Five rats trained on a ‘dad’ pitch discrimination task for two weeks, followed by a ‘dad’ gender categorization task for two weeks, followed by a ‘dad’ and ‘tad’ gender categorization task for two weeks (Figure C3.3). The final categorization task in each group used the exact same stimuli, ‘dad’ and ‘tad’ spoken by multiple male and female speakers. One group of rats was trained to categorize these stimuli by VOT, while the other group of rats was trained to categorize these stimuli by gender.

Anesthetized recordings

We recorded multi-unit responses ($n = 441$) to each of the ‘dad’ and ‘tad’ stimuli in the right primary auditory cortex of 11 anesthetized experimentally naïve female Sprague-Dawley rats. Rats were initially anesthetized with pentobarbital (50 mg kg^{-1}), and received dilute pentobarbital (8 mg ml^{-1}) as needed. Four Parylene-coated tungsten microelectrodes (1–2 MΩ, FHC) were used to record action potentials 600 μm below the cortical surface.

Thirty 60 dB speech stimuli were each presented 20 times, every 2 seconds at each recording site. Additionally, to determine the characteristic frequency of each site, 25 ms tones were presented at 81 frequencies (1 to 32 kHz) at 16 intensities (0 to 75 dB) at each site. Stimulus generation and data acquisition were performed with Tucker-Davis hardware (RP2.1 and RX5) and software (Brainware). The University of Texas at Dallas Institutional Animal Care and Use Committee approved all protocols and recording procedures.

Data analysis

To quantify neural performance, we used a nearest-neighbor classifier that made a PSTH template based on 19 of the 20 repeats for each sound at each recording site (Foffani and Moxon, 2004; Schnupp et al., 2006). The remaining repeat for each sound was compared to the PSTH templates based on the previous nineteen repeats, and the classifier chooses the PSTH template that has the smallest Euclidean distance from the repeat currently being considered. Responses included the entire 400 ms duration of the neural response, either with temporal precision (using 1 to 10 ms bins) or without temporal precision (a single 400 ms bin).

RESULTS

Rats accurately categorize voicing

Previous studies have shown that rats are able to easily discriminate differences in voicing (Engineer et al., 2008), and chinchillas can categorize voicing even when there are differences in vowel context and speaker (Kuhl and Miller, 1975). We predicted that rats would also be able to categorize by voicing on the first day of testing on a novel set of sounds. Six rats were trained to discriminate ‘dad’ and ‘tad’ spoken by a single female speaker. After two weeks of training on this task, performance was significantly better than discrimination performance on the first day of training (d’=2.5 ± 0.3 vs. d’=0.7 ± 0.3, p=0.0002, Figure C3.2a,b,g). These well trained rats were then required to categorize voicing differences when sounds were temporally compressed or spoken by multiple speakers.

The first categorization task involved temporally compressed versions of the same ‘dad’ and ‘tad’ sounds used for the initial discrimination training. The two sounds were temporally compressed in 10% increments, down to 10% of the original stimulus length to create a set of 9 novel ‘dad’ sounds and 9 novel ‘tad’ sounds. Rats were able to accurately categorize 16 of the 18 novel sounds on the first day of training (85 ± 2% vs. 35 ± 7%, d’=1.5 ± 0.2, Figure C3.2c,g). Only the most compressed stimuli (10% of the original length) required additional training to distinguish (Figure C3.2c,d). These results are consistent with observations in chinchillas and indicate that rats perceive the ‘dad’ and ‘tad’ sounds categorically even in the presence of a nearly ten-fold variation in stimulus duration.

After testing with temporally compressed versions of the original ‘dad’ and ‘tad’ sounds, rats were tested with a novel set of ‘dad’ and ‘tad’ sounds produced by three male and two female speakers (Figure C3.1). Rats accurately categorized these sounds on the first day of training (d’=1.4 ± 0.1, Figure C3.2e,g), and did not improve significantly with two weeks of additional testing (d’=1.6 ± 0.3, p=0.23, Figure C3.2f,g). These results indicate rats can also perceive ‘dad’ and ‘tad’ categorically in the presence of natural variation in VOT, formant transition, burst amplitude and burst spectrum.

Consistent with previous observations that males produce voiceless sounds with shorter voice onset times than females (Robb et al., 2005), the false alarm rate was significantly higher for male ‘tad’ stimuli compared to female ‘tad’ stimuli (56 ± 13% vs. 11 ± 5%, p=0.04). VOT duration was highly correlated with behavioral categorization (R^2=0.75, p=0.0002, Figure C3.7a). This result suggests that rats use VOT when it is helpful in classifying sounds produced by multiple speakers. The earlier observation that rats can ignore significant VOT variability when differences in burst acoustics reliably distinguish the temporally compressed stimuli suggests that rats, like humans, can make use of multiple acoustic features to categorize speech sounds (Summerfield and Haggard, 1977; Holt et al., 2001).

Rats accurately categorize gender

Five rats were trained to categorize speech stimuli based on the gender of the speaker. Rats were first trained to discriminate ‘dad’ with a high pitch from ‘dad’ with a low pitch. The target ‘dad’ stimulus used was identical to the ‘dad’ stimulus from the previous voicing discrimination, while the distracter ‘dad’ stimulus had a pitch one octave lower. On the first day of discrimination training, rats were unable to discriminate pitch, and pressed equally to both the high and low pitched ‘dad’ stimuli (78% vs. 78%, d’=0.01 ± 0.07, p=0.87, Figure C3.3a,g). After two weeks of training on this task, performance improved, and rats pressed the lever significantly more to the high pitched ‘dad’ than the low pitched ‘dad’ (78% vs. 48%, d’=1.05 ± 0.19, p=0.003, Figure C3.3b,g).

Following the pitch discrimination task, rats began the ‘dad’ gender categorization task where they were rewarded for pressing the lever in response to a female ‘dad’, and punished for pressing in response to a male ‘dad’. Rats were able to accurately categorize gender on the first day of training (d’=1.28 ± 0.34, Figure C3.3c,g), and performance improved over two weeks of training (d’=2.79 ± 0.17 on the last day of training, Figure C3.3d,g). On the first day of training, rats pressed the lever 82 ± 3% (mean ± se) to the female ‘dad’ stimuli, and pressed 38 ± 9% to the male ‘dad’ stimuli.

After two weeks of training on the ‘dad’ gender task, rats were switched to a gender categorization task using both the ‘dad’ and ‘tad’ multiple speaker stimuli, the exact same sounds used in the previous voicing categorization task. On the first day of training on this task, rats were able to successfully categorize by gender (d’ = 2.1 ± 0.19, Figure C3.3e,g), and after two weeks of training on the task, performance improved slightly (d’=2.6 ± 0.33, Figure C3.3f,g). On the first day of training, rats pressed 81 ± 8% to the female ‘dad’ and ‘tad’ stimuli, and pressed 13 ± 2% to the male ‘dad’ and ‘tad’ stimuli. Rats were able to successfully categorize each of the male and female stimuli except for the ‘tad’ stimulus produced by male #1, and they were unable to accurately categorize this sound after two weeks of training (Figure C3.3f).

The final categorization task for each group of rats used identical stimuli, ‘dad’ and ‘tad’ spoken by multiple male and female speakers. Rats were able to accurately categorize this set of sounds by both voicing (‘dad’ vs. ‘tad’) and gender (female vs. male).

A1 responses to speech sounds

We recorded the primary auditory cortex responses to each of the 15 ‘dad’ and 15 ‘tad’ stimuli at 441 multi-unit sites in 11 experimentally naïve rats. Neurograms for each of the ‘dad’ and ‘tad’ multiple speaker stimuli were constructed, with each of the 441 recording sites arranged by characteristic frequency (CF) from low to high frequency on the *y* axis (Figure C3.4).

To quantify the differences in the response to each sound and identify potential neural mechanisms for categorization, we calculated the peak firing rate to each of the sounds at each recording site. Female sounds evoked on average more than twice as many spikes as male sounds in high frequency neurons between 16 and 32 kHz (281 Hz vs. 118 Hz, p=0.002, Figure C3.5a). There was no difference between the firing rate evoked by female and male sounds in low frequency neurons between 1 and 2 kHz (p=0.49). In contrast to gender firing differences, ‘dad’ sounds evoked almost twice as many spikes as ‘tad’ sounds in low frequency neurons between 1 and 2 kHz (328 Hz vs. 182 Hz, p=0.03, Figure C3.5b), but there was no difference between the firing rate evoked by ‘dad’ and ‘tad’ sounds in high frequency neurons (p=0.54). For sounds that were temporally compressed, ‘dad’ sounds evoked a higher firing rate than ‘tad’ sounds in low frequency neurons (p=0.0001), while ‘tad’ sounds evoked a higher firing rate than ‘dad’ sounds in high frequency neurons (p=0.000006, Figure C3.5c).

We also calculated the total number of spikes at each recording site over the 400 ms duration of the response for each of the ‘dad’ and ‘tad’ stimuli. Using this method, female sounds evoked more spikes than male sounds in high frequency neurons between 16 and 32 kHz (4.8 spikes vs. 2.9 spikes, p=0.02, Figure C3.5d). As with the peak rate analysis, ‘dad’ sounds evoked more spikes than ‘tad’ sounds in low frequency neurons between 1 and 2 kHz (7.7 spikes vs. 6.6 spikes, p=0.06, Figure C3.5e). For the temporally compressed sounds, however, ‘tad’ sounds evoked more spikes than ‘dad’ sounds in high frequency neurons (5.3 spikes vs. 2.9 spikes, p=0.0000001), but there was no difference in the number of evoked spikes in low frequency neurons (p=0.32, Figure C3.5f).

The peak rate differences between sound categories are more distinct than the total number of spikes evoked over the entire duration of the response, suggesting that spike timing provides more information than mean rate, which is consistent with earlier reports (Engineer et al., 2008).

Neural responses predict categorization ability

To categorize A1 neural responses, the Euclidean distance was determined between each of the categorization sounds and the template sounds that rats were initially trained to discriminate. Responses included the 40 ms onset of the neural response, using 20 ms smoothing. When temporal precision was preserved, the Euclidean distance between the CS+ and CS- templates on the gender categorization task strongly predicts behavioral performance ($R^2 = 0.61$, $p = 0.004$, Figure C3.6a). The Euclidean distance for male sounds was low, and rats performing the gender categorization task rarely hit to these sounds. In contrast, the Euclidean distance for female sounds was high, and rats performing this task almost always pressed the lever in response to these sounds. The Euclidean distance between the CS+ and CS- templates strongly predicts behavioral performance on the voicing multiple speaker task ($R^2 = 0.75$, $p = 0.001$, Figure C3.6b), and on the temporal compression voicing task ($R^2 = 0.86$, $p = 0.00000004$, Figure C3.6c).

DISCUSSION

In this study, rats were able to successfully categorize speech sounds by both voicing and gender on the first day of training. This adds to previous findings that animals are able to categorize speech sounds (Kluender et al., 1987; Reed et al., 2003), including consonants which differ in VOT (Kuhl and Miller, 1975). Categorization performance was strongly predicted using the temporal pattern of A1 activity, but was not predicted by the mean firing rate.

This study helps to bridge the gap between the acoustic input and the category directed behavior. A1 responds to features of the input sound that are relevant and necessary for a higher association area to categorize the sound. Based on the spatiotemporal pattern of activation in A1, a higher area would be able to classify the sound and relay the category information to prefrontal cortex. This study indicates that the A1 representation of the acoustic input serves as a template of the features involved in the categorization of the sound, and we hypothesize that this area of the brain is approximately midway along the categorization pathway. A1 does not faithfully represent all possible acoustic features of the sound, as the auditory nerve might, but instead represents the features actually used by the animals to categorize the sounds. A1 also does not display category selective neurons, like the prefrontal cortex (Freedman et al., 2001), but instead offers a processed version of the stimulus that includes all the information necessary to accurately categorize the sounds.

APPENDIX C

FIGURES

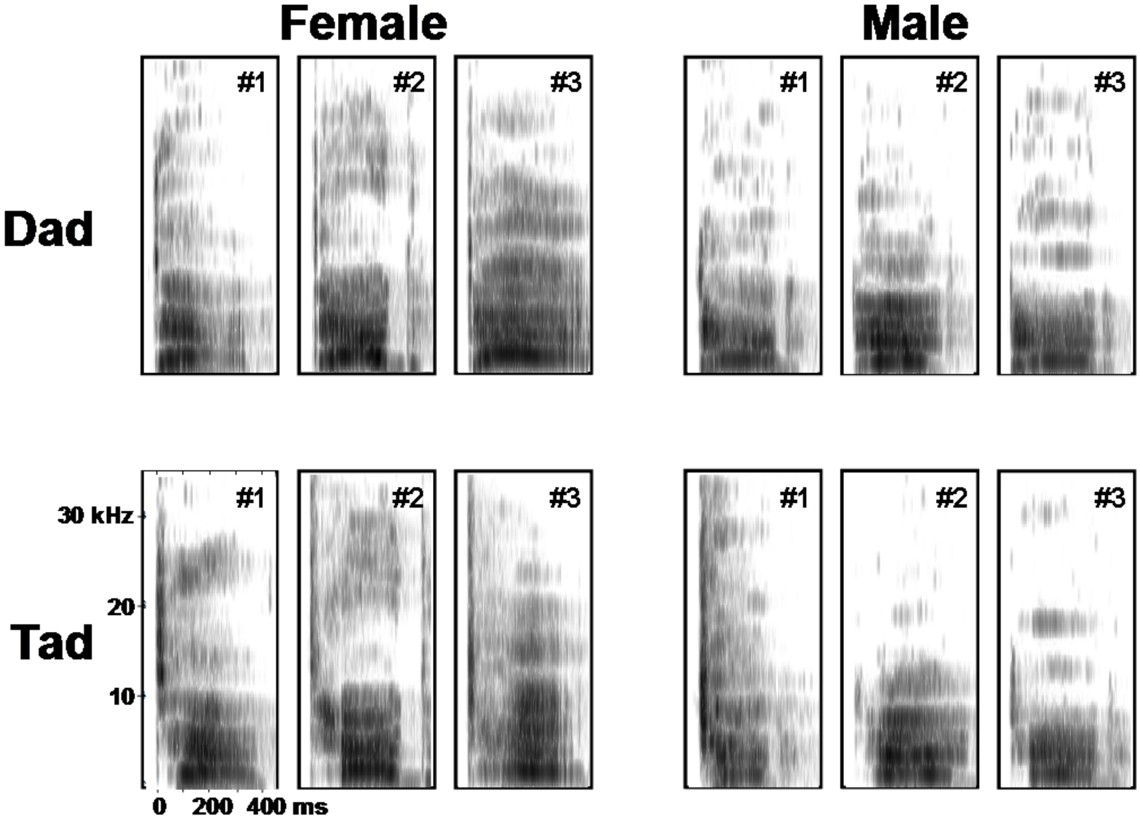


Figure C3.1. Spectrograms of each speech sound grouped by voicing (voiced ‘dad’ or voiceless ‘tad’) and speaker gender (female or male). Three female and three male speakers each produced the words ‘dad’ and ‘tad’ (n=6 speakers total). Frequency is represented on the *y* axis (0–35 kHz) and time on the *x* axis (–50 to 500 ms). Speech sounds were shifted one octave higher to accommodate the rat hearing range.

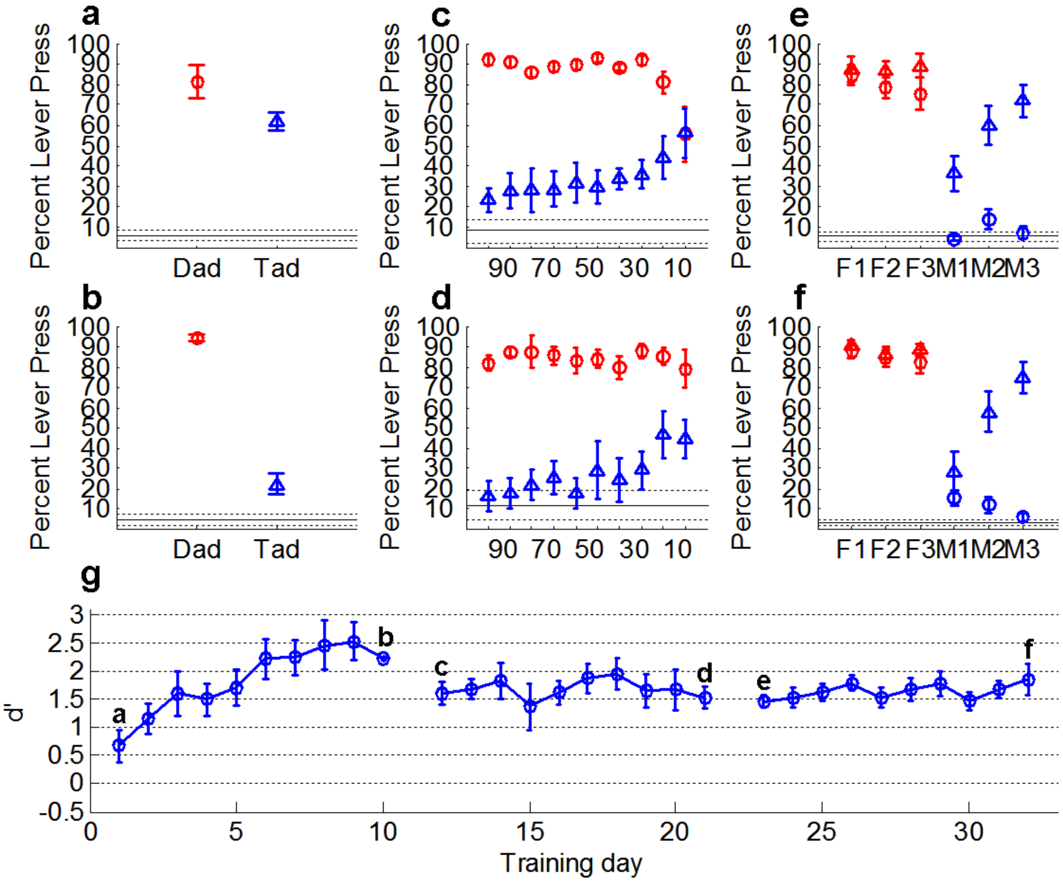


Figure C3.2. Behavioral performance time course for six rats trained on voicing categorization tasks. The top row shows the percent lever press on the first day of training for target sounds (red) and distractor sounds (blue) during (**a**) a single exemplar ‘dad’ versus ‘tad’ discrimination task, (**c**) a 20 stimuli temporal compression voicing categorization task, and (**e**) a 12 stimuli multiple speaker voicing categorization task. The middle row shows the percent lever press on the last day of training during (**b**) ‘dad’ versus ‘tad’ discrimination, (**d**) temporal compression categorization, and (**f**) multiple speaker categorization. The bottom row shows the time course of training on the three tasks over 30 days using the performance measure *d*’ (**g**).

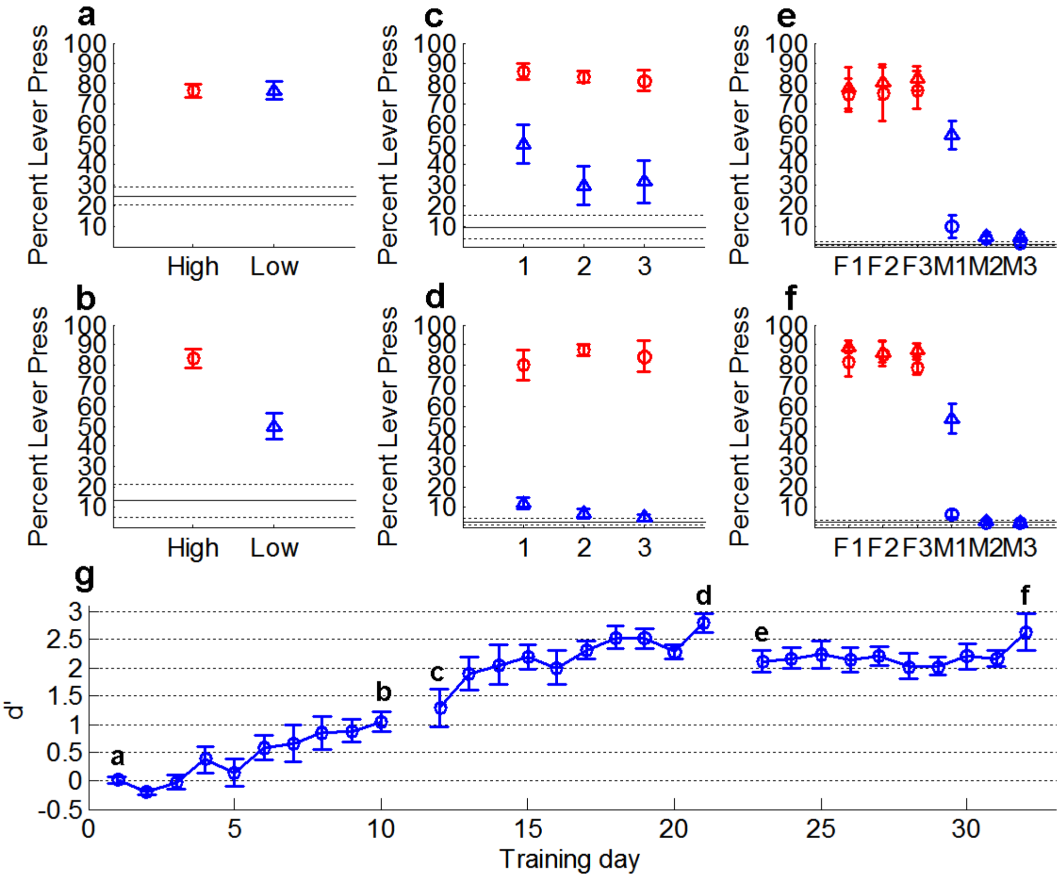


Figure C3.3. Behavioral performance time course for five rats trained on gender categorization tasks. The top row shows the percent lever press on the first day of training for target sounds (red) and distractor sounds (blue) during (**a**) a high versus low pitch discrimination task, (**c**) a 6 stimuli (‘dad’) gender categorization task, and (**e**) a 12 stimuli (‘dad’ and ‘tad’) gender categorization task. The middle row shows the percent lever press on the last day of training during (**b**) pitch discrimination, (**d**) ‘dad’ gender categorization, and (**f**) ‘dad’ and ‘tad’ gender categorization. The bottom row shows the time course of training on the three tasks over 30 days using the performance measure *d*’ (**g**).

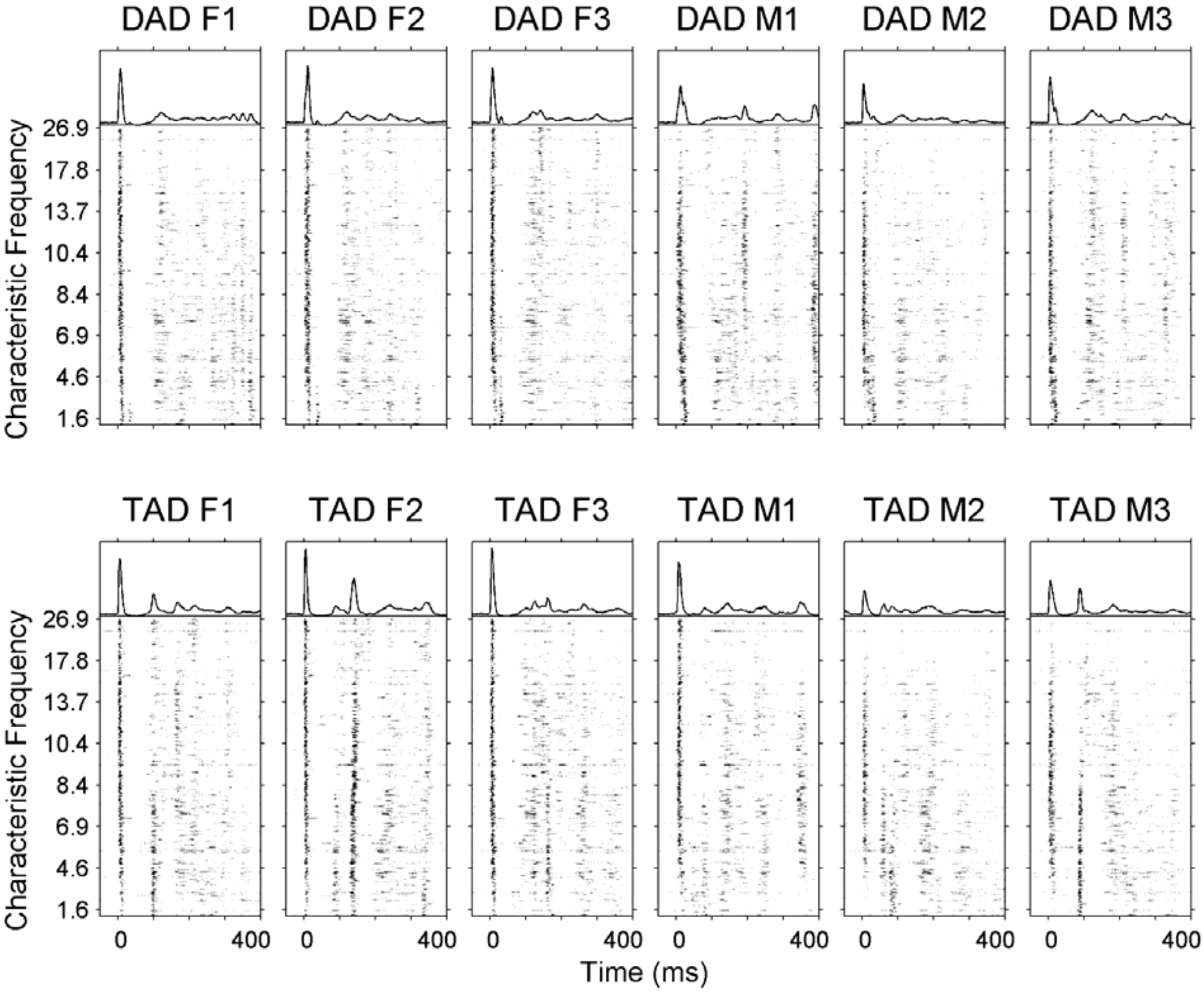
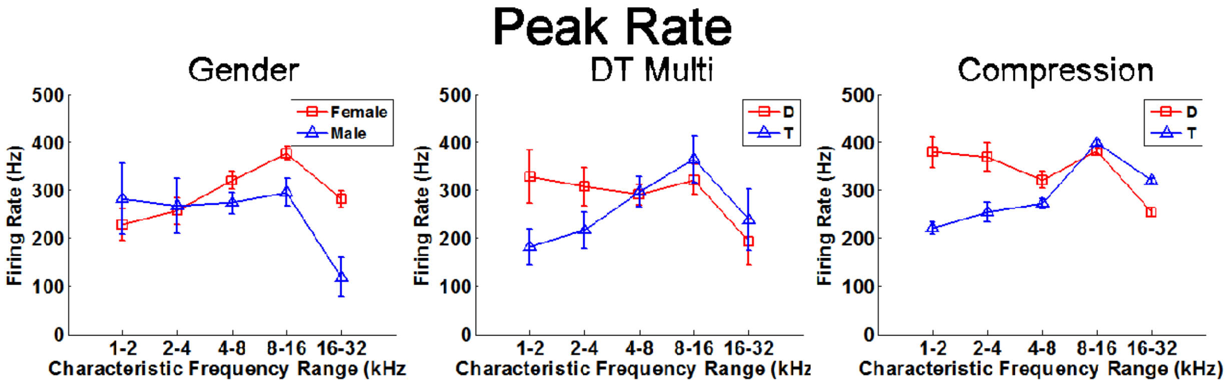


Figure C3.4. Neurograms depicting the response of rat A1 neurons to twelve speech sounds. Multi-unit data was collected from 441 recording sites in eleven anesthetized experimentally naïve adult rats. Average post-stimulus time histograms (PSTH) derived from twenty repeats were ordered by the characteristic frequency (kHz) of each recording site (*y* axis). Time is represented on the *x* axis (–50 to 400 ms). The firing rate of each site is represented in grayscale, where black indicates 400 Hz. For comparison, the mean population PSTH evoked by each sound is plotted above the corresponding neurogram. As in Figure C3.1, rows differ in voicing, while columns differ in speaker gender.



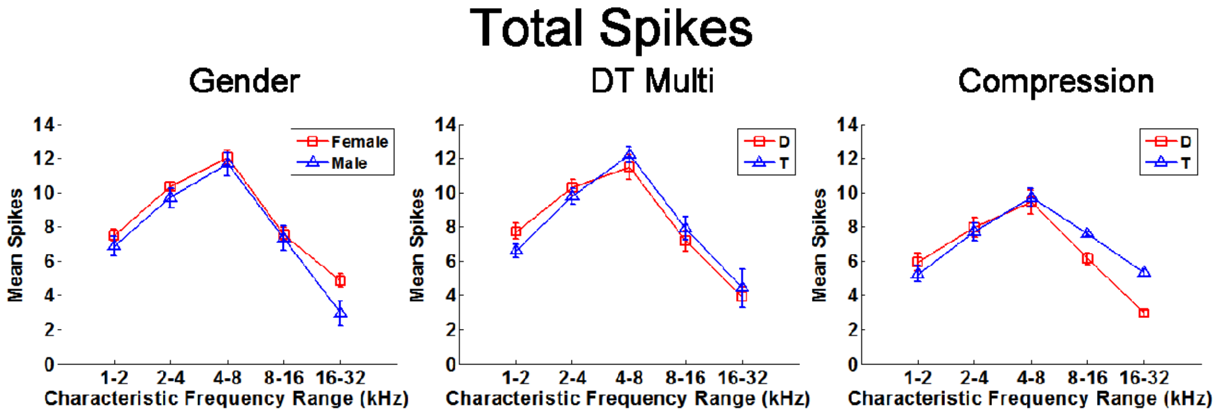


Figure C3.5. Neurogram quantification using peak firing rate and total number of spikes. Peak firing rate (Hz) is shown in five characteristic frequency bins from low (1-2 kHz) to high (16-32 kHz) frequency for target (red) and distractor (blue) sounds for each of the three categorization tasks: (**a**) gender, (**b**) multiple speaker voicing, and (**c**) temporal compression voicing. The total number of spikes is shown in five characteristic frequency bins from low (1-2 kHz) to high (16-32 kHz) frequency for target (red) and distractor (blue) sounds for each of the three categorization tasks: (**d**) gender, (**e**) multiple speaker voicing, and (**f**) temporal compression voicing.

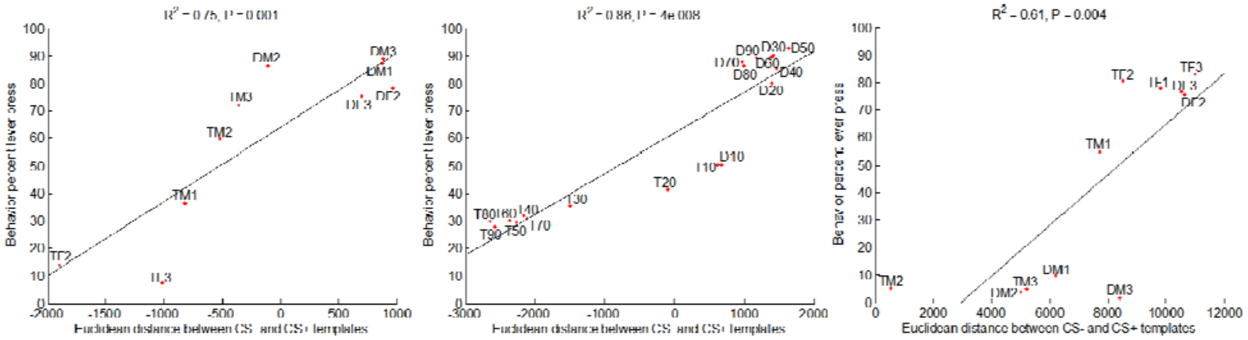


Figure C3.6. Trial-by-trial neural discrimination predicted categorization ability when temporal information was maintained. The average neural discrimination was correlated with behavioral categorization when 5 ms bins were used for the (**a**) gender, (**b**) multiple speaker voicing, and (**c**) temporal compression voicing categorization tasks. Neural discrimination was performed by a nearest-neighbor classifier (see methods) using a single sweep of neural activity. The stimulus names are printed next to each data point. Error bars indicate s.e.m. for behavioral and neural discrimination performance. Solid lines indicate the best linear fits.

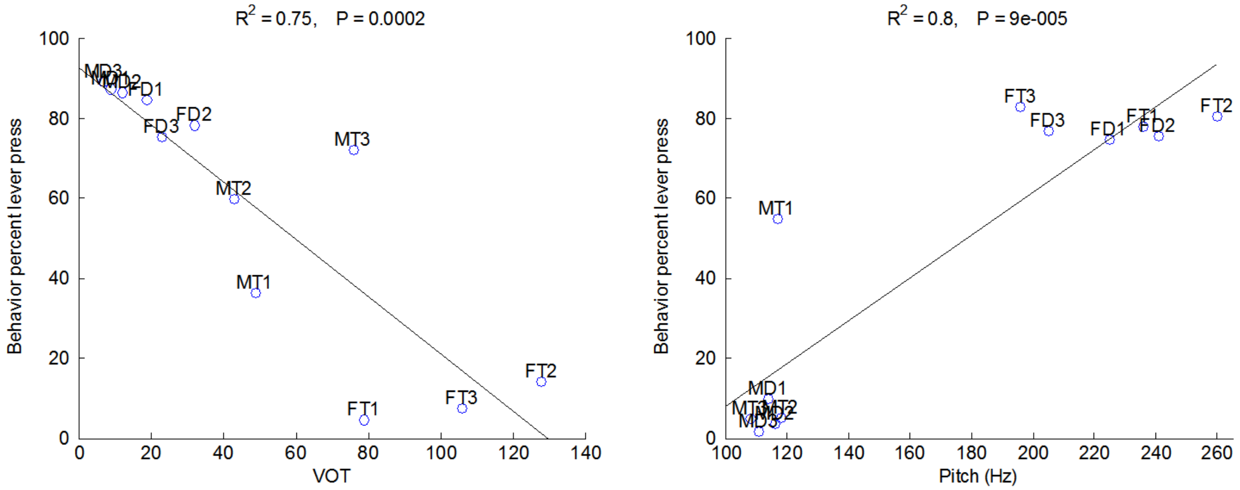


Figure C3.7. The voice onset time (VOT) and pitch acoustic features of the stimuli predict behavioral categorization performance. (**a**) VOT length (ms) was correlated with behavioral performance on the multiple speaker voicing categorization task. (**b**) Pitch (Hz) was correlated with behavioral performance on the gender categorization task.

REFERENCES

Cohen YE, Hauser MD, Russ BE (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. Biol Lett 2:261-265.

Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP (2008) Cortical activity patterns predict speech discrimination ability. Nat Neurosci 11:603-608.

Foffani G, Moxon KA (2004) PSTH-based classification of sensory stimuli using ensembles of single neurons. J Neurosci Methods 135:107-120.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312-316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23:5235-5246.

Gifford GW, 3rd, Hauser MD, Cohen YE (2003) Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. Brain Behav Evol 61:213-224.

Gifford GW, 3rd, MacLean KA, Hauser MD, Cohen YE (2005) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. J Cogn Neurosci 17:1471-1482.

Holt LL, Lotto AJ, Kluender KR (2001) Influence of fundamental frequency on stop-consonant voicing perception: a case of learned covariation or auditory enhancement? J Acoust Soc Am 109:764-774.

Kawahara H (1997) Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. Proceedings of the ICASSP 2:1303-1306.

Kluender KR, Diehl RL, Killeen PR (1987) Japanese quail can learn phonetic categories. Science 237:1195-1197.

Kuhl PK, Miller JD (1975) Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science 190:69-72.

Reed P, Howell P, Sackin S, Pizzimenti L, Rosen S (2003) Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. J Exp Anal Behav 80:205-215.

Robb M, Gilbert H, Lerman J (2005) Influence of gender and environmental setting on voice onset time. Folia Phoniatr Logop 57:125-133.

Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203-218.

Schnupp JW, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26:4785-4795.

Seyfarth RM, Cheney DL, Marler P (1980) Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210:801-803.

Steinschneider M, Fishman YI, Arezzo JC (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, Volkov IO, Noh MD, Garell PC, Howard MA, 3rd (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol 82:2346-2357.

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA, 3rd (2005) Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb Cortex 15:170-186.

Summerfield Q, Haggard M (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J Acoust Soc Am 62:435-448.

Toro JM, Trobalon JB, Sebastian-Galles N (2003) The use of prosodic cues in language discrimination tasks by rats. Anim Cogn 6:131-136.

Toro JM, Trobalon JB, Sebastian-Galles N (2005) Effects of backward speech and speaker variability in language discrimination by rats. J Exp Psychol Anim Behav Process 31:95-100.

Vogels R (1999) Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur J Neurosci 11:1223-1238.

Wong SW, Schreiner CE (2003) Representation of CV-sounds in cat primary auditory cortex: intensity dependence. Speech Communication 41:93-106.

Wright AA, Rivera JJ, Hulse SH, Shyan M, Neiworth JJ (2000) Music perception and octave generalization in rhesus monkeys. J Exp Psychol Gen 129:291-307.

Wyttenbach RA, May ML, Hoy RR (1996) Categorical perception of sound frequency by crickets. Science 273:1542-15

CHAPTER 4

SPEECH TRAINING INDUCES MAP REORGANIZATION BUT NOT SPEECH SPECIFIC PLASTICITY IN RAT PRIMARY AUDITORY CORTEX

Crystal T. Engineer, Claudia A. Perez, Ryan S. Carraway, Kevin Q. Chang, Jarod L. Roland, Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

ACKNOWLEDGEMENTS

The authors would like to thank Helen Chen, Dwayne Listhrop, Vikram Jakkamsetti, Roshini Jain, Amanda Reed, and the Kilgard lab behavior training team for their help with training and mapping for these experiments.

ABSTRACT

Discrimination training with speech sound contrasts results in improved behavioral performance and an increase in the cortical representation of the trained stimuli in humans. Due to the technical challenges and ethical issues involved in recording auditory neurons from human cortex, it is not known how primary auditory cortex (A1) neurons process speech sounds after behavioral training. Multi-unit recordings from animal studies allow us to study the fine spatial and temporal resolution that is lacking in brain imaging and evoked potential studies. In this study, groups of rats were trained to discriminate either a single speech sound contrast or multiple contrasts while another group was passively exposed to the speech sounds. After behavioral training, we recorded responses from A1 neurons to tones and speech sound stimuli.

Our results demonstrate that training on multiple speech contrasts increased response strength to tones, receptive field size, response latencies and the proportion of neurons responding to low frequencies compared to naïve rats. Auditory cortex responses were not significantly different from naïve rats after training on a single speech task. In addition, A1 plasticity was not specific to the behaviorally relevant speech sounds after speech sound training. This absence of stimulus specific plasticity for speech sounds in primary auditory cortex suggests that higher cortical areas may be responsible for encoding speech specific plasticity after behavioral training.

INTRODUCTION

Several weeks of discrimination training is sufficient to induce plasticity in visual, somatosensory, and auditory cortex (Karni and Sagi 1991; Recanzone et al. 1993; Xerri et al. 1996). Numerous studies have shown that speech sound discrimination training improves behavioral performance in humans. These improvements were accompanied by enhanced responses of auditory cortex potentials or activation of cortical areas during functional brain imaging (Callan et al. 2003; Kraus et al. 1995; Tremblay et al. 2001; Wang et al. 2003). While these studies give some indication of the spatiotemporal activation of the areas involved, they are still limited in their ability to resolve the fine spatial and temporal cues of speech sounds. Many studies have demonstrated how speech sounds are represented in the auditory cortex of rats, cats, monkeys and humans (Engineer et al. 2008; Steinschneider et al. 2003; Steinschneider et al. 1995; Steinschneider et al. 1999; Wong and Schreiner 2003). However, it is not known how auditory cortex neurons respond to tones and speech sounds after speech discrimination training.

We have previously demonstrated that rats can discriminate different speech features, and that A1 responses to these sounds are highly predictive of the rats behavioral performance on consonant discrimination tasks (Engineer et al. 2008). Here, we trained groups of rats to perform multiple speech discrimination tasks involving both consonant and vowel differences. Separate groups of rats were trained on a single speech discrimination task, or were passively exposed to speech sounds. In this study, we evaluated multiple forms of plasticity in A1 that have been documented in previous discrimination training studies, including changes in cortical maps, receptive field size, combination selectivity, response strength, and latency (Gentner and Margoliash 2003; Recanzone et al. 1993; Schnupp et al. 2006; Wong and Schreiner 2003).

METHODS

We trained 27 rats to discriminate and categorize speech sounds, after which primary auditory cortex (A1) responses to tones as well as speech sounds were recorded from 1199 A1 sites. Five groups of rats were trained; the first group was trained to discriminate consonants as well as vowels (n=4 rats), the second group was trained to discriminate consonants and gender differences (n=5 rats), the third group was trained to discriminate vowels (n=8 rats), the fourth group was trained to categorize sounds by voicing (n=5 rats), and the fifth group was trained to categorize sounds by gender (n=5 rats). A sixth group of rats was passively exposed to the same stimuli used in the fourth and fifth groups for the voicing and gender categorization tasks (n=5 rats, 244 A1 sites). A1 responses in these trained and passively exposed rats were compared to the responses recorded in experimentally naïve rats (n=11 rats, 514 A1 sites). Speech sounds, training, and anesthetized recording procedures are identical to the methods used in our previous study (Engineer et al. 2008).

Speech stimuli

Speech stimuli were produced in a CVC (consonant-vowel-consonant) context and were recorded in a soundproof booth. Twenty English consonants were produced in a '_æd' context (as in 'dad'). Additionally, the four vowels /æ/, /ɛ/, / ℘/, /i/, /u/ were produced in both a 'd_d' and 's_d' context (/ae/ 'dad', /ɛ/ 'dead', /Λ/ 'dud', /i/ 'deed', /u/ 'dood'). The words 'dad' and 'tad' were also recorded spoken by multiple speakers, 3 male and 3 female. As in our previous study, all words were shifted up an octave using the STRAIGHT vocoder (Engineer et al. 2008;

Kawahara 1997) in order to better match the rat hearing range and presented so that the intensity of the loudest 100 ms of the vowel was 60 dB.

Training timeline

Rats were trained using an operant go/no-go procedure in double-walled booths for two 1-hour sessions a day, 5 days a week. Each booth contained a speaker (Optimus Bullet Horn Tweeter), house light, and cage (8" length x 8" width x 8" height) which included a lever, lever light, and pellet dish. A pellet dispenser was mounted outside of the booth to minimize the sound of the pellet dispensing. Rats received a 45 mg sugar pellet reward for pressing the lever in response to the CS+ sounds within 3 seconds. If they pressed the lever in response to the CS- sounds, rats received a time out where the house light was extinguished for approximately 6 seconds.

To shape behavior, rats were initially trained to press the lever to receive a sugar pellet reward. After reaching the criteria of obtaining 100 pellets for 2 sessions, rats were moved to the detection stage. In this stage, rats were trained to press the lever in response to the CS+ sound ('dad') to receive a sugar pellet. This stage lasted until the rat was able to obtain a *d*' performance value of 1.5 or greater for 10 sessions. Following the detection stage, rats began the discrimination stage of training. Each discrimination task lasted for a fixed 2 week period.

Training groups

Of the five groups of trained rats, groups 1 through 3 were long-term experiments where rats trained on multiple discrimination tasks. Group 1 rats were trained on seven discrimination tasks: /d/ (target) vs. /s/ (distracter); /d/ vs. /t/; /d/ vs. /t/ when sounds were temporally compressed; /d/ vs. /t/ when sounds were spoken by multiple male and female speakers; /r/ vs. /l/; 'dad' vs. 'dead', 'dud', 'deed', and 'dood'; and /d/ vs. /b/ and /g/. Group 2 rats were trained on six discrimination tasks: pitch discrimination; gender discrimination of the word ‘dad’ spoken by multiple male and female speakers; gender discrimination of the words ‘dad’ and ‘tad’ spoken by multiple male and female speakers; /m/ vs. /n/, ‘shad’ vs. ‘fad’, ‘sad’, and ‘had’; and ‘shad’ vs. ‘chad’ and ‘jad’. Group 3 rats were trained on three vowel discrimination tasks: ‘dad’ vs. ‘dead’, ‘dud’, ‘deed’, and ‘dood’; ‘sad’ vs. ‘said’, ‘sud’, ‘seed’, and ‘sood’; and a combination of the two previous tasks, where the CS+ was the vowel /æ/ in either the /d/ or /s/ context, and the CS- sounds were the vowels /ɛ/, / ℘/, /i/, and /u/ in either the /d/ or /s/ context.

Groups 4 and 5 were short-term experiments where rats trained for 2 weeks on a single discrimination task. Group 4 rats were trained on a voicing discrimination task with a target of ‘dad’ and a distracter of ‘tad’. Both ‘dad’ and ‘tad’ were spoken by multiple male and female speakers. Group 5 rats were trained on a gender discrimination task using the exact same stimuli as Group 4, but with a target of ‘female’ and a distracter of ‘male’. A 6th group was passively exposed to the same stimuli used for groups 4 and 5, and was exposed to the sounds for 2 weeks.

Anesthetized recordings

Following the last day of training, rats were anesthetized and multi-unit and local field potential responses were recorded from the right primary auditory cortex. Rats were anesthetized with pentobarbital (50 mg kg^{-1}), and received dilute pentobarbital (8 mg ml^{-1}) as needed. Responses were recorded 600 µm below the cortical surface using four Parylene-coated tungsten microelectrodes (1–2 MΩ, FHC) simultaneously.

In order to determine the characteristic frequency at each recording site, 25 ms tones were presented at 81 frequencies (1 to 32 kHz) at 16 intensities (0 to 75 dB). A total of 80 speech stimuli were presented at each site for 20 repeats each. Each of the trained speech sounds was presented, and other novel speech sounds were presented at each site in order to determine if response changes were specific to trained stimuli. Stimulus generation and data acquisition were performed with Tucker-Davis hardware (RP2.1 and RX5) and software (Brainware). The University of Texas at Dallas Institutional Animal Care and Use Committee approved all protocols and recording procedures.

Data analysis

Behavioral performance was quantified using the measure d-prime (d'). This measure is based on signal detection theory, and uses the percent hit to CS+ and CS- sounds.

We quantified the receptive field properties for each group, including threshold, bandwidth, latency, spontaneous firing, and response strength to tones. These properties were quantified by a blind observer using customized MATLAB software. Threshold was defined as the lowest intensity that evoked a response at the characteristic frequency for each site. Bandwidth was measured in octaves as the frequency range that evokes a response 10, 20, 30, and 40 dB above the threshold. The onset, peak, and end of peak latencies were also determined for each site.

The percent of primary auditory cortex responding to each tone at each intensity was calculated for each group and compared to the percent of cortex responding in experimentally naïve rats, as in previous studies (Puckett et al. 2007). Briefly, we assume unsampled regions of cortex respond like the nearest sampled point.

Responses to speech stimuli were quantified using a nearest-neighbor classifier to determine neural discrimination performance (Engineer et al. 2008; Foffani and Moxon 2004; Schnupp et al. 2006). The classifier formed a PSTH template for each sound based on 19 of the 20 repeats from each recording site. The remaining repeat for each sound was then presented to the classifier, where it determined which sound the neural response was evoked by based on which PSTH template had the minimum Euclidean distance to the PSTH evoked by the remaining repeat.

RESULTS

Behavioral performance

Numerous previous studies have demonstrated the ability of animals to discriminate speech sounds (Engineer et al. 2008; Kluender et al. 1987; Kuhl and Miller 1975; Ramus et al. 2000; Reed et al. 2003). In this study, rats were trained to perform either a single speech discrimination task (groups 4 and 5, see methods), or multiple speech discrimination tasks (groups 1 through 3). After two weeks of training on each task, rats were able to accurately discriminate most speech tasks (Figure D4.1). In addition to the eleven consonant tasks that have been previously published (Engineer et al. 2008), rats were also trained on vowel, voicing, and gender categorization tasks. Rats were unable to successfully discriminate two of the consonant tasks: /m/ vs. /n/ (d'= -0.03, p=0.72) and /r/ vs. /l/ (d'=0.07, p=0.62, Figure D4.1) after two weeks of training. Rats were easily able to discriminate the consonant pairs /d/ vs. /s/ (d'=2.92, p=0.00009) and /d/ vs. /b/ (d'=2.99, p=0.0005, Figure D4.1). Rats were highly accurate at speech discrimination, and had better than chance performance on 22 out of the 24 discrimination tasks.

Receptive field changes

Previous studies have shown sharper tuning, longer latencies, and an increase in the cortical area of representation of the trained stimuli in A1 following long term frequency discrimination training (Recanzone et al. 1993). Multiple task trained rats exhibited a 21% increase in the threshold of A1 neurons following the completion of speech training (p=0.0000000007, Figure D4.2a). There was no significant change in threshold seen in rats who were trained on a single task compared to naïve rats (2% increase, p=0.49, Figure D4.2a). In rats passively exposed to speech sounds, there was a 12% decrease in the threshold of A1 neurons (p=0.006). The bandwidth of frequencies responded to only increased significantly in the multiple task group of rats (8% increase, p=0.004, Figure D4.2b). Spontaneous firing increased in both the multiple task and single task trained rats (45% increase, p=0.00000000002; 17% increase, p=0.003, respectively), but not in the passively exposed rats (8% increase, p=0.27, Figure D4.2c). After multiple task speech training, the response strength to tones increased by 9% (p=0.009, Figure D4.2d), while the response strength to tones did not change after training on a single task or passive exposure. Training on multiple tasks resulted in a larger threshold, wider bandwidth, and greater response strength to tones compared to naïve rats.

Latency changes

There was a 7% increase in onset latency for long-term trained rats (p=0.000003), while short-term trained rats did not exhibit a significant change in onset latency (p=0.42, Figure D4.3a). Rats who were passively exposed to speech showed a significant decrease in onset, peak, and end of peak latencies (6%, p=0.002; 3%, p=0.04; 4%, p=0.0009). Peak latency significantly increased for both long- and short-term trained rats (7%, p=0.000001; 3%, p=0.006; Figure D4.3b), while end of peak latency only significantly increased for long-term trained rats (7%, p=0.00008, Figure D4.3c). A1 neurons in rats who trained on multiple tasks were slower after training, while neurons were faster in rats passively exposed to speech sounds. This result is consistent with previous studies which show longer response latencies after long-term training on a frequency discrimination task (Recanzone et al. 1993).

Cortical map changes

Long-term trained rats had a low frequency map expansion for high intensity sounds, and a mid frequency map contraction for low intensity sounds (Figure D4.4). Short-term trained rats had a slight low frequency map expansion for high intensity sounds, a slight mid frequency map contraction for low intensity sounds, and a slight high frequency map expansion for low intensity sounds (Figure D4.5). Passive exposure rats had a slight low frequency map expansion for low intensity sounds, and a high frequency map expansion for sounds of all intensities (Figure D4.6).

Local field potential changes

Humans trained to discriminate speech sounds have exhibited an increase in the N1-P2 peak-to-peak amplitude following training (Tremblay et al. 2001). In naïve rats, the word 'dad' evoked negativities 20 ms and 125 ms after sound onset (N1 and N2) and evoked a positivity 53 ms after sound onset (P1, Figure D4.7). Rats who were trained on multiple speech discrimination tasks showed an increase in the onset latency of the N1 response evoked in response to the word 'dad' (Figure D4.7, top row). Additionally, N1 amplitude was decreased after training on multiple tasks compared to naïve control rats. The same increase in N1 onset latency and decrease in N1 amplitude was seen in response to the word 'tad' in rats trained to discriminate multiple tasks (Figure D4.8, top row). However, in response to the untrained sound 'wad', all three multiple task trained groups of rats exhibited an increase in the N1 onset latency (Figure D4.9, top row). Local field potential latencies and amplitudes were altered for both trained and untrained stimuli in rats trained to discriminate multiple speech tasks.

DISCUSSION

Following speech training on multiple tasks, there is an increase in the proportion of neurons responding to low frequency sounds. Additionally, response strength, threshold, bandwidth, and latencies are increased after multiple speech discrimination task training. Passive exposure to speech sounds causes an increase in the proportion of neurons responding to high frequency sounds, and leads to a decrease in the threshold and latencies of A1 neurons.

Although tone training changes A1 in a trained frequency specific manner in rats and monkeys (Polley et al. 2006; Recanzone et al. 1993), speech training does cause stimulus specific enhancement in A1. Our findings are consistent with Schnupp et al., which does not show evidence of specificity, as well as with previous studies in birds which do not show stimulus specific changes in primary fields (Gentner and Margoliash 2003; Tremblay et al. 2001). This suggests that the A1 representation of speech sounds is sufficiently precise so that there is no need to change it. While changes do occur in A1 after speech training, there is no stimulus-specific enhancement, which is consistent with previous studies showing generalization to untrained stimuli following speech training (Tremblay et al. 1997). Given the spectrotemporal complexity of speech sounds, it is likely that higher cortical areas extract higher order features of speech sounds and may show stimulus specificity following extensive speech training.

APPENDIX D

FIGURES

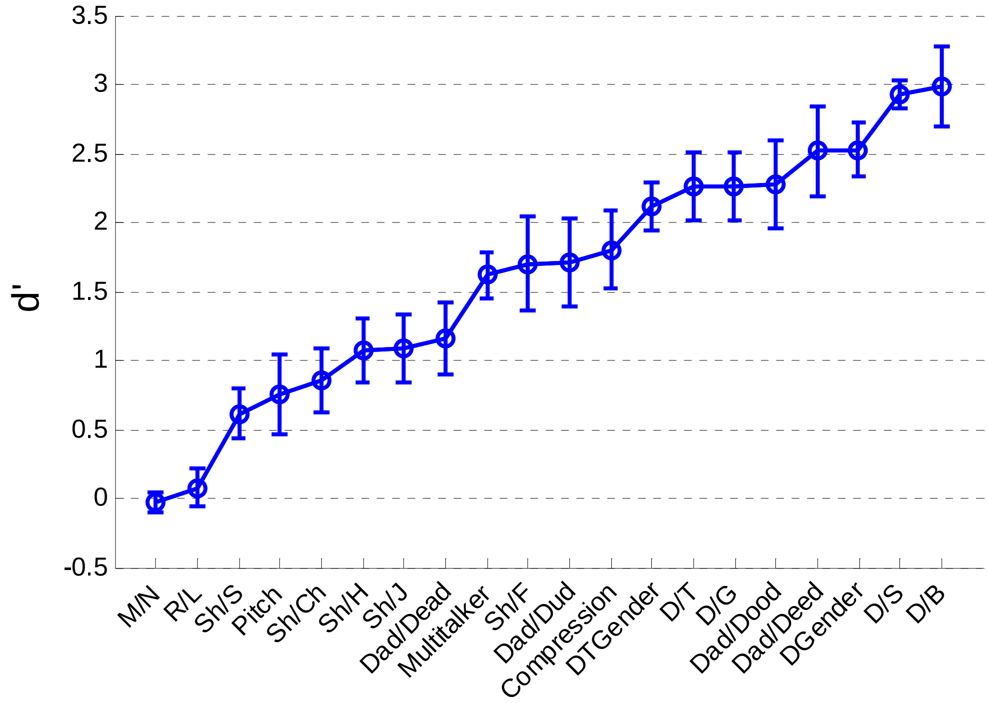


Figure D4.1. Behavioral performance on days 7/8 of 24 speech discrimination tasks.

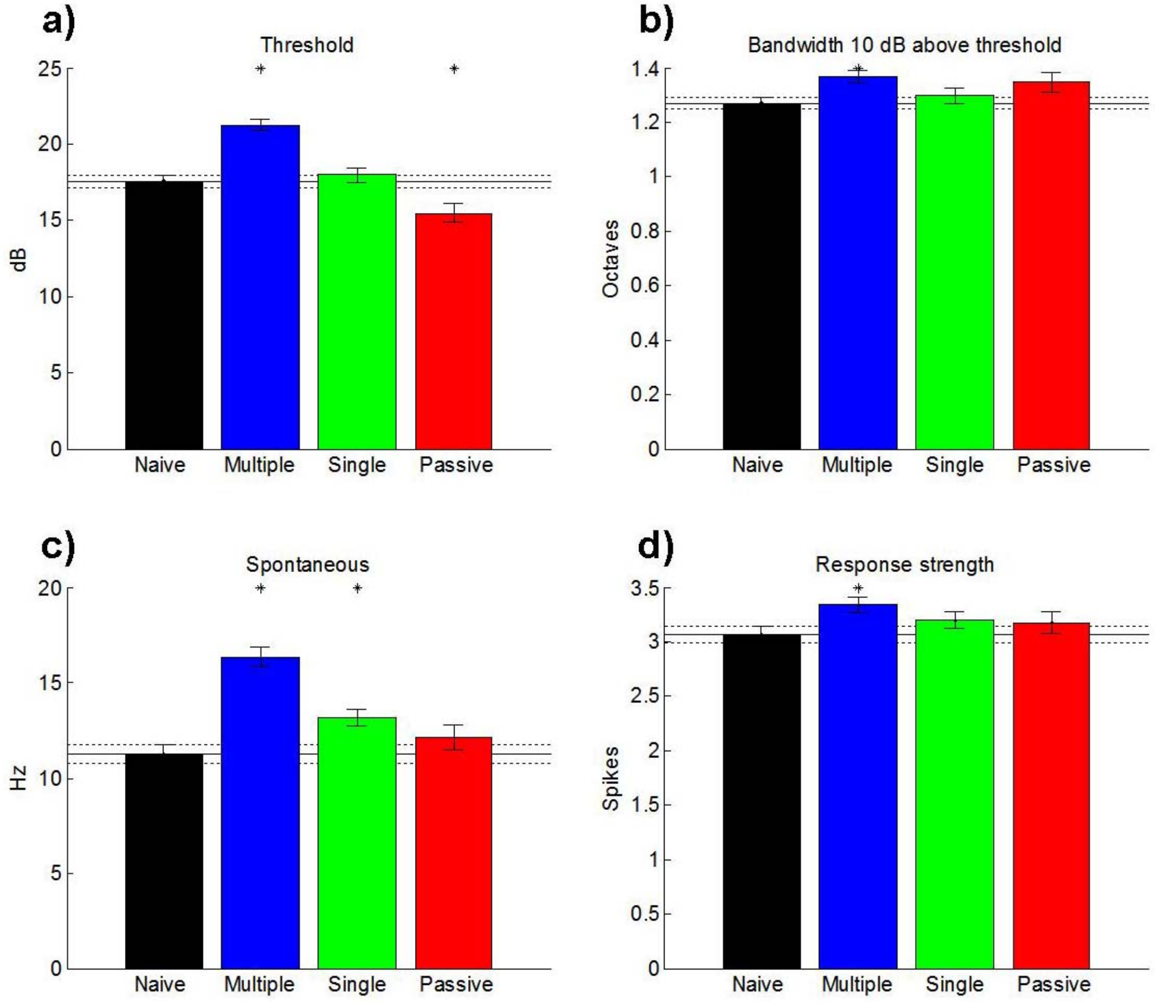


Figure D4.2. Changes in a) threshold, b) bandwidth 10 dB above threshold, c) spontaneous firing rate, and d) response strength in A1 after speech training and passive exposure.

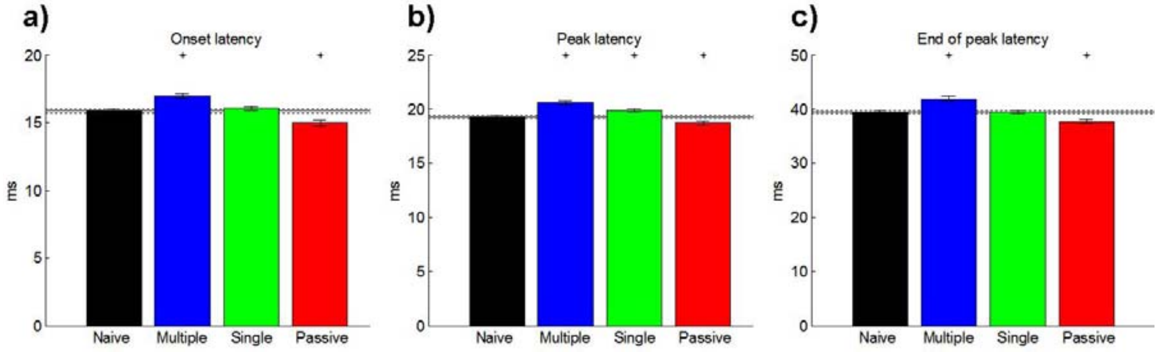


Figure D4.3. Changes in a) onset latency, b) peak latency, and c) end of peak latency in A1 after speech training and passive exposure.

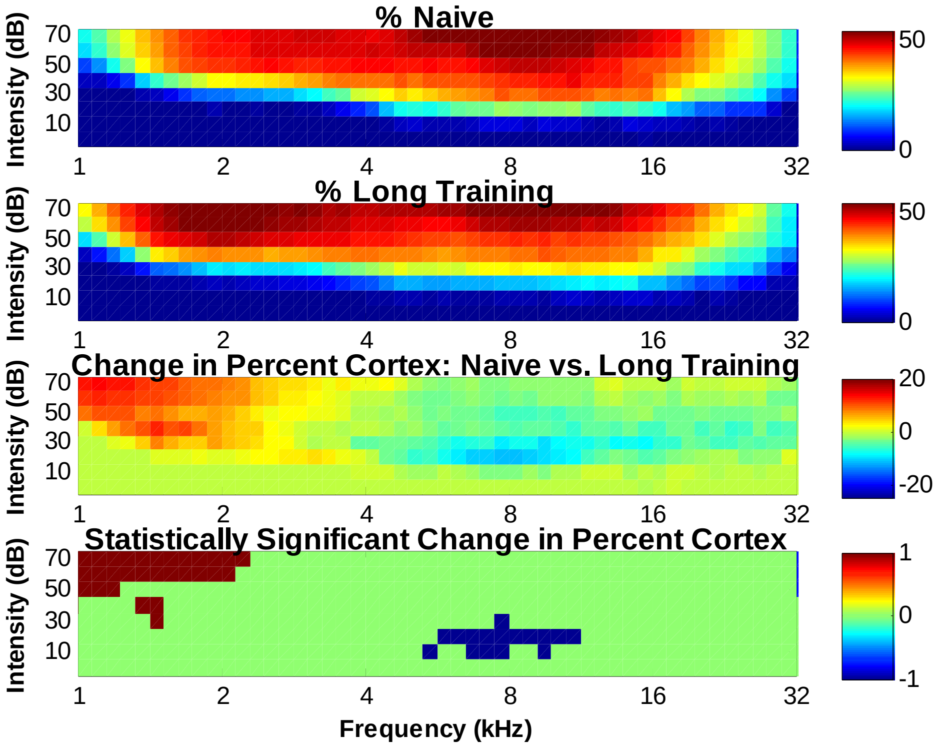


Figure D4.4. Percent cortex changes after training on multiple speech discrimination tasks.

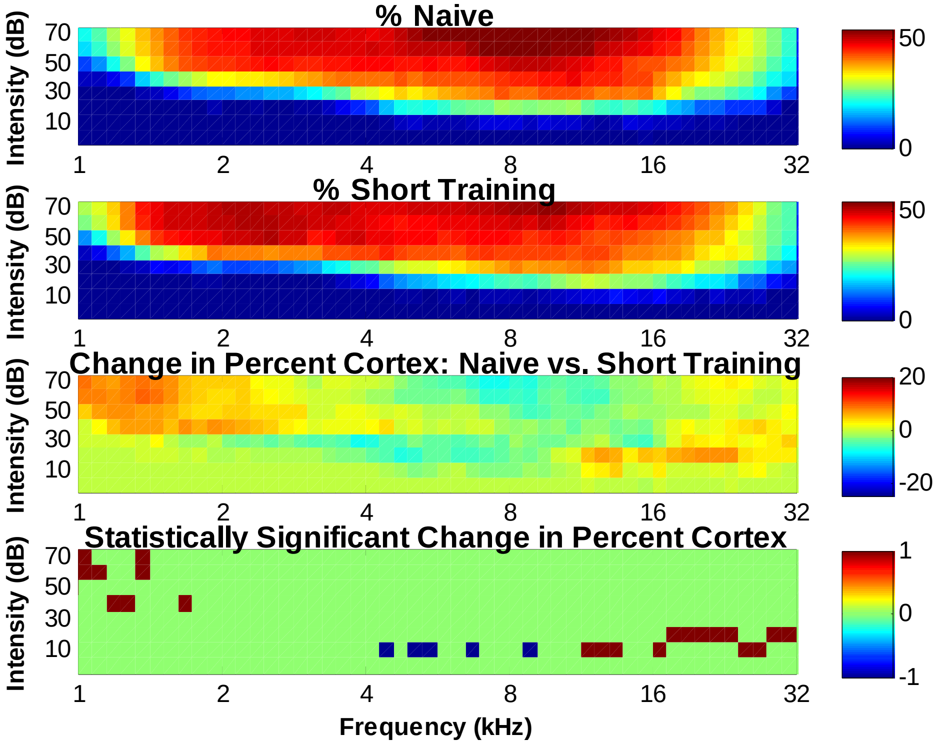


Figure D4.5. Percent cortex changes after training on a single speech discrimination task.

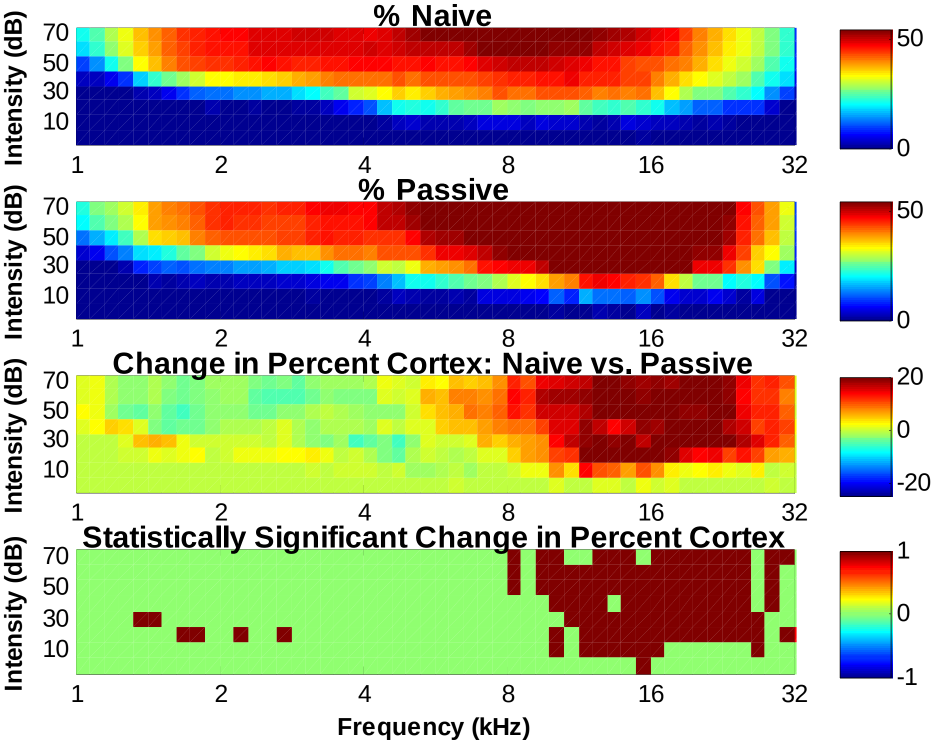


Figure D4.6. Percent cortex changes after passive exposure to speech sounds.

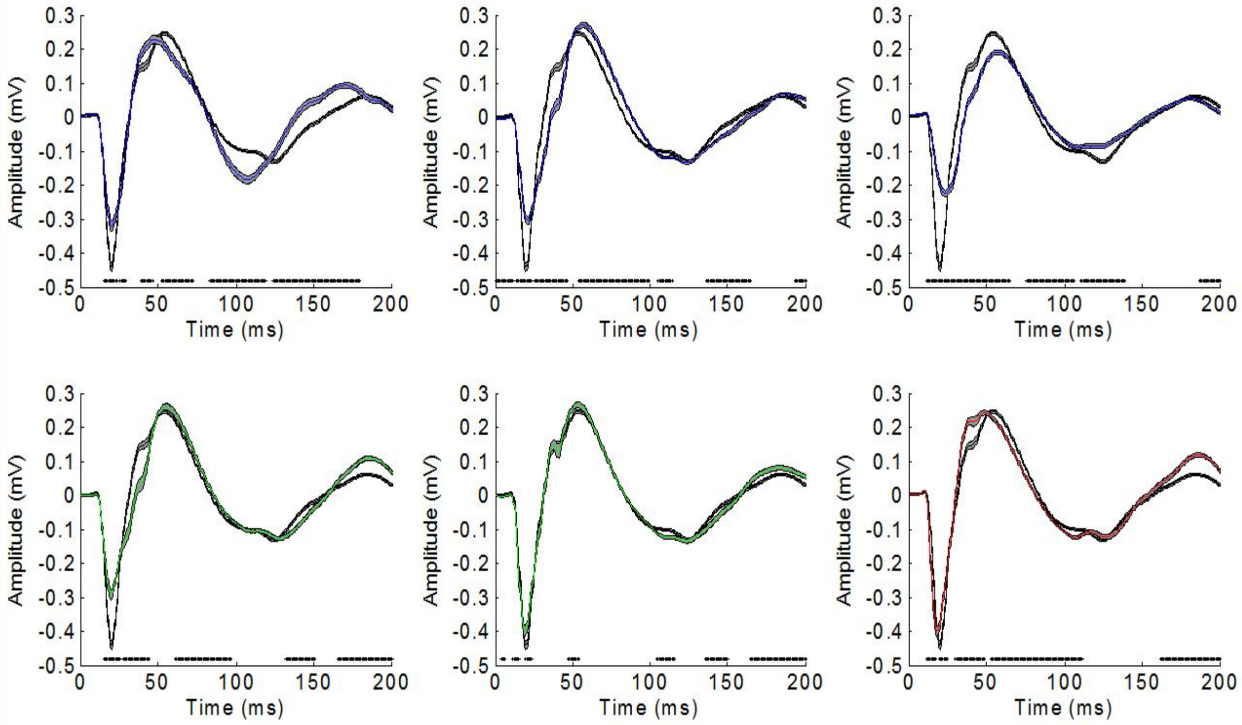


Figure D4.7. Local field potential response to the onset of the word 'dad'. a) multiple task trained group 1 (blue) and naïve (black) rats; b) multiple task trained group 2 (blue) and naïve (black) rats; c) multiple task trained group 3 (blue) and naïve (black) rats; d) single task trained group 4 (green) and naïve (black) rats; e) single task trained group 5 (green) and naïve (black) rats; f) passive exposure group 6 (red) and naïve (black) rats. Standard error is shown behind the colored lines in gray. The black line at the bottom of the figure indicates time points where the two groups of rats are significantly different.

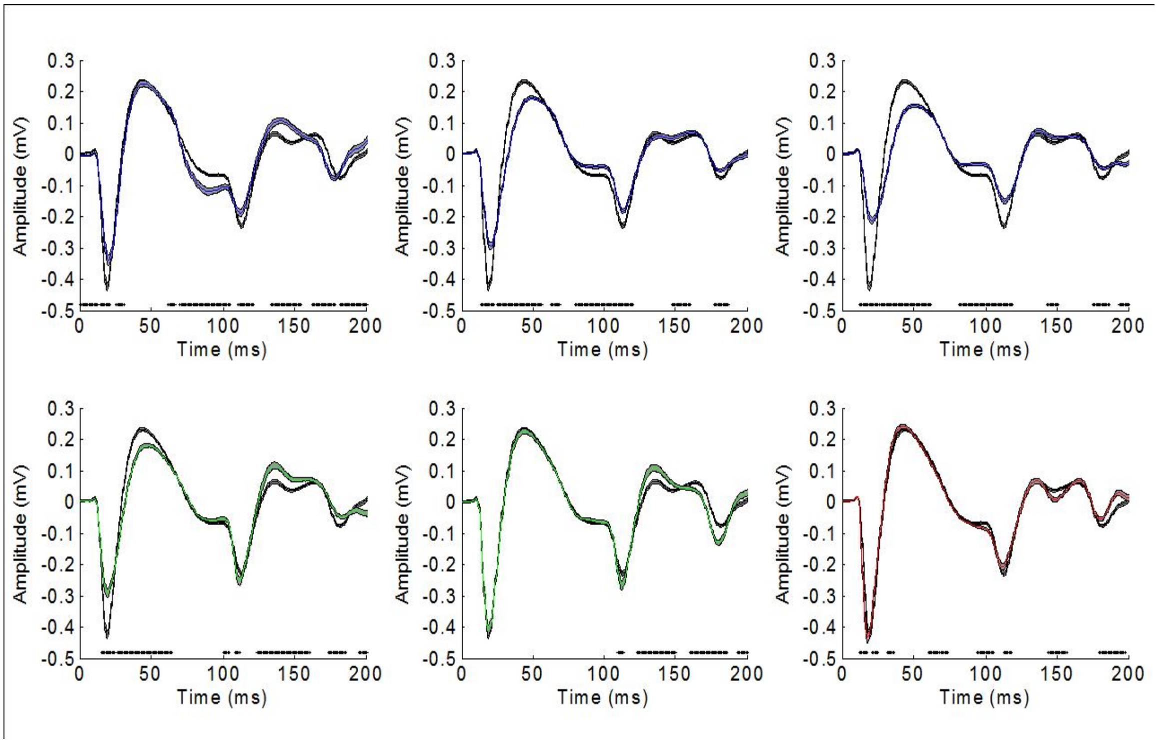


Figure D4.8. Local field potential response to the onset of the word ‘tad’. a) multiple task trained group 1 (blue) and naïve (black) rats; b) multiple task trained group 2 (blue) and naïve (black) rats; c) multiple task trained group 3 (blue) and naïve (black) rats; d) single task trained group 4 (green) and naïve (black) rats; e) single task trained group 5 (green) and naïve (black) rats; f) passive exposure group 6 (red) and naïve (black) rats. Standard error is shown behind the colored lines in gray. The black line at the bottom of the figure indicates time points where the two groups of rats are significantly different.

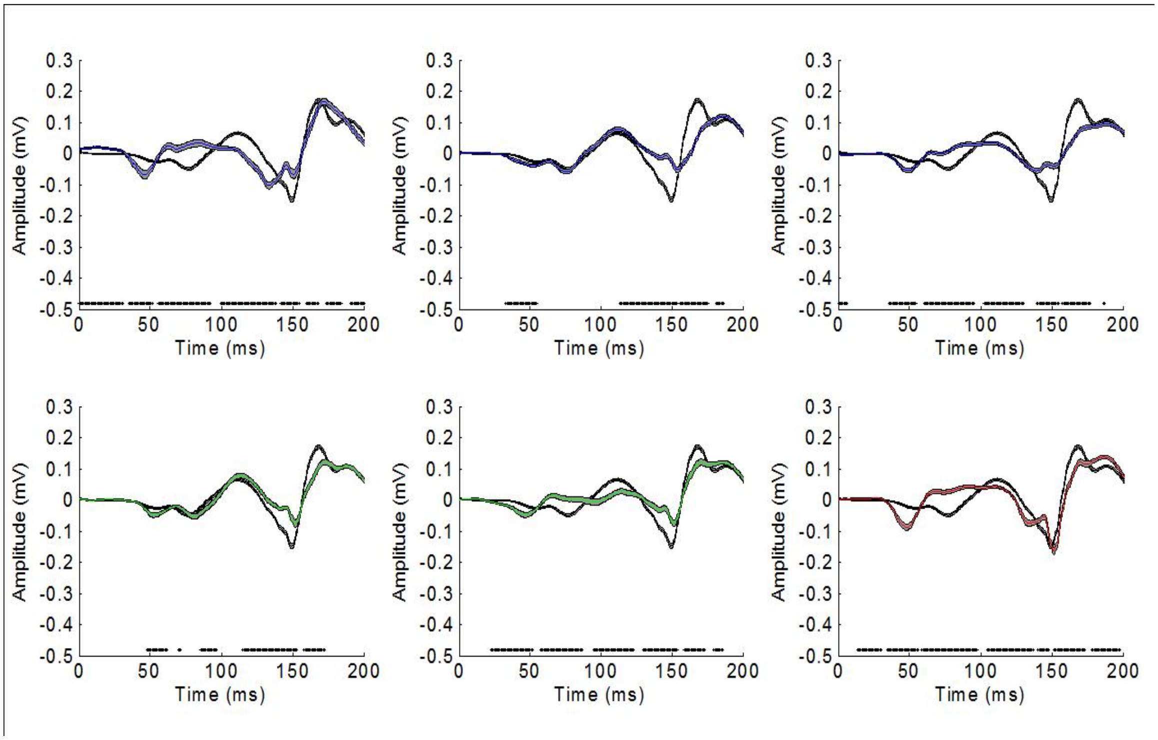


Figure D4.9. Local field potential response to the onset of the word 'wad'. a) multiple task trained group 1 (blue) and naïve (black) rats; b) multiple task trained group 2 (blue) and naïve (black) rats; c) multiple task trained group 3 (blue) and naïve (black) rats; d) single task trained group 4 (green) and naïve (black) rats; e) single task trained group 5 (green) and naïve (black) rats; f) passive exposure group 6 (red) and naïve (black) rats. Standard error is shown behind the colored lines in gray. The black line at the bottom of the figure indicates time points where the two groups of rats are significantly different.

REFERENCES

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, and Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19: 113-124, 2003.

Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, and Kilgard MP. Cortical activity patterns predict speech discrimination ability. *Nature neuroscience* 11: 603-608, 2008.

Foffani G, and Moxon KA. PSTH-based classification of sensory stimuli using ensembles of single neurons. *Journal of neuroscience methods* 135: 107-120, 2004.

Gentner TQ, and Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669-674, 2003.

Karni A, and Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America* 88: 4966-4970, 1991.

Kawahara H. Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. *Proceedings of the ICASSP* 2: 1303-1306, 1997.

Kluender KR, Diehl RL, and Killeen PR. Japanese quail can learn phonetic categories. *Science (New York, NY* 237: 1195-1197, 1987.

Kraus N, McGee T, Carrell T, King C, Tremblay K, and Nicol T. Central Auditory System Plasticity Associated with Speech Discrimination Training. *Journal of cognitive neuroscience* 7: 25-32, 1995.

Kuhl PK, and Miller JD. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science (New York, NY* 190: 69-72, 1975.

Polley DB, Steinberg EE, and Merzenich MM. Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26: 4970-4982, 2006.

Puckett AC, Pandya PK, Moucha R, Dai W, and Kilgard MP. Plasticity in the rat posterior auditory field following nucleus basalis stimulation. *Journal of neurophysiology* 98: 253-265, 2007.

Ramus F, Hauser MD, Miller C, Morris D, and Mehler J. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science (New York, NY* 288: 349-351, 2000.

Recanzone GH, Schreiner CE, and Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci* 13: 87-103, 1993.

Reed P, Howell P, Sackin S, Pizzimenti L, and Rosen S. Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. *Journal of the experimental analysis of behavior* 80: 205-215, 2003.

Schnupp JW, Hall TM, Kokelaar RF, and Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26: 4785-4795, 2006.

Steinschneider M, Fishman YI, and Arezzo JC. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *The Journal of the Acoustical Society of America* 114: 307-321, 2003.

Steinschneider M, Reser D, Schroeder CE, and Arezzo JC. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain research* 674: 147-152, 1995.

Steinschneider M, Volkov IO, Noh MD, Garell PC, and Howard MA, 3rd. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of neurophysiology* 82: 2346-2357, 1999.

Tremblay K, Kraus N, Carrell TD, and McGee T. Central auditory system plasticity: generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America* 102: 3762-3773, 1997.

Tremblay K, Kraus N, McGee T, Ponton C, and Otis B. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing* 22: 79-90, 2001.

Wang Y, Sereno JA, Jongman A, and Hirsch J. fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of cognitive neuroscience* 15: 1019-1027, 2003.

Wong SW, and Schreiner CE. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Communication* 41: 93-106, 2003.

Xerri C, Coq JO, Merzenich MM, and Jenkins WM. Experience-induced plasticity of cutaneous maps in the primary somatosensory cortex of adult monkeys and rats. *Journal of physiology, Paris* 90: 277-287, 1996.

CHAPTER 5

CONCLUSION

These experiments have shown that behavioral performance on speech discrimination tasks can be predicted using the pattern of A1 responses. Rats are able to discriminate many consonant pairs, and this performance is predicted when the spike timing information of A1 neurons is preserved, but is not predicted by the mean firing rate, when timing information is eliminated. Rats are also able to perform speech categorization tasks, and are able to categorize sounds spoken by multiple speakers by voicing or gender cues on the first day of training. Following speech discrimination training, A1 responses have altered receptive fields and response latencies, and there is an altered representation of the percent of cortex responding to tones.

Rate vs. temporal coding

Chapter 2 showed that temporal coding strongly predicts behavioral performance while rate coding does not. In addition to the temporal firing pattern, this chapter also determined that the spatial pattern of activation was necessary to accurately discriminate consonants. For example, the consonant /b/ activates low frequency neurons first, and later activates high frequency neurons, while the opposite is true for the consonant /d/. This study highlights the limitations of fMRI and EEG methods, which do not have the high temporal or spatial precision needed in order to see these patterns of activation. This study also shows the importance of studying complex sounds, since many real-world sounds, like speech, are spatially and temporally complex. Previous studies using simple stimuli that varied on a single dimension revealed that rate coding was sufficient to discriminate these sounds, while the current study shows that temporal precision is necessary to discriminate real-world stimuli. Our lab has also studied vowels, which are steady-state sounds without the rapid transitions that consonants have, and found that rate coding is sufficient to predict vowel discrimination (unpublished results; Perez et al., 2008).

The exemplar theory of categorization

The categorization study in Chapter 3 demonstrated that behavioral performance on the 3 voicing or gender categorization tasks was predicted based on the neural distance between each of the novel sounds and the template sounds that the rats had been previously trained to discriminate. Our hypothesis from this result is that the neural distance metric mimics a higher cortical area that is receiving input from A1 and using this input to categorize the stimuli. Since behavioral performance was predicted using the distance from the template sounds, we believe that the higher cortical area has stored templates for the categorization task, rather than retaining a representation of each of the individual stimuli.

Primary auditory cortex necessity

It is possible that although A1 has unique representations of each speech sound that are predictive of behavioral ability, this area may not be necessary for speech sound discrimination. The plasticity study in Chapter 4 did not show speech-specific changes in A1 or changes that were specific to the trained stimuli. This finding is consistent with previous studies that do not exhibit feature specific changes in primary fields (Schnupp et al., 2006; Gentner and Margoliash, 2003; Tremblay et al., 2001). Preliminary unpublished studies from our lab indicate that rats are still able to perform speech discrimination and categorization tasks with bilateral primary auditory cortex lesions (Porter et al., 2007; Porter et al., 2008). Future lesion and behavioral studies are needed to determine what areas are necessary for accurate speech sound discrimination.

Future studies

Many future studies are needed to expand upon the experiments presented here. Plasticity studies in awake, behaving animals are necessary to confirm the current findings. Speech training and physiology experiments conducted in the presence of background noise are needed since in the real world, comprehending speech often occurs in noisy environments, and not in a sound-shielded laboratory environment. Studies involving stimuli that have been spectrally degraded are necessary to determine how the cortical response in cochlear implant patients differs from the cortical response when all spectral information is available. Additionally, physiology experiments need to be conducted in other auditory areas, both higher and lower in the auditory pathway to determine how the representation of speech changes along the pathway, and which specific areas are necessary for accurate speech discrimination.

REFERENCES

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, and Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19: 113-124, 2003.

Gentner TQ, and Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669-674, 2003.

Gifford GW, 3rd, Hauser MD, and Cohen YE. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. *Brain, behavior and evolution* 61: 213-224, 2003.

Kent RD, and Read C. *The acoustic analysis of speech.* (Singular Publishing, San Diego, CA, 2002).

Kluender KR, Diehl RL, and Killeen PR. Japanese quail can learn phonetic categories. *Science* 237: 1195-1197, 1987.

Kuhl PK, and Miller JD. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69-72, 1975.

Perez CA, Engineer CT, Puckett AC, Jakkamsetti V, Chang KQ, Carraway RS, Chen YH, Perry MS, Kilgard MP. Double dissociation of consonant and vowel encoding in the rat inferior colliculus. UTMB, Galveston, Texas, June 2008.

Polley DB, Steinberg EE, and Merzenich MM. Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26: 4970-4982, 2006.

Porter, B. A., Floody, O. R., Ouda, L., Skillern, C., Sharma, T., & Kilgard, M. P. (2007, November). Unimpaired speech discrimination in rats following bilateral auditory cortex lesions. Poster session presented at the annual meeting of the Society for Neuroscience, San Diego, CA.

Porter, B.A., Jasti, T. K., Kavikova, D., Frazier, R.C., Kilgard, M.P. Speech sound discrimination in rats with primary auditory cortex lesions. Program No. 850.20. 2008 Neuroscience Meeting Planner. Washington, DC: Society for Neuroscience, 2008. Online.

Recanzone GH, Schreiner CE, and Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci* 13: 87-103, 1993.

Reed P, Howell P, Sackin S, Pizzimenti L, and Rosen S. Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. *Journal of the experimental analysis of behavior* 80: 205-215, 2003.

Rosen BR, Buckner RL, and Dale AM. Event-related functional MRI: Past, present, and future. *Proc Natl Acad Sci U S A* 95: 773-780, 1998.

Schnupp JW, Hall TM, Kokelaar RF, and Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26: 4785-4795, 2006.

Seyfarth RM, Cheney DL, and Marler P. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210: 801-803, 1980.

Steinschneider M, Fishman YI, and Arezzo JC. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *The Journal of the Acoustical Society of America* 114: 307-321, 2003.

Steinschneider M, Reser D, Schroeder CE, and Arezzo JC. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain research* 674: 147-152, 1995.

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, and Howard MA, 3rd. Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb Cortex* 15: 170-186, 2005.

Steinschneider M, Volkov IO, Noh MD, Garell PC, and Howard MA, 3rd. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of neurophysiology* 82: 2346-2357, 1999.

Tremblay K, Kraus N, McGee T, Ponton C, and Otis B. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing* 22: 79-90, 2001.

Wong SW, and Schreiner CE. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Communication* 41: 93-106, 2003.

VITA

Crystal Tasha Engineer was born in Etobicoke, Ontario, Canada on May 5, 1982 to Dave and Doreen Novitski. She moved to Houston when she was 9, and graduated from Cy-Fair High School with honors in 2000. She entered the University of Texas at Dallas in 2000, and was the recipient of a UTD Academic Excellence Award as an undergraduate. She joined Dr. Kilgard's lab in 2003 and completed an undergraduate honor's thesis titled *Sequence exposure in adult rat primary auditory cortex*. She received a Bachelor of Science degree in Neuroscience in 2003 with School of Behavioral and Brain Science honors. She began graduate school in 2003, and received a Master of Science degree in Applied Cognition and Neuroscience in 2005. In August 2006, she married Navzer Engineer.