CORTICAL PROCESSING OF SPEECH SOUNDS IN ADVERSE CONDITIONS

by

Jai Anand Shetake

APPROVED BY SUPERVISORY COMMITTEE:

_____

Dr. Michael P. Kilgard, Chair

_____

Dr. Peter F. Assmann

_____

Dr. Marco Atzori

_____

Dr. Christa K. McIntyre

To all my teachers

CORTICAL PROCESSING OF SPEECH SOUNDS IN ADVERSE CONDITIONS

by

JAI ANAND SHETAKE, B.E., 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

May, 2011

ACKNOWLEDGEMENTS

I have been extremely fortunate to pursue an education that was full of challenges and excitement. As a computer science undergrad I had learnt the workings of the most amazing machine invented by mankind. Yet, I was intrigued by how even the fastest machines on earth cannot keep up with some of the most basic tasks our brains can perform efficiently and effortlessly. I knew nothing about neuroscience so I started reading neuroscience textbooks. I was awed by how our brains represent information and how our brains develop to learn everything around it. I wanted to learn more about how the brain does what it does, but was uncertain whether I would be able to meet the challenge of switching fields. With high hopes of learning this amazing science, I contacted Dr. Kilgard.

My journey in Dr. Kilgard's lab has been nothing less of fabulous. I remember my first meeting with my advisor, Dr. Kilgard. He had said that by the time I am done, I will have answered the question: How do neurons represent complex stimuli? With no background experience in neuroscience, this task seemed impossible!! The people in Dr. Kilgard's lab made this possible! I want to thank everyone who has helped me accomplish my goals and been a part of this exciting journey!

To my advisor Dr. Kilgard- Thank you for being such an amazing teacher. Your contagious energy and passion for neuroscience has provided me a great intellectual boost in the past few years. You have been one of the most influential teachers in my life and I will always be extremely grateful for giving me this incredible opportunity to work in your lab. Working in your

lab gave me the ability to harvest my computer programming skills so that I could interpret neural signals and understand a part of the big puzzle of how the brain represents the complex environment around it!

To my Committee Members- for their guidance and for always pushing me to perform better. Dr. Assmann has been a great source of information for understanding speech processing. His class in speech perception helped me understand the gravity of the problem of speech perception in adverse conditions and laid the groundwork to start my experiments. He has always reminded me the importance of being rigorous and drawing careful conclusions from my data- a capability that every scientist must possess. Dr. Atzori and Dr. McIntyre have always been patient through all the long meetings and help me better communicate scientific information through presentations.

To all my lab friends-I would like to thank the entire army of undergrads who helped me complete my projects. Some of the undergrads have been an integral part of my projects from the start to the end and are authors on the papers I am working on. Among many important things, they helped me design experiments, collect equipment, behaviorally train animals, and perform electrophysiology mappings. Jordan Wolf is an author on 2 of the papers I am working on, and has been an un-replaceable asset. He cheerfully accepted all the night-time mapping shifts, collected and analyzed behavioral data. Ryan Cheung, Kinsey Ram and Will Vrana are authors on one of the papers I am working on and collected behavioral training data and performed electrophysiology mapping experiments and were always ready to help out at the even most unreasonable hours often required for these projects!! Hibah Mumtaz and Elise Tran recorded a number of electrophysiology sessions in awake rats. Their help made it possible to collect data

for one of the published papers. Thank you everyone-these experiments would not have been possible without you.

To my colleagues and friends Crystal Engineer, Amanda Reed, Vikram Jakkamsetti , Navzer Engineer, Ben Porter, Kamalini Ranasinghe, Navid Khodparast, Tracy Rosen, Mike Borland, Will Vrana who have helped me grow intellectually and have been a great source of support throughout this degree.

To the administrative staff of the school, Bonnie Dougherty, Mary Felipe, Abbie Bailey, Jo Valcik, and Susie Milligan—thank you for helping me with all the minute details and staying on track with all the paperwork so that I could concentrate on the science.

Getting a PhD was not easy. The path to successfully completing a doctorate required hard work and patience, the ability to not give up and yet the ability to let go of unfruitful endeavors without bitterness. A number of people in my life have helped me learn these qualities. To my parents and all the elders in my family who have instilled the importance of hard work in me. Coming from a strictly didactic educational culture, it was hard to realize the importance of being creative. To my uncle Sudhakar Asawale, who taught me how much fun out of box thinking can be. To all my fun uncles and aunts who have been extremely generous and helped me financially to take my first steps towards this career. To my dearest siblings, Indrani Shetake and Omkar Shetake, who have been a great emotional support and also a provided great comic relief through all the tough times.

To all my dear friends who make life my so rich. Of many friends whom I have been lucky to have shared a common path with, I am especially thankful to Niyati Jain and Nissan Mehtab- Thank you for listening to my how-hard-PhD-is rant, sharing so many dinners and all those

evenings filled with laughter! To Mitali Bose, thank you for teaching me how to drive a car. Having this ability meant I could go to the grocery store even after working in the lab till 1 in the morning and ensured that I was fed while pursuing my PhD! To Kaustubh Nehe, words can't even express the gratitude I feel for you being there! Amongst so many other things that I can't even start listing, thank you for teaching me how to not get bogged down by work! Thank you for bearing with all my crazy hours and always putting a smile on my face at the end of every day. I will never forget how kind and giving you all have been.

April, 2011

CORTICAL PROCESSING OF SPEECH SOUNDS IN ADVERSE CONDITIONS

Publication No. _____

Jai Anand Shetake, Ph.D.
The University of Texas at Dallas, 2011

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

The brain has an amazing capacity to accurately represent information about myriad complex stimuli from its surrounding environment. Problems with maintaining accurate representations of stimuli have harmful effects and are reflected behaviorally. Speech is one such complex stimulus. Normal hearing humans are adept at understanding speech in adverse situations. However hearing and learning impaired populations have great difficulty processing speech sounds in noise. Examining neural responses to speech sounds may provide a better insight into how the auditory system represents speech sounds. Previous studies with EEG and fMRI techniques have revealed much about how the brain responds to speech sounds. However, these techniques lack the spectral and temporal precision necessary to study the underlying neural representation of individual speech sounds. In this dissertation, microelectrode recordings of speech responses in rat primary auditory cortex provide high spatial and temporal precision. Previous studies have shown that animals have similar behavioral and neural responses to speech

sounds as humans. Primary auditory cortex (A1) responses in rats are correlated with their behavioral speech discrimination ability in quiet situations. Our study shows that rats can discriminate between speech sounds in noisy conditions and that this ability can be explained by the A1 spatiotemporal patterns observed in noisy conditions. Maintaining temporal resolution of neural responses is necessary to accurately predict behavior in noise. A number of previous studies suggest that speech processing disorders arise from underlying neural deficits in temporal processing. Given the importance of temporal information for the representation of speech stimuli, we tested whether changes in temporal response properties could affect neural responses to speech sounds. Vagus nerve stimulation (VNS) was paired with tone train stimuli to change the temporal response properties of A1 neurons. Results from this study indicate that increased temporal following capacity of neurons also increased neural discrimination ability of rapidly presented speech sounds.

TABLE OF CONTENTS

LIST OF FIGURES

Number Page

LIST OF TABLES

Number Page

CHAPTER 1

INTRODUCTION

Speech is an integral part of day to day human interaction. In everyday life, we are faced with the task of understanding speech sounds in a wide variety of adverse situations. Effective speech communication can occur only when the relevant information is picked out of myriad other disturbances. Difficulty in speech perception can arise from external sources like speech in noise, or from internal sources like rapidly presented speech in conversations or accented speech. Whereas small amounts of noise, like the sound of a computer fan in the background does not significantly degrade speech perception, large amounts of noise can cause significant interference in accurate speech communication even in normal hearing populations. Difficulty in perception of speech sounds is exacerbated in hearing and learning impaired populations and after stroke or traumatic brain injury. An estimated 38 million people suffer from hearing disorders and another 4 million children from language learning disorders in the United States alone. An understanding of how auditory cortex neurons represents sounds in adverse conditions will help in determining the nature of neural encoding deficits and may further lead to better treatment options for improving speech perception ability in these populations. This dissertation covers 2 interrelated topics in field of neural processing of speech sounds: 1) Processing of speech sounds in noisy situations. 2) Effect of vagus nerve stimulation (VNS) induced temporal plasticity on rapid speech sound processing.

Although technological advancement has made it possible to understand which areas of the brain are affected in the pathological nervous system, the underlying neural processing of speech sounds is still not well understood. For example, fMRI and EEG recordings do not provide the spatial and temporal precision necessary to understand neural processing of speech signals (Callan et al. 2003; Rosen et al. 1998). Microelectrode recordings in animals provide better spatial and temporal precision than EEG and fMRI recordings. Using this technique, a number of previous studies have shown that auditory cortex responses in animals respond well to a wide variety of speech sounds (Steinschneider et al., 1999; Wong and Schreiner, 2003; Steinschneider et al., 2004; Engineer et al., 2008; Skoe and Kraus, 2010). Neural responses to speech sounds in many species of animals are very similar to neural responses in humans. Behavioral discrimination performance in many species of animals is similar to humans for a variety of speech contrasts in quiet situations. It is not yet known whether animals, like humans can discriminate reliably between speech sounds even when the speech and noise signals are of equal intensities. In our study we tested neural and behavioral discrimination ability of rats in different noise intensity levels and noise types.

A previous study showed that precise spatiotemporal activity pattern in the primary auditory cortex (A1) of rats can predict their behavioral performance on a variety of speech discrimination tasks in quiet (Engineer et. al, 2008). Spike timing information is required for accurate prediction of speech discrimination ability in quiet. Using the rat model for speech processing in noise, I asked the following question. Can spike timing predict behavioral speech sound discrimination ability in noisy situations? Results from this study illuminate the brain mechanisms that make it possible to effortlessly distinguish complex textures, sights and sounds even in highly noisy

situations (Chapter 2). Chapter 2 discusses the behavioral speech discrimination ability of rats in different intensities and spectral compositions of noise. This chapter further examines neural correlates of speech discrimination in these noisy conditions. This study provides the first evidence that animals can discriminate between most of the tested speech contrasts in high levels of background noise at a level similar to humans. Neural responses analyzed using spike timing mechanisms in the rat A1 were highly correlated with their behavior speech discrimination ability in variety of noise conditions. Although previous studies have shown that animals can discriminate between speech sounds to a similar level as humans, the question of whether these results hold under adverse conditions had not yet been answered (Kuhl and Miller 1975; Tallal et al., 1993; Cunningham et al., 2002). These results suggest common speech sound discrimination mechanisms between humans and animals and support the use of animal models in understanding the neurophysiologic representation of speech sounds and speech sound processing impairments in humans.

A large number of learning impaired populations like dyslexics, autistics and poor readers have problems understanding rapidly presented stimuli like conversational speech (Tallal and Piercy 1973; Farmer and Klein 1995; Denckla and Rudel 1976; Tallal et al. 1985; 1981; Elliot et al. 1989; Reed 1989; Wolff et al.; 1990). A number of studies have indicated that the underlying problem lies in the basic auditory temporal processing (Merzenich et al., 1996; Kraus et al. 1996; Stark and Heinz 1996a; Tallal et al., 1998). Having established the importance of temporal processing in accurate predictions of behavioral speech sound discrimination, I sought to understand whether changes in temporal response properties of neurons can affect speech sound discrimination in adverse situations. The aim of this project was twofold: 1) Find a

clinically viable tool that is capable of inducing plasticity in the central nervous system and test whether this tool can induce temporal plasticity in the auditory cortex. 2) If this tool could induce temporal plasticity, then test effect of improved temporal plasticity on rapidly presented speech sounds.

VNS has been used in 50,000 clinically to treat epilepsy and chronic depression (Groves & Brown, 2005; Albert et al., 2009). Pairing VNS with tones has been shown generate spatial plasticity and has been a valuable tool in treating tinnitus equivalent in animals (Engineer et al., 2011).

Repeatedly pairing electrical stimulation of the nucleus basalis (NBS) or the vagus nerve (VNS) with a tone induces spatial plasticity by increasing the area of primary auditory cortex (A1) that responds to the paired tone frequency (Engineer et al., 2011). The similarity of spatial plasticity effects seen with NBS and VNS indicate similar underlying plasticity mechanisms with both methods. Previous studies have shown that pairing NBS with rapid stimuli can also induce temporal plasticity (Kilgard & Merzenich, 1998b; Kilgard, et. al, 2001). These results led me to test the hypothesis that vagus nerve stimulation (VNS) can induce temporal plasticity similar to NBS (Chapter 3). Chapter 3 & 4 discusses the temporal plasticity effects induced by VNS and the effect of this plasticity on neural processing of rapid speech stimuli. The key findings from this project are that: 1) VNS is capable of generating highly input specific temporal plasticity. 2) Animals which received VNS have increased responses to rapidly presented speech stimuli.

Chapter 5 discusses the relevance to previous literature, clinical implications of these results and directs to possible future studies.

CHAPTER 2

CORTICAL ACTIVITY PATTERNS PREDICT ROBUST SPEECH DISCRIMINATION ABILITY IN NOISE

Jai A. Shetake, Jordan T. Wolf, Ryan J. Cheung, Crystal T. Engineer, Satyananda K. Ram, Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

2.1 ACKNOWLEDGEMENTS

The authors would like to thank K. Ranasinghe, W. Vrana and J. Riley for their assistance with microelectrode mappings. We would like to thank B. Porter, A. Reed for their help with behavioral experiment set-up and T. Rosenthal, W. Vrana and K. Fitch for their help with behavioral training. We would also like to thank P. Assmann, H. Abdi, M. Atzori, C. McIntyre, B. Porter, K. Ranasinghe, A. Reed, T. Rosen and M. Borland for their suggestions about earlier versions of the manuscript. This work was supported by grants from the US National Institute for Deafness and Other Communicative Disorders (R15DC006624 and R01DC010433).

2.2 ABSTRACT

The neural mechanisms that support speech discrimination in noisy conditions are poorly understood. In quiet conditions, spike timing information appears to be used in the discrimination of speech sounds. In this study, we evaluated the hypothesis that spike timing is also used to distinguish between speech sounds in noisy conditions that significantly degrade neural responses to speech sounds. We tested speech discrimination performance in rats and recorded primary auditory cortex (A1) responses to speech sounds in background noise of different intensities and spectral compositions. Our behavioral results indicate that rats, like humans, are able to accurately discriminate consonant sounds even in the presence of background noise that is as intense as the speech signal. Our neural recordings confirm that speech sounds evoke degraded but detectable responses in noise. Finally, we developed a novel neural classifier that mimics behavioral performance. The classifier discriminates between speech sounds by comparing the A1 spatiotemporal activity patterns evoked on single trials with

the average spatiotemporal patterns evoked by known sounds. Unlike classifiers in most previous studies, this classifier is not provided with the stimulus onset time. Neural activity analyzed using relative spike timing was well correlated with behavioral speech discrimination ability in quiet and in noise. Spike timing information appears to be integrated over somewhat longer temporal intervals in noisy conditions. The similarity of behavioral speech discrimination and neural responses between humans and rats in noise suggests that similar brain mechanisms may be used to solve this problem behaviorally.

2.3 INTRODUCTION

Previous studies suggest that animals and humans may share speech sound processing capabilities (Kuhl and Miller 1975; Tallal et al., 1993; Cunningham et al., 2002; Mesgarani et al., 2008). Monkeys, cats, birds and rodents can accurately discriminate between various speech sounds (Kuhl and Miller 1975; Kluender et al, 1987; Ramus et al., 2000; Wong and Schreiner, 2003; Engineer et. al., 2008). Human listeners are adept at discriminating speech sounds in background noise that is as loud as the speech signals (Miller and Nicely, 1955; House et al., 1965; Wang and Bilger, 1973; Dubno and Levitt, 1981; Phatak and Allen, 2007). It is not known whether animals can discriminate between speech sounds in noise levels as high as humans. A finding that animal speech discrimination is more sensitive to background noise than human speech discrimination would support the earlier hypothesis that humans are uniquely equipped to process speech sounds (Liberman and Mattingly, 1985; Wilkins and Wakefield, 1995). A finding that animal speech discrimination is as robust to noise as human speech discrimination would support the hypothesis that humans and animals share common speech sound processing mechanisms.

Neurophysiologic studies suggest similar neural mechanisms for representation of speech sounds in humans and animals (Steinschneider et al., 1999; Wong and Schreiner, 2003; Engineer et al., 2008; Mesgarani et al., 2008). For example, in both humans and animals, voiced stop consonants evoke a single peak of cortical activity, while unvoiced stop consonants evoke two peaks (Steinschneider et al., 1999; Wong and Schreiner, 2003; Engineer et al., 2008; Mesgarani et al., 2008). In noisy situations, degraded human evoked-potential responses to speech in noise are correlated with impairments in behavioral speech discrimination ability (Whiting et al., 1998; Martin et al., 1999; Binder, et al., 2004). The neurophysiologic correlates of animal speech discrimination in noise are not yet known. Examining neural correlates of speech processing in noise may strengthen the hypothesis that basic neural processing of speech sounds is similar in humans and animals.

Although little is known about the neural mechanisms that allow speech discrimination to remain robust in high levels of noise, previous studies in the visual and auditory cortex have suggested that in noisy situations neurons accumulate information over longer durations (Binder et al., 2004; Huk and Shadlen, 2005). We tested the hypothesis that speech discrimination remains robust in noise by increasing the degree of temporal integration used when neural activity patterns are decoded. To test our hypothesis, we quantified behavioral and neural discrimination of consonant sounds in noise of differing intensity and spectral composition. Our previous study of speech discrimination in quiet suggested that spike timing with precision of 1-10 ms is used to discriminate speech sounds. We developed a novel classifier to compare behavior with neural discrimination using different levels of temporal precision. The classifier was not provided with information about when the sounds would occur because high levels of

noise were expected to eliminate the onset response to speech sounds (Martin et al., 1999; Cunningham et al., 2002; Russo et al., 2004).

2.4 MATERIALS AND METHODS

Stimuli

Speech Stimuli: We used 11 of the 20 English consonant-vowel-consonant (CVC) words, ending in ‘ad’ (as in 'tad') used in the Engineer et al. (2008) study. These sounds were ‘bad’, ‘dad’, ‘gad’, ‘pad’, ‘tad’, ‘sad’, ‘yad’, ‘rad’, ‘lad’, ‘shad’, and ‘chad’. A detailed description of the recording and processing on these sounds can be found in Engineer et al. (2008). In brief, we recorded these sounds in a double-walled, sound-proof booth. All speech sounds were produced by a female speaker. 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) to better match the rat hearing range (Sally and Kelly, 1988). The intensity of the speech sounds was adjusted so that the intensity during the most intense 100 ms is 60 dB SPL. Speech sounds were approximately 500 ms long.

Noise Stimuli: White noise was generated using a random number generator in MATLAB and covered a frequency range of 0.2-50 kHz. The speech-shaped noise was generated by passing white noise through first order Butterworth filters. Butterworth filters available in the FDATool function in MATLAB were used. Since the speech stimuli in our experiment were produced by a female speaker, we used the long term average speech spectrum (LTASS) of the female voice. LTASS for female voices has been shown to be flat from ~200 Hz to ~700Hz and falls off by approximately 7 dB SPL per octave on the either side. This shape for speech-shaped noise has been shown to be standard across almost all the languages in the world (Byrne et al., 1994). Like

speech stimuli, the frequency limits of speech-shaped noise from Byrne et al. (1994) were shifted up by a factor of two in order to adjust for the rat frequency hearing range. The speech-shaped noise used in this study has a flat spectrum between ~400 Hz and ~1400 Hz and slopes by 7 dB SPL per octave on either side. Both white noise and speech-shaped noises were calibrated to 48, 60 and 72 dB SPL.

Behavior Training Procedure

We trained 12 female Sprague-Dawley rats to discriminate speech sounds in quiet and noise. An operant go-no-go training procedure was used for speech discrimination training. Half of the rats (n=6) were trained to press a lever in response to the target sound 'dad' and ignore the distracter sounds 'bad', 'gad', 'tad' and 'sad'. The other half of the rats (n=6) were trained to press the lever on the target sound 'shad' and ignore the distracter sounds 'bad', 'dad', 'yad', 'sad', 'pad', 'tad', and 'chad'. The rats weighed an average of 275 ± 15 g and were housed in 12:12 hr reversed light cycle at a constant humidity and temperature. Animals were food deprived to provide motivation for food reward during the behavior training but were maintained to no less than 85% of their normal body weight. Access to water was free at all times except during the behavior training session which was an hour long. Rats were trained for two training sessions each day for 5 days a week in a sound-shielded operant-training booth. The booth contained a video camera for monitoring, a house light, a cage (8 inches length X 8 inches width X 8 inches height), and a speaker. The cage contained a lever, lever light, and pellet receptacle. The pellet dispenser was mounted outside the training booth to minimize noise. The speaker was

mounted approximately 10 cm away from the midpoint between the lever and the pellet dispenser receptacle since rats generally stayed in this area during behavior training sessions.

Rats progressed through the following behavior training stages: 1. Shaping, 2. Detection of sounds, 3. Discrimination in Quiet, and 4. Discrimination in Noise. The training protocol used up to Stage 3 i.e. Discrimination in quiet, was similar to our previous study, Engineer et al. (2008). During the shaping stage, the rats were trained to press the lever for food reward. Any time the rats were in the proximity of the lever, they were rewarded with a 45 mg sugar pellet. The target sound was played every time the rat received the food pellet, so that the rats learned to associate the sound with the food reward. We increased the reward criteria gradually until the rat was only rewarded for pressing the lever. After the rats learned to independently obtain approximately 100 pellets for 3 consecutive sessions, they were moved to the detection of sounds stage where they were trained to press the lever only in response to the target sound.

During the initial detection of sounds training stage, the target speech sound was presented every 10 seconds. Silent catch trials were randomly interleaved 25-30% of the time and there were no speech sound distracters in this phase. Rats were rewarded if they pressed the lever within a 8 second window after presentation of the target sound. Pressing the lever during the silent catch trials resulted in a time out during which the house light was extinguished, the training program was paused for 8 seconds, and the rats did not receive any food pellets. As performance improved over sessions, the sound presentation interval was gradually reduced to 6 seconds, and the lever press window was gradually reduced to 3 seconds. We calculate *d'* to determine the hit rates compared to the false alarm rates (Engineer et al., 2008). A *d'* of 0 indicates the rat is pressing the lever equally often to both target and distracter, while a positive

d' indicates that the rat has a higher hit rate than false alarm rate. A *d'* of 1.5 or higher indicates that the rats were reliably detecting the target sound. Once rats reached the performance $d' \geq 1.5$ for ten sessions, they were advanced to discriminate between consonant sounds in quiet.

During each consonant discrimination task, rats learned to discriminate the target sound from the distracter sounds. Initially, the rats learned to discriminate sounds in quiet. Trials began every 6 seconds and rats were only rewarded for lever presses to the target (conditioned) stimulus which were randomly interleaved 50% of the time. Silent catch stimuli and distracter sounds were randomly interleaved 50% of the time and pressing the lever on these trials resulted in a time-out of 6 seconds. After 20 sessions of discrimination training in quiet, the rats were trained to discriminate between speech sounds in background noise. During each session, rats were trained to discriminate speech sounds in both speech-shaped noise and white noise of 48, 60 and 72 dB SPL intensities. We also continued to test rat behavioral speech discrimination ability in quiet to allow direct comparison between discrimination ability in quiet and noise on the same days. Each session started with 13 trials of discrimination in quiet. For future reference, we will refer to 13 trials played in a particular noise intensity and noise type as a noise block. Sounds presented in each block were chosen randomly to maintain a 1:1 presentation ratio of target sound and the distracters. After discrimination of trials in quiet, either white noise or speech-shaped noise of 48 dB SPL was randomly chosen to be played in the background. Noise intensity was gradually increased every noise block to the maximum level, so that the animal could habituate to the noise, and then gradually decreased. After this sequence, the noise was changed to a random noise intensity and noise type for every block. Noise had a 3 second ramp on to avoid startling the rat with abrupt noise bursts. After this sequence, noise blocks were played

randomly. During all stages of training, lever press or the absence of lever press was recorded for each trial along with the time of lever press response. Training performance was quantified using percent correct and d prime measure. Percent correct performance was used to correlate neural performance which was obtained from a neural classifier, which is explained later in the section. The percentage of trials on which the rats pressed the lever to the target sound and the percentage of trials on which the rats withheld from pressing on the lever for the distracter sounds were averaged together to calculate the overall behavioral percent correct score. Discrimination performance of rats in noise improved consistently up to the fiftieth session after which their performance reached a plateau. Behavioral data collected on the last 20 training sessions was averaged together to determine the speech discrimination ability of rats. In order to compare the performance of our neural classifier with a wide variety of behavior tasks, we included behavioral data of 5 additional tasks from our earlier study in quiet and averaged behavioral discrimination ability of 6 tasks common to both studies. The discrimination tasks common to both studies were ‘dad’ versus ‘bad’, ‘dad’ versus ‘gad’, ‘dad’ versus ‘tad’, ‘dad’ versus ‘sad’, ‘shad’ versus ‘sad’, and ‘shad’ versus ‘chad’. The discrimination tasks used only in our previous study were ‘shad’ versus ‘fad’, ‘shad’ versus ‘jad’, ‘shad’ versus ‘had’, ‘rad’ versus ‘lad’, and ‘mad’ versus ‘nad’. For behavior analysis, we used two-factor repeated measures like analysis of variance i.e. ANOVA and Tukey *t*-tests for *post-hoc* analysis with significance set to $p < 0.05$. All handling, housing, and testing of the animals was approved by the University of Texas Institutional Animal Care and Use Committee.

Neural Data -Electrophysiology Recordings

We recorded multiunit responses to speech in quiet and in six noise conditions from 133 sites in right primary auditory cortex (A1) of 8 anesthetized naïve rats. Neural responses from our earlier study (n=445 A1 sites from 11 rats; Engineer et al., 2008) recorded in quiet were used to evaluate neural discrimination for the five additional behavior tasks that were only tested in quiet. Rats were anesthetized with pentobarbital for surgery (50 mg/kg) and a state of areflexia was maintained throughout the experiment with supplemental doses of dilute pentobarbital (0.2–0.5 ml; 8 mg/ml). Anesthesia depth was monitored by heart rate, breathing rate, corneal reflexes, and response to toe pinch. Nourishment was provided using 1:1 mixture of dextrose (5%) and standard Ringer's lactate solution and body temperature was maintained at 37 degrees C. A tracheotomy was performed to minimize breathing problems and breathing sounds, and a cisternal drain was made to minimize cerebral edema. A part of the skull over the temporal ridge was removed to expose the right primary auditory cortex. The dura was removed and the cortex was maintained under a thin film of silicone oil to prevent desiccation. Four parylene coated tungsten microelectrodes (FHC, 1-2 MΩ) were lowered simultaneously to a depth of 600 µm so that they were in layer IV/V of the primary auditory cortex. Blood vessels were used as landmarks to mark each of the electrode recording sites. To determine the characteristic frequency at each site we played ninety logarithmically spaced tones ranging from 1- 47 kHz at 16 intensities ranging from 0-75 dB SPL. The tones were 25 ms long and their presentation was randomly interleaved. We placed the speaker 10 cm away from the left ear. Neural response characteristics such as start latency, end latency and characteristic frequency at each recoding site were used to determine whether the electrodes were placed in the primary auditory cortex.

After all the tones, speech sounds were played. The speech stimulus set was comprised of the same 11 monosyllabic consonant-vowel-consonant (CVC) words used in behavior training, a silence stimulus, and two noise types i.e., white noise and speech-shaped noise. Each speech sound was separated by 2300 ms and presented 20 times in quiet and in 3 noise intensity levels viz., 48, 60 and 72 dB SPL of both white noise and speech-shaped noise. The different noise conditions i.e. noise levels and noise types were interleaved to avoid state of anesthesia from affecting neural responses in any particular noise condition. For example, speech sounds were first played 10 times in different intensities of speech-shaped noise, then 10 times in different intensities of white noise, then 10 times again in different intensities of speech-shaped noise and finally 10 times in different intensities of white noise resulting in all the speech sounds being presented 20 times in both speech-shaped noise and white noise. Within each noise type, the noise intensities were interleaved as follows. All the sounds (11 speech sounds and silent stimulus) were repeated 5 times at one noise intensity. From here on, we will refer to this set of sounds i.e. 12 sounds repeated 5 times played in a particular noise intensity and noise type as one block. Within each block, the speech sounds were randomly interleaved. The sequence in which the blocks were played is as follows. First, one block of sounds was played in quiet. Either white noise or speech-shaped noise was added in a linearly increasing order of intensity and then in a linearly decreasing order. This sequence resulted in each of the sounds being played 10 times in one noise type. As mentioned earlier, this sequence was repeated with the other noise type, and then again with the first noise type and lastly with the second noise type. Noise intensities within each noise type were played in a gradual increasing and decreasing order to avoid sudden changes in neural responses that could result from random presentation of noise. Stimulus

generation and data acquisition was performed with Tucker-Davis hardware (RP2.1 and RX5) and software (Brainware). Surgery protocols and recording procedures were approved by the University of Texas at Dallas Institutional Animal Care and Use Committee.

Neural Data-Electrophysiology Data Analysis

Neural response characteristics: Total number of spikes was calculated over the first 100 ms of the neural response onset. Onset start latency (ms) is defined as the time from stimulus onset to the earliest reliable neural response and was determined as the time when average neural responses from all recording sites were at least 3 standard deviations above spontaneous activity. End of peak latency (ms) is the time when the average neural responses return to baseline and was determined as the time when neural responses after the peak driven response were not significantly different than spontaneous activity. Distinctness between neural response patterns was calculated using City-block and Euclidean distance to test the correlation between neural distinctness and behavioral discrimination performance. For neural response analysis, we used two-factor repeated measures ANOVA and Tukey *t*-tests for *post-hoc* analysis with significance set to $P < 0.01$.

Classifier: To quantify neural discrimination performance, we modified a well studied nearest neighbor classifier (Engineer et. al, 2008; Schnupp et. al, 2006; Foffani et. al, 2004). Unlike our earlier classifier that was explicitly given the stimulus start time (Engineer et. al, 2008), the version of classifier used in the current study was not given information about the stimulus start time.

The neural classifier identifies sounds based on activity produced by a single

presentation. The classifier attempts to identify which of two possible sounds was presented by looking for the spatiotemporal activity patterns generated by each sound. The two activity patterns are derived from the average response from a set of recordings sites distributed across A1. The average spatiotemporal activity patterns are generated from 19 trials and the current trial is not included. In this study, the average spatiotemporal pattern evoked by each sound was usually 100 or 400 ms in duration and was binned with 10 or 60 ms precision. For example, Figure 2.6a shows the average response of 60 A1 sites to the word 'dad'. In this case, the duration of the pattern is 100 ms and the bin size is 10 ms. As previously reported, neurons tuned to low frequency tones respond to the onset of 'dad' approximately 20 ms **after** neurons tuned to high frequency tones (**Figure 2.6a**; Engineer et al., 2008; see Figure 2 and Video 1 from Engineer et al., 2008). The average spatiotemporal pattern evoked by the word 'bad' is distinct from the pattern evoked by 'dad'. In response to 'bad', neurons tuned to low frequency tones tend to respond **before** neurons tuned to high frequency tones (**Figure 2.6b**).

The classifier determines which of the two patterns was more likely to have occurred during the single trial response period that is 750 ms long (**Figure 2.6c**). The classifier computes the City-block distance (**Figure 2.6c**, top) between the average patterns and the single trial activity assuming all possible start times. City-block distance is simply the average difference between the firing rate at each bin of the single trial activity and the average pattern. All results were also tested using Euclidian distance and were similar to results with City-block distance. The classifier guesses that the sound that was presented on each trial was the sound whose average spatiotemporal pattern was closest (**Figure 2.6c**, asterisk) to the single trial activity. Neural discrimination performance for a group of sites was determined by calculating the

percentage of trials on which the classifier correctly guessed the presented sound. Behavioral performance was compared with classifier performance using average spatiotemporal patterns produced from 1-133 A1 sites using durations of 10-700 ms and a bin size of 1-700 ms. In the most common form of the classifier, activity patterns were derived from groups of 60 recordings sites that were randomly selected from the set of 133 A1 sites from which speech in noise responses were collected. $Classifier_{100/10ms}$ refers to the classifier when 100 ms of activity from 60 sites binned every 10 ms was used for the neural discrimination. $Classifier_{400/60ms}$ refers to the classifier when 400 ms of activity from 60 sites binned every 60 ms was used. The hybrid classifier uses 100 ms of activity binned every 10 ms in quiet conditions and 400 ms of activity binned every 60 ms in noise.

We also analyzed neural responses to sounds from our previous study in quiet using the classifiers used in this study and compared whether our new neural classifier could predict behavior discrimination ability of tasks from our previous study (Engineer et al., 2008).

2.5 RESULTS

Behavior Results

Our results provide the first evidence that animals, like humans, can discriminate between numerous speech sounds even when the speech signals and background noise are of equal intensity (**Figure 2.1**). We trained twelve rats to discriminate between different consonant sounds in noise. Half of the rats (n=6) were first trained to discriminate the target word ‘dad’ from the distracter words ‘bad’, ‘gad’, ‘tad’, ‘sad’ in quiet. The other half were first trained to discriminate the target word ‘shad’ from the distracter words ‘bad’, ‘dad’, ‘pad’, ‘tad’, ‘yad’, ‘sad’, ‘chad’ in quiet. All the speech sounds were presented such that the loudest 100 ms was at

60 dB SPL. After the rats learned to accurately perform the tasks, both groups were required to discriminate the same stimuli in white noise and speech-shaped noise of 48, 60, and 72 dB SPL.

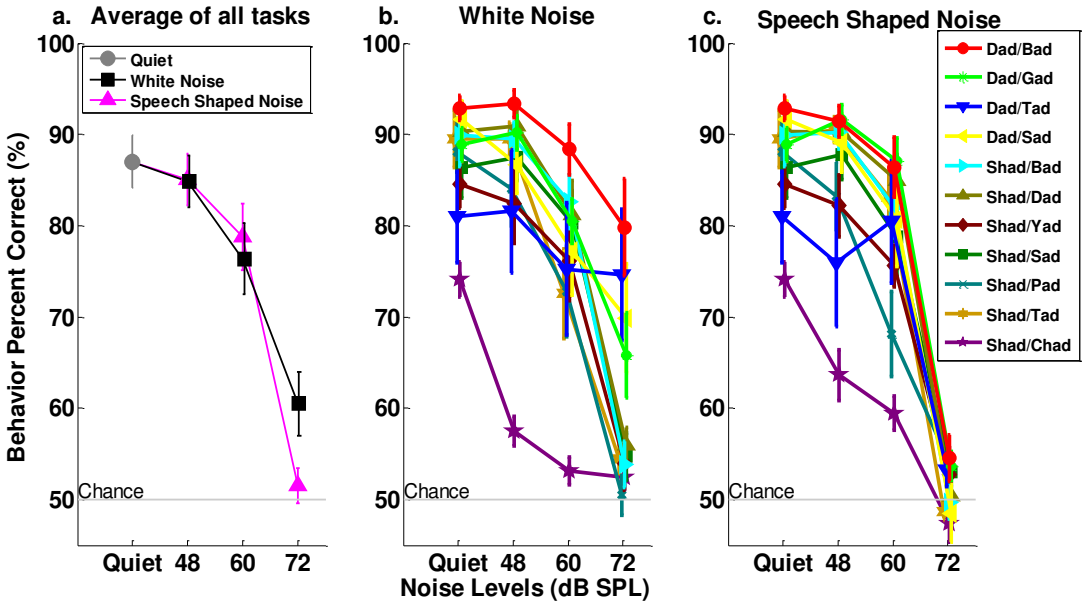

Figure 2.1. Behavioral discrimination of consonant sounds in different intensities of white noise and speech shaped noise.
(a) Average speech discrimination ability on all tasks. Rats could discriminate between consonant sounds well above chance level even when speech and noise were of equal intensity i.e. 60 dB SPL (p = 0.005; Tukey *posthoc*). (b) Discrimination performance of 11 consonant discrimination tasks in quiet and different intensities of white noise. (c) Discrimination performance of 11 consonant discrimination tasks in quiet and different intensities of speech-shaped noise. Chance level performance is shown in light gray lines. Speech-shaped noise of 72 dB SPL was more impairing than white noise of 72 dB SPL ($F_{(1, 66)}$ =10.20, MSE = 244.02; p = 0.001).

Consistent with our earlier report, the rats were able to discriminate the target from the distracter sounds on more than 85% of trials when tested in quiet (**Figure 2.1a**; Engineer et al., 2008). As expected, performance could be impaired by adding background noise and the extent of impairment depended primarily on the intensity of the noise. Moderate levels of background noise (i.e. 48 dB SPL) did not significantly impair speech discrimination performance compared

to quiet for most speech sound contrasts ($F_{(1, 66)} = 0.54$, MSE = 13, $p = 0.5$). Only the discrimination of 'shad' from 'chad' was significantly impaired by 48 dB SPL noise (**Figure 2.1b, c**; $p = 0.005$; Tukey *posthoc*). In 60 dB SPL noise, rats were able to perform each of the discrimination tasks at well above chance levels, but performance was significantly impaired compared to quiet ($p = 0.005$; Tukey *posthoc*). In 72 dB SPL noise, discrimination performance on all tasks, except 'dad' versus 'tad', was significantly impaired compared to quiet ($p = 0.005$; Tukey *posthoc*). These results demonstrate that rats can discriminate speech sounds at a minimum signal to noise ratio that is similar to humans (Miller and Nicely, 1955; House et al., 1965; Horii et al., 1971; Wang and Bilger, 1973; Dubno and Levitt, 1981; Phatak et al., 2008).

Behavioral discrimination ability on most tasks was impaired to a similar extent by white noise and speech-shaped noise when the noise intensities were 48 or 60 dB SPL (by 2±1.9% in 48 dB SPL and 9.5±2% in 60 dB SPL noise, respectively). However, in 72 dB SPL noise, speech discrimination ability was significantly more impaired by speech-shaped noise (average percent correct 51±2%) compared to white noise (average percent correct 60±3%) (**Figure 1a**; $F_{(1, 66)} = 10.20$, MSE = 244.02, $p = 0.001$). Rats could discriminate 6 out of 11 tasks significantly above chance in 72 dB SPL white noise and only 3 of 11 tasks in 72 dB SPL speech-shaped noise (**Figures 2.1b, c**; $p = 0.005$, Tukey *posthoc*). The observation that speech-shaped noise generally causes greater impairment in speech discrimination ability than white noise is consistent with psychophysical studies in humans (Busch and Eldredge, 1967; Dubno and Levitt, 1981).

There were significant differences in the degree to which different speech contrasts were impaired by the different noise conditions. In quiet, 'dad' versus 'bad' and 'dad' versus 'gad' were the easiest tasks and 'dad' versus 'tad' and 'shad' versus 'chad' were the hardest

discrimination tasks, consistent with our previous report (Engineer et. al, 2008). In the presence of background noise, ‘dad’ versus ‘bad’ and ‘dad’ versus ‘gad’ continued to be amongst the easiest discrimination tasks; ‘shad’ versus ‘chad’ continued to be the hardest discrimination task (**Figures 2.1b, c**). The ability of rats to discriminate ‘dad’ versus ‘tad’ in white noise, however, was quite robust and had the least impairment in performance as compared to all the other discrimination tasks (**Figures 2.1b, c;** $p = 0.005$, Tukey *posthoc*). For example, behavioral discrimination ability of ‘dad’ versus ‘gad’ fell by 23±3% from quiet to 72 dB SPL white noise, while that of ‘dad’ versus ‘tad’ fell by 6±2%. This result is consistent with results from psychophysical studies in humans, which show that voicing tasks are the most robust to white noise (Miller and Nicely, 1955; Wang and Bilger, 1973).Discrimination performance of ‘dad’ versus ‘tad’ was not as robust in speech-shaped noise and fell by 27±3% when 72 dB SPL speech-shaped noise was added. This result is consistent with previous results in humans which show that voicing discrimination is not as robust in speech-shaped noise as in white noise (Dubno and Levitt, 1981)

Discrimination tasks with ‘dad’ as the target stimulus were significantly more robust to 72 dB SPL white noise compared to discrimination tasks with target stimulus ‘shad’ ($p = 0.01$; Tukey *posthoc*), but were not more robust in 72 dB SPL speech-shaped noise ($p = 0.5$; Tukey *posthoc*). The greater masking of ‘shad’ by white noise is consistent with previous results in humans which show that fricatives and affricates are more sensitive than stop consonants to white noise (Miller and Nicely, 1955; Busch and Eldredge, 1967; Horii et al., 1971; Phatak et al., 2008). The only contrasts that were more impaired in white noise compared to speech-shaped noise were ‘dad’ versus ‘sad’ and ‘shad’ versus ‘chad’ in 48 dB SPL noise and ‘dad’ versus

'gad', 'shad' versus 'tad' and 'shad' versus 'chad' in 60 dB SPL noise (p = 0.005; Tukey *posthoc*). We expected that A1 responses to speech sounds in each of the noise types and intensities would clarify the auditory mechanisms that support robust speech processing in noisy environments.

Neural Responses to speech sounds in background noisy situations:

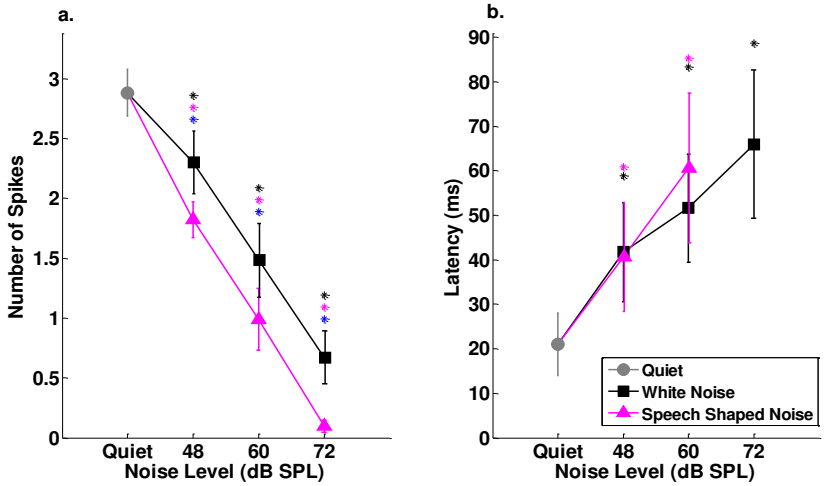

Figure 2.2. Degradation of neural responses in different intensities of white noise and speech-shaped noise.
Neural responses were obtained from 133 primary auditory cortex (A1) sites in 8 anesthetized naïve adult rats. (a) Average number of spikes evoked by all consonant sounds in the first 100 ms. None of the sounds had significantly driven activity in 72 dB SPL speech-shaped noise. (b) Average start latency evoked by all consonant sounds. Since there were no significantly driven spikes in 72 dB SPL speech-shaped noise, there was no latency in this condition. Black and magenta asterisks indicate neural responses in white noise and speech-shaped noise (respectively) are significantly degraded compared to quiet (p = 0.0001; Tukey *post-hoc*). Blue asterisks indicate that neural responses in white noise are significantly less degraded than responses in speech-shaped noise at that intensity (p = 0.0001; Tukey *post-hoc*).

Neural responses were recorded from 133 multiunit clusters of A1 neurons in barbiturate-anesthetized rats. Neural responses were obtained in response to the same 11 consonant sounds behaviorally tested in quiet and in 48, 60 and 72 dB SPL speech-shaped noise and white noise. Responses recorded in silence were consistent with results from previous studies in humans and animals (Steinschneider et al., 1999; Wong and Schreiner, 2003; Steinschneider et al., 2004;

Engineer et al., 2008; Skoe and Kraus, 2010). For example, voiced stop consonants (b/, /d/, /g/) evoked a single burst of activity whereas unvoiced stop consonants (/p/, /t/) resulted in a second peak of activity corresponding to the voicing onset. Stop consonants i.e. /b/, /d/, /g/, /p/, /t/, evoked a greater neural onset response compared to fricatives and affricates i.e. /s/, /sh/, /ch/ (average of 3.0±0.1 spikes versus 2.3±0.4 spikes, p = 0.0001; Tukey *posthoc*; **Figure 2.3**, grey lines).

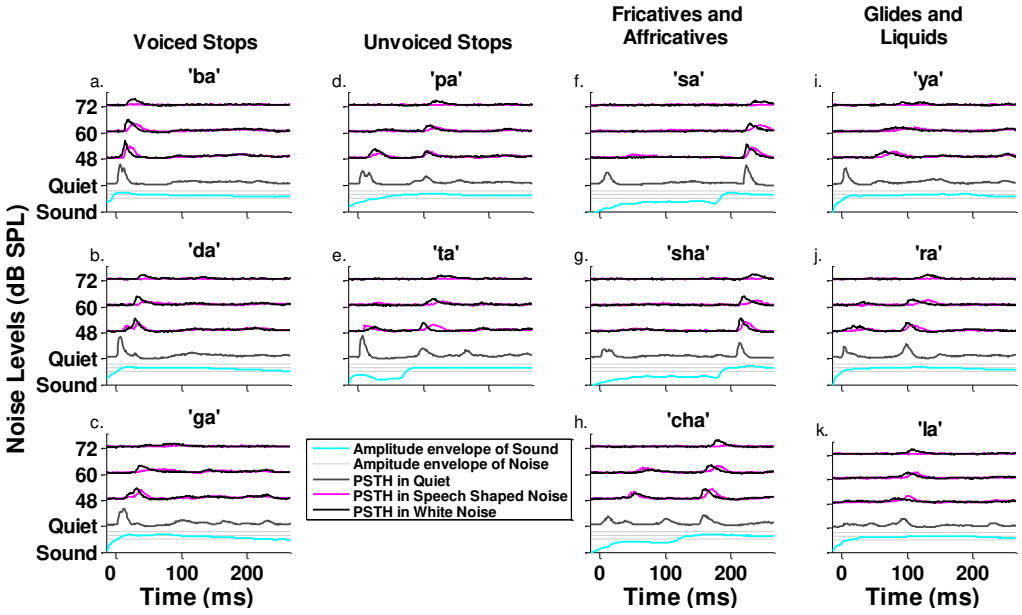

Figure 2.3. Average post-stimulus time histogram (PSTH) responses evoked by each speech sound in three different intensities of white noise and speech-shaped noise.
Sounds are grouped according to manner of articulation. Cyan and light gray lines show amplitude envelope of speech and noise signals. Speech signal was calibrated so that loudest 100 ms is at 60 dB SPL and noises were at 48, 60, 72 dB SPL. In quiet, most sounds, except voiced stop consonants evoke a 2 peaked response, the first peak is evoked by the consonant part of the sound, and the second peak is evoked by the vowel part of the sound. Stop consonants evoked the strongest onset response in quiet. For reference, sound /t/ evoked the strongest neural onset response at 377Hz. Voiced stop consonants (a-c.) were the most robust in all noise conditions (p = 0.0001; Tukey *post-hoc*). Fricatives and affricatives evoked the weakest response in quiet and noise (f-h.). For most sounds neural responses in white noise were significantly more robust than speech-shaped noise of equal intensities (p = 0.0001; Tukey *post-hoc*).

Neural responses to consonant sounds were degraded in the presence of background noise. The amount of degradation of the neural onset response depended on the intensity of the noise, type of speech sound, spectral composition of the noise and the characteristic frequency at each recording site. Increasing the intensity of background noise caused significant degradation of the neural onset response to all consonant sounds. Specifically, increasing the intensity of background noise reduced the total number of spikes and increased both start and end latency of the neural onset response to consonant sounds (**Figure 2.2**; $F_{(3, 396)} = 141.79$, MSE = 5163.12, $p = 0.0001$; $F_{(3, 396)} = 29.14$, MSE = 1686.21, $p = 0.0001$; $F_{(3, 396)} = 15.72$, MSE = 3202.99, $p = 0.0001$; respectively). For example, the average number of spikes evoked by consonant sounds in the first 100 ms was 1.7 ± 0.2 spikes less in 60 dB SPL noise, compared to quiet ($p = 0.0001$; Tukey *posthoc*). The average start and end latency of sounds increased by 35± 14 ms and 62.5± 14.5 ms respectively in 60 dB SPL noise compared to quiet ($p = 0.0001$; Tukey *posthoc*). The neural onset response to most sounds was prominent in 48 dB SPL noise and still present when the background noise and speech signal were of the same intensity (i.e. 60 dB SPL; **Figure 2.3**). The onset response to most speech sounds was eliminated by the presence of 72 dB SPL background noise. The severe reduction of neural activity by 72 dB SPL noise is consistent with our behavioral results that speech discrimination is severely impaired in this noise intensity. Neurophysiology studies in humans show that the cortical responses to some sounds are more degraded by noise than other sounds (Whiting et al., 1998; Martin et al, 1999; Billings et al., 2010). Our recordings in rats also show a significant effect of stimulus on cortical responses in noise (**Figure 2.3**; $F_{(30, 3960)} = 68.67$, MSE = 22133.63, $p = 0.0001$). Voiced stop consonants (/b/, /d/, /g/) were least affected by presence of background noise (**Figure 2.3a-c**; $p = 0.0001$; Tukey

posthoc). The number of spikes in the onset response evoked by voiced stop consonants was significantly greater than all the other consonant groups (i.e. unvoiced stop consonants, fricatives, affricates and glides). For example, sounds /b/, /d/, and /g/ evoked an average of 1.67±0.10 spikes and were significantly above spontaneous activity, even in 72 dB SPL white noise (p = 0.0001; Tukey *posthoc*). On the other hand, the neural onset response to fricative sound /s/ was almost completely eliminated, even in the presence of 48 dB SPL white noise (**Figure 2.3f**; p = 0.0001; Tukey *posthoc*). The onset responses to unvoiced sounds were eliminated in 72 dB SPL white noise. The prominence of neural response for /d/ and absence of neural response for /sh/ in 72 dB SPL white noise supports our behavior result of robust speech discrimination of tasks with target sound /d/. As in previous studies, in both humans and animals, neural responses to vowel sounds were more robust than consonant sounds (Cunningham et al., 2002; Russo et al., 2004; Song et al., 2010). For example, the neural response to vowel onset was present even in 72 dB SPL white noise for most sounds, whereas response to the unvoiced consonant was eliminated (**Figure 2.3d-k**). This robustness is likely due to the fact that vowel sounds are generally louder than consonants (Brinton, 2000).The differential effect of noise on neural responses to different sounds may clarify the greater behavioral sensitivity of certain tasks to noise.

The spectral composition of background noise significantly altered the degree of degradation of the neural response. Speech-shaped noise had a greater impact on most sounds compared to white noise (**Figure 2.3**; $F_{(1, 132)}$ = 391.14, MSE = 60346.02, p = 0.0001). For example, sound /p/ evoked 1.18±0.12 spikes in 60 dB SPL white noise, and 0.27±0.08 spikes in 60 dB SPL speech-shaped noise. The average number of spikes evoked by sounds /b/, /d/, /g/, /p/, /y/, /r/, and /l/ was

reduced by 0.94±0.14 spikes in presence of speech-shaped noise as compared to white noise of equal intensity (60 dB SPL, p = 0.0001; Tukey *posthoc*). The sounds /b/, /d/, and /g/ had significantly driven responses in 72 dB SPL white noise, whereas neural responses to all sounds were indistinguishable from spontaneous activity in 72 dB SPL speech-shaped noise (p = 0.0001; Tukey *posthoc*).These neural results provide support for our behavioral results and also results from previous human psychophysical studies which show that speech-shaped noise is more impairing than white noise sounds for most speech sounds (Busch and Eldredge, 1967; Dubno and Levitt, 1981).

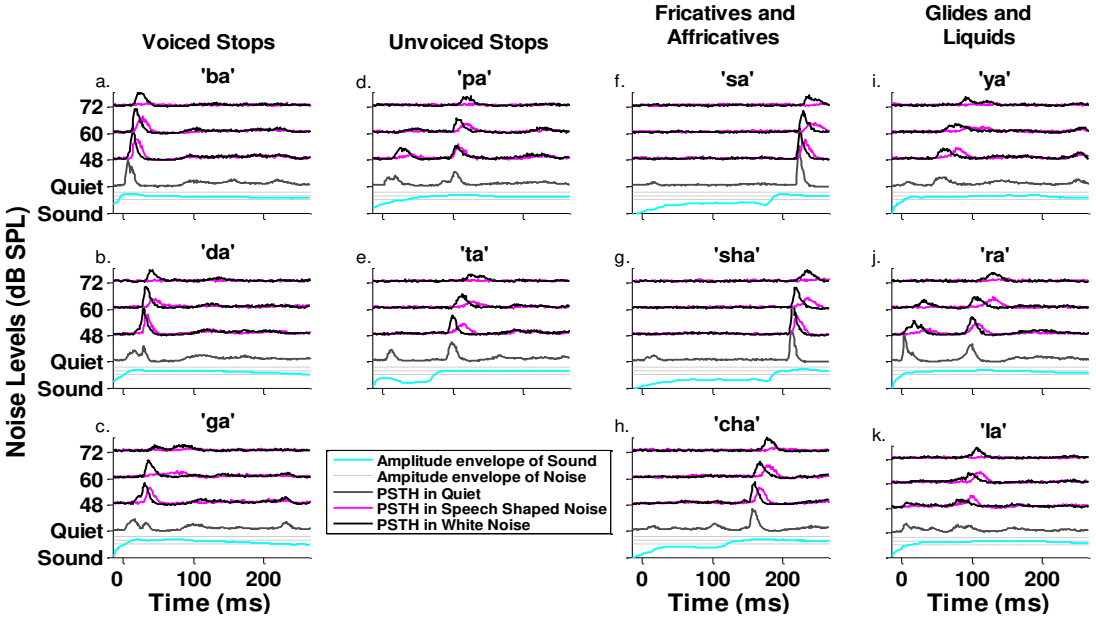

Figure 2.4. Average post-stimulus time histogram (PSTH) responses evoked by different speech sounds in the 45 low frequency A1 sites.
Neural responses in low frequency neurons (characteristic frequency of neurons<4 kHz) are more robust in white noise than speech-shaped noise (p = 0.0001; Tukey *post-hoc*). High frequency sounds (e.g. /t/, /s/, /sh/ and /ch/) evoke weak responses in low frequency neurons even in quiet, which are completely eliminated in 48 dB SPL noise (e, f, g, h).

The neural responses evoked by the sounds /s/, /sh/, /ch/, and /t/, which primarily contain high frequency energy (> 10 kHz), were more degraded in white noise compared to speech-shaped noise. For example, sound /t/ evoked 0.20±0.08 spikes in 60 dB SPL white noise and 0.76±0.01 spike in 60 dB SPL speech-shaped noise. The average number of spikes evoked by /s/, /sh/, /ch/, and /t/, was reduced by 0.31±0.12 spikes in white noise as compared to speech-shaped noise (60 dB SPL, $p = 0.0001$, Tukey *posthoc*). Our neural results show that white noise impairs high frequency speech sounds more than low frequency sounds, as seen in previous human psychophysical studies (Miller and Nicely, 1955; Busch and Eldredge, 1967; Wang and Bilger, 1971; Phatak et al., 2008).

The high spatial resolution of neural responses obtained from our study allowed us to explain the differential effect of noise type by comparing responses of neurons tuned to different frequencies. Although both noise types were presented at the same overall intensities, white noise contains approximately 9 dB more high frequency (>10 kHz) energy compared to speech-shaped noise which contains approximately 9 dB more low frequency (< 4 kHz) energy. As a result, speech-shaped noise degraded responses of low frequency neurons (characteristic frequency < 4 kHz) more than white noise (**Figure 2.4**), while white noise degraded neural responses of high frequency neurons (characteristic frequency >10 kHz) more than speech-shaped noise. The number of spikes evoked in the low frequency neurons by most speech sounds was 1.84±0.26 spikes less in speech-shaped noise than in white noise (60 dB SPL; $p = 0.0001$, Tukey *posthoc;* **Figure 2.4**). The number of spikes evoked in the high frequency neurons by high frequency speech sounds (i.e. /s/, /sh/, /ch/, /t/) was 0.80±0.09 spikes less in white noise than in speech-shaped noise (60 dB SPL; $p = 0.0001$, Tukey *posthoc*). Most speech sounds in our study

contain mostly low frequency energy and evoke greater responses in low frequency neurons. Since speech-shaped noise degraded neural responses in the low frequency region more than white noise, it is not surprising that most speech sounds were degraded to a greater extent by speech-shaped noise than white noise (**Figure 2.4**). The consonant sounds which primarily contain high frequency energy, i.e sounds /s/, /sh/, /ch/, and /t/, evoked the most activity in the high frequency group of neurons. Since white noise degraded neural activity in the high frequency neurons more than speech-shaped noise, it seems reasonable that white noise degraded neural responses of high frequency sounds to a greater extent as compared to speech-shaped noise. Our results are consistent with earlier reports in humans that spectral content of both the speech signal and the background noise influence neural responses to speech sounds (Whiting et al., 1998; Martin et al., 1999; Kozou et al., 2005; Martin and Stapells, 2005; Billings et. al, 2010).

The different noise conditions had similar effects on behavioral and neural responses. We found a high correlation between the average number of spikes evoked by sounds in each discrimination task and the behavioral discrimination ability on that task ($R^2 = 0.63$, $p = 10^{-18}$, **Figure 2.5a**). These results are consistent with EEG and imaging (BOLD) studies in humans which show degradation of behavioral and neural responses proportional to the noise intensity (Whiting et al, 1998; Muller-Gass et al., 2001; Binder et al., 2004).

Our observation that the average number of spikes evoked by speech sounds in noise is correlated with the ability of rats to discriminate between them should not be taken as evidence that spike timing is not important. When we quantified the difference in the number of spikes evoked by pairs of sounds in each speech contrast, we found that the difference was poorly

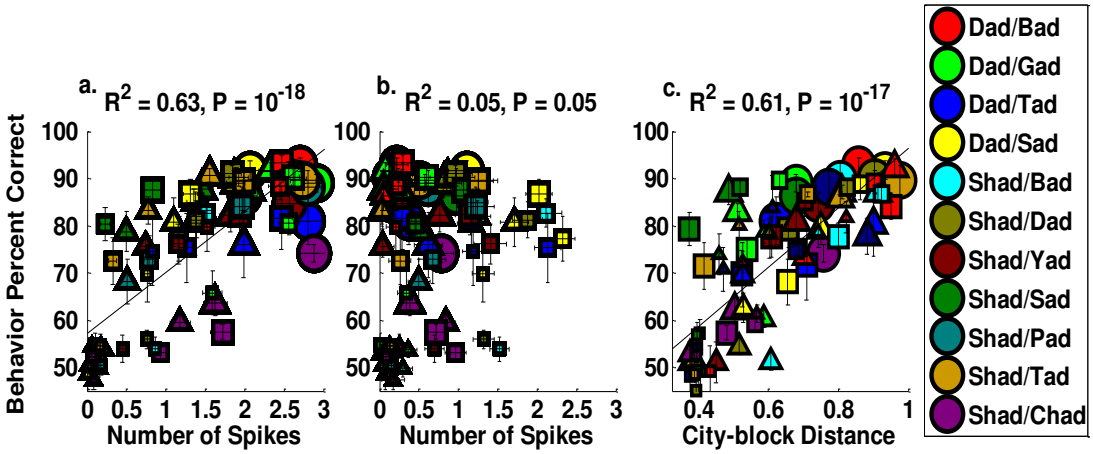

Figure 2.5. Correlation between neural and behavioral responses of all tasks. (a) Average number of spikes evoked by 2 sounds in the discrimination task is well correlated with behavioral discrimination ability. (b) Difference in the average number of spikes evoked by 2 sounds in the discrimination task is poorly correlated with behavioral discrimination ability. Number of spikes was averaged together over the first 100 ms. (c) Normalized City-block distance between the average spatiotemporal patterns evoked by each sound involved in the discrimination tasks were well correlated with behavioral discrimination ability when the neural responses were analyzed over 100 ms duration and binned with 10 ms spike timing precision. Circles, triangles and squares represent performance in quiet, speech shaped noise and white noise, respectively. Size of the symbols indicates noise level, smaller symbols indicating greater amount of noise (i.e. lower signal to noise ratio).

correlated with behavioral discrimination ability ($R^2 = 0.05$, $p = 0.05$, **Figure 2.5b**). This observation is consistent with our earlier report that speech discrimination in quiet is not correlated with the difference in the total number of spikes evoked by each sound at each multiunit recording site (Engineer et al., 2008). In our previous study, speech discrimination in quiet was only correlated with A1 activity when spike timing was used. In this study, we also found that distinctness of neural activity patterns is well correlated with behavioral speech discrimination ability when neural responses were analyzed using spike timing.

While the earlier study only considered the activity evoked within 40 ms of speech onset, in this study we used activity occurring during the first 100 ms duration. This is because background

noise significantly delayed the neural response to speech sounds. Neural activity was best correlated with behavioral discrimination when 100 ms of neural activity was binned with 10 ms spike timing precision ($R^2 = 0.61$, $p = 10^{-17}$; **Figure 5c**). Neural activity was also significantly well correlated when neural responses were analyzed over 70-180 ms duration with 1-30 ms spike timing precision ($R^2 > 0.55$; $P < 10^{-12}$). Correlation between behavioral and neural responses was high ($R^2 > 0.55$) when either City-block distance or Euclidean distance were used to quantify distinctness of neural responses. Although this result extents our earlier report to noisy conditions, this method requires the stimulus onset time to be known. In the following section we will explain the disadvantage of specifying the stimulus onset time and explain a neural analysis method that does not require the stimulus onset to be specified.

Quantifying Neural Discrimination without Reference to Sound Onset

Previous studies of neural coding using spike timing strategies generally assume that the decoding mechanism knows the precise stimulus start time (Gawne et al., 1996; Furukawa et. al, 2000; Ahissar et al., 2001; Panzeri et al., 2001; Schnupp, et. al 2006; Wang et al., 2007; Foffani et al., 2004, 2008; Engineer et. al, 2008). The stimulus start time is often calculated from the average response of a large group of neurons (Chase and Young, 2007; Engineer et. al, 2008). Although it is reasonable to expect that stimulus onset time information is available in quiet situations, loud background noise significantly degrades neural responses making it difficult to determine the stimulus onset time even based on the average activity of many neurons. Therefore in this study, we developed a new form of neural classifier that is able to determine which speech sound was presented by analyzing neural activity without precise knowledge of stimulus onset

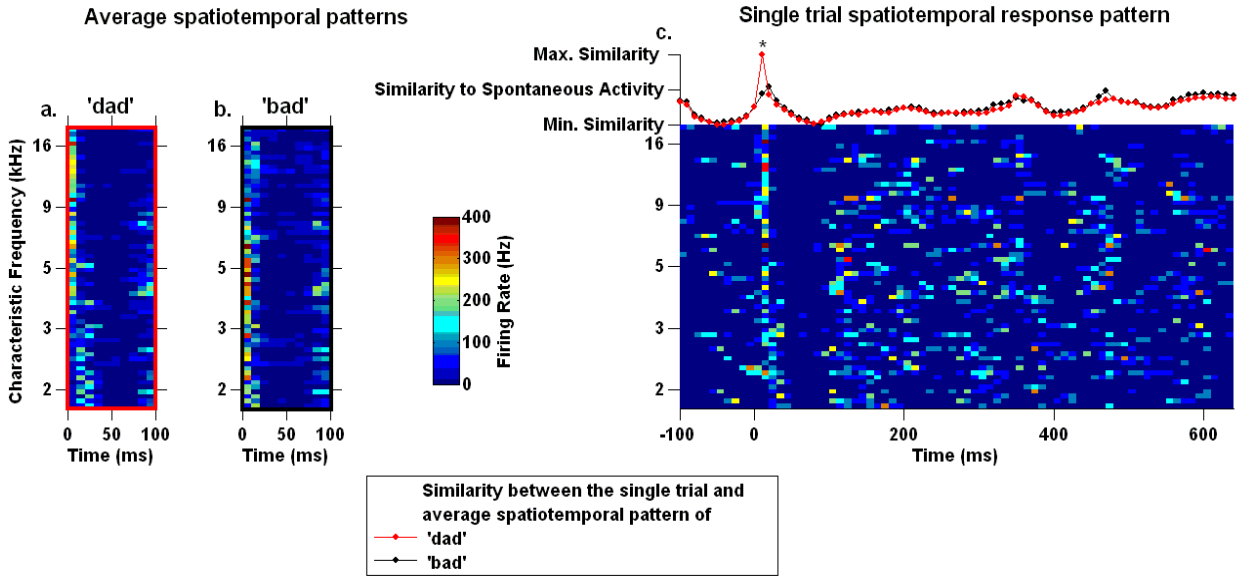

Figure 2.6. Consonant discrimination using relative spike timing.
This figure illustrates how the classifier evaluates which of two speech sounds ('dad' or 'bad' in this case) was presented from a single trial of neural activity recorded at 60 A1 sites. (a, b) In this example, the average spatiotemporal patterns that the classifier was looking for were 100 ms in duration and binned with 10 ms spike timing precision. This classifier is referred to as $\text{Classifier}_{100/10\text{ms}}$. Neural activity from different A1 sites is arranged according to the characteristic frequency (CF) of each site. (c) Single trial spatiotemporal activity from the same 60 sites is shown from 100 ms before the stimulus onset to 150 ms after the stimulus end. The similarity of the single trial to each of the average activity patterns is shown as red and black lines at the top of the figure, each point indicating the different possible stimulus start times. The point of greatest similarity occurred immediately after the word 'dad' was presented (asterisk). The classifier correctly identified that 'dad' had been presented, because the similarity of the single trial was highest to the average pattern generated by 'dad'. Neural discrimination performance is determined by calculating the percentage of trials on which the classifier correctly guessed the presented sound (see Methods). The set of A1 sites shown in this example was able to correctly identify whether 'dad' or 'bad' was presented on 39 out of 40 presentations in quiet (97.5% correct).

time. We compared the classifier performance with behavioral speech discrimination ability to determine the neural analysis methods that are correlated with behavior.

In brief, the neural classifier examines activity from a single trial collected from a set of A1 neurons and attempts to identify which of two possible sounds was presented by looking for the

spatiotemporal activity patterns generated by each sound (see Methods). Neural discrimination ability is determined by calculating the percentage of trials on which the classifier correctly guessed the speech sounds.

We first tested whether the new classifier could mimic the behavioral discrimination ability on the eleven consonant tasks reported in our earlier study (Engineer et al, 2008). Using single trial data from groups of sixty A1 sites, the classifier was able to discriminate 'dad' from 'bad' in quiet 99±1% of the time, which is comparable to behavioral discrimination ability. Classifier performance on the eleven tasks was highly correlated with behavioral discrimination ability in quiet when the average spatiotemporal patterns were composed of 100 ms of neural activity binned with 10 ms precision ($R^2 = 0.75$, $p = 0.0005$). This correlation was similar to the correlation observed using the classifier from our previous study, which was given the exact start time of each stimulus ($R^2 = 0.66$, $p = 0.005$; Engineer et al., 2008).

The new classifier required neural activity from at least twenty-five recording sites to achieve performance that was comparable to the old classifier using one site. Performance of the new classifier was almost at chance performance when only one A1 site was provided (51±0%). This poor performance when given neural data from only one recording site is because the classifier could not distinguish spontaneous activity from driven activity without knowledge of the stimulus onset time. When the spatiotemporal activity patterns included many sites, the new classifier was able to reliably discriminate between speech sounds using data from a single trial. Using a large number of sites provides the classifier with additional information about the spatial pattern of activity which is not available when only one recording site is used. When neural data from large number of sites were analyzed together, classifier performance improves because the

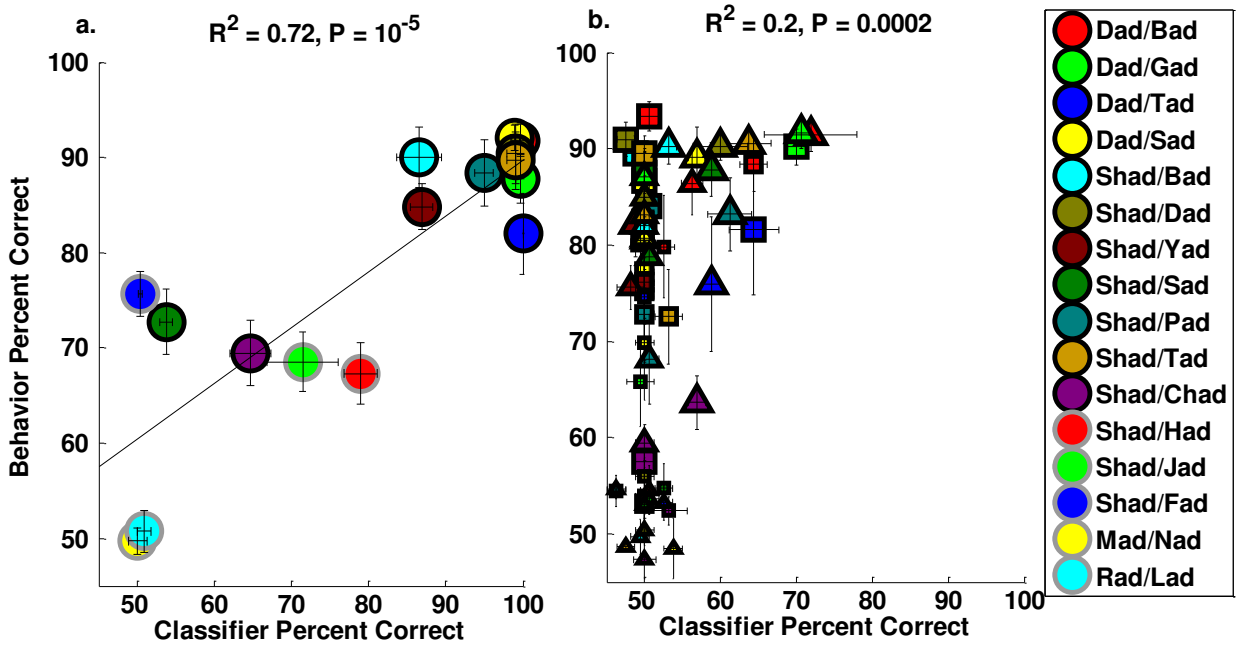

Figure 2.7. Correlation between Classifier$_{100/10ms}$ performance and behavioral discrimination ability in quiet and noise.
(a) Classifier performance was significantly well correlated with behavioral discrimination ability in quiet, when average spatiotemporal patterns were analyzed using 100 ms duration and binned with 10 ms precision. (b) In noise, classifier performance analyzed using the same parameters was poorly correlated with behavioral discrimination performance. Symbols with black border are the tasks used in this study. Symbols with gray border are extra tasks used from our previous study (Engineer et. al, 2008; see methods). Circles, triangles and squares represent performance in quiet, speech shaped noise and white noise, respectively Size of the symbols indicates noise level, smaller symbols indicating greater amount of noise (i.e. lower signal to noise ratio).

spatiotemporal pattern of driven activity across sites was distinct from spontaneous activity patterns. The sound 'dad', for example, causes high frequency neurons to fire 5-10 ms before low frequency neurons, while the sound 'bad' causes the neurons to fire in the opposite order (**Figure 2.6a, b**). When activity from many sites was analyzed, this pattern could be easily distinguished from spontaneous activity occurring during the 750 ms analysis window (**Figure 2.6c**). This analysis (using published behavioral and neural responses) suggests that it is not necessary to know the precise stimulus onset time to decode A1 patterns.

In the current study, we tested five additional consonant discrimination tasks that were not included in our earlier study (Engineer et al, 2008), resulting in a total of 16 consonant discrimination tasks tested in quiet. Behavioral and neural discrimination from each data set were averaged together for the six tasks that were examined in both studies. We observed a high correlation between behavior and classifier performance ($R^2 = 0.72$, $p = 10^{-5}$**; Figue 2.7a**) when the classifier analyzed neural activity using 100 ms of activity binned with 10 ms precision. The correlation between neural and behavioral discrimination was also high when the average patterns had a duration of 70-150 ms and were binned with 5-20 ms precision ($R^2>0.55$, $P < 0.0001$). For future reference, we will refer to the classifier that analyzes activity from 60 sites using 100 ms durations binned with 10 ms precision as $Classifier_{100/10ms}$. In our previous study in quiet, spike timing information was well correlated with behavioral discrimination ability and average spike count was not correlated with behavioral consonant discrimination ability. Neural responses analyzed using our new classifier also were very poorly correlated with behavioral consonant discrimination ability on the 16 tasks when average spike count over 100 ms duration was used ($R^2 = 0.11$, $p = 0.3$). These results support the hypothesis from our earlier study that spike timing information is required to discriminate between consonant sounds in quiet conditions.

After confirming that the new classifier was correlated with behavioral performance in quiet, we tested whether the classifier performance was correlated with speech discrimination in noise (n=11 tasks). $Classifier_{100/10ms}$ was unable to discriminate between speech sounds presented in even moderate noise. For example, $Classifier_{100/10ms}$ discrimination of ‘dad’ versus ‘bad’ in 48

dB SPL white noise was at chance (**Figure 2.7b**). Behavioral discrimination performance in this condition was at 91±1%. This discrepancy between neural and behavioral discrimination performance was seen across all noise conditions and discrimination tasks, leading to a poor correlation between behavior and Classifier$_{100/10ms}$ performance in noise (R^2 = 0.2, **Figure 2.7b**). None of the spike timing analysis ranges which could predict behavior discrimination ability in quiet could predict behavioral discrimination ability in noise. This result led us to hypothesize that neural responses in noise are analyzed differently than in quiet.

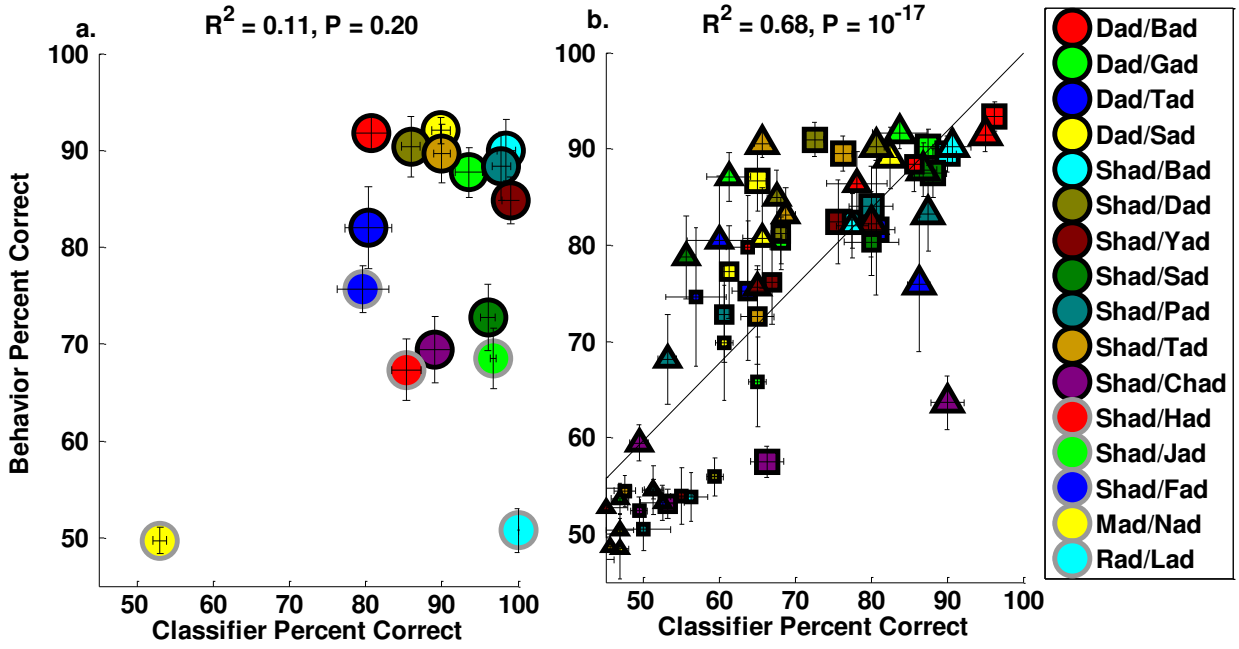

Figure 2.8. Correlation between Classifier$_{400/60ms}$ performance and behavioral discrimination ability in quiet and noise.
(a) In quiet, classifier performance analyzed using 400 ms duration binned with 60 ms precision was poorly correlated with behavioral discrimination performance. (b) In noise, classifier performance was significantly well correlated with behavioral discrimination ability, when the average spatiotemporal patterns were analyzed using 400 ms duration binned with 60 ms precision. Symbols with black border are the tasks used in this study. Symbols with gray border are extra tasks used from our previous study (Engineer et. al, 2008; see methods). Circles, triangles and squares represent perfromance in quiet, speech shaped noise and white noise, respectively. Size of the symbols indicates noise level, smaller symbols indicating greater amount of noise (i.e. lower signal to noise ratio).

Previous studies in the visual and auditory systems suggested that the brain uses longer neural integration windows when stimuli are presented in low signal to noise situations (Roitman and Shadlen, 2002; Binder et. al, 2004; Huk and Shadlen, 2005). We tested whether a classifier with longer durations of activity (for example, 400 ms instead of 100 ms) binned with coarser precision (for example, 60 ms instead of 10 ms) would compensate for the delayed and degraded speech responses in noise and result in a better correlation with behavior. Classifier performance was best correlated with behavioral performance on all tasks in the different noise conditions when the average spatiotemporal patterns from 60 A1 sites were analyzed using a duration of 400 ms and binned with temporal precision of 60 ms ($R^2 = 0.68$, $p = 10^{-17}$; **Figure 2.8b**). For example, classifier performance on 'dad' versus 'bad' in 48 dB SPL white noise was at 96±1%. For future reference, we refer to the classifier that analyzes activity from 60 sites using 400 ms durations binned with 60 ms precision as $Classifier_{400/60ms}$. In 48 to 60 dB SPL noise, $Classifier_{400/60ms}$ performance was significantly better on all eleven tasks compared to $Classifier_{100/10ms}$ performance (**Figure 2.7b and Figure 2.8b**; $F_{(1,10)} = 12.95$, MSE = 0.11, $p = 0.001$). $Classifier_{400/60ms}$ performance remained close to chance for most tasks in 72 dB SPL noise ($p = 0.005$ for 16 of 22 tasks), which is similar to the behavior discrimination performance. $Classifier_{400/60ms}$ performance was well correlated with behavioral performance in noise when neural responses were analyzed using 25 or more sites together ($R^2>0.55$). Classifier performance was well correlated with behavioral performance in noise provided that neural activity was analyzed over 300-600 ms durations and binned with 50-100 ms temporal precision ($R^2>0.55$, $P < 10^{-10}$). None of the analysis ranges that were well correlated with behavioral discrimination ability in noise were well correlated with behavioral discrimination ability in

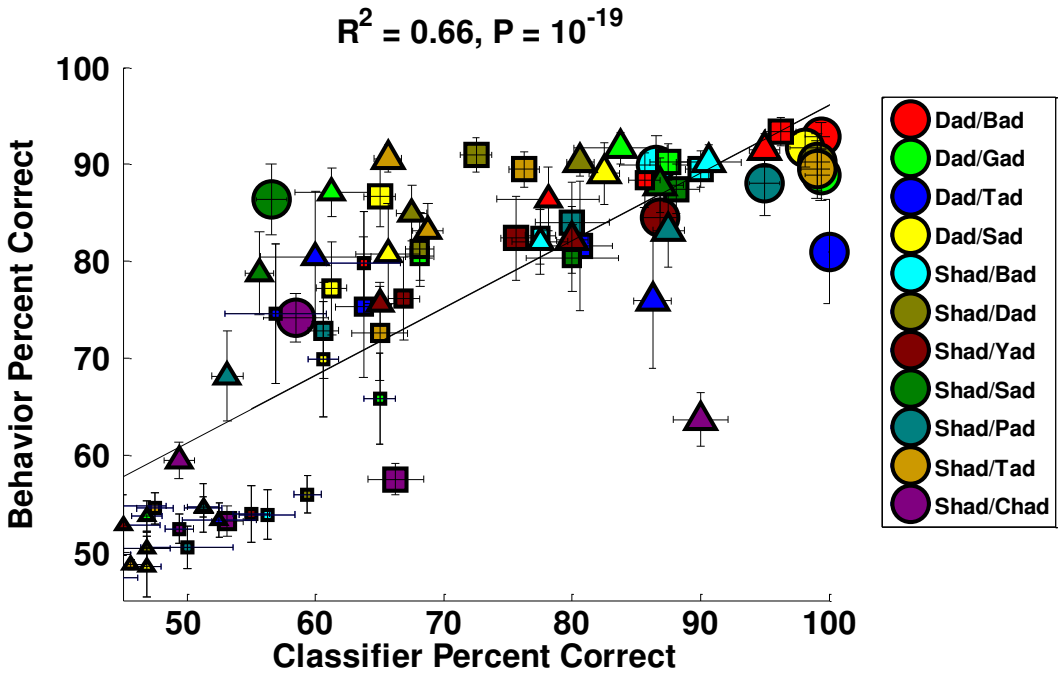

Figure 2.9. Single trial responses analyzed using longer integration timescales as compared to quiet were well correlated with behavioral discrimination ability in all conditions.
Hybrid classifier performance when the average spatiotemporal patterns in quiet were analyzed using 100 ms duration and binned with 10 ms precision and the average spatiotemporal patterns in noise were analyzed using 400 ms duration and binned with 60 ms precision are correlated with behavioral discrimination ability in quiet all conditions. Circles, triangles and squares represent performance in quiet, speech shaped noise and white noise, respectively. Size of the symbols indicates noise level, smaller symbols indicating greater amount of noise (i.e. lower signal to noise ratio).

quiet. For example, $\text{Classifier}_{400/60\text{ms}}$ was poorly correlated with behavior in quiet because of a ceiling effect observed for many tasks ($R^2 = 0.11$, $p = 0.2$; **Figure 2.8a**). Our observations that 1) neural responses analyzed over small integration windows can predict behavior in quiet but not in noise and 2) responses analyzed using longer integration windows can predict behavior in noise but not in quiet are consistent with earlier proposals that neurons use greater temporal integration in noisy conditions (Roitman and Shadlen, 2002; Binder et. al, 2004; Huk and Shadlen, 2005).

A hybrid classifier that uses the 100/10 ms parameters in quiet and the 400/60 ms parameters in noise was highly correlated with behavior under all conditions tested ($R^2 = 0.67$, $p = 10^{-19}$; **Figure 2.9**). Hybrid classifier performance was well correlated with behavior when neural responses of more than 25 sites were grouped together ($R^2>0.55$, $P < 10^{-11}$; **Figure 2.10a**). Classifier accuracy was closest to behavioral performance (**Figure 2.1a**) when it was provided with neural activity recorded from sixty sites (**Figure 2.10b**). Classifier performance was well correlated with behavioral discrimination ability when the average patterns in quiet were analyzed using 70-150 ms durations and binned with 5-20 ms precision and when the average patterns in noise were analyzed using 300-600 ms durations and binned with 50-100 ms precision ($R^2>0.55$, $P < 10^{-11}$). These results support our hypothesis that neural responses can explain behavioral discrimination provided they are analyzed over longer integration periods and with lesser temporal precision in noise.

Although the prediction of the hybrid classifier could be off by as much as 30% for a specific task in a specific noise condition (i.e. 'shad' vs. 'chad' in 48 dB speech shaped noise), the hybrid classifier was highly accurate in predicting the average performance of rats across eleven tasks in different noise conditions ($R^2 = 0.89$, $p = 0.002$, **Figure 2.11a**). If relative spike timing information was not used (i.e. spike count only), the hybrid classifier was not able to match this level of correlation regardless of the number of sites or duration of activity examined ($R^2<0.35$).For example, Classifier performance on most tasks was poorly correlated with behavior discrimination ability even when noise dependent integration windows (i.e. 100 ms in quiet and 400 ms in noise) for spike count were used ($R^2 = 0.03$). Classifier performance in both quiet and noise was poorly correlated with behavior discrimination ability when spike count over

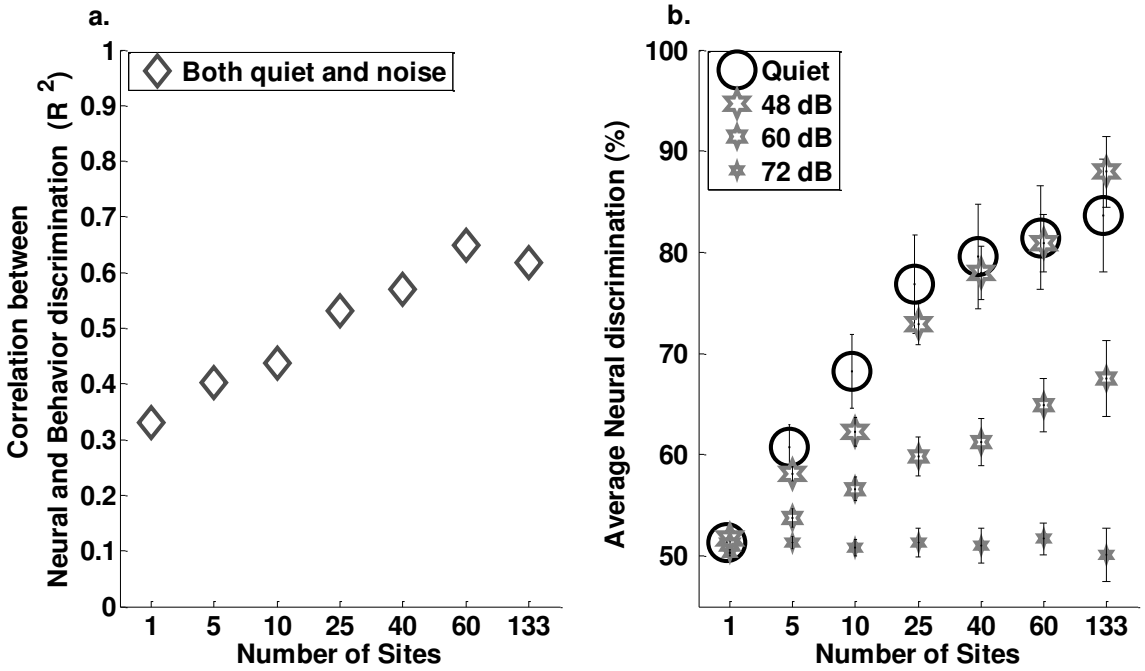

Figure 2.10. Average neural discrimination and correlation between the hybrid classifier and behavioral performance as a function of sites
(a) Correlation (R^2) between neural and behavioral discrimination performance improved significantly ($F_{(1, 10)} = 12.95$, MSE = 0.11, $p = 0.001$) as a function of sites. Neural discrimination ability in quiet and noise were significantly well correlated with behavior when neural responses from more than 25 sites were analyzed together ($R^2 > 0.55$; $P < 10^{-11}$). Neural discrimination from 60 multiunit clusters was best correlated with behavior ($R^2 = 0.68$, $p = 10^{-17}$). (b) Average neural discrimination ability of all tasks in quiet, 48dB SPL and 60 dB SPL noise increased as a function of sites and was comparable to behavioral discrimination ability. Increasing the number of sites did not affect neural discrimination ability in 72 dB SPL noise which is also comparable with behavioral discrimination ability at this noise level.

the entire duration of the sound was used ($R^2 = 0.1$). Collectively, our results suggest spike timing information is required to explain behavioral speech discrimination ability and that the precision of the spike timing information depends on whether significant background noise is present.

2.6 DISCUSSION

Our study provides the first evidence that rats, like humans, can discriminate between most speech contrasts in high levels of background noise. Neural responses to speech in noise in the rat were similar to neural responses recorded in humans. The observations that behavioral and neural responses are similar for rats and humans supports the hypothesis that human speech sound processing employs the same basic neural mechanisms as other mammals. More specifically, our results suggest that rapid speech sound transitions are encoded in spatiotemporal activity patterns at least up to the level of A1, even without explicitly providing the stimulus onset time to the neural classifier. These responses are likely decoded using longer temporal integration windows in noisy conditions compared to quiet conditions. Collectively, our results

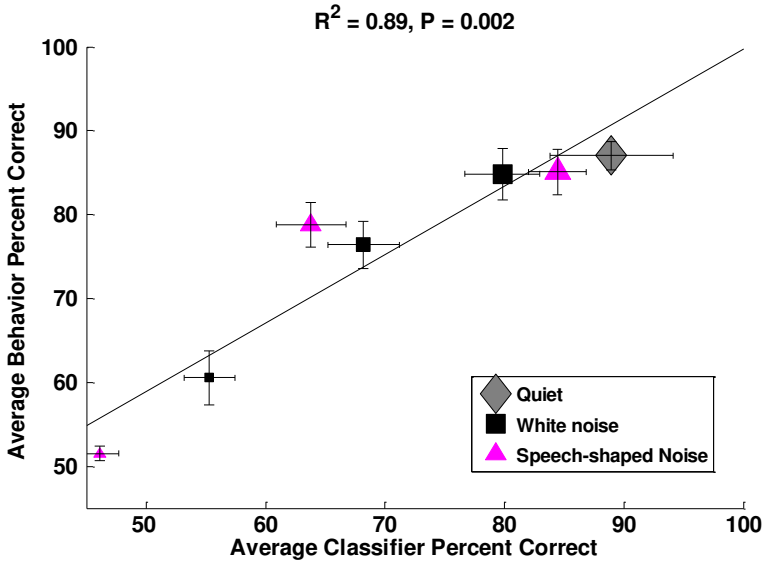

Figure 2.11. Hybrid classifier percent correct can explain variance of behavioral discrimination caused by either noise conditions or discrimination tasks alone.
The hybrid classifier can explain 89% of the variance (p = 0.002) caused by the seven noise conditions when the classifier performance is averaged across all eleven tasks. Size of the symbols indicates noise level, smaller symbols indicating greater amount of noise (i.e. lower signal to noise ratio).

provide support for the development of animal models to better understand speech processing and speech processing impairments in humans

Comparison with human psychophysical and human neurophysiology literature:

Our behavioral results reveal a number of important similarities between human and rat speech discrimination ability in noise. Normal hearing humans can discriminate between consonant vowel stimuli even when speech and noise are of equal intensity (Miller and Nicely, 1955; House et al., 1965; Horii et al., 1971; Wang and Bilger, 1973; Dubno and Levitt, 1981; Phatak et al., 2008). Our study shows that rats can also accurately discriminate between speech sounds even when speech and noise are of equal intensity. The average behavioral discrimination ability of rats is still significantly above chance in -12 dB signal to noise ratio (SNR) i.e. 72 dB SPL white noise. This result is similar to human psychophysics studies which show almost chance performance in white noise between -10 and -16 dB SNRs (Wang and Bilger, 1973; Miller and Nicely, 1955; Phatak et al., 2008).

Although the speech discrimination threshold in speech-shaped noise appears to be lower in rats than in humans, methodological differences can account for the apparent difference. The average speech discrimination ability of rats was at chance in -12 dB SNR speech-shaped noise, while human performance does not fall to chance in speech-shaped noise until -20 dB SNR. Two factors could account for the 8 dB difference in threshold. First, due to a subtle difference in how intensity is defined, our consonant sounds were approximately 4 dB quieter than most human studies. Intensity is traditionally defined in human studies as the average intensity throughout the duration of the stimulus. In our studies, speech sound intensity was quantified based on the most intense 100 ms (Engineer et al., 2008). Second, subjects in the human studies were provided

additional information that is expected to further lower the threshold. In Phatak and Allen (2007), the subjects were allowed to repeat the speech stimuli as many times before responding. Repetition of stimuli provides approximately 2.5 dB advantage in noise (Miller et al., 1951). In House et al., 1965, the speech stimuli had semantic meaning associated with them which provides additional information that can aid identification in noise. Native speakers have approximately 6 dB advantage over non-native speakers in noise when words have linguistic meaning associated with them (Cooke et al., 2007). The rats in our study did not have either form of additional information available to them. After accounting for these methodological differences, the performance of rats in speech-shaped noise is found to be similar to human performance. The similarity of rat and human speech discrimination thresholds supports proposals that the early stages of speech sound processing employ the basic auditory processing mechanisms common to all mammals (Kuhl and Miller 1975; Merzenich et al., 1993; Tallal et al., 1993; Cunningham et al., 2002; Mesgarani et al., 2008).

Human psychophysical studies have observed that certain speech sounds are more easily masked by continuous noise than others and that some types of noise mask certain sounds more than others. These differences provide another opportunity to evaluate the similarity of speech discrimination in rats and humans. The hypothesis that rats and humans employ similar brain mechanisms would be strengthened if rats exhibit similar impairments in noise as humans (like similar threshold of noise intensity where speech sounds were still intelligible). Our behavioral results in rats confirm the following key findings from human psychophysics studies. In both rats and humans, most speech sounds are more impaired in speech-shaped noise compared to white noise (Busch and Eldrege, 1967; Dubno and Levitt, 1981) and high frequency sounds are much

more impaired in white noise as compared to speech-shaped noise (Miller and Nicely, 1954; Wang and Bilger, 1971). Voicing contrasts are much more robust than other speech contrasts in white noise, whereas their impairment is very similar to other contrasts in presence of speech-shaped noise in both rats and humans (Miller and Nicely, 1954; Wang and Bilger, 1971; Dubno and Levitt, 1981). Our neuro-physiological results reveal a number of important similarities between neural responses to speech in noise recorded in rats and humans. The addition of background noise increases the latency and decreases the amplitude of consonant responses in both rats and humans (Martin et al., 1999; Whiting, et al., 1998; Martin and Stapells, 2005). Our recordings mimic human imaging results which show that the neural representation of vowel sounds is more robust to noise compared to consonant sounds (Russo et al., 2004; Song et al., 2010). The higher resolution view of auditory cortex responses provided by microelectrode recordings in rats provides a possible neurobiological explanation for the behavioral observation that the spectral content of both speech and noise influence speech discriminability (Miller and Nicely, 1954; Busch and Eldrege, 1967; Dubno and Levitt, 1981; Phatak et al., 2008). The similarity of behavioral and neural responses between rats and humans suggests that the rat is a reasonable animal model of speech sound processing and that the rat model could be useful in understanding the greater noise sensitivity of certain clinical populations, including individuals with hearing loss and learning impairments (Cunningham et al., 2002; Ziegler et al., 2005; Ziegler et al., 2009; Anderson et al., 2010).

Comparison with neural decoding literature:

An extensive literature on neural decoding has shown that spatiotemporal patterns in the cortex provide significant information about the sensory world (Abeles et al., 1996; Villa et al.,

1999; Butts et al., 2007; Kayser et al., 2009; Huetz et al., 2010). Although the amount of information is almost invariably greater when spike timing is included in the analysis, there is relatively little behavioral evidence that spike timing information can predict behavior (Reinagel and Reid, 2000; Panzeri et al., 2001; Foffani et al., 2004; Butts et. al, 2007; Panzeri and Diamond 2010; Huetz et al., 2010). There is a high correlation between behavioral consonant discrimination ability in quiet and the distinctness of cortical activity patterns, but only when spike timing information is included (Engineer et al., 2008). This result differed from earlier studies in primates showing that cortical decoding based on spike count was best correlated with behavior (Romo and Salinas, 2003; Liu and Newsome, 2003; Lemus et al., 2009). The most likely explanation for the apparent contradiction is that the primate studies used continuous or periodic stimuli that lack the kind of temporal transitions that are present in consonant stimuli. In this study, we tested consonant sounds embedded in background noise to determine the effect of degrading stimulus onset timing on behavioral and neural discrimination. Because background noise makes it harder to determine when the sounds occur, we developed a method of decoding neural responses without providing the classifier with the stimulus onset time. Our results provide the first behavioral evidence that a code based on relative spike timing can account for behavior over a wide range of conditions without the need for precise knowledge of stimulus start time. The observation that neural decoding is well correlated with behavior provides strong support for the long held hypothesis that sensory scenes are encoded in highly distributed spatiotemporal activity patterns (Abeles et al., 1996; Villa et al., 1999; Butts et al., 2007; Kayser et al., 2009; Huetz et al., 2010).

The precision of spike timing needed to accurately explain consonant discrimination behavior seems to depend on whether the sounds are presented in quiet or noise. Neural activity analyzed with longer temporal integration compared to quiet is necessary to account for consonant discrimination in noise. This result is consistent with previous reports in the visual and auditory systems which indicate neural responses are averaged over longer durations in low signal to noise ratio situations (Roitman and Shadlen, 2002; Huk and Shadlen, 2005; Binder et al., 2004). We know of no study that directly related neural discrimination in cortex based on different levels of temporal integration with behavioral discrimination in high and low noise situations. Our combined neural classifier model uses spike timing activity integrated over a longer window in noise and results in neural discrimination that is comparable to observed behavior discrimination.

CHAPTER 3

PAIRING TONE TRAINS WITH VAGUS NERVE STIMUMATION INDUCES TEMPORAL PLASTICITY IN THE PRIMARY AUDITORY CORTEX

Jai A. Shetake, Navzer D. Engineer, Will A. Vrana, Jordan T. Wolf and Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

3.1 ACKNOWLEDGEMENTS

The authors would like to thank K. Ram and R. Cheung for their assistance with microelectrode mappings. We would like to thank S. Sudanagantha, M. Fink, H. Rasul, D. Vuppalla, A. Kuzu, C. Walker, C. Omana and M. Borland for their help with daily VNS pairing sessions.

3.2 ABSTRACT

The selectivity of neurons in sensory cortex can be modified by pairing neuromodulator release with sensory stimulation. Repeatedly pairing electrical stimulation of the nucleus basalis (NB) with a tone increases the number of neurons in primary auditory cortex (A1) that respond to the paired tone frequency. Pairing NB stimulation with tone trains alters the ability of neurons in A1 to respond to rapidly presented tones. Pairing vagus nerve stimulation (VNS) with a single tone altered spectral tuning in the same way as NB-tone pairing. In this study, we tested whether pairing VNS with temporally modulated acoustic stimuli can change the temporal response properties of A1 neurons. Rat A1 neurons respond strongly to tone trains at repetition rates up till 10 pulses per second (pps). VNS paired with 15 pps tone trains increased the temporal following capacity of A1 neurons and pairing VNS with 5 pps tone trains decreased the temporal following capacity of A1 neurons. Pairing of these temporal modulations with VNS did not affect the frequency selectivity or tonotopic organization of auditory cortex neurons. VNS is well tolerated in patients. As a result, VNS-tone pairing represents viable method to direct temporal plasticity in a variety of conditions associated with temporal processing deficits.

3.3 INTRODUCTION

Acoustic experience can modify response properties of neurons in the auditory cortex. Neural response properties can change depending on the relevance of acoustic stimuli used for behavior training (Recanzone et al., 1993; Bao & Merzenich, 2004; Polley et al., 2006). For example, pairing fast tone trains with food reward increases the temporal following capacity of auditory cortex neurons. Behavioral relevance of either intensity or frequency features causes frequency specific or intensity specific plasticity without affecting other spatially organized feature representations. Pairing electrical stimulation of the cholinergic nucleus basalis (NB) with acoustic stimuli also causes stimulus feature specific plasticity (Kilgard & Merzenich 1998a, Kilgard & Merzenich 1998b; Kilgard et al., 2001). For example, pairing fast or slow tone trains with NB stimulation can increase or decrease the maximum following rate of auditory cortex (A1) neurons, respectively. These studies show that specific features of acoustic stimuli are responsible for directing cortical plasticity.

A recent study showed that vagus nerve stimulation (VNS) paired with tones of a particular tone frequency can induce map plasticity specific to the paired frequency (Engineer et. al, 2011). The degree and direction of spatial plasticity induced by VNS tone pairing parallels the spatial plasticity with NB stimulation tone pairing. For example, as with NB stimulation tone pairing, VNS tone pairing causes an increase in the percent of A1 neurons responding to the paired tone and decrease in the percent of neurons responding to the unpaired tones. If similar neural mechanisms are responsible for driving plasticity observed with VNS and NB stimulation, then VNS should be capable of inducing stimulus feature specific plasticity as observed with NB stimulation. In this study, we tested whether pairing VNS with rapid tone trains can increase the

temporal following rate of A1 neurons. We paired VNS with tone trains of random carrier frequencies but a fixed repetition rate. The maximum following rate of many rat A1 neurons is approximately 10 pulses per second. We hypothesized that pairing VNS with 15 pps tone trains would increase the maximum following rate of A1 neurons and pairing VNS with 5 pps tone trains would decrease the maximum following rate of A1 neurons.

3.4 MATERIALS AND METHODS

Subjects and Experimental groups:

Nineteen female Sprague Dawley rats were used in these experiments. The rats weighed an average of 286 ± 21 g and were housed in 12:12 hr reversed light cycle at a constant humidity and temperature. Thirteen rats were randomly assigned to experimental groups which were implanted with vagus nerve stimulators and paired with tone trains of either 15 pps (n=7) or 5 pps (n=6). The remaining rats (n=6) were used as naïve controls. All handling, housing, and surgery and testing of the animals was approved by the University of Texas Institutional Animal Care and Use Committee.

Vagus Nerve Cuff Implantation Surgery:

Custom-made cuff electrodes were implanted on the left vagus nerve of the experimental rats. Construction of this electrode and surgical procedures are the same as described in our previous study (Engineer et al., 2011). In brief, the cuff was 4mm long and made from Micro-Renethane tubing (8 mm diameter). The bipolar stimulating electrodes were made from Teflon coated multi-stranded platinum-iridium wires and placed 2 mm apart inside the cuff. An 8 mm region of the Teflon coating was stripped to provide contact with the vagus nerve. A cut was

made lengthwise along the tubing to allow the cuff to be wrapped around the nerve and then closed with silk threads.

The surgery was conducted under pentobarbital anesthesia (50 mg/kg) and a state of areflexia was maintained throughout the experiment with supplemental doses of dilute pentobarbital (0.2– 0.5 ml; 8 mg/ml). Anesthesia depth was monitored by heart rate, breathing rate, corneal reflexes, and response to toe pinch. Nourishment was provided using 1:1 mixture of dextrose (5%) and standard Ringer's lactate solution and body temperature was maintained at 37 degrees C. Doses of cefotaxime sodium (2 x 10 mg, s.c.), 2: 3 mixture of atropine and dexamethazone (4mg/kg) s.c solution were given to the rats before and after the surgery to prevent infection and reduce fluid accumulation in lungs. Doses of dextrose/Ringer's (10 x 1 ml total) were given throughout and after the surgery to provide nourishment. An incision was made to expose the lambda landmark on the skull. Four to five bone screws were drilled into the skull at points close to the lambda suture and over the cerebellum. These screws provided an anchor for a custom made skull cap which contained 4 electrodes. An incision and blunt dissection of the muscles in the neck exposed the left cervical branch of the vagus nerve. Two of these electrodes were used to deliver current to stimulating electrode, one for monitor EEGs and one for electrical grounding during the daily VNS sessions. The leads from cuff electrodes were tunneled under the skin and attached to the skull cap. All incisions were sutured and the exposed two-channel connector encapsulated in acrylic. A topical antibiotic cream was applied to both incision sites. Immediately after the surgery, rats were given amoxicillin (5 mg) and carprofen (1 mg) to prevent infection and facilitate recovery.

VNS stimulation and tone-train pairing:

After a week of recovery from surgery, tone-train stimuli were paired with VNS. Rats were un-anaesthetized, unrestrained and were placed in a 25 cm X 25 cm X 25 cm cage which was inside a 50 cm X 60 cm X 70 cm sound- shielded booth with 3 in acoustic foam lining. The speaker was placed inside the sound shielded booth and was 20 cm above the cage. VNS stimulation parameters used were similar those used in Engineer et al. (2011) and were identical for each rat in the study.

The acoustic tone-train stimuli presented in this study were similar to stimuli presented in previous nucleus basalis stimulation studies (Kilgard & Merzenich 1998b; Kilgard et al., 2001). The tone-trains consisted of six 25-ms long tones which were paired with 500 ms of vagus nerve stimulation (**Figure 3.1**). The electrical stimulation was delivered as 100 us charge-balanced biphasic pulse with 0.8 mA current and consisted of a 0.5 s long train of 30 Hz (500 ms train duration). The carrier frequency of the tone was equally distributed across the frequency range represented in A1 and was either one of seven frequencies (1.3, 2.2, 3.7, 6.3, 10.6, 17.8, 29.9kHz). The frequencies used from train to train were randomly varied but the frequency of the tones within each train was constant. Tone amplitude was 20–30 dB above the minimum rat hearing threshold. Interstimulus intervals varied randomly between 10 to 30 seconds to prevent rats from anticipating the stimulus timing. The tone pips in stimulus trains were presented in a given rat at 5 or 15 pps. Electrical stimulation began with the onset of the third tone. Impedance of the cuff electrodes was monitored every day and was at 4.4±1.5 kOhm. The impedance was stable across the entire training duration for all the rats. VNS tone-train pairing was delivered

300-320 times per day for 20 days (5 days a week), with each pairing sessions lasting 2.5 hours (**Figure 3.1**).

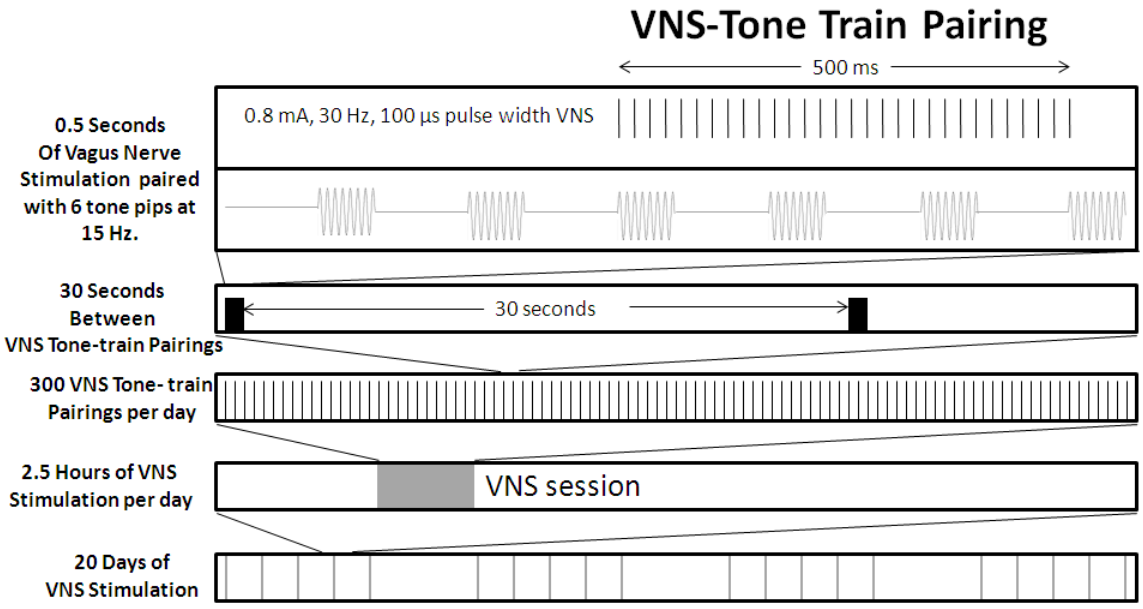

Figure 3.1. Schematic of the timing of VNS tone train pairing.
A 0.5 second, 30Hz train of 0.8 mA 100 µs pulse width was delivered to the left vagus nerve via a platinum iridium bipolar cuff electrode. VNS was paired with the third tone in the tone train.Each tone in the tone train was 25 ms long with a 5 ms ramp on and 5 ms ramp off. The tone train comprised of 6 tone pips and was 400 ms long when presented at 15 pps and 1200 ms long when presented at 5 pps. Rats received VNS on an average every 30 seconds, 300 times during each 2.5 hour session. VNS was delivered on 20 weekdays. Note: Carrier frequency of the tones shown in the figure is not to scale.

Electrophysiology data recording and analysis:

Twenty-four hours after the last pairing, animals were anesthetized with sodium pentobarbita (50 mg/kg) and a state of areflexia was maintained throughout the experiment with supplemental doses of dilute pentobarbital (0.2– 0.5 ml; 8 mg/ml). Anesthesia depth was monitored by heart rate, breathing rate, corneal reflexes, and response to toe pinch. Nourishment was provided using 1:1 mixture of dextrose (5%) and standard Ringer's lactate solution and body temperature was maintained at 37 degrees C. A tracheotomy was performed to minimize

breathing problems and breathing sounds, and a cisternal drain was made to minimize cerebral edema. A part of the skull over the temporal ridge was removed to expose the right primary auditory cortex. The dura was removed and the cortex was maintained under a thin film of silicone oil to prevent desiccation. Four parylene coated tungsten microelectrodes (FHC, 1-2 MΩ) were lowered simultaneously to a depth of 600 μm so that they were in layer IV/V of the primary auditory cortex. Blood vessels were used as landmarks to mark each of the electrode recording sites. To determine the characteristic frequency at each site we played ninety logarithmically spaced tones ranging from 1- 47 kHz at 16 intensities ranging from 0-75 dB SPL. The tones were 25 ms long and their presentation was randomly interleaved. We placed the speaker 10 cm away from the left ear. Neural response characteristics such as start latency, end latency and characteristic frequency at each recoding site were used to determine whether the electrodes were placed in the primary auditory cortex.

After all the tones were presented, we played tone-trains of variable repetition rates (3, 5, 7, 9, 10, 11, 14, 15, 17, 20, 25 pps). To determine the RRTF for each site, six tones (25 ms with 5-ms ramps, 70 dB SPL) were presented twelve times at each of eleven repetition rates. To minimize adaptation effects, repetition rates were randomly interleaved, and two seconds of silence separated each train. The tone frequency used to determine the RRTFs was one of the seven paired frequencies closest to the characteristic frequency of the site. To reduce the variability resulting from different numbers of neurons at recording site, response amplitude was normalized using the number of spikes evoked at each site to an isolated tone. The normalized RRTF was defined as the average number of spikes evoked for each of the last five tones in the train divided by the number of spikes evoked by the first tone in the train. Thus, a normalized

spike rate of one indicates that, at the given repetition rate, each of the tones in the train, on average, evoked the same number of spikes as the first tone. Paired pulse depression i.e. normalized response for the second tone was determined by calculated the average number of spikes evoked by the second tone and normalizing it by spikes evoked by the first tone. Values greater than one indicate facilitation; values less than one indicate response adaptation. Only spikes occurring from 5–40 ms after each tone onset were used to calculate the RRTF.

3.5 RESULTS

As in previous studies, most A1 neurons in naïve animals responded well to tones presented at 10 pps (Kilgard et al., 1998b; Boa et al., 2004). Each successive tone typically evoked the same number of spikes as the first tone for up to 10 pps (**Figure 3.2a**). For repetition rates higher than 10 pps, the successive tones evoked lesser spikes than spikes evoked by the first tone.

Animals which received VNS paired with tone trains had significantly altered cortical temporal response properties. Pairing VNS with 15 pps tone trains increased the strength of response to tone trains faster than 9 pps (**Figure 3.2b**). Compared to naives, these animals had 34 % more spikes for tone trains faster than 9 pps. For example, the representative neuron in Figure 2 b had 60% more spikes than naïve neuron at 10 pps. In contrast to naives, many neurons in these animals responded strongly to repetition rates between 10 pps and 15 pps (**Figure 3.2b**). A1 neurons in these animals typically showed a stronger facilitation response for fast tone trains. A facilitation response is seen when response to the repetitive stimuli is equal to or greater than response to the first tone. For example, neuron in Figure 2b had a facilitation response up till

tone trains as high as 15 pps and had 405% increase in the number of spikes than a typical neuron in the naïve animal at this repetition rate. In contrast, pairing VNS with 5 pps tone trains

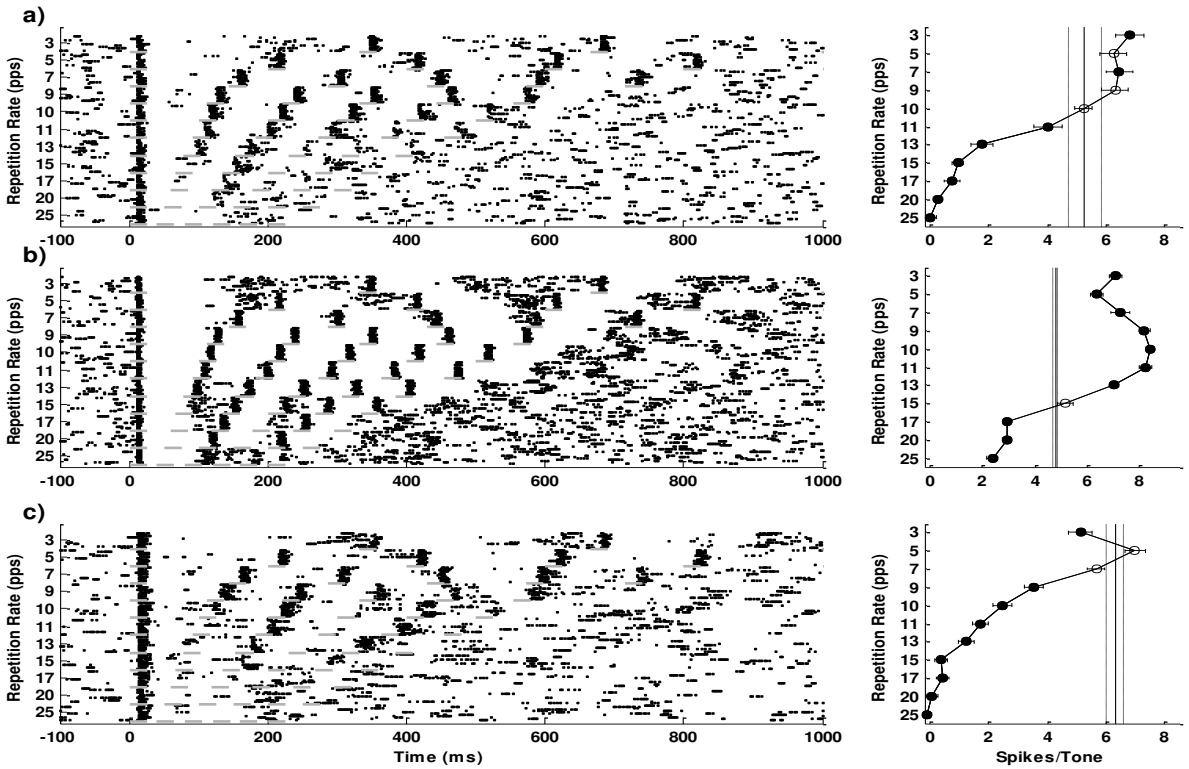

Figure 3.2. Temporal plasticity induced by pairing vagus nerve stimulation (VNS) with tone trains.

Dot rasters and repetition rate transfer function are shown for representative A1 sites in naives (a) and after pairing with 15 pps tone trains(b) and 5pps tone trains (c). Carrier frequency varied from train to train and was equally distributed over the octave space. Each dot represents a single action potential. The short horizontal gray lines indicate the occurrence of each tone in the tone train. Neural responses occurring in these periods were used to make the RRTF function on the right of each dot raster. The solid vertical line shows the average number of spikes evoked by the first tone in the tone train and the dotted vertical lines show standard error of mean for the first tone response. (a) RRTF of naïve A1 neurons shows a general decrease in response strength to stimuli faster than 10 pps. B) VNS paired with 15 pps tone trains increased the temporal following capacity of A1 neurons. c) VNS paired with 5 pps tone trains decreased the temporal following capacity of A1 neurons. The filled circles indicate responses significantly different than those evoked by the first tone.

decreased the strength of response to tone trains faster than 9 pps. Compared to naives, these animals had 14 % less spikes above 9 pps. For example, the representative neuron in Figure 3.2c had 52% less spikes at 10 pps than the naïve neuron in Figure 3.2a. Neurons in these animals typically showed a suppression response above repetition rates of 9 pps. Even at a slow repetition rate of 9pps, this neuron had a strong suppression response and evoked 43% less spikes than the naïve neuron.

To ensure accurate quantification of the following capacity of neurons, we used response evoked by the first tone to normalize the response evoked by the successive 5 tone trains pips. Up to 10 pps, the normalized cortical response in naives was 1, showing that neurons respond equally well to successive tone trains as to the first tone (**Figure 3.3a**). For repetition rates higher than 10pps, the normalized cortical response in naives dropped drastically, and around 15pps the A1 neurons had a very weak temporal following capacity. This shape of the RRTF is similar to naives from previous studies (Kilgard & Merzenich, 1998b; Kilgard, et. al, 2001).

Pairing VNS with 15 pps tone trains increased the temporal following capacity of A1neurons. Compared to naives, these animals had significantly stronger normalized cortical responses above 9 pps. For example, at 10 pps, the normalized cortical response in these animals was 22% larger than in naives. The shape of RRTF at fast repetition rates (> 10 pps) shifted up showing the higher temporal following capacity after pairing (**Figure 3.3a**). A strong facilitation response was seen up to 11 pps. The maximum following rate after pairing increased to 11 pps as compared to 10 pps in naives. The normalized response at this repetition rate was significantly higher than the normalized response in naïve animals (**Figure 3.3a**; unpaired ttest $p<0.05$). These

results show that pairing VNS with fast tone trains increases the temporal following capacity of A1 neurons.

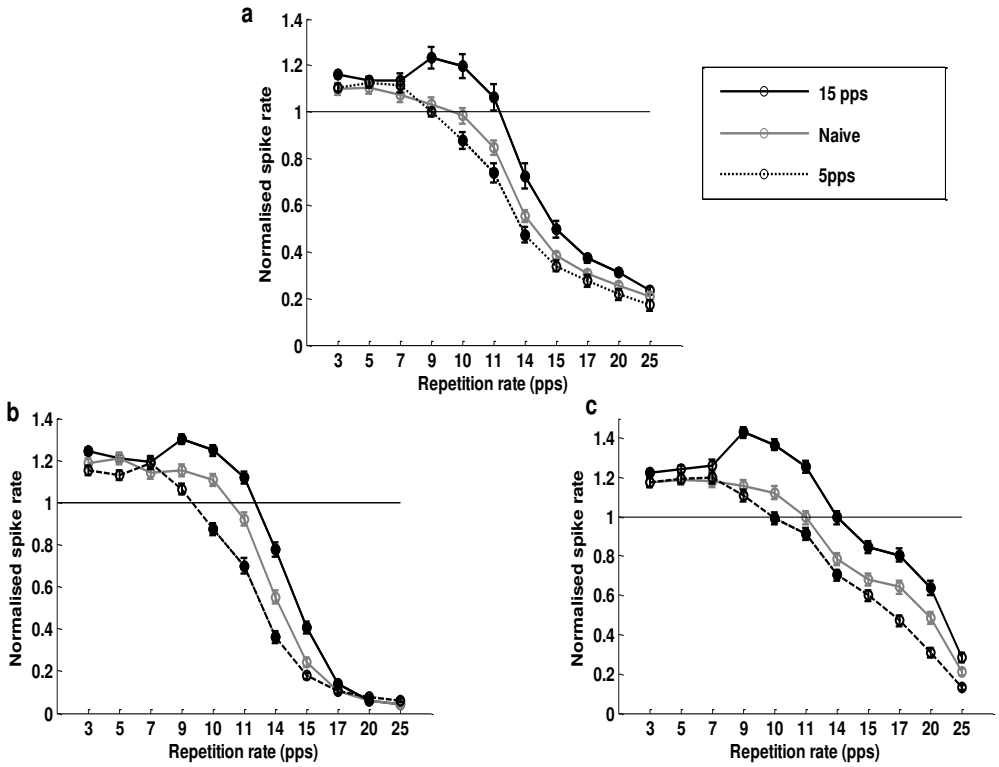

Figure 3.3. Mean normalized RRTF for all groups.
a) Normalized cortical following response recorded from naives and animals that had VNS paired with 15 and 5 pps tone trains. Cortical response was calculated by averaging spikes from tones 2 to 6 in the tone train and normalized by the spikes evoked by the first tone. Normalized cortical response showed a right-ward shift for animals that had VNS paired with 15 pps tone trains indicating increased temporal following capacity in these animals as compared to naives. Normalized cortical response showed a left –ward shift in animals which had VNS paired with 5 pps tone trains indicating decreased temporal following capacity compared to naives. This pattern was also seen when neural responses were normalized using just the 2nd or the 3rd tone. b) Normalized cortical response was calculated by considering spikes evoked by the 2nd tone in the tone train. The 2nd tone generally evoked the least of all spikes in the tone train indicating a paired pulse depression effect. The normalized response was also weaker than when all tones were considered together. c) Normalized cortical response was calculated by considering spikes evoked by the 3nd tone in the tone train. Of all tones in the tone train, the third tone evoked the strongest response and therefore resulted in the strongest normalized spike rate.

In contrast, pairing slow tone trains with VNS decreased the temporal following capacity of A1 neurons. The overall RRTF function in these animals showed a significant decrease in responses to stimuli higher than 9 pps (**Figure 3.3a**). The normalized cortical following response above 9 pps was significantly lower than in naives and in animals which had VNS paired with 15 pps tone trains(**Figure 3a**; unpaired ttest $p<0.05$). For example, at 10 pps, the normalized cortical response in these animals decreased by 11% compared to naïves and by 27% compared to animals which had VNS paired with 15 pps tone trains. On an average, for all repetition rates above 9 pps, the normalized responses in these animals were 14% smaller than naives and 31% smaller than in animals which had VNS paired with 15 pps tone trains. The average maximum following rate after pairing decreased to 7 pps. The normalized response at this repetition rates was significantly lower than the normalized response of naïve animals and also lower than normalized response of animals which had VNS paired with 15 pps tone trains (**Figure 3.3a**). These results provide the first evidence that VNS can drastically alter the temporal response properties of A1 neurons to refine or degrade the capacity of the cortex to respond to rapidly successive stimuli.

Evidence from previous studies suggests that the specificity of temporal plasticity to the paired repetition rates is due to altered cortical excitability (Kilgard et al., 2001; Boa et al., 2004). Naïve animals showed a significant paired pulse depression effect at 11 pps (**Figure 3.3b**). PPD is reported when the average responses evoked by the 2^{nd} tone is significantly lower than to average responses evoked by the 1^{st} tone, i.e. when the normalized response is significantly below 1 (paired ttest $p<.05$). We observed a significant change in paired pulse depression with VNS pairing. Animals which had VNS paired with fast tone trains showed a

significant paired pulse depression effect only at 14 pps. For example, the number of spikes evoked by the 2nd tone was 1.1 ±.1 lower than the number of spikes evoked by 1st tone. On the other hand, animals which had VNS paired with slow tones, had a significant paired pulse depression effect at a low repetition rate of 10 pps. These results suggest that VNS pairing with fast tone trains decreases paired pulse depression and pairing VNS with slow tone trains increases paired pulse depression. These changes in PPD could result in altered temporal plasticity effects when VNS is paired with temporally modulated stimuli.

Of all the tones in the tone train, the third tone generally evoked the strongest response (**Figure 3.3c**). Although this result seems surprising at first, it can be explained by the paired pulse depression seen for the second tone. PPD for the second tone eventually resulted in a facilitation response for the 3rd tone. This was especially true for fast repetition rates i.e. faster than 10 pps. The threshold of following tone trains seems higher when responses from only the third tone in the tone train are considered. Compared to a facilitation response upto10 pps when all tones in the tone train were considered, the facilitation response is seen up to 11 pps in naïve animals when only the 3rd tone is considered (**Figure 3.3c**). After pairing VNS with 15 pps tone trains, the facilitation response for the 3rd tone was present up to 15 pps instead of 14 pps as seen when normalized response was calculated from all tones in the tone trains. Similarly, in animals which had VNS paired with 5 pps tone trains, a facilitation response for the 3rd tone was seen up till 10 pps. At most repetition rates, a ripple like effect of alternating paired pulse depression and facilitation was seen with the 4th tone evoking lesser spikes than the 2nd tone and the fifth tone evoking lesser spikes than the 3rd tone but more spikes than the 4th tone. The 6th tone evoked the weakest response compared to all other responses in the tone train.

Spectral plasticity:

Since VNS was paired with tone trains of variable carrier frequencies we did not expect any spectral map plasticity. Examining frequency representations showed that animals of all groups had tonotopically organized A1. The size of A1 was similar across all groups (Naives: 1.77 ±0.4; 15 pps with VNS: 1.82±0.51; 5pps with VNS 1.72±0.2). There was no significant change across any groups in the percent of sites responding to tones of different frequencies. Analyzing tuning curve properties showed that VNS with tone train pairing did not significant change receptive field sizes. Bandwidth of neural responses was similar across all groups (Bandwidth at 30 dB above threshold was 2.6±0.05 octaves for naives, 2.65±0.05 octaves for fast group and 2.58±0.08 octaves for the slow group). These results affirm the hypothesis of feature specific plasticity and show that only temporal, not spatial features of neurons are altered when VNS is paired with temporally modulated stimuli.

Some of the basic response properties like driven spikes (2.07±.05, 2.17±.04, 2.14±.07), start latency (11.4±.2 ms naives, 11.1 ± .1 ms in fast, 11.6± .1 ms in slow), and peak latency did not change significantly. Spontaneous activity and end of response latency changed significant with VNS pairing. Spontaneous activity decreased when VNS was paired with 5 pps tone trains but was unaffected when VNS was paired with 15 pps tone trains. End of response latency in animals which had VNS paired with slow tone trains was significantly longer than naives (39±0.8 ms in slow vs 36.3±0.4 ms in naives unpaired ttest $p<0.05$). End of response latency of neurons in animals which had VNS paired with 15 pps tone trains was not significantly different than naïve animals (35.7±0.4 ms). The broadening of response to tones in animals with VNS

paired with slow tone trains suggests slower recovery of response in these animals and is similar to results from previous studies (Kilgard et al., 2001; Schriener et al, 1997).

3.6 DISCUSSION

A recent study showed that VNS with tone pairing can cause map plasticity in the auditory cortex that is specific to the paired tone. The changes in spatial map representation observed in this study were similar to those seen with NB stimulation tone pairing. NB stimulation with temporally modulated tones has been shown to induce temporal plasticity in A1. Based on these studies we predicted that VNS with temporally modulated stimuli would induce temporal plasticity in A1. To test this prediction we paired VNS with fast tone trains i.e. at a rate faster than the rate naïve animals can generally keep up with. Pairing VNS with fast tone trains increased the capacity of A1 neurons to follow fast tone trains. Pairing VNS with slow tone trains decreased the temporal following capacity of A1 neurons showing the VNS is induces input feature specific plasticity in A1. This result is similar to results with NBS tone train pairing.

Further comparison with NBS:

NB stimulation with tone train pairing has been shown to significantly alter neural response properties like spontaneous activity, latency and receptive field sizes (Kilgard et al., 2001). We saw similar changes in spontaneous activity, end of response latency after VNS pairing (**Table 3.1**). Spontaneous activity decreased in animals which had VNS stimulation paired with 5 pps tone trains but did not change in animals which had VNS paired with 15 pps tone trains. This result is similar to results from previous studies with NB stimulation and tone train pairing (Kilgard et. al, 2001). Increase in end of response latencies was seen with both

VNS and NB in animals which had VNS paired with 5 pps tone trains. The number of driven spikes did not increase in either group. This result too is similar to results seen with NB stimulation. As with NB stimulation tone train pairing, VNS tone train pairing did not induce any spatial map plasticity.

Table 3.1. Degree and direction of VNS induced plasticity and comparison with effects seen with NB stimulation from previous study.
Direction and the number of arrows indicate the direction and magnitude of observed plasticity for the tabulated neural response parameters. Zeros indicate no significant difference from naïve controls. Comparison is made with results from previous study where NBS was paired with similar tone trains in 2 groups i.e. VNS was paired with 15pps in one group and 5 pps in the other group. Green and red color indicates whether the direction of plasticity seen with VNS and NB pairing was same or different. Black indicates parameters not studied in NB.

Group	CF Shift	A1 Area	RF Size	Driven Spikes	Start Latency	End Latency	Spont-aneous	Threshold	Maximum Following Rate
VNS paired with 15 pps tone trains of random carrier frequency (n=7)	0	0	0	0	0	0	0	↑	↑
VNS paired with 5 pps tone trains of random carrier frequency (n=6)	0	0	0	0	0	↑	↓	↑↑	↓

Although most response properties changed in similar direction and extent in VNS and NB pairing, some of response properties did not. NB stimulation with tone trains of both 5 pps and 15 pps increased the bandwidths at all intensity levels. VNS pairing with tone trains however did not significantly change bandwidths at any intensity level. The non similarity of this property

with NB stimulation, however suggests that plasticity effect of VNS could be more input specific.

Potential mechanisms:

The exact mechanisms responsible for inducing VNS directed plasticity is unknown. Previous studies have shown that VNS realizes a number of neuromodulators like acetylcholine, norepinephrine, serotonin and other neuromodulators (Albert et al., 2009; Hassert et al., 2004; Dorr & Debonnel, 2006; Follesa et al., 2007). The similarity in the degree and direction of plasticity effects observed with NBS and VNS for most neural response properties suggest the possibility that release of acetylcholine may be play a role in VNS directed plasticity (**Table 1**). Some neural response properties changed in a different direction than seen with NVS tone train pairing. For example, NBS pairing with tone trains increases receptive field sizes of A1 neurons. However there was no significant change in receptive field sizes when VNS was paired with the same tone train stimuli. The similarity of most plasticity but not all plasticity effects between VNS and NBS pairing indicate that these techniques share some common but not same neural mechanisms. It is highly likely that the synergistic action of the different neuromodulators together is responsible for directing VNS regulated plasticity.

Clinical relevance:

VNS is FDA approved and is a safe and effective treatment for certain types epilepsy and chronic depression (Groves & Brown, 2005; Albert et al., 2009). Pairing VNS with tones has been shown to be a valuable tool in treating tinnitus equivalent in animals (Engineer et al., 2011). Our study showed that VNS with tone train pairing can change temporal response dynamics of A1. Deficits with temporal processing have been commonly seen in a number of learning

impaired populations. Dyslexics have significantly smaller EEG responses when presented with rapid stimuli compared to normal subjects (McAnally and Stein, 1997). Primary auditory cortex responses of poor readers' exhibit significantly more paired pulse depression compared to normal readers (Nagarajan et al., 1999). VNS paired with 15 pps tone trains increased the strength of response to fast tone trains and also reduced the paired pulse depression as compared to naïve animals. Autistic individuals exhibit significantly less paired pulse depression than normal individuals which might increase their propensity for sensory overload (Buchwald et al., 1992; O'Neill et al., 1997). Vagus nerve stimulation paired with 5 pps tone trains increased paired pulse depression as compared to naives. Vagus nerve stimulation paired with temporally modulated stimuli has the capability to change cortical temporal response properties in an input specific manner. Together these results indicate the potential of VNS pairing with temporal stimuli to possibly find treatment options for disorders primarily related to temporal processing deficits.

CHAPTER 4

TEMPORAL PLASTICITY IN AUDITORY CORTEX IMPROVES NEURAL DISCRIMINATION OF SPEECH SOUNDS

Jai A. Shetake, Navzer D. Engineer, Will A. Vrana, Jordan T. Wolf and Michael P. Kilgard

Department of Behavioral and Brain Sciences, GR41

The University of Texas at Dallas

800 West Campbell Road

Richardson, Texas 75080-3021

4.1 ACKNOWLEDGEMENTS

The authors would like to thank K. Ram and R. Cheung for their assistance with microelectrode mappings. We would like to thank S. Sudanagantha, M. Fink, H. Rasul, D. Vuppalla, A. Kuzu, C. Walker, C. Omana and M. Borland for their help with daily VNS pairing sessions.

4.2 ABSTRACT

A large number of studies show that speech perception problems arise from underlying cortical temporal processing deficits. Intensive behavioral training with non-speech temporally complex stimuli is capable of increasing speech discrimination capability in a pathological nervous system. Behavioral training with these temporal stimuli has been shown to increase temporal processing capability of the cortex in animals. Together these results suggest that improved temporal processing of the auditory cortex is responsible for improved speech perception. Based on these studies we hypothesized that animals with increased cortical temporal following capacity in response to tone trains will also have better neural discrimination ability of compressed speech stimuli. Repeated pairing of VNS with fast tone trains increases temporal following capacity of primary auditory cortex (A1) neurons. To test our hypothesis we paired VNS with fast tone trains for 20 days and recorded A1 responses to contiguously presented compressed speech sounds from A1 of barbiturate-anesthetized rats. After pairing, neural activity patterns evoked by compressed speech sounds were more distinct compared to naïves and control animals. Neural discrimination ability quantified using spike timing based classifier showed increased neural discrimination ability for compressed speech sounds in animals which had VNS paired with fast tone trains.

4.3 INTRODUCTION

A number of biologically relevant sounds, like speech sounds, animal vocalizations, etc. have a distinct temporal structure which is utilized by the auditory system for accurate perception of these sounds. The underlying temporal processing capability of the auditory cortex can predict behavioral discrimination ability on rapidly presented speech sounds (Ahissar et. al, 2001). An influential hypothesis suggests that neural deficits in general auditory temporal processing of rapid stimuli are directly responsible for speech processing deficits in learning impaired populations (Merzenich et al., 1996; Kraus et al. 1996; Stark and Heinz 1996a; Tallal et al., 1998).This hypothesis arises from a large body of work in dyslexic and LLI children and is based on the following evidence.

Dyslexic and LLI populations display behavioral and neurophysiologic deficits in temporal processing. For example, LLI subjects require processing time at least an order of magnitude longer to discriminate or remember any brief stimulus if followed in rapid succession (tens of ms) by another stimulus (Tallal and Piercy 1973; Farmer and Klein 1995; Tallal et al. 1985; 1981; Elliot et al. 1989; Reed 1989; Wolff et al.; 1990). Neural responses recorded to rapid stimuli in these populations are weaker than normal subjects and have been associated with speech perception deficits (McAnally and Stein, 1997; Protopapas et al. 1997; Nagarajan et al., 1999). Intensive behavioral training with rapid non speech stimuli has been shown to improve identification non speech and speech stimuli in these populations (Merzenich et al.1996; Tallal et al. 1996). Plasticity studies in animals have demonstrated that intensive behavior training with rapid stimuli increases temporal processing capacity of the auditory cortex (Merzenich et al., 1993, 1995; Bao & Merzenich, 2004; Zhou & Merzenich, 2009).Together these studies

hypothesize that improvement in behavioral processing of rapid stimuli is a result of increased temporal processing capacity of the auditory cortex. The effect of increased temporal following capacity of the auditory cortex on processing of rapidly presented speech stimuli has not been directly studied in the same species.

In this study we investigated how change in temporal response properties of auditory cortex neurons affects neural representation of rapidly presented speech stimuli. Specifically, we hypothesized that improvement in temporal response properties of neurons will improve neural representation of rapidly presented speech stimuli. Repeated pairing of VNS with fast tone trains has been shown to increase temporal following capacity of primary auditory cortex (A1) neurons (Chapter 3). To test our hypothesis, we paired VNS with fast tone-trains and recorded A1 responses to strings of speech stimuli compressed at different levels in both naïve and in animals which has VNS paired with fast tone trains.

4.4 METHODS

Stimuli

We used 5 of the 20 English consonant-vowel-consonant (CVC) words, ending in ‘ad’ (as in 'tad') used in the Engineer et al. (2008) study. These sounds were ‘bad’, ‘dad’, ‘sad’, ‘wad’, ‘yad’. A detailed description of the recording and processing on these sounds can be found in Engineer et al. (2008). In brief, we recorded these sounds in a double-walled, sound-proof booth. All speech sounds were produced by a female speaker. 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) to better match the rat hearing range (Sally and Kelly, 1988). The intensity of the speech sounds was adjusted so that the intensity during the most intense 100 ms

is 60 dB SPL. Speech sounds ‘sad’, ‘wad’ and ‘yad’ were approximately 600 ms, 590 ms and 560 ms long respectively and sounds ‘bad’ and ‘dad’ were 500 ms long.

Speech sounds were concatenated together without any inter-stimulus interval between them to generate stimuli ‘sad-wad-dad-yad’ and ‘sad-wad-bad-yad’. These stimuli are here on referred to as ‘word strings’. These word strings were then compressed to 70%, 50%, 30%, 20% and 10% of their original length using the STRAIGHT vocoder (**Figure 4.1**). Speech sounds ‘bad’ and ‘dad’ were also compressed to 50% and 30% of their original length and presented in isolation and serve as control sounds for compressed sounds presented word strings.

Subjects and electrophysiology recordings

All methods regarding subjects, surgery procedures for implantation of vagus nerve stimulator, surgery procedures for electrophysiology recordings and analysis, tone and tone train stimuli for this Chapter are the same as in Chapter 3. As a brief overview, nineteen female Sprague Dawley rats were used in these experiments; thirteen of which rats were randomly assigned to experimental groups implanted with vagus nerve stimulators and paired with tone trains of either 15 pps (n=7) or 5 pps (n=6). The remaining rats (n=6) were used as naïve controls. The experimental groups of animals was implanted with a vagus nerve stimulation cuff electrode and were placed in sound- shielded booth in which stimulation of the vagus nerve was paired with tone train stimuli of either 15 pps or 5 pps repetition rate depending on the experimental group. The carrier frequency of the tone was equally distributed across the frequency range represented in A1 and was either one of seven frequencies (1.3, 2.2, 3.7, 6.3, 10.6, 17.8, 29.9 kHz). After pairing VNS with tone-trains for 20 days, the primary auditory cortex (A1) in these animals was

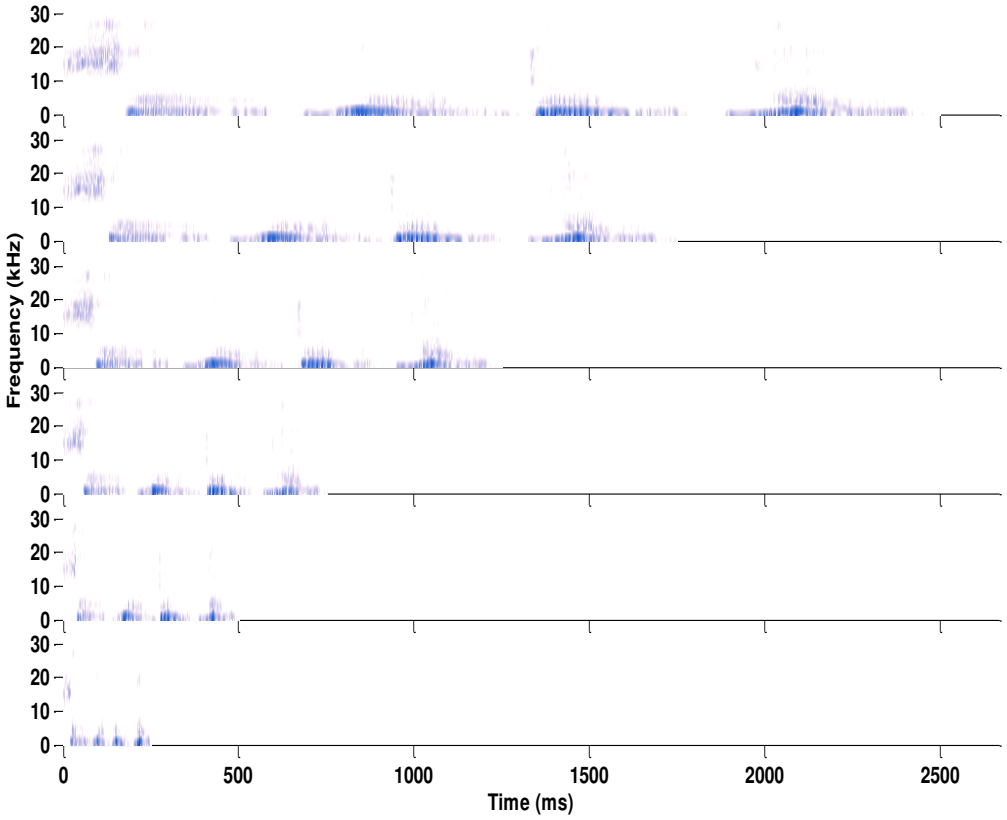

Figure 4.1. Spectrogram of word string ‘sad-wad-dad-yad’.
Spectrogram of word string ‘sad-wad-dad-yad’ is shown without compression and when the word string is compressed to 70%, 50%, 30%, 20%, 10% of its original length. The word strings were generated by concatenating the speech words without any interstimulus interval between individual speech words. These sounds are shifted up by an octave to match for the frequency hearing range of rats.

mapped under pentobarbital anesthesia (50 mg/kg). After playing tones and tone trains, neural responses were recorded to the compressed speech stimuli.

Neurophysiology Data Analysis

Classifier: To quantify neural discrimination performance, we used the nearest neighbor classifier from our previous study (Engineer et al., 2008). Analysis using a neural classifier allows direct correlation of neural and behavioral discrimination in units of percentage correct. Another advantage of using a neural classifier is that it analyses neural data on individual trial basis. This is because, in real life, speech sounds often need to be identified based on single trial basis. For a given recording site, the classifier compares each trial to the memory template of both the sounds. The memory template is average neural activity of the remaining 19 trials and is based on the assumption that the brain has heard the sound 19 times. The current trial is excluded from this average to prevent artifact. The sound whose memory template was closest to the trial activity (smallest Euclidean Distance) was guessed by the classifier to have generated the neural activity of the trial under consideration. In case both the Euclidean Distance's (E.D.) are same, the classifier randomly chooses which sound the trial belongs to. For a given recording site, neural discrimination performance (% correct) between 2 speech sounds is calculated by comparing the classifier sound guesses of all the trials to the correct sounds presented during those trials. Total neural discrimination for a discrimination task was obtained by averaging neural discrimination performance of all the recording sites. The neural activity is binned into several neural integration windows over a specific duration. Neural integration windows were in the range from 1ms to 500ms. The duration time was also in a range starting from the total duration of the onset response (40ms) to the total duration of the sound (500ms). Maximum window size tested was 500ms because this is the duration of sounds 'bad' and 'dad' when they were uncompressed. Euclidean distance is the square root of the sum of squared differences

between firing rate at each neural integration time. Other measures like Minkowski Distance and Cityblock distance were also used to test the results.

4.5 RESULTS

We used speech 'sentences' to record neural responses from primary auditory cortex of naïve (n=6), control (n=6) and experimental rats (n=6). One of the speech sentences was 'sad-wad-dad-yad' and the other one was 'sad-wad-bad-yad'. These 2 speech sentences are referred to here on as word strings. We recorded neural responses to these word strings compressed to70%, 50%, 30%, 20% and 10% of its original length. Neural responses were also recorded to sounds individual sounds 'bad' and 'dad' compressed at different levels which served as control neural responses for comparison when these stimuli were compressed in word strings. Neural responses to word strings have not been previously studied in animal cortex. Therefore, the first half of the results section describes neural responses to compressed and non compressed word strings in naives. The second half of the results section describes how neural responses to the same word strings were affected in animals which had VNS paired with fast tone-trains.

Neural responses to word strings in naives

Neural responses to most speech sounds in the word strings were similar to neural responses of these sounds in isolation from previous studies in humans and animals (Steinschneider et al., 1999; Wong and Schreiner, 2003; Steinschneider et al., 2004; Engineer et al., 2008; Skoe and Kraus, 2010). For example, voiced stop consonant 'ba' evoked only one peak of response whereas unvoiced consonants 'sa', 'ya' and 'wa' evoked a second peak of response corresponding to the voicing (**Figure 4.2**). Neural responses to sound 'dad' had an altered

responses than reported in previous studies where it were presented in isolation. Neural response of voiced stop ‘da’ evoked 2 peaks of responses in close succession (~ 25 ms apart) as opposed to only one peak when presented in isolation. The second peak of response is likely due to the forward masking effects of sounds presented closely before. These results are consistent with previous studies which show frequency specific suppressed responses due to forward masking (Schriener 1997, Condon and Weinberger, 1991).

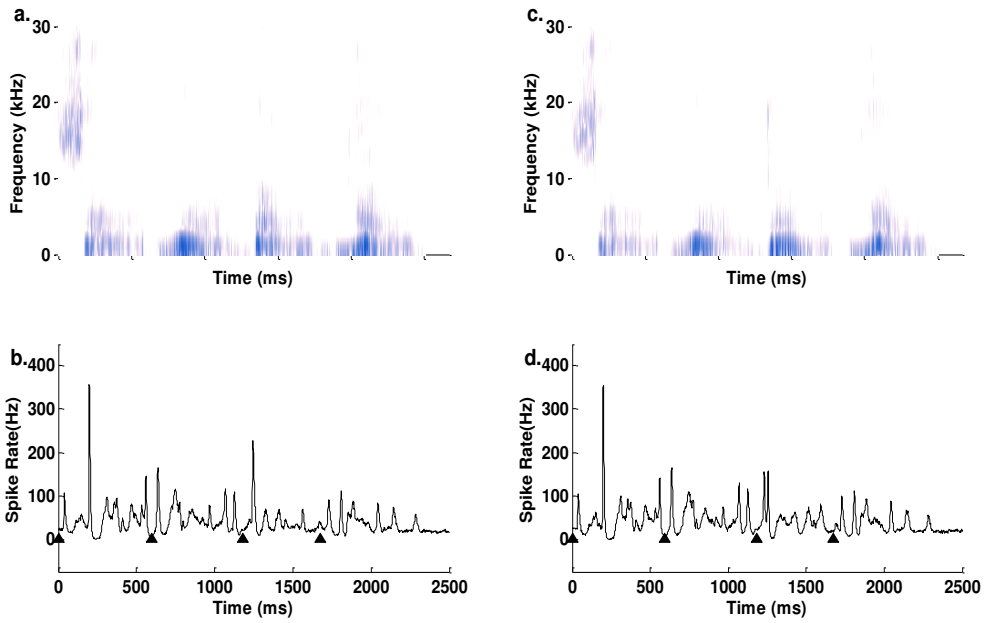

Figure 4.2. Spectrograms of word strings and PSTH response of naïve animals. Spectrograms of word strings ‘sad-wad-bad-yad’.
(a) and word string ‘sad-wad-dad-yad’ (c) without any compression. Neural response to word string ‘sad-wad-bad-yad’ (b) and word string ‘sad-wad-dad-yad’ (d) without any compression. Neural responses were recorded from 170 multiunit clusters of primary auditory cortex (A1) neurons of anesthetized naïve rats (n=6). Most neural responses to speech sounds in word strings were similar to neural responses of these sounds in isolation. Filled triangles show where each speech sound in the word string begins.

Neural response cues to most sounds were preserved even when the word stings were compressed. For example, neural response to unvoiced consonant ‘sa’ continued to evoke 2

peaks of response even when compressed to 10% of its original length (**Figure 4.3**). Most temporal cues of sounds were prominent even up till compression level of 50%. Sounds 'dad' and 'yad' continued to evoke 2 peaks of response until compression level of 50 % and sound 'wad' continued to have a 2 peaked response up till 30% compression, after which they evoked only one peak of response. Temporal response properties of neurons were significantly degraded when sounds were compressed to more than 20%. At 20 % compression, the middle sounds 'bad', 'dad' and 'wad' only had a single peak of response for the entire sound. Compression of sounds significantly decreased the number of driven spikes (paired ttest $p<.01$). For example, stimulus string 'sad wad dad yad' evoked 34% less driven spikes when compressed to 50% of its original length. This result was expected since compression of speech sounds provided with lesser spectral energy than available without compression. These results are similar to results from previous human neurophysiologic results which show that temporal pattern of neural responses to speech sentences are degraded but still present even when sounds are compressed to 20% of their original length (Ahissar et al., 2001). Our results demonstrate the similarity of neural responses to compressed speech sounds in humans and animals and support the long debated hypothesis that underlying auditory processing of complex sounds is similar in humans and animals.

Although compression reduced the number of driven spikes, the spatiotemporal patterns between sounds were preserved even at high compression levels. As in our previous study, sound's' evoked high frequency neurons (Engineer et al., 2008; Chapter 2). This pattern was seen even when the word strings were compressed to 10% of its original length. The primary

distinction between spatiotemporal patterns of sounds ‘ba’ and ‘da’ is the relative timing at which they evoke activity in the different groups of neurons. Sound ‘ba’ evoked low frequency

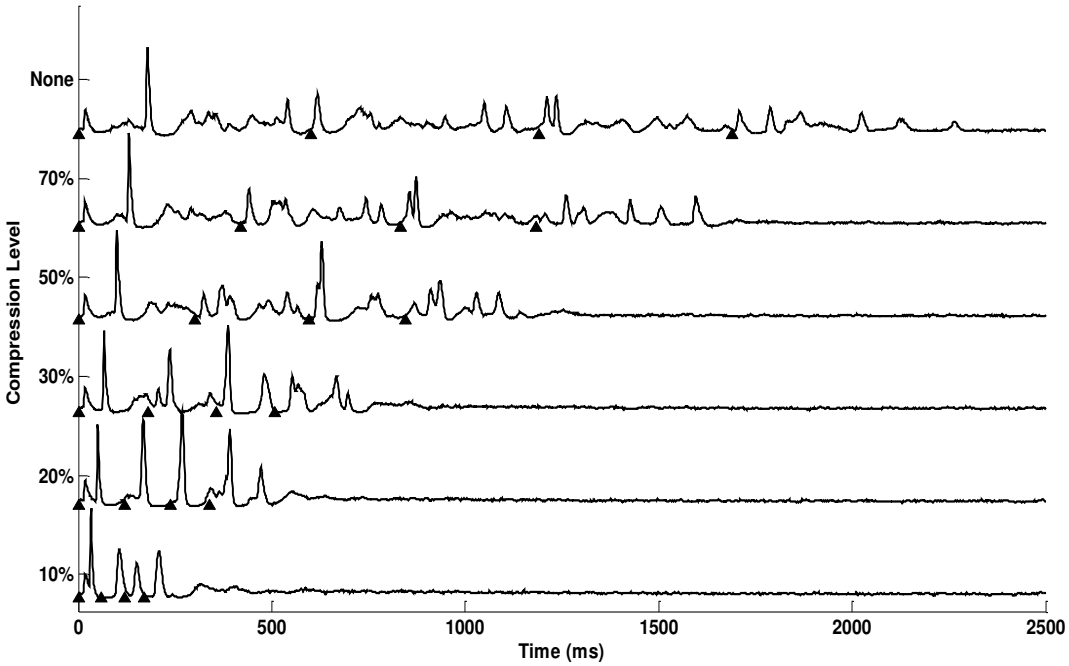

Figure 4.3. PSTH responses to word string ‘sad-wad-dad-yad’ compressed to different levels from naives.
Although most temporal patterns of neural responses to the word string were degraded with compression, the distinguishing temporal patterns were still present even under compression. Filled triangles show where each speech sound in the word string begins.

neurons followed by high frequency neurons and sound ‘da’ evoked high frequency neurons first followed by low frequency neurons a few milliseconds later (**Figure 4.4**). This distinguishing pattern was seen even when the sounds were presented in a word string that was compressed to 20% of its original length. This result is similar to results from our previous study where speech sounds were presented in isolation and without any compression (Engineer et al.,

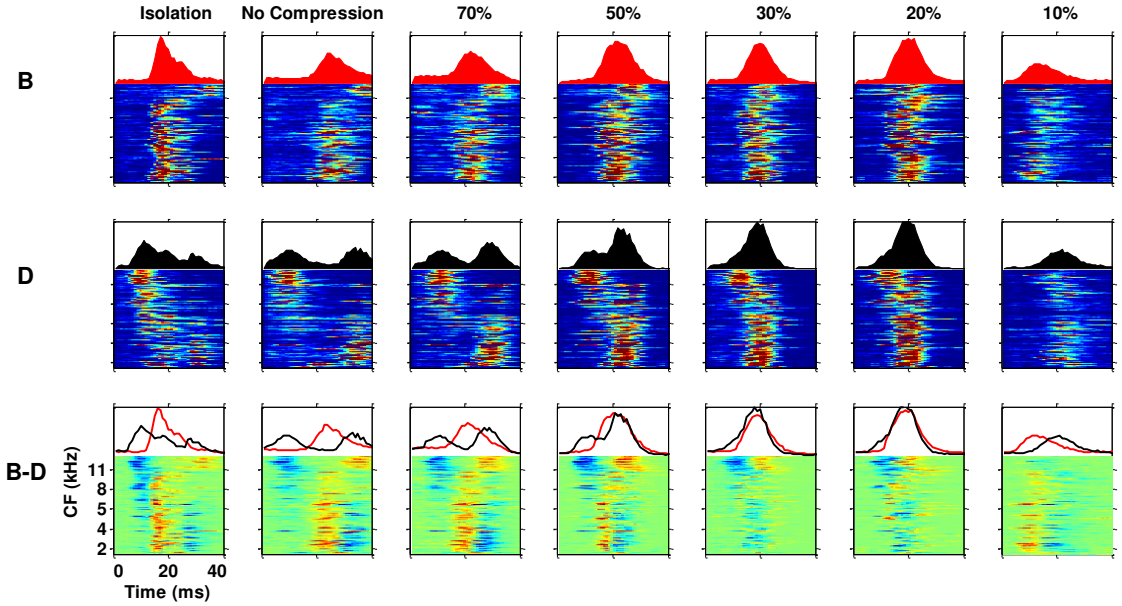

Figure 4.4. Neurograms and difference neurograms for speech sounds in naives. Multiunit data was collected from 170 recording sites in 6 anesthetized, experimentally naive adult rats. Neurograms to the speech sounds 'bad' and 'dad' when they were presented in isolation and in word strings where they were compressed to different levels. Neurograms show the average poststimulus time histograms (PSTH) responses across all the A1 sites. Average PSTHs were derived from 20 repeats and are ordered by the characteristic frequency (kHz) of each recording site (y axis). Difference neurograms are obtained by subtracting the neurogram responses of the two speech sounds and show distinctness between neurograms. Time is represented on the x axis (0 to 40 ms).

2008; Chapter 2). These results show that spatiotemporal patterns of speech sounds are preserved even under highly adverse situations like rapidly presented speech sounds.

Previous studies have shown that A1 neurons can accurately discriminate between spatiotemporal patterns of speech sounds when spike timing information is preserved (Engineer et al., 2008; Chapter 2). For example, A1 neurons can discriminate between speech sounds 'bad' and 'dad' up to 79.5 ± 0.8% when spike timing with 1ms precision is used over 40 ms of neural onset response (Engineer et al., 2008). Neural responses analyzed using relative spike timing

information between speech sounds has been shown to accurately predict behavioral speech discrimination ability. A1 responses are poor predictors of behavioral discrimination ability when temporal information is removed, i.e. only average firing rate is used. We used same neural classifier as used in our previous study to determine neural discrimination words ‘dad’ vs ‘bad’ when presented in a word string (Engineer et al., 2008). As in our previous study, neural discrimination ability was quantified using 1 ms spike timing precision over a duration of 40 ms. Neural discrimination ability of sound ‘dad’ vs ‘bad’ without any compression when presented in a word string was at 81 ± 1% (**Figure 4.5**). This result is comparable to neural and behavioral discrimination performance from our previous study when speech stimuli were presented in isolation.

Neural discrimination performance was decreased but was significantly above chance level as increased compression (**Figure 4.5**). Neural discrimination performance of sounds ‘dad’ vs ‘bad’ consistently degraded with increased compression but still significantly above chance even when stimuli were compressed to 10% of their original length. This is because the spatiotemporal patterns between sounds were still distinct at this level of compression (**Figure 4.4**). We did not have behavioral discrimination ability between word strings compressed to different levels. However, neural discrimination ability of A1 neurons in our study is comparable with previous human psychophysical studies which demonstrate speech sentence discrimination at similarly high levels of compression (Ahissar et. al, 2001). Together these results support the hypothesis of common underlying auditory processes between humans and animals even in highly adverse conditions like rapidly presented speech.

Neural responses after pairing VNS with tone trains

A large number of studies suggest that temporal processing capability of the auditory cortex is responsible for limitations over speech processing capability. These studies suggest that increased speech perception capability in pathological nervous systems is caused by the increased temporal following capacity of auditory cortex neurons. Based on these studies we

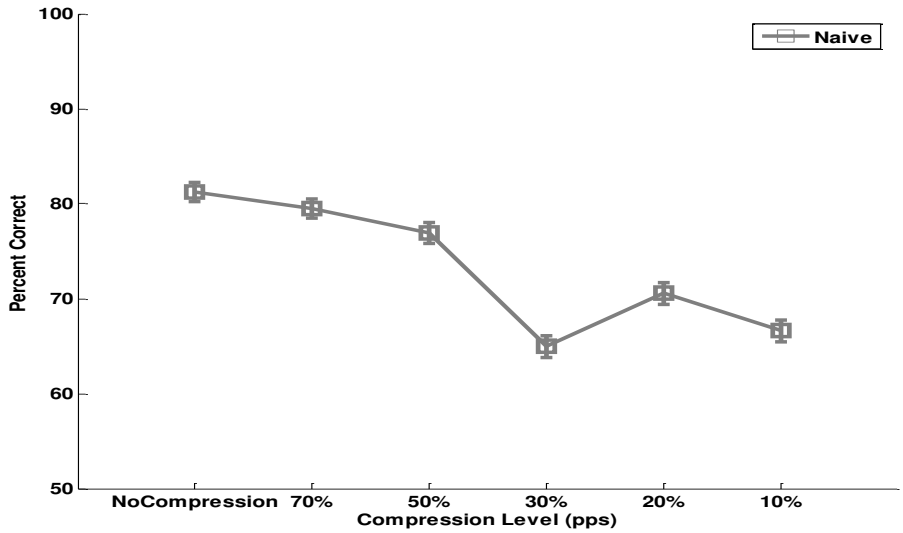

Figure 4.5. Neural discrimination performance in naives.
Neural discrimination performance between sounds 'bad' and 'dad' which were presented in word strings 'sad-wad-bad-yad' and 'sad-wad-dad-yad respectively. A nearest neighbor classifier (see methods) with 1 ms spike timing precision over 40 ms duration was used to determine the neural discrimination ability.

hypothesized that A1 neurons which have increased temporal following capacity in response to tone trains will also have better speech discrimination ability on compressed speech stimuli.

To test our hypothesis we paired VNS with fast tone trains for 20 days and recorded A1 cortex responses to contiguously presented compressed speech sounds. Repeated pairing of VNS with fast tone trains increases the temporal following capacity of A1 neurons (Chapter 3). We

quantified neural discrimination ability of sounds ‘dad’ and ‘bad’ presented in a word string and compressed to different levels. As in naives, neural responses were quantified using 1 ms spike timing over 40 ms duration. Neural discrimination ability in animals that received VNS pairing

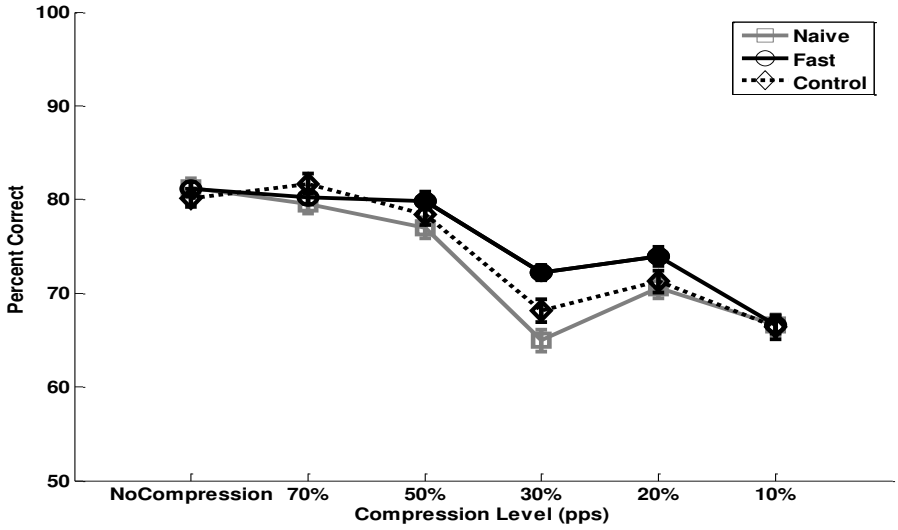

Figure 4.6. Improvement in neural discrimination performance of animals which had VNS paired with fast tone trains.
Pairing VNS with fast tone trains increased significantly increased neural discrimination performance of sounds ‘bad’ vs ‘dad’ which were compressed to from 20-50% of the original length of the word string as compared to naives(unpaired ttest $p<.05$). Neural responses in the ‘fast group’ were recorded from 181 multiunit A1 clusters of 7 anesthetized experimental rats in the fast group. The same neural classifier using 1 ms spike timing precision over 40 ms duration was used to quantify neural discrimination ability of the speech sounds. Neural discrimination ability in the ‘control’ group of rats recorded from 160 multiunit A1 sites of 6 anesthetized rats was not significantly different from naives. Filled symbols show whether neural discrimination ability was significantly different from neural discrimination ability of naives at the same compression level.

with fast tone trains was significantly higher compared to naives (**Figure 4.6**). The largest benefit of pairing VNS with fast tone trains was seen from compression level of 20%-50%. For example, neural discrimination ability when word strings were compressed to 30% of their original length was at 10% greater than naives. Our results support the hypothesis that increased

temporal following capacity of neurons underlies the increased neural discrimination ability of contiguously presented compressed speech.

The advantage of increased neural discrimination ability after pairing appears to be specific to the level of compression and the context in which the speech sounds were presented.

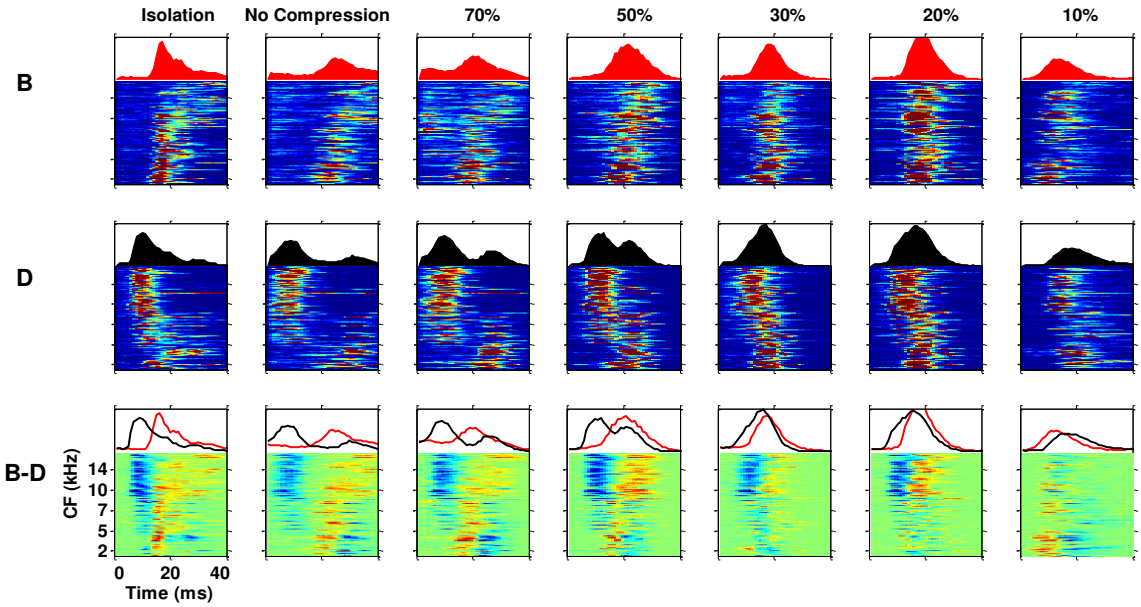

Figure 4.7. Neurogram and difference neurograms evoked by speech sounds after VNS pairing with fast tone trains.
Multiunit data was collected from 180. Neural responses evoked by the speech sounds 'bad' and 'dad' when they were presented in isolation and in word strings compressed to different levels in animals which had VNS paired with fast tone trains. From compression level 50%-20%, neurograms in these animals were more distinct than neurograms in naïve animals. Average poststimulus time histograms (PSTH) placed on top of the neurograms were derived from 20 repeats are ordered by the characteristic frequency (kHz) of each recording site (y axis). Time is represented on the x axis (0 to 40 ms).

Neural discrimination after pairing was not significantly different from naïves when word strings were presented without compression and at 70% compression. These results suggest that after pairing, A1 neurons have an increased propensity for stimuli with high degree of compression

than A1 neurons in naives. A1 neurons in animals that received pairing also had significantly worse neural discrimination ability than naives when the compressed sounds were presented in isolation. This result shows that pairing increases neural discrimination ability only when sounds are presented in a word string. Together these results show that the beneficial effects of VNS

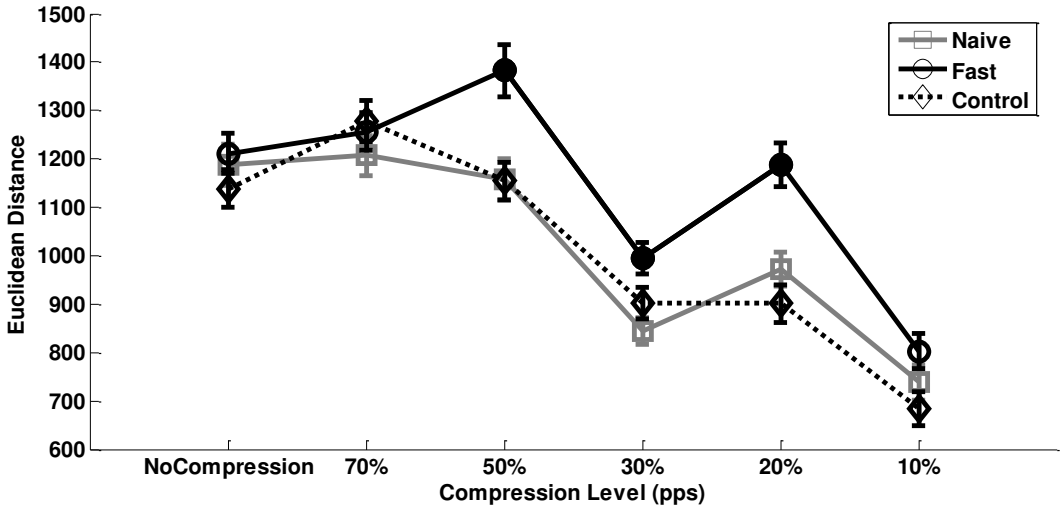

Figure 4.8. Euclidean distance between speech sounds for all groups.
Euclidean distance between speech sounds 'bad' and 'dad' presented in word strings at different compression levels. Euclidean distance was used to quantify the distinctness of two neurograms with 1ms spike timing precision over 40 ms duration. Euclidean distance between the speech sounds was significantly more distinct in animals which had VNS paired with fast tone trains than in naïve animals for compression levels 50%-20% (unpaired ttest; $p<.05$)

pairing with fast tone trains are only seen when speech sounds are presented in compressed word strings.

The improvement in neural discrimination ability after VNS is paired with fast trains can be explained by the enhanced spatiotemporal differences at high compression level. When no compression is applied the distinction between spatio-temporal patterns of sounds 'bad' and 'dad' are similar in both naïve and VNS group (**Figure 4.4, 4.7**). We used Euclidean distance to quantify distinction between spatiotemporal patterns these sounds. Euclidean distance between

the neurograms of sounds at this level of compression is similar in both groups (**Figure 4.8**). At compression levels 50% - 20%, the distinctions in spatio-temporal patterns between the speech sounds was much greater in the VNS group than in naives. Euclidean distance between the neurograms of these sounds at this level of compression is different across the 2 groups (**Figure 4.8**). These results show that VNS pairing with fast tone trains results in distinct spatiotemporal patterns which help in better neural discrimination ability of compressed speech sounds.

To test whether the effect of pairing VNS could only be induced only by fast tone trains pairing, we paired tone trains of a different frequency. These animals served as a control group since the only difference between this group and the group which had VNS paired with rapid tone trains was the frequency of the tone train (5 pps vs 15 pps). Neural discrimination ability in these animals was not significantly different than naives at any levels of compression (**Figure 4.6**). As quantified by Euclidean distance between the groups, the spatiotemporal patterns of sounds in these animals were also similar to spatiotemporal patterns in naives (**Figure 4.8**). Together these results show that animals which did not have improved temporal following capacity did not have improved neural discrimination ability of compressed speech and animals which had improved temporal following capacity had improved discrimination of speech sounds.

Our results demonstrate that neural discrimination ability of compressed speech sounds increases when rapid tone trains are paired with VNS and support the hypothesis that improvement in the underlying temporal processing capability of auditory cortex can improve neural discrimination ability of contiguously presented compressed speech sounds.

4.6 DISCUSSION

In this study we investigated how change in temporal response properties of auditory cortex neurons affects neural representation of rapidly presented speech stimuli. Specifically we hypothesized that improve temporal response properties will result in better neural discrimination of compressed stimuli than naives. Vagus nerve stimulation was paired with fast tone trains to improve the temporal response properties of primary auditory cortex neurons (Chapter 3). A string of words compressed at different levels was presented in both naïve and in animals which has VNS paired with fast and slow tone trains. Our results show that animals which had VNS paired with fast tone trains show a significant increase in the neural discrimination ability of rapidly presented compressed speech sounds.

A number of studies suggest that underlying neural temporal processing deficits are responsible for speech processing deficits in learning impaired populations. For example, Dyslexics have significantly smaller EEG responses when presented with rapid stimuli compared to normal subjects (McAnally and Stein, 1997). Primary auditory cortex responses of poor readers' exhibit significantly more paired pulse depression compared to normal readers (Nagarajan et al., 1999). Intensive behavioral training with rapid non speech stimuli can increase speech discrimination ability in these populations (Merzenich et al.1996; Tallal et al. 1996). Our study shows that the degree of temporal plasticity seen in our study when VNS is paired with fast tone trains is similar to behavior training techniques (Bao & Merzenich, 2004). VNS paired with 15 pps tone trains increased the strength of response to fast tone trains, reduced the paired pulse depression and increased speech discrimination ability of compressed speech sounds as compared to naïve animals. These results support the hypothesis that general auditory temporal

processing mechanisms could underlie processing of complex sounds like conversational speech. Our results show the similarity of neural response properties between humans and animals and propagate the use of animal models for studying speech sound processing disorders.

Results from previous work in our lab show that rats can behavioral discriminate isolated speech sounds even when they are compressed to 10% of their original length. Previous studies have shown that humans can discriminate between speech sentences even when they are compressed to 20% of their original length (Ahissar, et al., 2001). Neural discrimination ability seen in our study results with a words ‘dad’ vs ‘bad’ in word strings shows a degraded but above chance performance by A1 neurons even when speech sounds are compressed. These results indicate the behavioral and neural capability of representing highly compressed speech sounds is similar to humans. Further studies which directly behavioral and neural performance on compressed and rapidly presented speech are needed to understand how the auditory cortex neurons represent rapidly presented speech.

BIBLIOGRAPHY

Abeles M, Bergman H, Margalit E, Vaadia E (1993) Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J Neurophysiol 70:1629-1638.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367-13372.

Albert GC, Cook CM, Prato FS, Thomas AW (2009) Deep brain stimulation, vagal nerve stimulation and transcranial stimulation: An overview of stimulation parameters and neurotransmitter release. Neurosci Biobehav Rev 33:1042-1060.

Anderson S, Skoe E, Chandrasekaran B, Kraus N (2010) Neural timing is linked to speech perception in noise. J Neurosci 30:4922-4926.

Bao S, Chang EF, Woods J, Merzenich MM (2004) Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nat Neurosci 7:974-981.

Billings CJ, Bennett KO, Molis MR, Leek MR (2010) Cortical Encoding of Signals in Noise: Effects of Stimulus Type and Recording Paradigm. Ear Hear.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295-301.

Brinton L (2000) The structure of modern English: a linguistic introduction. 1:42.

Buchwald JS, Erwin R, Van Lancker D, Guthrie D, Schwafel J, Tanguay P (1992) Midlatency auditory evoked responses: P1 abnormalities in adult autistic subjects. Electroencephalogr Clin Neurophysiol 84:164-171.

Busch AC, Eldredge D (1967) The effect of differing noise spectra on the consistency of identification of consonants. Lang Speech 10:194-202.

Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, Alonso JM, Stanley GB (2007) Temporal precision in the neural code and the timescales of natural vision. Nature 449:92-95.

Byrne D, Harvey D, Khanh T, Stig A, Keith W, Robyn C, Bjorn H, Raymond H, Joseph K, Lui C, Jurgen K, Kotby MN, Nasser HAN, Wafaa AHEK, Yasuko N, Herbert O, Richard P, Dafydd S, Rhys M, Tony S, George T, Gregory IF, Soren W, Carl L (1994) An

international comparison of long-term average speech spectra. In, pp 2108-2120: J. Acous. Soc. Am.

Chase SM, Young ED (2007) First-spike latency information in single neurons increases when referenced to population onset. Proc Natl Acad Sci U S A 104:5175-5180.

Condon CD, Weinberger NM (1991) Habituation produces frequency-specific plasticity of receptive fields in the auditory cortex. Behav Neurosci 105:416-430.

Cooke M, Lecumberri MLG, Barker J (2008) The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. The Journal of the Acoustical Society of America 123:414-427.

Cunningham J, Nicol T, King C, Zecker SG, Kraus N (2002) Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hear Res 169:97-111.

Detari L, Juhasz G, Kukorelli T (1983) Effect of stimulation of vagal and radial nerves on neuronal activity in the basal forebrain area of anaesthetized cats. Acta Physiol Hung 61:147-154.

Dorr AE, Debonnel G (2006) Effect of vagus nerve stimulation on serotonergic and noradrenergic transmission. J Pharmacol Exp Ther 318:890-898.

Dubno JR, Levitt H (1981) Predicting consonant confusions from acoustic analysis. J Acoust Soc Am 69:249-261.

Elliott LL, Hammer MA, Scholl ME (1989) Fine-Grained Auditory Discrimination in Normal Children and Children with Language-Learning Problems. In, pp 112-119.

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.

Engineer ND, Riley JR, Seale JD, Vrana WA, Shetake JA, Sudanagunta SP, Borland MS, Kilgard MP (2011) Reversing pathological neural activity using targeted plasticity. Nature 470:101-104.

Foffani G, Tutunculer B, Moxon KA (2004) Role of spike timing in the forelimb somatosensory cortex of the rat. J Neurosci 24:7266-7271.

Foffani G, Chapin JK, Moxon KA (2008) Computational role of large receptive fields in the primary somatosensory cortex. J Neurophysiol 100:268-280.

Follesa P, Biggio F, Gorini G, Caria S, Talani G, Dazzi L, Puligheddu M, Marrosu F, Biggio G (2007) Vagus nerve stimulation increases norepinephrine concentration and the gene expression of BDNF and bFGF in the rat brain. Brain Res 1179:28-34.

Furukawa S, Xu L, Middlebrooks JC (2000) Coding of sound-source location by ensembles of cortical neurons. J Neurosci 20:1216-1228.

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. J Neurophysiol 76:1356-1360.

Groves DA, Brown VJ (2005) Vagal nerve stimulation: a review of its applications and potential mechanisms that mediate its clinical effects. Neurosci Biobehav Rev 29:493-500.

Hassert DL, Miyashita T, Williams CL (2004) The effects of peripheral vagal nerve stimulation at a memory-modulating intensity on norepinephrine output in the basolateral amygdala. Behav Neurosci 118:79-88.

Horii Y, House AS, Hughes GW (1971) A masking noise with speech-envelope characteristics for studying intelligibility. J Acoust Soc Am 49:1849-1856.

House AS, Williams CE, Hecker MHL, Kryter KD (1965) Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set. The Journal of the Acoustical Society of America 37:158-166.

Huetz C, Gourevitch B, Edeline JM (2010) Neural codes in the thalamocortical auditory system: From artificial stimuli to communication sounds. Hear Res.

Huk AC, Shadlen MN (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. J Neurosci 25:10420-10436.

Kawahara H (1997) Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. In: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, pp 1303-1306 vol.1302.

Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61:597-608.

Kilgard MP, Merzenich MM (1998) Plasticity of temporal information processing in the primary auditory cortex. Nat Neurosci 1:727-731.

Kilgard MP, Merzenich MM (1998) Cortical map reorganization enabled by nucleus basalis activity. Science 279:1714-1718.

Kilgard MP, Pandya PK, Vazquez J, Gehi A, Schreiner CE, Merzenich MM (2001) Sensory input directs spatial and temporal plasticity in primary auditory cortex. J Neurophysiol 86:326-338.

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

Kozou H, Kujala T, Shtyrov Y, Toppila E, Starck J, Alku P, Naatanen R (2005) The effect of different noise types on the speech and non-speech elicited mismatch negativity. Hear Res 199:31-39.

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

Lemus L, Hernandez A, Romo R (2009) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A 106:9471-9476.

Liberman P (1970) Towards a Unified Phonetic Theory. Linguistic Inquiry 1:307-322.

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1-36.

Liu J, Newsome WT (2005) Correlation between Speed Perception and Neural Activity in the Middle Temporal Visual Area. In, pp 711-722.

Martin BA, Kurtzberg D, Stapells DR (1999) The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/and/da. J Speech Lang Hear Res 42:271-286.

Martin BA, Stapells DR (2005) Effects of low-pass noise masking on auditory event-related potentials to speech. Ear Hear 26:195-213.

McAnally KI, Stein JF (1997) Scalp potentials evoked by amplitude-modulated tones in dyslexia. J Speech Lang Hear Res 40:939-945.

Merzenich MM, Schreiner C, Jenkins W, Wang X (1993) Neural mechanisms underlying temporal integration, segmentation, and input sequence representation: some implications for the origin of learning disabilities. Ann N Y Acad Sci 682:1-22.

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271:77-81.

Mesgarani N, David SV, Fritz JB, Shamma SA (2008) Phoneme representation and classification in primary auditory cortex. J Acoust Soc Am 123:899-909.

Miller GA, Heise GA, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. J Exp Psychol 41:329-335.

Miller GA, Nicely PE (1955) An Analysis of Perceptual Confusions Among Some English Consonants. J Acous Soc Am 27:338-352.

Muller-Gass A, Marcoux A, Logan J, Campbell KB (2001) The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. Neurosci Lett 299:197-200.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM (1999) Cortical auditory signal processing in poor readers. Proc Natl Acad Sci U S A 96:6483-6488.

O'Neill M, Jones RS (1997) Sensory-perceptual abnormalities in autism: a case for more research? J Autism Dev Disord 27:283-293.

Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME (2001) The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29:769-777.

Panzeri S, Diamond ME (2010) Information carried by population spike times in the whisker sensory cortex can be decoded without knowldege of stimulus time. Frontiers in Synaptic Neuroscience 2.

Phatak SA, Allen JB (2007) Consonant and vowel confusions in speech-weighted noise. J Acoust Soc Am 121:2312-2326.

Phatak SA, Lovitt A, Allen JB (2008) Consonant confusions in white noise. J Acoust Soc Am 124:1220-1233.

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

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

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

Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. J Neurosci 20:5392-5400.

Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci 22:9475-9489.

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

Rudel RG, Denckla MB, Spalten E (1976) Paired associate learning of Morse code and Braille letter names by dyslexic and normal children. Cortex 12:61-70.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clin Neurophysiol 115:2021-2030.

Sally SL, Kelly JB (1988) Organization of auditory cortex in the albino rat: sound frequency. J Neurophysiol 59:1627-1638.

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.

Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. Ear Hear 31:302-324.

Song J, Skoe E, Banai K, Kraus N (2010) Perception of Speech in Noise: Neural Correlates. J Cogn Neurosci.

Stark RE, Heinz JM (1996) Vowel perception in children with and without language impairment. J Speech Hear Res 39:860-869.

Stark RE, Heinz JM (1996) Perception of stop consonants in children with expressive and receptive-expressive language impairments. J Speech Hear Res 39:676-686.

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.

Tallal P, Piercy M (1973) Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241:468-469.

Tallal P, Stark RE (1981) Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J Acoust Soc Am 69:568-574.

Tallal P, Stark RE, Mellits D (1985) The relationship between auditory temporal analysis and receptive language development: evidence from studies of developmental language disorder. Neuropsychologia 23:527-534.

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann N Y Acad Sci 682:27-47.

Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM (1996) Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271:81-84.

Tallal P, Merzenich MM, Miller S, Jenkins W (1998) Language learning impairments: integrating basic science, technology, and remediation. Exp Brain Res 123:210-219.

Villa AE, Tetko IV, Hyland B, Najem A (1999) Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. Proc Natl Acad Sci U S A 96:1106-1111.

Wang MD, Bilger RC (1973) Consonant confusions in noise: A study of perceptual features. J Acous Soc Am 54:1248-1266.

Wang L, Narayan R, Grana G, Shamir M, Sen K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? J Neurosci 27:582-589.

Whiting KA, Martin BA, Stapells DR (1998) The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da. Ear Hear 19:218-231.

Wilkins WK, Wakefield J (1995) Brains evolution and neurolinguistic preconditions. In, pp 161-182: Cambridge Journals Online.

Wolff PH, Michel GF, Ovrut M (1990) The timing of syllable repetitions in developmental dyslexia. J Speech Hear Res 33:281-289.

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

Zhou X, Merzenich MM (2009) Developmentally degraded cortical temporal processing restored by training. Nat Neurosci 12:26-28.

Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C (2005) Deficits in speech perception predict language learning impairment. Proc Natl Acad Sci U S A 102:14110-14115.

Ziegler JC, Pech-Georgel C, George F, Lorenzi C (2009) Speech-perception-in-noise deficits in dyslexia. Dev Sci 12:732-745.

VITA

Jai Anand Shetake was born in Kolhapur, Maharashtra, India. After completing her schoolwork in Pune, Maharashtra, India at St. Mira's English Medium School in 1998, she entered Modern College, Pune to complete her higher secondary education in 2001. She then joined the Government College of Engineering, Pune where she earned her Bachelor of Engineering in Information technology in 2005. During the following ten months she was employed as Quality Assurance Engineer at Cybage Software Systems in Pune. She then joined the graduate program University of Texas at Dallas in Fall 2006 where she worked with Dr. Michael Kilgard. Jai has since earned a Master of Science in Cognition and Applied Neuroscience in 2009.

Publications:

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.

Engineer ND, Riley JR, Seale JD, Vrana WA, Shetake JA, Sudanagunta SP, Borland MS, Kilgard MP (2011) Reversing pathological neural activity using targeted plasticity. Nature 470:101-104.