Summer 8-15-2016

Neural Representation of Vocalizations in Noise in the Primary Auditory Cortex of Marmoset Monkeys

Ruiye Ni
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science
Department of Biomedical Engineering

Dissertation Examination Committee:
Dennis L. Barbour, Chair
ShiNung Ching
Daniel W. Moran
Baranidharan Raman
Mitchell S. Sommers

Neural Representation of Vocalizations in Noise

in the Primary Auditory Cortex of Marmoset Monkeys

by
Ruiye Ni

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

There are many people without whom this dissertation might not have been written and to whom I am deeply indebted.

Firstly, I would like to express my sincere gratitude to my advisor Dr. Dennis Barbour for his continuous support of my Ph.D. study and related research, and for his patience, motivation, and immense knowledge. His intellectual curiosity about the unknown greatly inspired me to keep making progress in my scientific research. His encouragement gave me the courage to move forward in the face of difficult situations. His guidance helped me through the research and writing of this dissertation. I cannot imagine having a better advisor and journey for my Ph.D. study.

Besides my advisor, I would like to thank the members of my thesis committee: Dr. Dan Moran, Dr. Baranidharan Raman, Dr. ShiNung Ching, and Dr. Mitchell Sommers, for their insightful comments and encouragement, but also for their hard questions which led me to widen my research from various perspectives. I was fortunate to be able to work over a semester in Dr. Moran's Lab and Dr. Raman's Lab, where I was first exposed to neurophysiology and developed a genuine interest in it. I would also like to thank Dr. Bruce Carlson, Dr. Alexandre Carter, and Dr. Vitaly Klyachko for time, talent and expertise they gave as committee members of my qualifying exam, thesis proposal, and thesis committee.

I would like to express my great gratitude to my fellow lab members Wensheng Sun, Jeffrey Gamble, and David Song for the relaxing and supportive working atmosphere they maintained, for stimulating discussion, and for all the fun we have had in the last five years. I especially thank Wensheng for her generous help with my neurophysiology experiment setup and for her dedicated work for caring the marmoset colony. My thanks also go to David Bender

for his assistance with data collection. Among former lab members, I first would like to thank Kim Kocher for her outstanding work as a lab technician and for all her fascinating stories and jokes. I am indebted to Drew Sinha, who was an undergraduate research assistant when I joined the lab. My computational work might not have been possible without Drew's excellent simulator framework. I thank former lab postdocs Dr. Noah Ledbetter and Dr. Ammar Hawaslia, who set good examples of researchers with critical thinking and great creativity for me to follow.

In the past six years, I have become acquainted with many friends both within and outside of Washington University. Their companion makes my time in St. Louis especially memorable. I will never forget the very first day I landed in St. Louis in August 2010 after flying a half-day overseas with my former roommate Jiami Wu. We've been family to each other ever since. I would like to thank Chen Zheng and Junye Zhang for being supportive neighbors for the first two years. I am thankful for my friends in different departments of Wash U: Lin Wang, Linjia Mu, Guoxi Xu, Hao Yang, Guannan He, Jin Hao, Jianqing Li, Yiqun Zheng, Fei Wang, Chao Li, Dongsu Du, Liren Zhu, Vynn Huh, Wandi Zhu, Debajit Saha and Kevin Leong. Special thanks go to staffs of Biomedical Engineering at Wash U: Karen Teasdale, Glen Reitz, and Amanda Carr, for taking care of the necessary logistics to make my work as smooth as possible. Among friends outside of Wash U, who made St. Louis feel like home to me, I would like to thank Master Miaohan, Xiaoli Gu, Huiping Dong, Ailian Liu, Wenfang Luo … and the list goes on and on.

This dissertation work was supported by a grant from the National Institutes of Health. I am beholden for its support to allow me to carry out my work. I am also grateful for the Cognitive, Computational and Systems Neuroscience pathway grant and a Biomedical Engineering Department grand for my conference travels.

Last but not least, I would like to thank my dear families. Thank you to my husband Ning Cheng for being my rock. We've grown together and witnessed each other's progress. I am greatly indebted to my parents Xiaodong Ni and Ni Lu. Words are powerless to express my gratitude. My parents have always encouraged me to be my best self and provided me with unconditional love and enormous support. I know that no matter what I do, I will never be able to repay even the half of what they've done for me. My great appreciativeness also goes to my in-laws, Weisong Cheng and Nong Lv. Their selfless help has allowed me to focus on study and research. Ultimately, I would like to thank my baby boy Wentao for bringing joy to my life.

Ruiye Ni

Washington University in St. Louis

August 2016

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Neural Representation of Vocalizations in Noise in the Primary Auditory Cortex of Marmoset Monkeys

by

Ruiye Ni

Doctor of Philosophy in Biomedical Engineering

School of Engineering and Applied Science

Washington University in St. Louis, 2016

Professor Dennis Barbour, Chair

Robust auditory perception plays a pivotal function in processing behaviorally relevant sounds, particularly when there are auditory distractions from the environment. The neuronal coding enabling this ability, however, is still not well understood. In this study we recorded single-unit activity from the primary auditory cortex of alert common marmoset monkeys (*Callithrix jacchus*) while delivering conspecific vocalizations degraded by two different background noises: broadband white noise (WGN) and vocalization babble (Babble).

Noise effects on single-unit neural representation of target vocalizations were quantified by measuring the response similarity elicited by natural vocalizations as a function of signal-to-noise ratio (SNR). Four consistent response classes (*robust*, *balanced*, *insensitive*, and *brittle*) were found under both noise conditions, with an average of about two-thirds of the neurons changing their response class when encountering different noises. These results indicate that the distortion induced by one particular masking background in single-unit responses is not necessarily predictable from that induced by another, which further suggests the low likelihood of a unique group of noise-invariant neurons across different background conditions in the

primary auditory cortex. In addition, for a relatively large fraction of neurons, strong synchronized responses can be elicited by white noise alone, countering the conventional wisdom that white noise elicits relatively few temporally aligned spikes in higher auditory regions.

The variable single-unit responses yet consistent population responses imply that the primate primary auditory cortex performs scene analysis predominately at the population level. Next, by pooling all single units together, pseudo-population analysis was implemented to gain more insight on how individual neurons work together to encode and discriminate vocalizations at various intensities and SNR levels. Population response variability with respect to time was found to synchronize well with the stimulus-driven firing rate of vocalizations at multiple intensities in a negative way. A much weaker trend was observed for vocalizations in noise. By applying dimensionality reduction techniques to the pooled single neuron responses, we were able to visualize the dynamics of neural ensemble responses to vocalizations in noise as trajectories in low-dimensional space. The resulting trajectories showed a clear separation between neural responses to vocalizations and WGN, while trajectories of neural responses to vocalization and Babble were much closer to each other together. Discrimination of neural populations evaluated by neural response classifiers revealed that a finer optimal temporal resolution and longer time scale of temporal dynamics were needed for vocalizations in noise than vocalizations at multiple different intensities. Last, among the whole population, a subpopulation of neurons yielded optimal discrimination performance.

Together, for different background noises, the results in this dissertation provide evidence for heterogeneous responses on the individual neuron level, and for consistent response properties on the population level.

Chapter 1: Introduction

1.1 Background and Motivation

Everyone's daily life is composed of a series of events, such as eating, walking, talking, and reading. While most people can successfully complete these events easily, we are usually not aware of the underlying complex computations being processed by our brain in order to generate the corresponding behaviors. Neuroscience is the field of study that helps us to understand such phenomena. The investigation can be done at different levels, ranging from basic molecules, synapses, neurons, networks, maps, and systems to the central neural system which scales from angstroms to meters (Churchland and Sejnowsky, 1991). Neurons, the fundamental components of the brain, are electrically excitable, generating events called action potentials. They connect with each other through synapses to form networks and transmit information with spike trains, which are temporal sequences of action potentials.

As a subfield of neuroscience, sensory neuroscience is mainly aimed at studying how different characteristics of an external stimulus, such as light, sound, or smell, are transformed by neural circuits into sequences of action potentials that will lead organisms to decide whether nor not to change their behaviors. Among all the sensory perceptions, auditory perception is crucial for our interaction with the surrounding environment. For instance, human babies are born with relatively mature hearing compared with other sensory modalities, which prepares them well for acquiring spoken language and bonding with their mothers (DeCasper and Fifer, 1980). When auditory stimuli are processed, sound is first transformed by the inner ear into spike trains and relayed along the ascending auditory pathway into the brain. The neocortex is a thin layered structure of mammalian brains. In large mammals, the neocortex has deep grooves and ridges.

Different regions of the neocortex perform different functions. The primary auditory cortex in the temporal lobe of the neocortex is mainly responsible for processing auditory information.

One typical way to get access to the electrical activities of individual neurons is single-unit recording (Boulton et al., 1990). By inserting a metal microelectrode or a glass micropipette with a fine tip and high impedance into the brain, we can measure the electrophysiological responses of individual neurons in response to a sensory stimulus across time. A single, firing neuron with a distinctive action potential shape, called a single unit, can be isolated from the recording microelectrode. Depending on where the microelectrode is placed relative to the cell, we can have intracellular or extracellular recording. Intracellular recording is implemented by inserting the electrode through the cell membrane, while extracellular recording only places the electrode close to the neuron so that spiking activity can be captured. Though intracellular recording allows us to gain more information with regard to a neuron's activity, such as postsynaptic potentials and resting membrane potentials, extracellular recording is more stable, especially in awake experimental subjects. Based upon the collected single-unit activity, data analysis can be done on a single-unit level.

Single-unit recording is not efficient if a large number of single units need to be recorded. Multichannel microelectrode recording techniques permit simultaneous recordings of neuron populations (Buzsáki, 2004). The distributed coding hypothesis that stimulus-related information is distributed over a large population of neurons can thus be tested. The related analysis is called neural population or neural ensemble analysis (Brown et al., 2004; Bartho et al., 2009). In some studies, when the multichannel recoding technique is not available, a pseudo-population analysis is implemented based upon the population of individual neurons recorded by single-unit recording (Gutnisky and Dragoi, 2008; Meyers et al., 2008).

1.1.1 Auditory Scene Analysis

In natural settings, behaviorally relevant acoustic signals are usually contaminated by ambient sounds, for example, the thrumming of the rain, the roar of rushing traffic, and background conversations among people attending a conference. Different sound sources mix together and arrive at our ears simultaneously. Do we perceive the mixed signal as a whole, or are we able to distinguish the different sound sources? In reality, both humans and animals exhibit reliable auditory detection in the presence of substantial amounts of noise. For instance, when you are talking with a friend at a party, you are still able to focus your listening attention and maintain the conversation with your friend although there are hundreds of other people talking around you. This is well known as the cocktail party effect (Cherry, 1953).

Though we behaviorally demonstrate the striking ability of reliable perception under adverse listening conditions, how our brain addresses this challenge is still a mystery. Psychologists and neuroscientists have strived to provide answers to this interesting phenomenon, and the term “auditory scene analysis” was coined and introduced by psychologist Albert Bregman in his book with the same title to define this research topic (Bregman, 1994). The underlying assumption is that sounds emitted by different sources have their own unique temporal and spectral features, and by integrating the information across time and frequencies belonging to the same source, one is be able to form a sound “stream” or “auditory object” corresponding to a perceptually meaningful sound unit. Meanwhile, the information about other sound sources is segregated from the target sound source to form additional “streams” and “auditory objects”.

Extensive studies on this topic have been carried out based upon different experimental models. The addition of distractive background sounds, called sound masking, is used to study

speech signal with human subjects. There are two main different types of masking: “energetic” masking and “informational” masking (Brungart, 2001; Scott et al., 2004). In energetic masking, the masking element simultaneously contains energy in the same critical band as the target element, and part or all of either the masking or the target element is not perceived. Speech intelligibility under energetic masking decreases monotonically with the signal-to-noise ratio (SNR) (French and Steinberg, 1947; Fletcher and Galt, 1950). In informational masking, subjects are able to hear both the target and masking elements, but are not able to tell them apart. This type of masking has a non-linear effect on the speech intelligibility with performance plateaued at SNRs below 0 dB (Egan et al., 1954; Dirks and Bower, 1969; Brungart, 2001).

Advances in technology have allowed researchers to monitor subjects’ brain activity under different masking conditions. Using magnetoencephalography, Ding and Simon found out that in the auditory cortex of humans, low-frequency brain activity seems to provide neural cues for stable speech recognition under both energetic and informational masking (Ding and Simon, 2012, 2013). With the application of invasive neural recording technologies, neuroscientists have revealed more detail of the underlying neural responses in animal models. For example, under interference from noise stimuli, single neurons in auditory brain area filed L of songbirds tend to suppress their activities corresponding to the informative elements in sound stimuli, while increasing their spike activities in response to the noise element (Narayan et al., 2007). In addition, individual neurons with robust resistance to background noise have been discovered in avian models (Moore et al., 2013; Schneider and Woolley, 2013).

Although we now have a more profound understanding of auditory scene analysis regarding the effects of different masking types on speech intelligibility and the associated neural activities, there are still many questions. One question being addressed in this dissertation is the

effects of different background noises on the functional identity of individual neurons in response to behaviorally relevant acoustic signals and on the discriminability of collective single units. In other words, if a neuron behaves as a robust vocalization encoder under one noise condition, does it still function in a robust way under another noise condition? Because we need to investigate the neural responses of individual neurons, which are typically acquired with invasive recording techniques, an appropriate animal model is essential.

1.1.2 Marmoset Monkey Vocalization

Vocalization is an essential communication channel used by humans and animals for social interaction. Common marmoset monkeys (*Callithrix jacchus*), a New World monkey, represent one of the closest evolutionary relative to human beings and exhibit a complex vocal communication system. Due to their densely vegetated living environment in nature, marmosets rely on vocalization to compensate for the lack of visual contact. Marmosets have a rich vocalization repertoire, and they use vocalizations for numerous purposes, such as to claim their own territory, to keep track of their group members, and to warn of the presence of a potential predator. Due to their complex social system, marmosets are highly vocal even within a captive colony, and the varieties of vocalizations are very similar to those produced by wild colonies (Bezerra and Souto, 2008). Marmosets are easy to handle and breed in the laboratory environment. Adult marmosets usually weigh between 300 to 500 grams, and they can give birth to twins or triplets twice a year. These traits make marmosets a good candidate primate model for neuroscience study.

Xiaoqin Wang's lab at the Johns Hopkins University has systematically quantified the vocalization repertoire of marmosets (Agamaite et al., 2015). In their studies, vocalizations were quantified by their temporal and spectral properties, such as length, frequency corresponding to

the maximum in the spectrum, and modulation rate. According to their complexity, there are two main kinds of vocalizations: simple and compound calls. Simple calls are the basic acoustic elements/phrases uttered by marmosets. There are four major types: twitters, phees, trills and trillphees. Theses vocalizations are named based upon their pronunciation. Compound calls are combinations of multiple simple calls with less than a 0.5 s interval between phrases, such as peep-string, peep-trill, and tsik-string. Marmoset vocalizations contain acoustic information over distributed frequencies and a wide range of time scales (DiMattina and Wang, 2006; Agamaite et al., 2015). The vocalization's acoustic energies are maximized around 6 KHz to 8 KHz, and their durations last from hundreds of milliseconds to several seconds. Very few studies, however, have systematically investigated the association between vocalization content and behaviors. Generally speaking, twitter is a between-group territorial call, trills and phees are within-group contact calls, and tsik is an alarm call (Marmosetcare.com, 2011).

Given their rich vocalization repertoire and easiness to house and handle in laboratory settings, marmoset monkeys are ideal animal models to study the auditory perception of communication sounds.

1.1.3 Marmoset Monkey Auditory Cortex

A reliable brain structure model of marmosets has been built (Hashikawa et al., 2015). The brain structure of marmosets shares many common characteristics with other primate species (Paxinos et al., 2012). The brain area involved in the auditory perception process is called auditory cortex, and it has been extensively studied. The auditory cortex is located on both the left and right brain hemispheres, at the upper side of the temporal lobes along the lateral sulcus. Researchers are able to further divide the auditory cortex into subareas according to their architectonic features, neuronal response properties, and input/output connections. Like that of

other primates and humans, the auditory cortex of marmosets is mainly composed of two parts: a "core" region and a "belt" region (Hackett et al., 2001). Three subdivisions are further identified within the core region: the primary auditory cortex, the rostral field, and the rostrotemporal field (Morel and Kaas, 1992; Petkov et al., 2006; Bendor and Wang, 2008). The three areas are all responsive to narrowband acoustic stimuli, for instance, tones, and are tonotopically organized. Among the three areas, the primary auditory cortex tends to have stronger responses and shorter response latencies (Bendor and Wang, 2008). Studies of neural responses to stimuli with complex spectral and temporal features, such as marmoset vocalizations, have been conducted in the primary auditory cortex. Neurons in the primary auditory cortex are selectively more responsive to natural vocalizations rather than other synthetic stimuli with the same spectral content but disrupted temporal features (Wang et al., 1995). Given the rich responsiveness to natural vocalizations, the primary auditory cortex is a reasonable marmoset brain area for studying auditory scene analysis.

1.2 Research Objectives

The goal of this dissertation is to enhance our understanding of the neural processing in the non-human primate auditory cortex related to vocal communication in noisy conditions. In particular, using the marmoset monkey model, I aim to study the neural representation of conspecific vocalizations embedded in noisy background. The activities of individual auditory neurons and neural ensembles evoked by vocalizations with different levels of noise will be investigated under two noise conditions: white Gaussian noise and marmoset vocalization babble. Marmoset babble is generated by mixing multiple marmoset vocalizations to simulate the situation of multiple callers in the background. I hypothesize that individual neurons that provide more informative spikes about clean vocalizations are less likely to suffer from spiking

suppression with increasing noise interference. Furthermore, I hypothesize that the spiking activities of neural ensembles are more informative than the spiking of single neurons with regard to the target vocalization masked with noise, since population coding has been demonstrated to be more robust in other sensory modalities.

Objective 1: Characterize single neuron responses to clean vocalizations and broadband white noise in the primary auditory cortex. Auditory cortex neurons generate more spikes in response to vocalizations than to other sound stimuli, but how differently they encode vocalizations from other stimuli (e.g., white noise) is not clear. I hypothesize that the activities of single auditory neurons are spectral-temporally modulated by vocalizations and systematically encode the stimulus with higher information rates, while single neurons' responses to pure wide-band noise are less structured, with lower information encoding rates. To test the hypothesis, 20 marmoset vocalizations and pure wideband white noise were delivered to awake marmosets (N = 2) and the corresponding single neuron activity was evaluated for information content.

Objective 2: Quantify the effects of different levels of broadband white noise and babble noise on the single-neuron representation of target vocalizations in the primary auditory cortex. Single neurons' computational strategy for processing noisy vocalizations in the primary auditory cortex is largely unknown. I hypothesize that the same neurons that provide more information about the target clean vocalizations, either in absolute terms or relative to pure noise, are also more resistant to noise interference and with less spiking suppression. To test the hypothesis, the responses of single neurons in the primary auditory cortex of awake marmosets (N = 2) were recorded while a set of stimuli was presented. Five marmoset vocalizations were presented to the animals, and were masked with different levels of background noise and

presented later in a random order. Clustering methods were used to identify neuron response patterns.

Objective 3: Quantify the effects of different levels of broadband white noise and babble noise on the neural ensemble coding of target vocalizations in the primary auditory cortex. Large variance exists between responses of neurons to the same stimulus; however, whether the variance provides more stimulus-related information to account for behavioral performance is not clear. In this aim, I explored the neural ensemble coding of clean vocalizations and noisy vocalizations. I hypothesize that neuron populations are more resistant to the contamination of useful sensory information by noise than single neurons. To test this hypothesis, sufficiently large ensembles of neurons' responses were recorded in the primary auditory cortex of awake marmosets. The stimuli were five marmoset vocalizations degraded by various levels of background noise, presented in random order. Dimensionality reduction techniques and discriminative analysis were used to analyze the population coding.

1.3 Research Approach and Overview of Dissertation

To study the single and population neuron responses to natural communication sounds, single neurons in the primary auditory cortex of four marmoset monkeys were recorded using invasive extracellular recording techniques. A neural population response study was conducted based upon a collection of single neurons. Data collection was conducted when animals were passively listening to the delivered acoustic stimuli, and more details about experimental setup are introduced Chapter 2. The results of the first objective are presented and discussed in Chapter 3. The single unit analysis of objective two is presented and discussed in Chapter 4. Chapter 5 presents the results of population analysis in the third objective. Data analysis is introduced

separately for Chapter 3, Chapter 4, and Chapter 5. Chapter 6 gives an overall summary of conclusions and recommends future work to extend the research of this dissertation.

Chapter 2: Experimental Methods

This chapter describes the methodologies that I used to investigate how neurons in the primary auditory cortex (A1) of awake marmoset monkeys represent and encode vocalizations delivered together with different masking sounds. It is composed of four parts: preparation of animal subjects and neurophysiology recording, acoustic stimulus generation, experimental procedure, and data analysis. The data analysis section includes a general summary, and details of analytic techniques are addressed in subsequent chapters.

2.1 Surgery and Recording Methodology

Adult common marmoset monkeys (*Callithrix jacchus*) were the subjects of this research. All training, recording, and surgical procedures complied with the US National Institute of Health Guide for the Care and Use of Laboratory Animals and were approved by the Animal Studies Committee of Washington University in St. Louis. Subjects were initially trained to sit upright in a custom, minimally restraining primate chair inside a double walled sound-attenuation booth (IAC 120a-3, Bronx, NY) for the same duration as would be used for later physiology recording. After they had become accustomed to this setup, a custom head cap for neural recording was surgically affixed to the skull of each subject. The animals were allowed to sufficient time to recover following surgery and were given pain medication to eliminate discomfort during recovery. The animals were able to feed themselves properly afterwards. The location of the vasculature running within the lateral sulcus was marked on the skull. Using the lateral sulcus as a guide, microcraniotomies (<1 mm diameter) were drilled through the skull over the temporal lobe with a custom drill, for physiology experiments. An active recording hole was partially filled with ointment and dental cement to prevent excess tissue growth and infections after each recording session. The hole was permanently sealed with dental cement

before the next craniotomy was drilled. In this way, we largely preserved the intactness of the bone and the landmarks. Daily recordings lasted about 4 hours for each animal and were continued for several months. The animal's awake state was monitored with a camera throughout the recording session. The location of A1 was identified anatomically based upon the lateral sulcus and bregma landmarks and confirmed with physiological mapping (Stephan et al., 2012).

Within each microcraniotomy, a single high-impedance tungsten-epoxy 125 µm electrode (~5 MΩ @ 1 kHz, FHC, Bowdoin, ME) was advanced perpendicularly to the cortical surface by a hydraulic system. Microelectrode signals were amplified using an AC differential amplifier (AM systems 1800, Sequim, WA) with the differential lead attached to a grounding screw on the animal's head. Initially, the electrode was advanced at a speed of 10 µm per step for the first 200 µm, which is a rather shallow level for single unit detection. As the electrode went deeper, a "hash" of background sounds gradually built up as the acoustic stimuli were delivered. The background sound would pause during the intervals between acoustic stimuli. The buildup of background sounds was an important indicator, confirming that the electrode was approaching populations of auditory neurons. Once the "hash" sound began, the electrode was slowly advanced at 1 µm per step to detect single units. Single-unit action potentials were sorted online using manual template-based spike-sorting hardware and software (Alpha Omega, Nazareth, Israel). Each single unit was confirmed by its consistent distinctive action potential shape. Single units were usually collected at depth between 400 µm and 2000 µm below the surface of the auditory cortex. The median signal-to-noise ratio for single units recorded with this technique was 24.5 dB. When a template match occurred, the spike-sorting hardware relayed a TTL pulse to the DSP system (TDTRX6, Alachua FL) that temporally aligned recorded spike times (2.5 µs

accuracy) with stimulus delivery. Recording locations within the head cap were varied daily, eventually covering all regions of interest.

2.2 Acoustic Stimuli

Table 2.1 Vocalization repertoire

No.	Vocalization Name
1	phee_m87_t449da10_415
2	pheestrg_m60107_t459b010_042
3	pheepeep_m335_t4530c11_454
4	peepphee_m290_t397ca10_414
5	trilphee_m70100_4580010_020
6	trill_m87_t449e011_112
7	peeptril_m60107_t459dd10_495
8	peeptril_87_t450da10_182
9	trilpeep_m70100_t462ad10_133
10	twitter_m363_t455da10_263
11	twitpeep_m87_t450ad10_181
12	triltwit_m70100_t448uc10_064
13	triltwit_m87_t449cb11_394
14	trtwpp_m87_t4490d11_492
15	twitphee_m87_t449a010_263
16	tsikbark_m60107_t4592d11_055
17	tsikstrg_m87_t499e011_245
18	tsikstrg_m335_t451ea10_003
19	dtwitter_m86dtwit1nat
20	peepstrg_m87_t450aa10_034

A vocalization repertoire of 20 vocalizations was used. These vocalizations were recorded from the marmoset colony maintained at The Johns Hopkins University School of Medicine, and were sampled across animals of different ages and genders (Agamaite et al., 2015). Vocalization types, along with animal caller identities, are listed in **Table 2.1**. All 20 vocalizations were used for studying the WGN effect on information encoding rate of vocalizations.

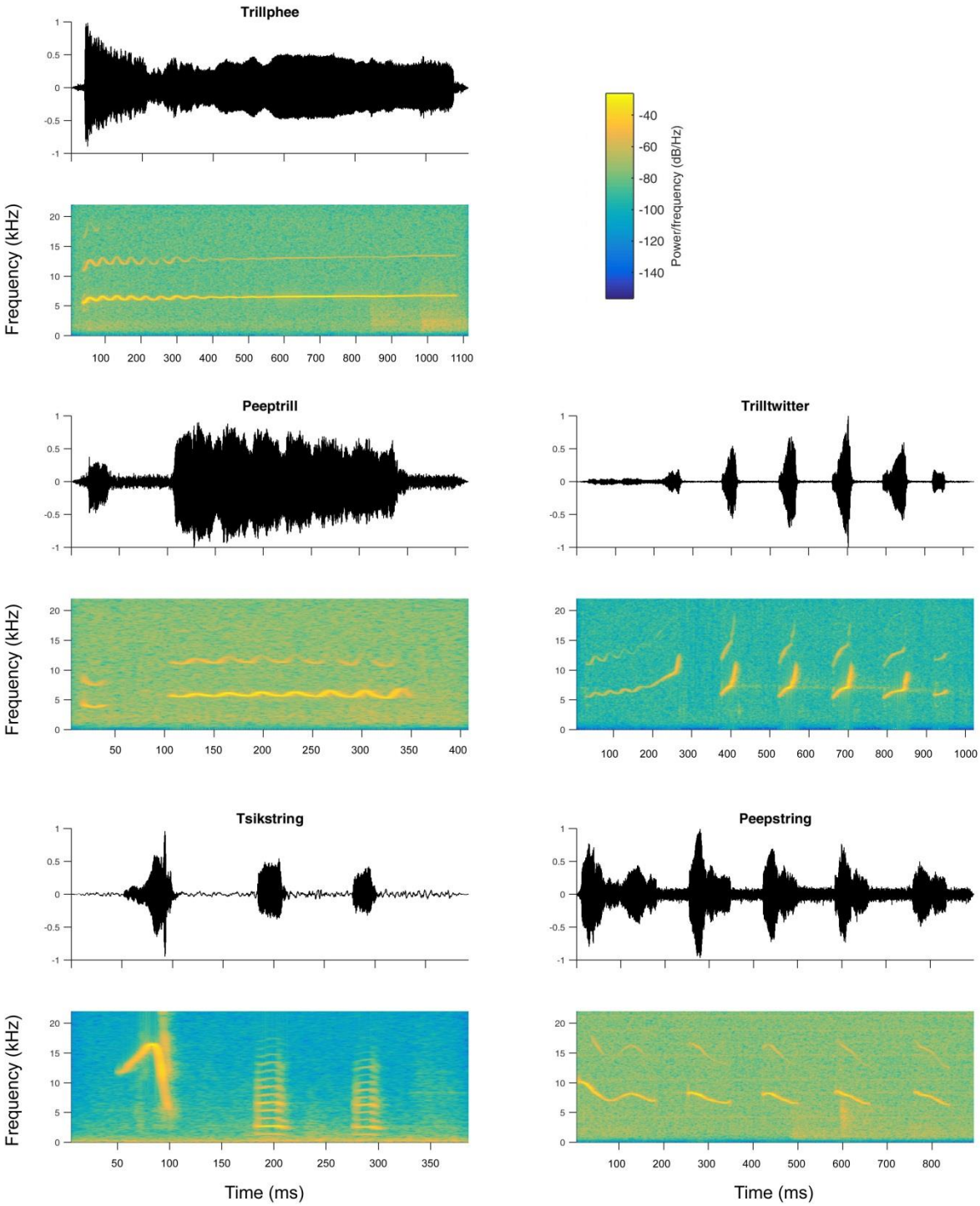

Figure 2.1 Temporal waveforms and spectrograms of five vocalizations: Trillphee, Peeptrill, Trilltwitter, Tsikstring, and Peepstring.

Five vocalizations were selected from the vocalization repertoire to further study the masking effect of different noises. They are No. 5 trilphee_m70100_4580010_020, No.7 peeptril_m60107_t459dd10_495, No.12 triltwit_m70199_t448uc10_064, No.18 tsikstrg_m335_t451ea10_003, and No.20 peepstrg_m87_t450aa10_034. These five vocalizations are called Trillphee, Peeptrill, Trilltwitter, Tsikstring, and Peepstring in the rest of the dissertation.

The temporal and spectral features of each vocalization are displayed in **Figure 2.1**. Among the five vocalizations, Trillphee, Trilltwitter, and Peepstring have durations around 1000 ms. Peeptrill and Tsikstring are two relatively shorter calls, with duration about 400 ms. In terms of complexity, Trillphee is a simple call, as it is a complete call without temporal gaps. Peeptrill, Trilltwitter, Tsikstring, and Peepstring are all considered compound calls, composed of multiple acoustic elements separated by gaps of less than 100 ms. The five vocalizations were selected to represent most of the acoustic features of the marmoset vocalization repertoire, so that the conclusions obtained from the dissertation could be generalized to the overall vocalization repertoire.

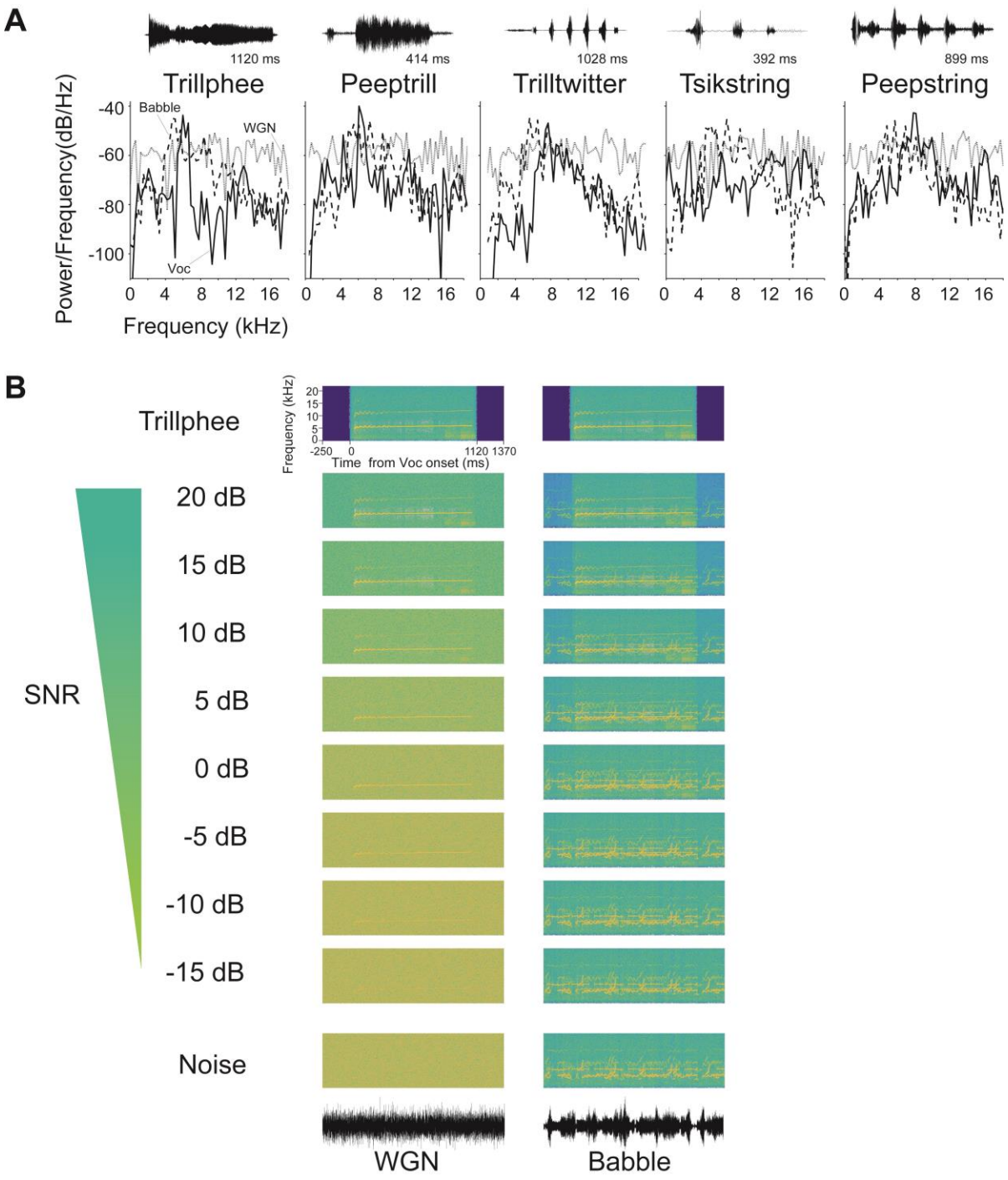

Figure 2.2 Acoustic stimuli used to investigate robust sound encoding in the primary auditory cortex. (A) Power spectrum of five vocalizations (solid lines), WGN (gray lines) and Babble noise (dashed lines). Background noises were truncated to have the same duration as each vocalization. The temporal waveform of each vocalization is displayed above each power spectrum. (B) Example spectrogram of vocalization Trillphee in noise at 10 different SNR levels, including pure noise and pure vocalization. The first column is Trillphee with WGN as background noise, and the second column is with Babble as background noise. The temporal waveforms of WGN and Babble are shown below each column.

Using MATLAB, two types of noise, WGN and Babble, were mixed individually with five natural marmoset conspecific vocalizations, generating noisy vocalizations at eight different signal-to-noise ratios (SNR; -15 dB to 20 dB at 5dB intervals, plus pure noise and pure vocalizations) as shown in **Figure 2.2**. The average spectral power of the noise at each SNR level $P_{noise(SNR)}$ was calculated relative to the average spectral power of pure vocalization P_{voc}, as in equation (2.1). The waveform of noise at each SNR $A_{noise(SNR)}$ was further scaled and added to the waveform of clean vocalization to generate the acoustic waveform of noisy vocalization A_{SNR} at each SNR level in equations (2.2) ~ (2.3). The resulting A_{SNR} was normalized between -1 and 1.

$$SNR = 10\log_{10}\left(\frac{P_{voc}}{P_{noise(SNR)}}\right), \tag{2.1}$$

$$\left(\frac{A_{noise(SNR)}}{A_{noise}}\right)^2 = \frac{P_{noise(SNR)}}{P_{noise}}, \tag{2.2}$$

$$A_{SNR} = A_{noise(SNR)} + A_{voc}. \tag{2.3}$$

In order to distinguish the onset responses induced by the components of noise and vocalization in the synthesized stimuli, a 250 ms interval of pure noise was concatenated to either end of each noisy vocalization. Babble sharing certain acoustic attributes of vocalizations was created by shuffling superimposed 50 ms-long pieces of four different randomly selected vocalization instances from the remaining 15 vocalizations in the repertoire (Trillpeep, Peeptrill, Twitterpeep, and Trillphee), which were different from the five test vocalizations. Both WGN

and Babble were synthesized with durations equivalent to the longest vocalization, the Trillphee. For the other four vocalizations, WGN and Babble were truncated to the same length as each vocalization.

2.3 Experimental Procedures

Acoustic stimuli were delivered in free-field through a loudspeaker (B&W 601S3, Worthing, UK) located 1 meter along the midline of and in front of the animal's head. The output of the speaker was calibrated so that the maximum sound level delivered was approximately 105 dB SPL with a flat frequency response from 60 Hz to 32 kHz (Watkins and Barbour, 2011). Single-unit activities in A1 were recorded from two alert adult marmoset monkeys while they passively listened to the playback of natural and synthesized conspecific vocalizations. Auditory neurons were detected based upon their responses evoked by pure tones and vocalizations. Once an auditory neuron was isolated, its characteristic frequency was estimated using random spectrum stimuli (RSS) (Barbour and Wang, 2003a) and/or pure tones to confirm that the response field of the neuron overlapped with at least some vocalization energy. Next, the rate-level function of each of the five vocalizations was measured at four intensities, ranging from 15 dB SPL to 75 dB SPL in 20 dB steps. Each of the 20 combinations was presented in random order at ten times. The intensity evoking the strongest responses to most of the vocalizations was selected to deliver noisy vocalizations, which were also randomly delivered at between five and ten repetitions. A rate-level function covering the same range of intensities as the vocalizations was also obtained for WGN.

In order to assess the neural responses as a function of SNR from the perspective of information theory, we also recorded a second data set from two additional alert adult marmoset monkeys using a similar recording procedure. The main difference is that each auditory neuron

in these additional experiments was first evaluated with rate-level functions of twenty marmoset vocalizations. The vocalization at a particular attenuation evoking the most modulated responses was identified visually. Neural responses to degraded versions of the vocalization were presented at eight SNR levels (-15 dB to 20 dB at 5 dB intervals, plus pure WGN and pure vocalization) and were further recorded at 30 to 50 repetitions. Given the limited time a single unit can be stably recorded in this procedure, we investigated only the information coding of WGN noisy vocalizations in this dissertation.

2.4 Data Analysis

Collected data were subjected to both single-unit analysis and population analysis. Single-unit analysis was emphasized to investigate the response properties of each individual neuron in response to acoustic stimuli, such as information coding rate, response reliability, vocalization intensity-invariance, and noise-invariance. Single-unit analysis revealed how neurons encode acoustic stimuli with sequences of action potentials. Detailed calculations for each single-unit analysis are explained in Chapter 3 and Chapter 4.

In contrast to single-unit analysis, population analysis investigated how individual neurons work together to discriminate vocalizations at different intensities/SNRs, without considering the particular identity of each individual neuron. Population analysis is mainly based upon building neural response classifiers to predict the identities of acoustic stimuli, which is a decoding process. How to build neural decoding models is described in Chapter 5.

Chapter 3: Feature-algined Responses to White Gaussian Noise

3.1 Introduction

The early stations of the auditory pathway, such as the auditory nerve fibers, can produce spike train patterns that faithfully follow the time-varying spectral features of complex stimuli, because auditory nerve fibers can encode fine temporal details of a stimulus (Delgutte and Kiang, 1984; Carney and Geisler, 1986). This kind of phase-locked response can be generated by auditory nerve fibers even in response to white Gaussian noise (WGN) (Ruggero, 1973). As the stimulus information encoded in spike trains is transmitted to higher stations of the auditory pathway, the neuronal responses are less likely to be faithful replicas of the acoustic stimuli. For instance, a majority of neurons at the level of the primary auditory cortex (A1) produce responses following the envelope of marmoset vocalizations, synchronized to the phrases of the vocalizations, instead of the fine temporal details (Wang et al., 1995). Neurons in A1 generate even sparser responses to encode WGN (de Boer and Kuyper, 1968; Aertsen and Johannesma, 1981; Valentine and Eggermont, 2004). Therefore, WGN is conventionally considered a poor stimulus to drive neurons in A1.

In our preliminary dataset, however, we discovered A1 neurons with phase-locked responses to WGN, as if they were synchronized to spectrotemporal features in the same way they often are to complex stimuli. This finding drove us to investigate the hypothesis that there exists a positive correlation between neurons' encoding properties for WGN and vocalizations. In addition, the proportion of A1 neurons with feature-aligned responses was quantified.

3.2 Data Analysis

In order to compare the distribution of single-unit response reliability to pure noise and pure vocalizations, we pooled neurons from the WGN noisy vocalization datasets of four monkeys.

We adopted a correlation metric proposed by Schreiber *et al.*, to evaluate neuronal response reliability across repetitions (Schreiber et al., 2003). Spike trains of neural responses to the same stimulus presented n times were first convolved with a Gaussian filter having a window length of 50 ms to obtain the vectors $\vec{S}_i$ (i = 1, …, n). Correlations between all pairs of filtered spike train vectors $\vec{S}_i$ and $\vec{S}_j$ were computed, and the resulting average correlation value was defined as the response reliability of that single unit to a particular stimulus, as displayed in equation (3.1). If either $\vec{S}_i$ or $\vec{S}_j$ was empty, the correlation between them was set to zero.

$$R = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\vec{S}_i \cdot \vec{S}_j}{\left\| \vec{S}_i \right\| \left\| \vec{S}_j \right\|}. \tag{3.1}$$

This correlation value ranges from zero to one, with higher values denoting more consistent responses across trials.

We implemented an information measurement in the extended dataset. Information theoretic measures can describe how much information about a stimulus is encoded in neural responses. As described in other studies (de Ruyter van Steveninck et al., 1997; Vinje and Gallant, 2002), the information content of the spike trains was calculated in the following way. First, the full complement of spike trains resulting from the same experimental condition for

each unit was digitized by counting the number of spikes within bins of width $\Delta\tau$, using non-overlapping rectangular windows. By specifying the number of letters constituting a word, we defined a K-letter word, which had a time length of $T = K \times \Delta\tau$. The total response variability of the words in the spike trains is termed total response entropy, and is given by the following equation:

$$H(r) = -\sum_{W} P(W) \log_2 P(W), \tag{3.2}$$

where $P(W)$ is the occurrence probability of word W through all the spike trains. Total response entropy quantifies the variations across time and the capacity of the spike train to carry information. By quantifying the variability of the responses across time from trial to trial, conditional response entropy can be defined as

$$H(r \mid s) = -\sum_{W} P(W \mid t) \log_2 P(W \mid t), \tag{3.3}$$

where $P(W \mid t)$ is the probability of obtaining the word W at time t. Finally, mutual information (called "information" for short in what follows) between the spike train and the stimulus quantifies the amount of variation in the spike train resulting from changes in the stimulus. It is simply the difference between the total response entropy and conditional response entropy:

$$I(r,s) = H(r) - H(r \mid s). \tag{3.4}$$

By normalizing the information in (3.4) with word time length T, we can obtain the information rate in bits/s. We can also calculate information per spike by dividing the information rate by the average number of spikes generated during word length T. In addition, information efficiency measures the fraction of available bandwidth that a neuron actually uses

to transmit information, i.e., the ratio of the amount of information actually transmitted divided by the theoretical maximum amount of information that could be transmitted, as in (3.5):

$$E = \frac{H(r) - H(r \mid s)}{H(r)} = \frac{I(r,s)}{H(r)}. \tag{3.5}$$

The information rate and efficiency of neurons with empty responses to a particular stimulus were set to 0. We implemented information calculations using $\Delta\tau$ = 3, 5, 10, 20, and 30 ms and K = 1, 2, and 3, yielding qualitatively similar trends for all combinations. The results presented in this study used $\Delta\tau = 3$ ms and K = 1.

3.3 Results

3.3.1 The Information Content of Neural Responses Decreases as a Function of SNR under the WGN Condition

Information theory provides a useful tool for neuroscientists to uncover important features of sensory processing and perception (Abolafia et al., 2013). We studied how much information was transmitted via spike trains of A1 neurons in response to vocalizations in WGN as a function of SNR. For each single unit, one out of twenty vocalizations was selected for further study because it evoked the greatest response. Twenty sample vocalizations spanning the marmoset vocalization repertoire were delivered in a WGN background at eight different SNRs (-15 to 20 dB, in 5 dB steps) and repeated for 30-50 repetitions. Altogether, 273 single units were isolated from two marmoset monkeys and 191 single units passing the response criteria were analyzed. The distribution of the neuron counts for each vocalization is shown in **Figure 3.1A**. Some vocalizations elicited neural activity more commonly than others, such as vocalization No. 7, Peeptrill.

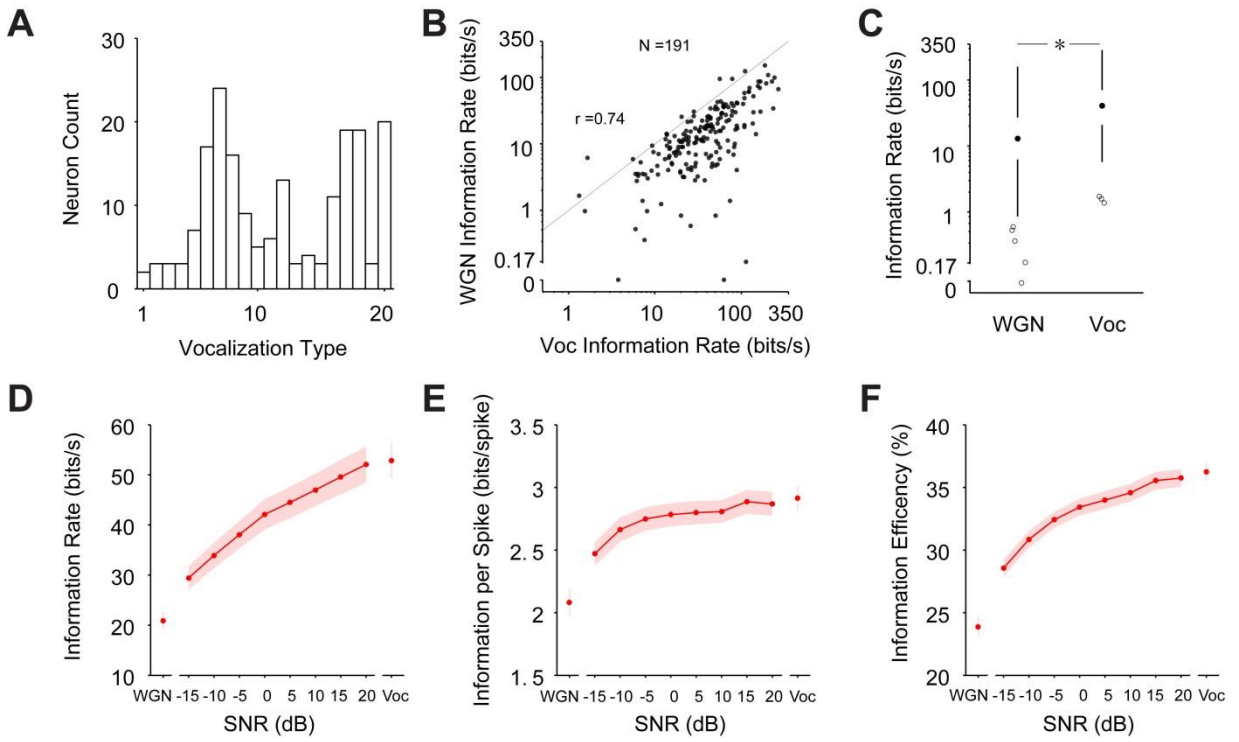

Figure 3.1 Vocalizations consistently elicit more informative responses than WGN. (A) Distribution of number of responsive units to twenty different vocalizations spanning the marmoset repertoire. (B) Scatter plot of information rate of single units to pure vocalization versus pure WGN. (C) Boxplot of information rate of single units to pure vocalization and pure WGN. The asterisk denotes that the information rate of WGN is significantly lower than pure vocalization (see text). (D) Population average information rate (mean ± s.e.m) of vocalizations under WGN background as a function of SNR. (E) Population average information per spike (mean ± s.e.m) of vocalizations under WGN background as a function of SNR. (F) Population average information efficiency (mean ± s.e.m) of vocalizations under WGN background as a function of SNR.

By calculating the information embedded within spike trains based upon the distribution of the occurrence of spiking "words", (i.e., the number of spike counts in a defined bin width of 3 ms), we determined the relationship between information rate (information normalized by time bin width) elicited by pure vocalizations and by pure WGN in **Figure 3.1B.** The result indicates that most single units in A1 encode natural vocalizations with a higher information rate than WGN, and these two values are fairly highly correlated (r = 0.74). The trend can be better observed in **Figure 3.1C**, with a paired Wilcoxon signed-rank test showing significantly higher

information rate of pure vocalizations (Z = -11.52, p = 1.0×10^{-30}). Extending the calculation of information along the SNR axis, a monotonically increasing information rate is revealed in the population average in **Figure 3.1D** as SNR increases. Interestingly, the information rate of WGN was still relatively high in absolute terms, rather than dropping close to zero. It is plausible that the high information rate originates primarily from a high discharge rate, given that vocalizations usually induce higher discharge rates than WGN.

What about the amount of information transmitted by each spike? Are vocalizations encoded at a faster information rate because each spike conveys more information than WGN? We address this question in **Figure 3.1E**, where the information rate normalized by the mean number of spikes occurring per word length is displayed against SNR. As expected, individual spikes carry more information about a stimulus at increasing SNR values. Although decreasing vocalization content leads to a nearly linear drop in information rate (Figure 3.1D), the decrease in the information amount transmitted by each spike is rather shallow across SNRs down to very low SNRs. This result suggests that auditory cortical neurons not only encode natural stimuli such as conspecific vocalizations with more spikes overall, but also are more efficient in transferring information by individual spikes for such stimuli compared to behaviorally irrelevant noise. The high encoding efficiency for vocalizations is further revealed in **Figure 3.1F**, which shows the efficiency of total entropy in neural responses transformed into stimulus-relevant information. Information efficiency followed the same trend as information rate.

3.3.2 Feature-aligned Response to White Gaussian Noise

The neural responses underlying the aforementioned non-trivial information content encoded for WGN are of potential interest. Units with strong temporal-structured responses to WGN were found in all four marmosets. Feature-aligned responses to WGN are well established

in the auditory periphery (de Boer and Kuyper, 1968; Ruggero, 1973), whereas the response to unmodulated WGN in auditory cortex has been deemed very inefficient, equivalent to the poor stimuli for high-level auditory neurons in animals such as songbirds and cats (Valentine and Eggermont, 2004; Theunissen and Elie, 2014). The original rationale for using WGN as one of the masking noises in our study was based upon the assumption that reliable A1 responses to WGN would be insignificant. In that case, information measures of responses to vocalizations could be evaluated on an absolute scale. Figure 3.1 clearly shows that substantial stimulus-specific information exists in spike trains of A1 neurons even when WGN is delivered, implying that at least some individual cortical neurons must be encoding WGN directly.

We therefore directly examined the feature-aligned WGN stimulus-encoding properties of A1 neurons. This temporal alignment can be viewed from repeated presentations of the same frozen WGN. Units with feature-aligned spiking activities in response to WGN were arbitrarily selected as having a response reliability value equal to or greater than 0.5. Forty-eight WGN-reliable neurons from one monkey with feature-aligned responses to the same WGN are displayed in **Figure 3.2A.** Structured responses to WGN are visually apparent in this subset of units because stimuli were repeated for 20-50 trials. Except for four units with only onset-aligned responses, the majority of neurons exhibited temporally aligned spikes at various time points of the WGN.

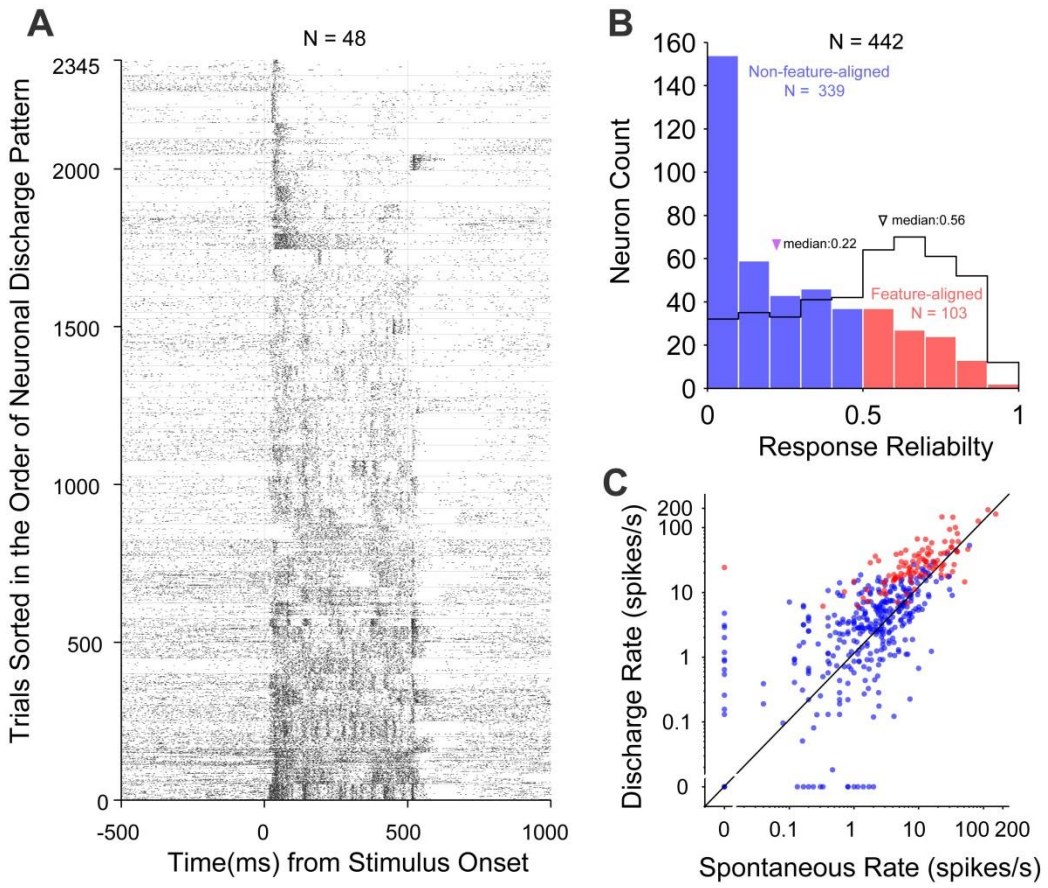

Figure 3.2 A large number of A1 neurons generate spikes that are feature-aligned to WGN. (A) Raster plots of example single units with feature-aligned responses to a single frozen WGN. (B) Histogram of single units' response reliability to pure WGN (in red and blue) and pure vocalizations (black line) from WGN noisy vocalization protocols. Feature-aligned (in red) and non-feature-aligned responses (in blue) to WGN were differentiated with a response reliability threshold value of 0.5. (C) Scatterplot of the spontaneous rate and the discharge rate to pure WGN of single units in (B).

By investigating the neural responses to pure WGN for all four monkeys (N = 442), we obtained the distribution of single unit response reliabilities shown in **Figure 3.2B.** In comparison to the WGN response reliability distribution with a median of 0.22, vocalization responses resulted in a right-shifted distribution with a median of 0.56. Around 20% of A1 units exhibited an ability to encode WGN in a feature-aligned fashion by our reliability criterion ≥ 0.5. In **Figure 3.2C**, we also plotted the corresponding spontaneous rates and discharge rates, to

evaluate any potential association between feature-aligned response property and neuron activity. It turns out that A1 neurons are more likely to have a structured response to WGN when their discharge rate is above 5 spikes per second. While feature-aligned units have spontaneous rates spanning a wide range, from highly inactive to extremely active, more units are identified as feature-aligned when their spontaneous rates are above 1.5 spikes per second.

3.3 Discussion

A widely accepted yet poorly documented viewpoint regarding the responsiveness of A1 neurons under awake conditions to unfiltered, unmodulated WGN is that this represents a poor stimulus class to drive these neurons, presumably because of the frequency selectivity of the neurons active under awake conditions and the relatively sluggish response of A1 neurons (Depireux et al., 2001; Elhilali et al., 2004; Nelken, 2004). In practice, filtered noise is often used when a simple stimulus having bandwidth wider than a pure tone is desired (Wang et al., 2005). We had predicted given this conventional wisdom that the average information rate of A1 spike trains in response to WGN would be near 0 bits/s. To our surprise, this rate was actually 20 bits/s, and many neurons had visible feature-aligned spiking responses to WGN that were robust across different stimulus intensities. A continuum of spiking reliability exists across the population studied, and we estimate that about 20% of our population could comfortably be classified as generating feature-aligned spikes in response to WGN.

Perceptually this finding raises interesting questions because it seems unlikely that individuals could reliably distinguish two distinct WGN stimuli with the same statistics, yet the neural population, even at high levels of the auditory system, would be able to do so. The possibility exists that training might influence top-down modulation in order to improve WGN discrimination, though with unknown significance. Furthermore, using pure noise stimuli

without requiring intermediate stimulus modifications, the original formulation of receptive field estimation using spike-triggered averages might be useful in a subset of central auditory neurons (de Boer and Kuyper, 1968; Aertsen and Johannesma, 1981; Eggermont et al., 1983), much as can be done in auditory nerve because of temporally aligned spiking to WGN (Ruggero, 1973).

Natural communication sounds contain more semantic information than typical experimental sounds. In the auditory nerve, for example, naturalistic stimuli are encoded at a higher information rate than WGN (Rieke et al., 1995). We have shown that in A1 neurons, as well, information in natural communication sounds was transmitted at a higher rate than WGN. Naturalistic noise has been shown to decrease information in different temporal codes as noise increases (Kayser et al., 2009). Additionally, overall information content in spike trains decreases along the ascending auditory pathway (Chechik et al., 2006), implying that the nature of vocalization coding changes at higher levels of processing, as well.

Chapter 4: Contextual Effects of Noise on Vocalization Encoding in the Primary Auditory Cortex

4.1 Introduction

In natural settings, behaviorally relevant acoustic signals usually co-occur with other acoustic sources. Therefore, the auditory system's ability to process multiple competing sound sources is closely linked with our ability to perceive individual sounds. Although humans and animals exhibit reliable auditory detection against substantial amounts of noise, the underlying neural representation of sound in such contexts is still not well understood.

Individual auditory neurons are believed to represent behaviorally relevant natural sounds effectively, particularly animal vocalizations and human speech. Animal call/song-selective neurons have been discovered in multiple sensory system models, such as crickets, frogs, songbirds, guinea pigs and non-human primates (Newman and Wollberg, 1973; Feng et al., 1990; Libersat et al., 1994; Wang et al., 1995; Grace et al., 2003; Grimsley et al., 2012). To cope with distortion induced by noise, mammalian auditory cortex appears to be actively involved in recovering the disrupted upstream neural representation (Anderson et al., 2010). Neurons in primary auditory cortex (A1) have been found to be sensitive to the masking component of complex stimuli (Bar-Yosef and Nelken, 2007) , and neuronal adaptation to stimulus statistics has been identified to be responsible for building noise-invariant responses (Rabinowitz et al., 2013; Willmore et al., 2014). In the auditory cortex of humans, low-frequency activity has been suggested to provide neural cues for stable speech recognition against both energetic and informational masking (Ding and Simon, 2012, 2013). This line of research is more advanced in

avian models: individuals neurons in higher avian auditory brain regions have been identified with robust encoding of noisy vocalizations (Moore et al., 2013; Schneider and Woolley, 2013) . Furthermore, multiple complex background maskers affect neuronal discriminability differently, but behavioral discriminability is degraded to the same degree regardless of masking type (Narayan et al., 2007).

Relatively few studies, however, have investigated reliable auditory discrimination of complex sounds in noise in nonhuman primates. One study has explored neural coding of degraded marmoset twitter calls in anesthetized marmosets, showing robust neural responses to vocalizations at medium signal-to-noise ratios (SNR) (Nagarajan et al., 2002). But how noisy vocalizations are encoded in awake marmoset auditory cortex remains uncertain. In addition, only one type of vocalization in a single type of noise was studied with limited levels of SNR; therefore no generalized conclusions can be drawn for the quantitative effects of noise on the neural representation of natural calls.

Marmoset vocalizations contain acoustic information over distributed frequencies and a wide range of time scales (DiMattina and Wang, 2006; Agamaite et al., 2015). Neurons in marmoset A1 exhibit multiple encoding strategies (Barbour and Wang, 2003b; Wang, 2007; Watkins and Barbour, 2011), with a majority of them responsive to vocalizations. Here we evaluate the effects of white Gaussian noise (WGN) and four-marmoset-talker babble (Babble) on the auditory cortical representation of conspecific vocalizations in alert adult common marmoset monkeys. Babble noise covers a similar frequency range as natural vocalizations, while WGN has considerably different statistics and has historically been considered to be a poor stimulus for driving neurons in higher auditory areas (Miller and Schreiner, 2000; Theunissen et al., 2000; Valentine and Eggermont, 2004). We predicted that Babble noise would generally

result in more disruption of the neural representation of targeted vocalizations than WGN and that a subclass of vocalization-responsive neurons would preferentially be responsible robust vocalization encoding in the face of either type of noise. In particular, we predicted that neurons robustly encoding vocalizations across intensity might be also encode them robustly across noise classes. While we did discover clear evidence of robust vocalization encoding in awake marmoset A1, the form this encoding takes appears to be considerably more complex than originally anticipated.

4.2 Data Analysis

Because average neural responses to vocalizations in A1 do not necessarily surpass mean spontaneous activity (Wang, 2007), we implemented a relatively loose criterion for defining responsive neurons in our dataset. Neurons generating at least one spike in the presence of a clean vocalization in at least 50% of the trials were defined to be responsive. A total of $N = 326$ single units were extracellularly isolated from A1 of two marmosets. After applying the responsiveness criterion for each of the five vocalizations, 216 (Trillphee), 191 (Peeptrill), 222 (Trilltwitter), 200 (Tsikstring) and 224 (Peepstring) single units were included for further data analysis. All data analyses were conducted in MATLAB R2014a (The MathWorks Inc, Natick, MA).

For each unit of the dataset, mean spontaneous rates and mean discharge rates (excluding the two concatenated noise portions) were measured for each stimulus. Neuronal response reliability across repetitions was calculated in the same way as in Chapter 3 (Schreiber et al., 2003). Basically, spike trains were first binned into vectors with a 50 ms window. Pairwise correlations of spike trains were computed for all trials of neuron responses to the same stimulus. The response reliability of a single unit to a particular stimulus is the averaged correlation.

Correlation between empty spikes trains were set to zero. This correlation value ranges from zero to one, with higher values denoting more consistent responses across trials.

To evaluate the influence of noise upon neural representation of vocalizations, we quantified the amount of vocalization encoded by single neurons at a particular SNR level by calculating an extraction index (EI) adapted from a similar study in songbirds (Schneider and Woolley, 2013). This metric is based upon the repetition-averaged peristimulus time histogram (PSTH) of neural response, a temporal sequence of spike counts, with a time bin of 50 ms. Different window bins of 5 ms, 10 ms, and 20 ms were also evaluated, which yielded qualitatively similar results. In this chapter, we only report results based upon 50 ms time bins. The initial and final 250 ms-long noise segments were again excluded from the PSTH during this analysis. The number of time bins used to calculate EI is 113 (Trillphee), 43 (Peeptrill), 104 (Trilltwitter), 41 (Tsikstring), and 91(Peepstring). EI is computed as in equation (4.1):

$$EI = \frac{D_{nsnr} - D_{vsnr}}{D_{nsnr} + D_{vsnr}}, \tag{4.1}$$

$$D_{nsnr} = 1 - \frac{\vec{P}_n \cdot \vec{P}_{snr}}{\left\|\vec{P}_n\right\| \left\|\vec{P}_{snr}\right\|}, \quad D_{vsnr} = 1 - \frac{\vec{P}_v \cdot \vec{P}_{snr}}{\left\|\vec{P}_v\right\| \left\|\vec{P}_{snr}\right\|}, \tag{4.2}$$

where D_{nsnr} is the distance between PSTHs $\vec{P}_n$ of noise and $\vec{P}_{snr}$ of vocalization at a particular SNR, while D_{vsnr} is the distance between $\vec{P}_v$ of pure vocalization and $\vec{P}_{snr}$. EI is bounded between -1 and 1: a positive value indicates the neural response to noisy vocalization is more vocalization-like, and a negative value implies the neural response is more noise-like. The EI profile for each single unit was determined by computing EI at every SNR level. The normalized

inner product was utilized to compute distance between $\vec{P}_n / \vec{P}_v$ and $\vec{P}_{snr}$, as shown in (4.2). For computational purposes, empty PSTHs were replaced by a vector generated from a Gaussian distribution with mean rate of zero and standard deviation of 0.001 so that we could report the distance between two empty PSTHs as 0. In order to reduce the artifact introduced by using an artificial PSTH vector, while calculating distance between an empty PSTH and non-empty PSTH, the non-empty PSTH was augmented with the same artificial vector used to replace the empty PSTH.

To probe the hidden response patterns, we further implemented an exploratory analysis based upon the calculated EI profiles. By applying k-means clustering on the blended EI profiles from both noise conditions together, we obtained subgroups of EI profiles, which divided single units into clusters according to the similarity of their EI profiles. Similarity was quantified by Euclidean distance. The number of clusters was determined by the mean squared error (MSE) of clustering as in equation (4.3), where N is number of neurons, EIP_i is the EI profile of a single neuron, and $\overline{EIP}_{cluster-i}$ is the mean EI profile of the cluster that this neuron is categorized into.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\overline{EIP}_{cluster-i} - EIP_i\right)^2 . \tag{4.3}$$

We selected the number of clusters based upon two criteria. We first narrowed down the candidates of number of clusters to the ones where the decrease of MSE flattens according to the elbow method (Tibshirani et al., 2001). We furthered selected the appropriate number of clusters from the candidates, which yielded response groups with distinguished functionality in terms of their resistance to noise. This analysis was performed by pooling EI profiles from all five

vocalizations together for each unit. Hierarchical clustering was also implemented and yielded very similar results; therefore, we report the results of k-means clustering only.

To examine if a link exists between neurons' intensity-invariance and noise-resistance, discriminative analysis of single units responding to vocalizations at varied intensities was implemented to compute an intensity-invariance index (Billimoria et al., 2008). This value represents the discriminability of neural representations of a particular vocalization at multiple intensities from all the other vocalization-intensity instances. Neural responses to five vocalizations at four intensity levels (15, 35, 55, and 75 dB SPL) were truncated to the same length as the vocalization with the shortest duration. To calculate the intensity-invariance index for a particular vocalization, a single trial of neural responses at 75 dB SPL was randomly selected for each of five vocalizations as master templates and all the remaining trials were classified into the vocalization type as the most similar master trial based upon normalized inner product metric. The process was repeated 100 times. Classification accuracies were obtained for the investigated vocalization at each intensity level. We measured the deviation of the classification accuracies C_i ($i = 1, 2, ..., N_{Atten}$) at all tested intensities from the classification accuracy C_{master} of the intensity of the master template, denoted as I in (4.4), where N_{Atten} is number of tested intensities. We further linearly scaled I so that it takes values between zero and one and therefore represents an intensity-invariance index.

$$I = \sqrt{\frac{1}{N_{Atten}} \sum_{i=1}^{N_{Atten}} \left(\frac{C_i - C_{master}}{C_{master}} \right)^2} \quad . \tag{4.4}$$

We quantified the alternation in the number of spiking activities induced by noisy vocalizations relative to clean vocalizations by measuring the firing rate within and between

vocalization phrases. For each vocalization, we manually segmented to mark the temporal boundaries of phrases and gaps between phrases. We furthered computed the firing rate change within each phrase and gap for noisy vocalization at 20 dB SNR relative to clean vocalization. The resulted differences in firing rates within and between phrases were averaged across trials, phrases/gaps and vocalizations.

Normality was verified by the Lilliefors test. Unless otherwise indicated, hypothesis testing was conducted using a two-sided Wilcoxon signed-rank test. The significance criterion was set to 0.05.

4.3 Results

4.3.1 Mean Discharge Rate and Response Reliability both Decreases as SNR Decreases

We recorded single-unit responses in A1 to five vocalizations embedded within two different background noises, WGN and Babble, at multiple SNR levels (–15 dB to 20 dB, 5 dB steps) from two marmosets. **Figure 4.1A** shows an example of a typical neuronal response. For this example unit, all vocalizations presented alone evoked neural responses that were locked to particular acoustic features, which can be observed from the temporal patterns formed by aligned spikes in the raster plots. As the amount of noise in the stimuli increased, the neural responses gradually deviated from the pure vocalization response. Neural encoding of vocalizations was particularly susceptible to the presence of Babble noise, given that spikes corresponding to the acoustic features of target vocalizations started to diminish even at 20 dB SNR, where the noise component was quite small. Responses at lower SNR values appeared to be mainly dominated by the Babble noise component. In comparison, this particular unit's responses to WGN were not as strong as responses to vocalizations. As a result, the temporal firing patterns to vocalization

components were maintained at lower SNR levels. Additionally, response nonlinearities can also be discerned. For example, responses to vocalization Peepstring between -5 dB and -15 dB SNR under the WGN condition elicited stronger activity than in response to vocalization or WGN present alone.

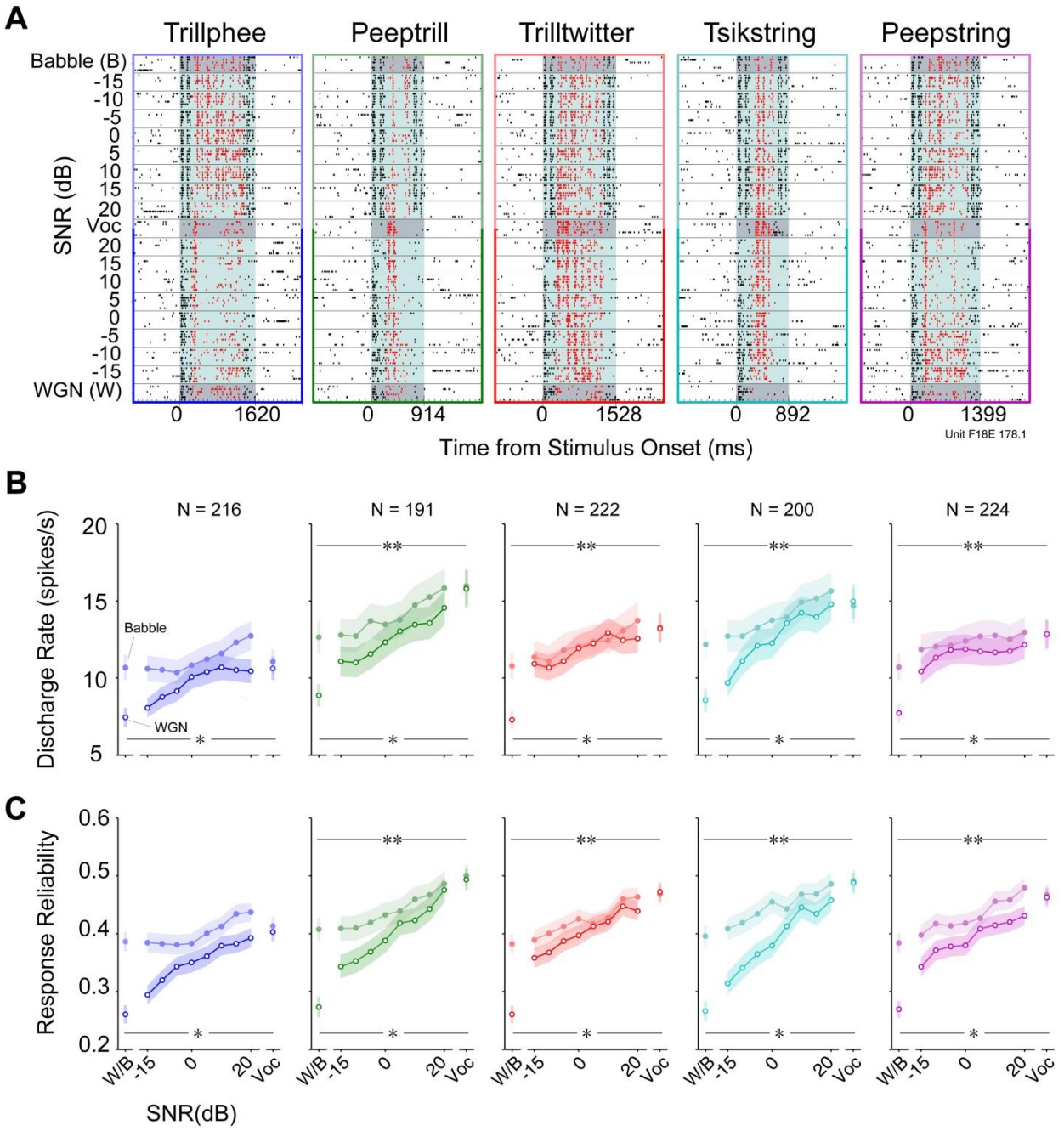

Figure 4.1 Clean vocalizations generally elicit the most spiking and the most reliable spiking. (A) Example spike raster plots of one unit's responses to five noisy vocalizations under two different noise conditions. Each point denotes an action potential generated by this unit. Black dots with white background are the neural activities in silence. Black dots with light blue backgrounds are neural activities occurring during two concatenated noise presentations. Red dots with light blue background are spikes driven by noisy vocalizations. Response to vocalizations and noises alone are highlighted with gray background for distinction. (B) Mean discharge rates to five vocalizations at multiple SNRs (mean ± s.e.m). Lines with filled circles are discharge rates of vocalizations in Babble. Lines with open circles are discharge rates of vocalizations in WGN. Single-asterisk/double-asterisk indicates average discharge rate to WGN/Babble alone is significantly lower than vocalization alone (see text).(C) Mean response reliabilities to five vocalizations at multiple SNRs (mean ± s.e.m). The same color and line type

denotations as in (B) are depicted. Single-asterisk/double-asterisk indicates mean discharge rate to WGN/Babble alone is significantly lower than vocalization alone (see text).

To quantify noise effects upon the discharge rates in response to noisy vocalizations as a function of SNR, we calculated the mean discharge rate (absolute spiking rate without subtracting spontaneous spiking rate) evoked during noisy vocalizations in **Figure 4.1B**. Generally, mean discharge rates of single neurons to vocalizations masked with both noises increased as SNR increased. Discharge rates of pure WGN were significantly lower than discharge rates of all pure vocalizations (Trillphee: $Z = -5.31$, $p = 1.10\times10^{-7}$; Peeptrill: $Z = -7.88$, $p = 3.40\times10^{-15}$; Trilltwitter: $Z = -8.91$, $p = 5.10\times10^{-19}$; Tsikstring: $Z = -7.95$, $p = 1.84\times10^{-15}$; Peepstring: $Z = -9.09$, $p = 9.04\times10^{-22}$). With regard to Babble, the same trend was observed for four out of five vocalizations (Trillphee: $Z = -0.438$, $p = 0.661$; Peeptrill: $Z = -5.71$, $p = 1.10\times10^{-8}$; Trilltwitter: $Z = -5.37$, $p = 7.74\times10^{-8}$; Tsikstring: $Z = -3.91$, $p = 9.10\times10^{-5}$; Peepstring: $Z = -4.28$, $p = 1.83\times10^{-5}$). It is noteworthy that under Babble noise conditions, more spikes were evoked on average at higher SNR levels than clean vocalizations. One extreme case is vocalization Trillphee, to which the neural responses were as strong as or even stronger at all SNR levels than the vocalization component alone. In addition, the mean responses to WGN alone were relatively weaker than responses to Babble, but were still comparable. This is surprising given the traditional view of WGN that it is a poor stimulus for activating auditory neurons at higher processing levels (Miller and Schreiner, 2000; Theunissen et al., 2000; Valentine and Eggermont, 2004).

Due to the stochastic nature of neural spike timing, neurons produce varied spike trains in response to a stimulus presented multiple times (de Ruyter van Steveninck et al., 1997; Oram et al., 1999). We also examined the noise-induced alteration in response reliability of neurons to

vocalizations. By computing the mean value of pair-wise correlation between individual spike trains, we quantified stability of single-neuron response as a function of SNR in **Figure 4.1C**. In the WGN condition, more noise resulted in decreased neural response reliability for all five vocalizations. Additionally, pure WGN induced a significantly lower response reliability than pure vocalizations (Trillphee: Z = -7.15, p = 8.6×10^{-13}; Peeptrill: Z = -9.05, p = 1.5×10^{-19}; Trilltwitter: Z = -10.34, p = 4.46×10^{-25}; Tsikstring: Z = -9.53, p = 1.54×10^{-21}; Peepstring: Z = -10.55, p = 5.04×10^{-26}). Although the response reliability degradation induced by Babble noise was weaker, the response reliability was still significantly lower than most pure vocalizations except Trillphee (Trillphee: Z = -1.88, p = 0.0608; Peeptrill: Z = -4.46, p = 8.34×10^{-06}; Trilltwitter: Z = -5.69, p = 1.26×10^{-8}; Tsikstring: Z = -4.97, p = 6.54×10^{-7}; Peepstring: Z = -5.08, p = 3.81×10^{-7}).

To summarize, vocalizations generally elicited more spikes and more reliable spikes than either noise alone in this neuronal population. A more prominent increase in neuronal discharge rate and response reliability was observed as SNR increased in the WGN case compared to Babble. Finally, rates and reliability elicited by WGN were both higher than anticipated at this high level of auditory processing.

4.3.2 Low Correlation between Single Units' Resistances to Different Noises

We next studied the ability of single units to consistently encode vocalizations despite the influence of background noises by calculating an EI profile for each neuron as a function of SNR. The EI measurement was previously implemented in a songbird study and was demonstrated to be able to reflect single neurons' vocalization-coding ability more accurately than average discharge rate, given that both vocalization and noise components elicited strong responses in that study (Schneider and Woolley, 2013). Essentially, EI is designed to quantify

whether a trial-averaged response is more vocalization-evoked or more noise-evoked. For example, EI of 1 suggests that the evaluated response is evoked by vocalization alone, and EI of -1 suggests the response is evoked by noise alone. By calculating EI at each SNR level, we obtained an EI profile for each neuron. We further computed the mean value of EI profile at SNR levels ranging from -15 dB to 20 dB to obtain an overall picture of the pairwise relationship between single-neuron responses to vocalizations mixed with WGN and Babble. The results are shown in **Figure 4.2**.

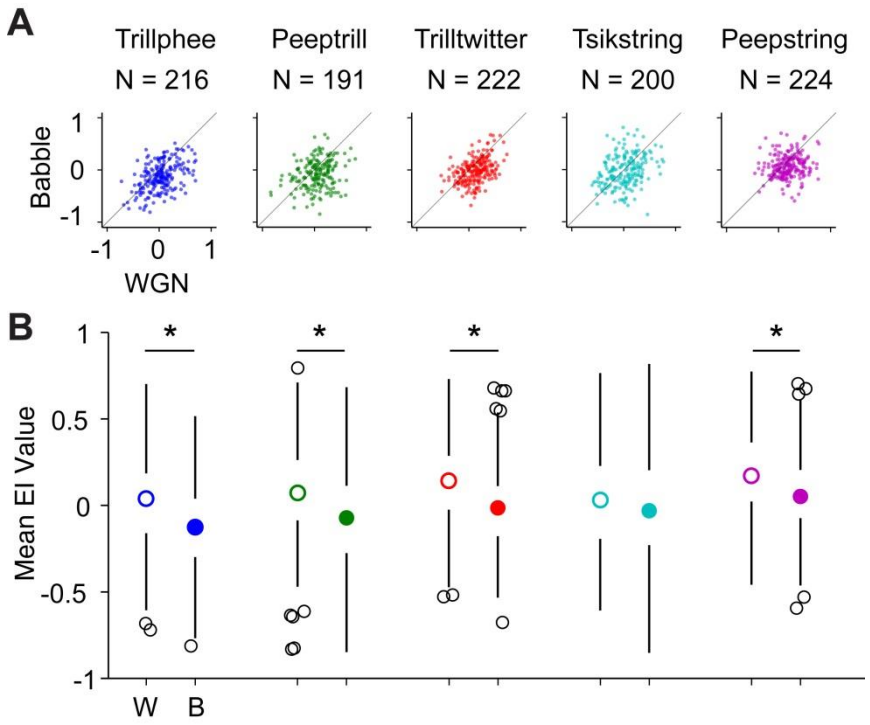

Figure 4.2 Babble tends to disrupt vocalization encoding more than WGN. (A) Scatter plot of mean of EI profile values under WGN and Babble conditions. Neuron counts are indicated above each plot.(B) Boxplots of medians of EI profile mean values under WGN and babble conditions. Neuron counts are the same as in (A). Boxplots with open circles display distribution of mean EI values under WGN condition, and boxplots with filled circles display distribution of mean EI values with background of Babble noise. Outliers are indicated by black open circles. Asterisks indicate mean EI values under WGN and Babble are significantly different (see text).

We visualized the data in two ways. In **Figure 4.2A**, EI profile mean values of single units under WGN and Babble were first scattered so that the pairwise relationship can be better observed and quantified by Pearson correlation. This analysis indicates that a weak positive correlation exists between single units' rejection of WGN and Babble for vocalization Trillphee, Trilltwitter and Tsikstring (Trillphee: $r = 0.413$, $p = 3.09\times10^{-10}$, Trilltwitter: $r = 0.424$, $p = 5.03\times10^{-11}$, Tsikstring: $r = 0.332$, $p = 2.41\times10^{-6}$), while the correlation is low for vocalization Peeptrill and Peepstring (Peeptrill: $r = 0.191$, $p = 8.70\times10^{-3}$, Peepstring: $r = 0.130$, $p = 4.41\times10^{-2}$). This result implies that a relatively large variability exists in the relationship of individual neurons' resistance to disruption by WGN and Babble. In other words, the ability to predict a single unit's responses to vocalizations in WGN based upon its responses to vocalizations in Babble is limited, and vice versa.

Additionally, Figure 4.2A demonstrates that more points are located below the diagonal than above. We can better observe this trend in **Figure 4.2B** across vocalizations**,** where medians of population EI mean values are mostly positive in WGN while medians of population EI mean values are mostly negative in Babble. A paired Wilcoxon signed-rank test shows that the mean EI values under the two noises are significantly different for four out of five vocalizations (Trillphee: $Z = 7.43$, $p = 1.13\times10^{-13}$; Peeptrill: $Z = 5.04$, $p = 4.55\times10^{-7}$; Trilltwitter: $Z = 7.27$, $p = 3.59\times10^{-13}$; Tsikstring: $Z = 1.69$, $p = 0.0900$; Peepstring: $Z = 5.67$, $p = 1.39\times10^{-8}$). Therefore, while there is large variability among single unit responses to vocalizations under different noise conditions, a majority of neurons is more resistant to vocalization degradation by WGN than by Babble.

4.3.3 Intensity-invariance is Insufficient to Account for Noise-resistance

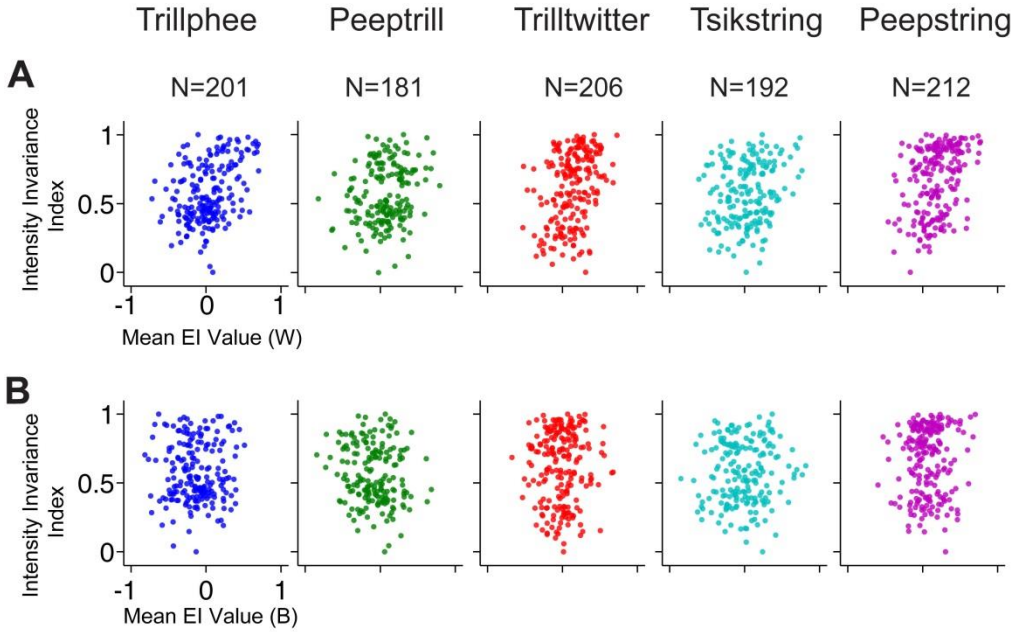

Figure 4.3 Intensity invariance correlates poorly with noise resistance. (A) Intensity-invariance versus noise-resistance in WGN condition. (B) Intensity-invariance versus noise-resistance in Babble condition.

Acoustic signals can be varied in terms of a rich set of parameters. Previous evidence points to the existence of intensity-invariant neurons that retain neural responses to natural stimuli delivered at multiple intensities (Billimoria et al., 2008; Sadagopan and Wang, 2008; Schneider and Woolley, 2010; Watkins and Barbour, 2011). We examined the relationship between neurons' intensity-invariance and noise-resistance to test the hypothesis that neurons whose responses are intensity-invariant are also noise-resistant (i.e., robust). Single units with rate-level functions measured using vocalizations were included in this portion of analysis. Neural responses were truncated to the length of the shortest vocalization in order to determine the intensity-invariance index. The intensity-invariance index was scaled for each vocalization separately, and was bounded between 0 and 1. The higher the index value, the better the unit is at discriminating vocalizations in an intensity-invariant manner. We associated the intensity-

invariance index of single units with their noise-resistance as reflected by EI profile values using Pearson correlation in **Figure 4.3**. Generally, a weak but significant positive correlation between intensity invariance and noise resistance exists in the WGN condition (Trillphee: $r = 0.376$, $p = 3.72\times10^{-8}$; Peeptrill: $r = 0.198$, $p = 7.60\times10^{-3}$; Trilltwitter, $r = 0.370$, $p = 3.81\times10^{-8}$; Tsikstring: $r = 0.198$, $p = 6.00\times10^{-3}$; Peepstring: $r = 0.349$, $p = 1.90\times10^{-7}$). On the other hand, no significant correlation exists between intensity invariance and noise resistance in the Babble situation (Trillphee: $r = 0.00860$, $p = 0.904$; Peeptrill: $r = -0.142$, $p = 0.0569$; Trilltwitter, $r = 0.0328$, $p = 0.638$; Tsikstring: $r = 0.0542$, $p = 0.455$; Peepstring: $r = 0.0626$, $p = 0.364$). The relatively low correlations in both noise cases imply that intensity invariance and noise resistance reflect two mostly separate processes. The weak but significant correlations for WGN raise the possibility that these processes are related and perhaps overlap in some way, at least under the conditions evaluated by WGN.

4.3.4 Selecting the Number of Neural Response Groups

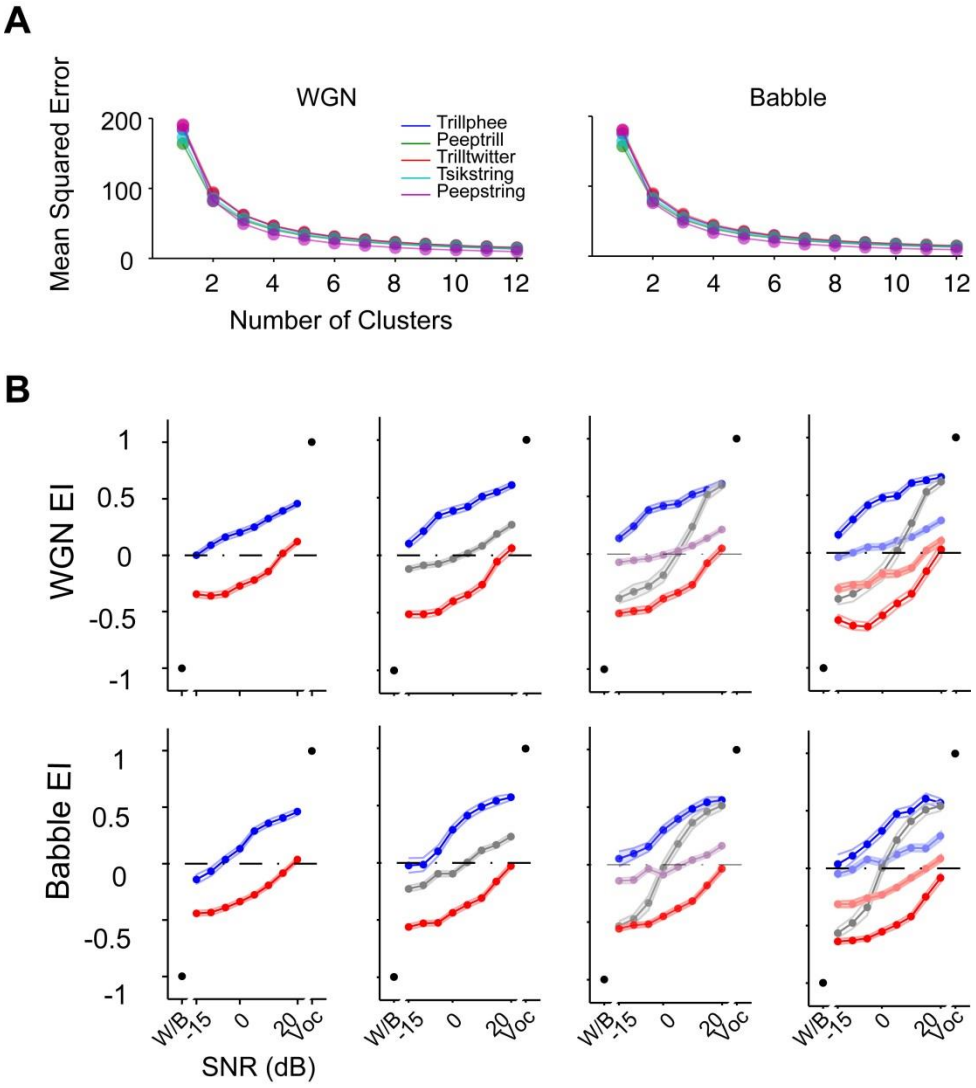

Figure 4.4 Selecting the number of response groups. (A) Mean squared error of EI clustering as a function of number of clusters for each vocalization under two noise conditions. (B) Population-averaged EI profile clusters of Trillphee vocalization with different number of response groups. Numbers of response groups vary between 2 and 5 from left to right.

Babble noise was demonstrated to induce more distortion in neural responses to vocalizations than WGN in terms of mean discharge rate and EI value. Next, we examined each single unit's EI profile more closely to elucidate in detail the large variability of neural responses to vocalizations in noise. To investigate the potential patterns embedded within EI profiles, we

implemented unsupervised k-means clustering on the raw EI profiles across SNR levels from -15 dB to 20 dB. EI profiles with similar shapes were grouped together automatically by this method in order to minimize the distance between group centroids and individual profiles. According to the MSE values for different numbers of clusters in **Figure 4.4A**, 2, 3 and 4 number of clusters all appear to be potential candidates.

We applied clustering with respect to cluster numbers of 2, 3, 4, and 5 respectively. The resulting population-averaged EI profiles for vocalization Trillphee are displayed in **Figure 4.4B**. In all cases, this analysis revealed clusters the appeared to preferentially encode the vocalization or the noise. Higher cluster numbers revealed intermediate responses that appeared to encode neither. Given that five clusters revealed two redundant clusters both tending to encode the noise, and visual inspection of raster plots revealed that the four clusters corresponded to easily discernible spiking patterns, we completed further analysis with four clusters. It is worth noting, however, that all the results that follow were also found for all the other cluster numbers that we considered (data not shown).

4.3.4 Constant Response Groups with Dynamic Neuron Membership

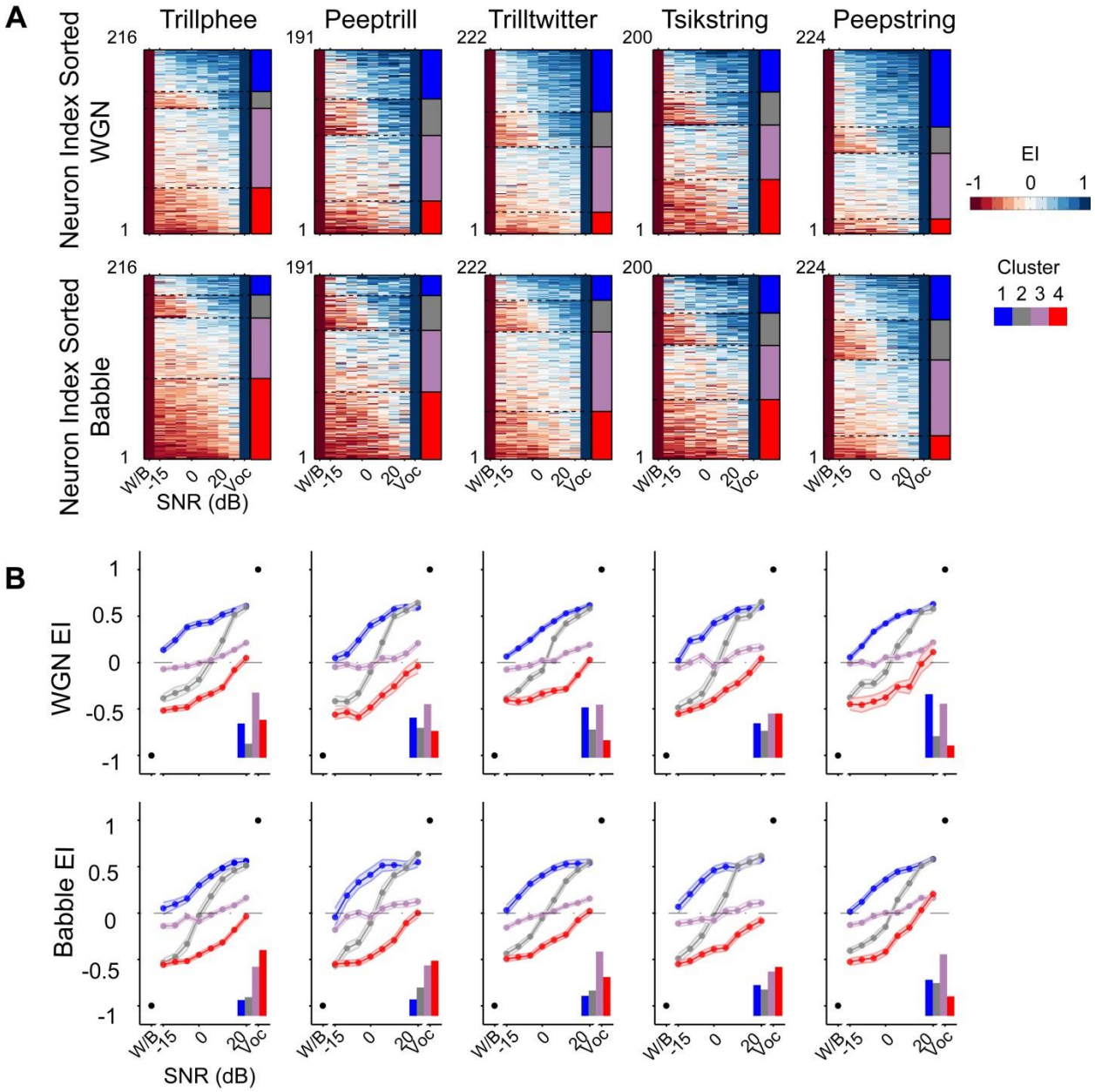

Figure 4.5 All noisy vocalization responses fall into a consistent set of classes. (A) Raw EI profiles of single units are sorted in the order of clusters they were classified into. Four similar EI profile clusters were identified automatically in all five vocalizations. Each row is an EI profile of a single unit. Four stacked colors on the right side of each panel indicate the identity of the neurons under each noise condition, with the top panels indicating the WGN condition and the bottom panels indicating the Babble condition. The robust group is in blue (cluster 1), balanced group in gray (cluster 2), insensitive group in purple (cluster 3), and brittle group in red (cluster 4). (B) Population-averaged EI profile of each cluster (mean ± s.e.m). The relative numbers of neurons classified into each group are displayed at the lower right of each panel. The cluster identities are indicated with the same color denotations as in (A).

The clustering methodology described above yielded constant EI profile groups across all five vocalizations under both noise conditions. As exhibited in **Figure 4.5A**, single units' EI profiles were sorted based upon their group identity to form a matrix for each vocalization and noise combination. The four groups of responses can be identified by the color transition in the matrix, which reflects their noise resistance ability. The first group (blue) exhibits positive EI values down to the lowest SNR. Neurons in this group can individually encode more vocalization than noise as long as some vocalization component exists in the auditory scene, even if it is small. We refer to this group as *robust*. The second group (gray) appears to have a more varied pattern in the EI profile matrix given that blue occupies the upper half of SNR values, through red dominates the lower half of SNR values. This trend indicates that neural responses in this group encode either noise or vocalization, depending upon which is more prominent in the auditory scene. We refer to this group as *balanced*. The third group (purple) is the least varied group in EI profile matrix. Instead of dominated by blue and red, most areas are filled with white mixed with little read and blue, indicating that this group of neural responses exhibits little preference for either vocalization or noise, having EI values deviating little from zero. We refer to this group as *insensitive*. The fourth group (red) exhibits a matrix mostly dominated by red. Neural responses in this group thus are more susceptible to the presence of noise, and they are more likely to mainly encode noise even if only a small amount of noise exists in the stimulus. We refer to this group as *brittle*.

These four distinctive encoding patterns are summarized graphically in **Figure 4.5B,** where EI profiles of individual neurons were averaged with others sharing the same group identity. Robust profiles can be identified by predominantly positive EI, brittle profiles can be distinguished by predominantly negative EI values, balanced profiles can be recognized by near-

equal positive and negative EI values and insensitive profiles show EI around zero. The fractions of profiles classified into one of the four groups are depicted in the lower right of each panel. Generally, more than 30% of EI profiles were classified into the insensitive group irrespective of vocalization-noise combination. The fraction of profiles categorized into the robust and brittle groups was vocalization-dependent. Consistent with the observation from the mean values of EI profiles in Figure 3B, more EI profiles were classified into the robust group in the WGN condition than Babble condition for all vocalizations. Still, considerable variance can be observed in the distribution of EI profiles across the four groups.

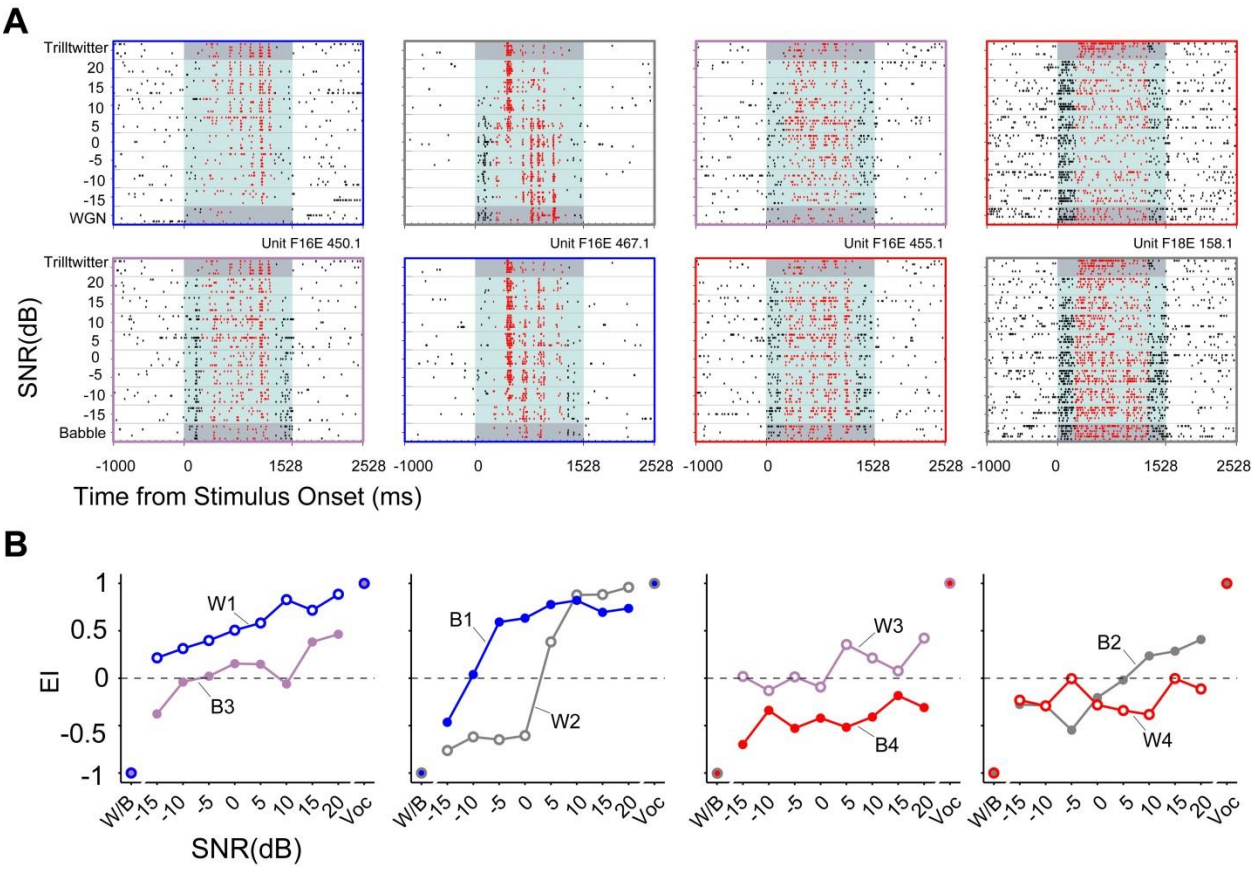

Figure 4.6 Exemplar neurons for each response group. (A) Exemplar units are displayed for each of the four groups of robust, balanced, insensitive and brittle. The top panels show raster plots of four different single units' responses to the Trilltwitter call in WGN and the bottom panels display the same unit's responses to the Trilltwitter call in Babble. Color denotations of dots in raster plots are the same as in Figure 2A. (B) EI profiles corresponding to the same four exemplar neural responses shown in (A).

One single unit from each response group with distinct response patterns to Trilltwitter in WGN/Babble is displayed in **Figure 4.6A**. The top panels show spike raster plots of four single-unit responses to Trilltwitter degraded by WGN, and they were classified into insensitive, robust, brittle and balanced groups, respectively, as indicated by the panel frames colors. The corresponding neural responses of the same four neurons to Trilltwitter degraded by Babble are displayed at the bottom panels in Figure 4.6A. It is worth noting that these neurons' response memberships were not the same in the WGN condition as in Babble. This phenomenon can be more easily observed directly from EI profiles corresponding to each raster plot, as shown in **Figure 4.6B**.

The previous observation in Figure 4.5B of consistent response group forms under all stimulus conditions tested, yet individual units' inconsistent response group identities in Figure 4.6B**,** led us to ask whether the majority of units retain their group identities under different noise conditions. To address this question, we evaluated the proportional distribution of constituent units in the four response groups with Babble masking given the neurons' group identity in WGN condition in **Figure 4.7A**. Two possible scenarios might be predicted, as illustrated in the top panel of Figure 4.7A. One is an invariant model, in which all units completely retain their response group identities across the two noise types. For example, units falling into the robust group under the WGN condition as W1 fully preserve their group response identity in Babble as B1. In a similar way, W2-W4 and B2-B4 share the same subgroup of units. This model predicts a

diagonal group switching matrix. Another model is a random model, in which all units randomly change their response group identities under different noise conditions. This model predicts a uniform group switching matrix.

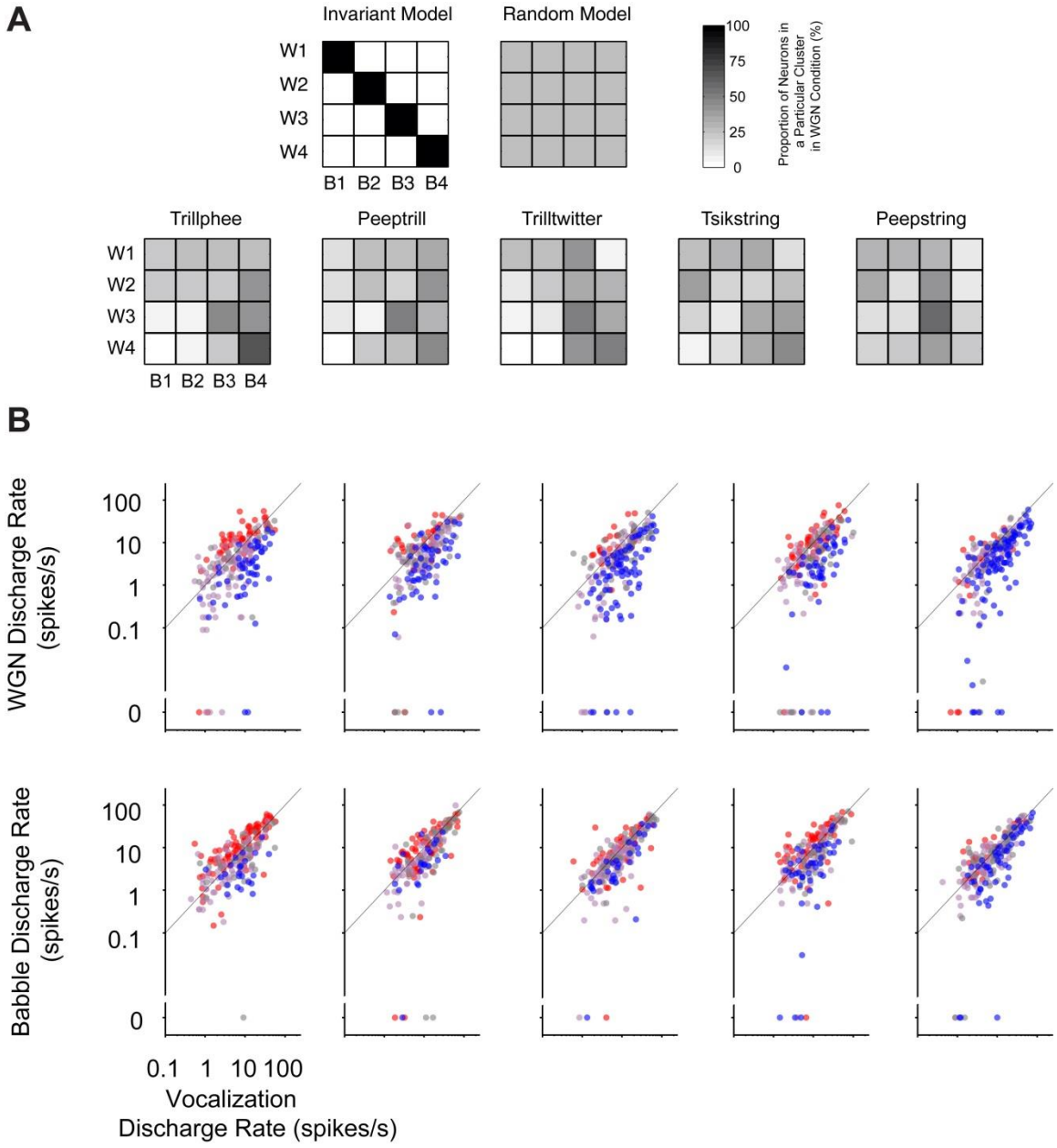

Figure 4.7 Noisy vocalization response types are not consistent for individual units. (A) Group switching matrices of unit membership in response clusters under WGN and Babble conditions. Two hypothetical models are depicted in the top panels: the invariant model and the random model. Group switching matrices of five vocalizations are displayed in the bottom panels. For each matrix, the abscissa indicates cluster identity in the Babble condition (B1-4), and the ordinate represents cluster identity in the WGN condition (W1-4). The grayscale value in each unit square denotes the proportion of units originally falling into a particular cluster identity in the WGN condition being reclassified into a particular cluster identity in the Babble conditions. (B) Scatterplots of mean discharge rate elicited by pure vocalizations and pure WGN noises (top panels) or Babble (bottom panels). Mean discharge rates of single units are colored with their corresponding cluster identity as determined by their EI profiles. Color conventions are the same as those used in Figure 4.

The group switching matrices of our neuronal population can be seen in the bottom panel of **Figure 4.7A**. These matrices lie between the invariant and random models, though considerably closer to random. As a consequence, units appear to change their response group identities as background noise is altered, and less than 40% (M = 0.37, SD = 0.04) of neurons in general retain the same response group when the background noise shifts from WGN to Babble. For example, with vocalization Tsikstring, we might naively expect that a majority of units in the brittle group under the WGN condition would still be classified into the same group under the Babble situation, given that Babble noise has a more disruptive effect upon vocalization encoding, as seen in Figure 4.2. In actuality, however, more than 50% of neurons in the WGN brittle group have response group identities as balanced, insensitive or even robust in the Babble condition. The proportion of neurons that retained their response group identity across masking noises was considerably less than expected, indicating strong noise-dependent response properties in A1. Under both noise conditions, about 5% of the 163 single units responsive to all five tested vocalizations retained their response group identity across all vocalizations. Among those neurons, most of them kept their identity as insensitive or brittle. About 80% of single units fell into 2 or 3 different response groups across vocalizations, and the remainder covered all four response group identities across vocalizations. Therefore, despite four consistent clusters of

neural responses for all vocalization-noise combinations tested, the individual units constituting each of these groups differed substantially.

We further examined the mean discharge rates of single units in response to pure vocalizations and noises with units' response groups being specified in colors in **Figure 4.7B**. We found that a majority of units belonging to robust groups in both noise scenarios were situated below the diagonal, indicating that units with stronger responses elicited by pure vocalization than pure noise are more likely to encode vocalizations at lower SNR values. The discharge rates of units belonging to other response groups, however, were more blended without a clear boundary. Neuronal response reliability to pure WGN and pure vocalizations was also compared in terms of response groups, and a similar trend was observed (data not shown). Therefore, neither discharge rate nor response reliability alone is sufficient to explain neuronal response type in the face of noise interference.

4.3.5 Suppression and Addition of Spiking Activity within and between Vocalization Phrases

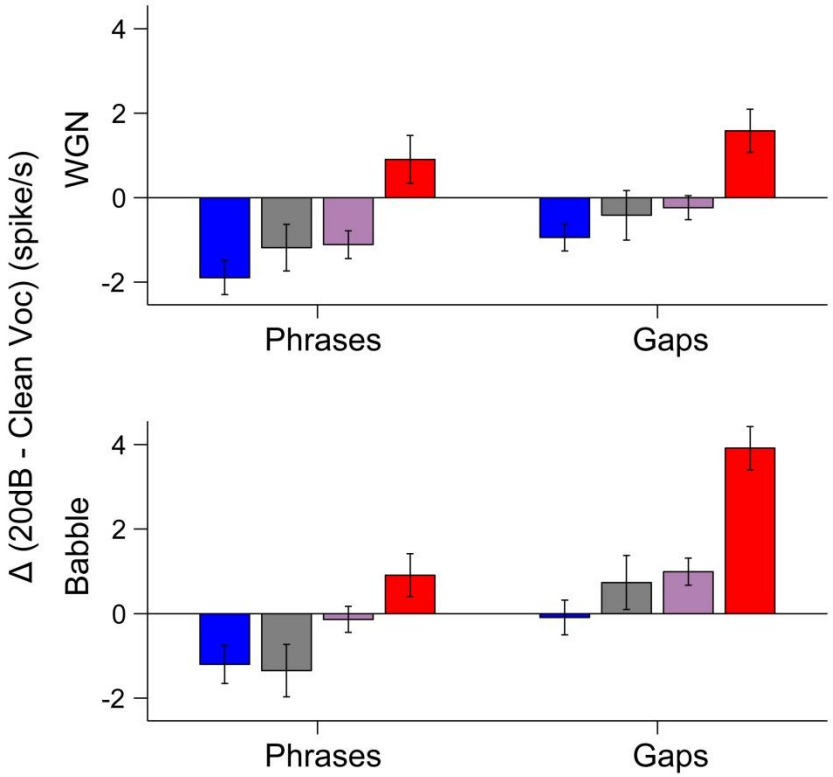

Figure 4.8 Difference of discharge rates to phrases and gaps of vocalizations. Discharge rates at 20dB SNR within and between gaps of vocalization phrases relative to clean vocalizations were displayed for each neural response type (mean ± s,e,m). The same color denotations as in Figure 4.4A were used.

To investigate how the addition of noise affected the spiking activity elicited by vocalizations in terms of acoustic features, we measured the discharge rate within vocalization phrases and between gaps of vocalization phrases in **Figure 4.8.** By comparing vocalizations at 20 dB SNR with clean vocalizations, the brittle response group showed increased spiking activity both within and between vocalization phrases under both noise conditions. This finding serves as a strong evidence for the existence of a group of neurons that can detect a barely audible noise stream once it is present in the auditory scene. In contrast to the brittle response group, the robust

response group, along with balanced and insensitive groups, showed suppression of neural activity within vocalization phrases under both noise conditions, indicating that the response patterns to vocalizations might be largely preserved but with smaller amplitude. While the robust response group also showed suppression of activities in gaps between vocalization phrases in both noise conditions, the activities of less robust groups (balanced and insensitive) during the gaps are more subjective to the interference of noise under Babble condition than WGN.

4.3 Discussion

Introducing either White Gaussian Noise or Babble noise reduced mean vocalization-induced discharge rates in A1 neurons. Similar observations were made in other preparations under different stimulus conditions (Gai and Carney, 2008; Schneider and Woolley, 2013). Both noise types also reduced mean reliability, which was correlated with the decreases in mean discharge rates. Higher average discharge rates alone are not the source of higher response reliability, however, because reliability calculations were based upon correlation of neuron responses across trials, unaffected by the absolute firing rate.

Nagarajan *et al.* observed that 15% of barbiturate-anesthetized A1 neurons respond more strongly to calls in white noise than to pure calls alone (Nagarajan et al., 2002). Similar neural responses were also noticed in our awake animals, such that on average 13% of units exhibited responses stronger than 1.5 times the response to pure vocalizations at 20 dB SNR in the WGN condition and 19% in the Babble condition. This phenomenon therefore appears to be independent of animal state (i.e., anesthetized vs awake). One potential explanation might be the widespread nonmonotonic rate-level functions of marmoset A1 (Sadagopan and Wang, 2008; Watkins and Barbour, 2011). A nonmonotomic, vocalization-selective neuron with best intensity somewhat lower than the one at which the noisy vocalizations were delivered might respond

better as some noise is added in power-normalized fashion, thereby reducing the intensity of the vocalization component. Another feasible explanation is that these neurons might dynamically adjust their contrast gains under this specific combination of acoustic signal and noise (Barbour and Wang, 2003b; Willmore et al., 2014).

We found that the single-unit neural representation of communication sounds in marmoset A1 is also context-dependent (Narayan et al., 2007). Four consistent response groups were identified by their responses to noisy vocalizations. We first determined that the group identity of a neuron's responses to a vocalization in WGN poorly predicted its responses to the same vocalization in Babble. A small subset of neurons demonstrated sustained high responsiveness to WGN but suppressed responses to Babble, implying a complex spectrotemporal integration. We further concluded that the responses of A1 neurons to a sound mixture were dominated by the stimulus that more efficiently induced a response when delivered alone, which is consistent with the "strong signal capture" observed with pure tones and broadband noise (Phillips and Cynader, 1985; Gai and Carney, 2008).

One critical question for auditory scene analysis is where the auditory segmentation initially occurs (Shamma and Micheyl, 2010). Formation of auditory objects has been widely studied both at cortical and subcortical levels using multiple recording techniques (Fishman et al., 2001; Fishman et al., 2004; Bar-Yosef and Nelken, 2007; Pressnitzer et al., 2008). A unifying principle from previous studies is that auditory streaming processing exists at least as late as A1 and possibly begins as early as cochlear nucleus. In our study, the presence of four noisy vocalization response groups supports A1 as a source of stream segregation processing.

Regarding functionality, the robust group represents a neural substrate for the vocalization stream, while the brittle group represents a neural signature of the noise stream. The balanced group provides another coding dimension reflecting the SNR between two auditory streams (i.e., indicating which stream is relatively more intense). The insensitive group responds equally to both streams and is not particularly useful for segregating either of them but may have other coding functions. Given that these response groups emerged consistently under every signal and noise condition tested, we speculate that this may be general coding mechanism for representing sounds. It is worth emphasizing again that individual A1 neurons can generate responses that fall into any of these classes depending upon stimulus context.

How can A1 project to downstream areas so that noise-invariant responses to sounds become possible? Feedforward suppression is suggested to be the underlying mechanism of noise-invariant responses(Schneider and Woolley, 2013). In the context of our dataset, the readout from the robust group would be strengthened while the readout from the brittle group is suppressed. Top-down regulation is probably needed for such enhancement (Jancke et al., 1999; Tervaniemi et al., 2009), and this control signal would also need to be contextually dependent and possibly under attentional control.

No evidence of a strong positive relationship between intensity-invariance and noise-resistance exists in our dataset. One potential explanation is our observation that if a neuron is very responsive to vocalizations alone, it is also likely to be actively driven by noise alone. When these two stimuli are combined we often see sensitivity to both. Alternatively, we tested a limited range of intensity levels, which may be insufficient to capture neurons' full intensity-invariance. Combinations of different auditory objects evoke some A1 neurons to respond predominantly to the weaker object (Bar-Yosef et al., 2002; Bar-Yosef and Nelken, 2007). This type of response

may be reflected in our brittle group; nevertheless, this response pattern in the population would lower the probability that simple context-independent rate-level response features induced by single stimuli can explain the present results.

Vocalizations in WGN yield a generally lower neuronal detection threshold than in Babble, in agreement with human listeners' susceptibility to different types of masking (Carhart et al., 1969; Brungart, 2001). Nevertheless, the lack of behavioral performance in the current study makes it challenging to infer this perception accurately. The auditory stream processing we observed occurred without requiring attention to a particular auditory stream. An important additional goal would be to determine whether this attention-modulation neural encoding affects neural representations of already formed auditory streams, or if it influences the representation formation process itself.

By segmenting the neural activities based upon the spectrotemporal phrases of vocalizations, we revealed the underlying firing rate alternation of different response groups at 20 dB SNR relative to the firing rates in response to clean vocalizations. It explicitly shows the distinctive difference between the robust and brittle group. The robust group suppressed spiking activities both within and between phrases, and the brittle group behaved oppositely. The previous songbird reported suppression of neural activities within song syllables and addition of spikes between song syllables on population level (Narayan et al., 2007). Our results, however, further exhibited that the influence of noise on the neural activities varies between different types of neural response group, and the suppression/addition of spikes was dominated by different subgroups of neurons. The emergence of different response groups who preferentially encoding individual auditory streams in the primary auditory cortex servers as the evidence of the underlying neural processing for the auditory scene analysis.

We have studied neural encoding of vocalizations presented in conjunction with two types of background noises in marmoset monkey A1. Subsets of single units with high discrimination performance existed under both noise conditions. The dynamic role of single units indicates there are relatively few individual neurons in primary auditory cortex that can robustly encode stimuli in the presence of different noises. Robust encoding clearly exists in A1 when considering population responses, however, and future studies should consider evaluating integrated population responses and the effects of top-down influences mediated by attention.

Chapter 5: Population Coding of Vocalizations at Multiple Intensities and SNRs

5.1 Introduction

Due to the inherent noise in the activity of individual neurons, multiple presentations of an identical sensory stimulus do not yield exactly the same spike trains. By computing the spiking rate averaged across trials to get rid of the response noisiness, researchers have generally expected to estimate the true firing rate driven by a stimulus. Large amounts of single-unit analysis have been conducted in this fashion. In studying neural responses to auditory stimuli, much insight has been gained from analyzing coding properties of individual neurons based upon the simplistic rate-coding hypothesis (Aitkin et al., 1986; Imig et al., 1990; Bendor and Wang, 2005; Woolley et al., 2006; Barbour, 2011). However, the information represented by the ensemble of individual neurons has typically been overlooked. This seems to be a minor concern for studies investigating relatively simple acoustic stimuli, such as pure tones, but there are studies showing that even stationary acoustic stimuli induce dynamic responses on a population level. For more complicated acoustic signals with rich temporal-spectral structures, such as marmoset vocalizations (Gehr et al., 2000; Nagarajan et al., 2002), an analytical method for inspecting the response properties among neural population is needed. Here in Chapter 5, subsets of dataset used in Chapter 4 are analyzed from a population perspective.

In contrast to single-unit coding, population coding hypothesizes that the stimulus information is encoded in the brain by a large population of neurons via distributed firing rate patterns (McIlwain, 2001). Over the past two decades, multiple population analyses have

emerged to reveal the neural encoding and decoding properties at the population level, such as population variability analysis and spatiotemporal coding analysis. Population responses have been demonstrated to vary meaningfully (Churchland et al., 2010). The onset of a sensory stimulus leads to an acute decrease in the population response variability, indicating that the brain prepares itself to be in a stable state to process the coming stimulus. Churchland et al. claimed this phenomenon to be universal because the driving down of population response variability by the stimuli exists in the visual cortex, the parietal reach region, the dorsal premotor cortex, and the orbitofrontal cortex independent of the behavioral state. The auditory cortex, however, is not explicitly addressed in their study. Here, a trivial hypothesis to test is that the onset of complicated acoustic signals suppresses the population response variability. More importantly, to study the dynamics of population variability in response to vocalizations at multiple intensities and SNR levels, I further ask whether acoustic features of vocalizations modulate the population variability of the ongoing neural activities. If the answer is yes, then it suggests that the response variance, in addition to the spiking rate, can potentially encode information about vocalizations.

Neocortical neurons generate time-varying firing patterns with particular temporal structures. In the sensory areas, even presentation of a temporally unstructured stimulus, such as a stationary odor or pure tone, is likely to induce a complex temporal pattern of spiking (Stopfer et al., 2003; Bartho et al., 2009). By unifying the temporally-structured responses of individual neurons, we can visualize the complex spatiotemporal patterns at the population level. The spatiotemporal patterns of the population responses vary with the stimulus when a particular feature of the stimulus is slightly changed, such as intensity (Stopfer et al., 2003). Such visualization analysis has revealed the dynamics of responses to relative simple stimuli, however,

little is known about the spatiotemporal patterns of complex vocalization stimuli. Here, by varying intensities and SNR levels, we also studied the alteration of spatiotemporal patterns of population neural responses to five marmoset conspecific vocalizations and tested the hypothesis that the population responses are not just a linear scaling of their amplitude.

Neurometric analysis is a useful tool for linking neural activity with perception to identify the underlying neural substrate that generates the sensory perception and behaviors of interest (Walker et al., 2008). Individual neurons vary widely in their ability to discriminate complex acoustic stimuli (Narayan et al., 2006; Wang et al., 2007; Schneider and Woolley, 2010). However, the degree to which a population of neurons can recognize different types of vocalizations at multiple intensities and SNR levels is not well known. Would distributed firing rate patterns across a whole population of recorded cells optimize the performance of the stimulus discrimination task, or is there a subpopulation of neurons that yields the best performance? With respect to the dynamics along the time course, does a population of neurons have a constant discriminability, or are the neural responses at certain time epochs better than other epochs? These questions were investigated by building population response decoding models that are sensitive to temporal discharge patterns. Using this decoding tool, we further inferred the perception intensity threshold of vocalizations and the detection threshold of vocalizations masked with WGN/Babble noise.

In summary, in this Chapter, by pooling the activities of individual together, we analyzed population responses to vocalizations at multiple intensities and SNR levels from three aspects: population response variability with respect to time, spatiotemporal structures of population responses, and the ability of population responses to identify stimuli in different experimental conditions.

5.2 Data Analysis

Two sets of data were studied using the population analysis in this chapter. The first dataset is single-unit responses to five vocalizations (Trillphee, Peeptrill, Trilltwitter, Tsikstring, and Peepstring) at four intensities (from 15 dB SPL to 75 dB SPL, in 20 dB SPL steps). In total, N = 326 single units were included in the analysis. For the second dataset, a subset of N = 172 single units was used to study noise interference with population responses to five vocalizations at 10 SNR levels (from -15 dB to 20 dB, in 5 dB steps) including pure vocalization and noise, delivered at 75 dB SPL.

For each unit of both datasets, a peristimulus time histogram (PSTH) for each response trial was calculated by binning spike trains into rate vectors with a 50 ms window in 10 ms steps. The following data analyses are based upon this preprocessing, unless otherwise stated. All data analyses were conducted in MATLAB R2014a (The MathWorks Inc, Natick, MA). Because most of the neurons in our dataset were recorded sequentially one at a time, we created pseudo-populations to substitute for simultaneous recordings. Creating these pseudo-populations potentially ignores the correlation between individual neurons that exists in a simultaneously recorded neuronal population, and may change the estimates of the absolute level of performance. The majority of conclusions drawn in this chapter, however, would most likely not be altered by the sequential recording, because previous studies show that similar conclusions are obtained by simultaneously recorded neurons and sequentially recorded neurons (Gochin et al., 1994; Baeg et al., 2003; Panzeri et al., 2003; Aggelopoulos et al., 2005; Nikolić et al., 2006; Anderson et al., 2007).

We used the Fano factor to quantify the population response variability to a particular stimulus with respect to time. For a time bin at time t, the mean of spike count $\mu_{bin,t}$ and variance

$\sigma^2_{bin,t}$ across trials were calculated for each neuron separately. The Fano factor for a single neuron at time t is just the ratio between the variance and the mean of the spike count as displayed in (5.1).

$$F_{bin,t} = \frac{\sigma^2_{bin,t}}{\mu_{bin,t}}. \tag{5.1}$$

With regard to the calculation of Fano factor for a population of neurons, a slightly different procedure was implemented. First, a scatterplot of the trial-averaged mean of the spike count and variance of a population of neurons at time *t* was obtained. A regression was later performed to relate the distribution of variances with the distribution of means of spike count. The resulting slope was the Fano factor of the population response at time *t*. The MATLAB code used to compute the Fano factor is available at (http://www.stanford.edu/~shenoy/GroupCodePacks.htm) (Churchland et al., 2010).

Visualization of population responses in 3D space could help us gain an intuitive understanding about highly complicated neural responses (Stopfer et al., 2003; Bartho et al., 2009; Saha et al., 2013). Such visualization can be realized by principal component analysis (PCA). PCA is a linear dimensionality reduction technique. It identifies a set of linearly uncorrelated variables, called "principal components", from an original dataset composed of a large number of possibly correlated variables and captures as much of the variability in the dataset as possible (Jolliffe, 2002). The principal components are ordered by the amount of variability that each component accounts for, and the first principal component has the largest variance. With regard to neural population responses, each single unit counts as one dimension in the population response space. Given the neural response of *n* single units in a neural population,

an n-dimensional response vector R^n can be generated. By applying PCA on the n-dimensional response space, we can obtain m virtual neurons to constitute an m-dimensional response space preserving as much of the variance in the original dataset as possible, where m <= n. If we keep only the first three virtual neurons' responses (m = 3), we can visualize the population responses as a trajectory in a 3D space by connecting the responses at a series of time points. In the following analysis, PCA was implemented for each vocalization type separately. The first three principal components accounted for about 30% of the original dataset's variance for each vocalization at multiple intensities, and for about 22% of the variance for each vocalization at multiple SNR levels, under either WGN or Babble noise. For visualization, trajectories were smoothed with a 10-point running window for pure vocalizations, and a 20-point running widow for noisy vocalizations.

Based upon the trajectory visualization analysis, we can further quantify the structure of trial-averaged population responses. To quantify the rotation of the population response vectors in response to a particular stimulus s_0 in 3D space, the angle between a response vector $\vec{r}_t$ at time t and a reference response vector $\vec{r}_0$ at time t_0 can be computed as in (5.2) (Bartho et al., 2009). By defining the first point of the spontaneous population response corresponding to stimulus s_0 as the reference vector, the angle evolution of population response vectors can be obtained by concatenating the angles calculated at all the available time points, $t = t_0, t_1, ..., t_n$, during pre-stimulus, stimulus, and post-stimulus. We calculated the intra-trajectory angle evolution for the five vocalizations at four intensities and ten SNR levels under WGN and Babble noise conditions as follows:

$$\theta(s_0,t;s_0,t_0)=\cos^{-1}\frac{\vec{r}_t\cdot\vec{r}_0}{\left\|\vec{r}_t\right\|\left\|\vec{r}_0\right\|} \tag{5.2}$$

Similarly, we also computed the inter-trajectory angle evolution of population response vector $\vec{r}_s$ corresponding to stimulus s relative to response vector $\vec{r}_{s_0}$, which in turn corresponds to stimulus s_0 across time, $t=t_0,t_1,...,t_n$, as displayed in (5.3).

$$\theta(s,t;s_0,t)=\cos^{-1}\frac{\vec{r}_s\cdot\vec{r}_{s_0}}{\left\|\vec{r}_s\right\|\left\|\vec{r}_{s_0}\right\|} \tag{5.3}$$

The angles of population responses to vocalizations at three softer (15 dB SPL, 35 dB SPL, and 55 dB SPL) levels relative to the population response to vocalization at 75 dB SPL were calculated. The noise effects on population response angle evolution were also investigated by computing the angles of responses to noisy vocalizations and pure noise relative to pure vocalizations.

To investigate the discriminability of population responses trial-by-trial (Meyers et al., 2008; Bartho et al., 2009), template-based stimulus identity predictive models were built based upon neural population responses. For predictive model decoding of vocalizations at multiple intensities, there are three types of models: single-bin based, sliding-bin based, and varying-cell-number based. To build a single-bin based model, for each stimulus condition (five vocalizations at four intensities), five trials of single-unit responses at a particular time bin were randomly sampled out of, at most, 10 trials for each neuron (N = 326, each has 5~10 trials). The five trials were concatenated to form a 100 × 326 population response matrix. The stimulus identity

corresponding to each response trial is called a label. There are a total of five labels, each representing a vocalization type (c = 5). Four trials of the neural population responses to each of five vocalizations delivered at the highest intensity, 75 dB SPL, were further randomly selected as the training templates. The remaining 80 trials of neural population responses were used as the testing dataset, and the label corresponding to a test trial was decoded by calculating the cosine distance between this trial and the 20 template trials. The vocalization type or stimulus label corresponding to the template trial with the shortest distance from the test trial was assigned as the predicted label. This whole process was repeated 50 times. The performance of the predictive model was evaluated by its overall accuracy and confusion matrix. The overall accuracy is defined as the percentage of stimulus labels that are correctly predicted, and the confusion matrix revealed the chance of a particular label being predicted as one of the five labels.

While a single-bin based model was built to investigate the neural discriminative performance at each time point, a sliding-bin based model was used to study the effect of time accumulation on neural discriminability. The process of building a sliding-bin based model was very similar to that of a single-bin based model, except for that the response bin from each single unit were varied from 1 to 41, where 41 is the number of bins that the shortest vocalization Tsikstring has. Performance as a function of temporal resolutions (time bin width) was investigated with predictive model using 41 bins, and the temporal resolution was varied from 5 ms, 10 ms, 20 ms, … , to 100 ms. Last, a varying-cell-number based model was created by changing the number of neurons in the population, and only models with 41 bins were studied.

The population neural discriminability of vocalizations at multiple SNR levels was investigated, much like that of vocalizations at multiple intensities. Predictive models under WGN and Babble conditions were built separately. Here, the training templates had six labels,

including five pure vocalizations and one type of noise (c = 6). In addition, to further study population decoding with a subpopulation of neurons, a predictive model for each vocalization type was built by using the number of time bins available for that particular vocalization. For each vocalization, a different subpopulation of neurons was included because of the contextual dependent effect in Chapter 4. There were only two labels, vocalization and pure noise (c = 2). Population neural responses from ten SNR levels were decoded as either vocalization-present or vocalization-absent. A linear support vector machine (SVM) classifier was used instead of the template-based predictive model to accomplish the binary classification task. The SVM classified the neuronal responses by training a separating hyperplane based upon the labeled training trials and had very good performance for binary classification (Van Gestel et al., 2002), while the template-based method did not have a particular training session. The generalizability of the classifier over lower SNR levels was studied by using different training data, for instance, neural responses to vocalizations at 20 dB SNR.

Normality was verified by the Lilliefors test. Unless otherwise indicated, hypothesis testing was conducted using a two-sided Wilcoxon signed-rank test. The significance criterion was set to 0.05.

5.3 Results

5.3.1 Population Response Variability of Vocalizations at Multiple Intensities

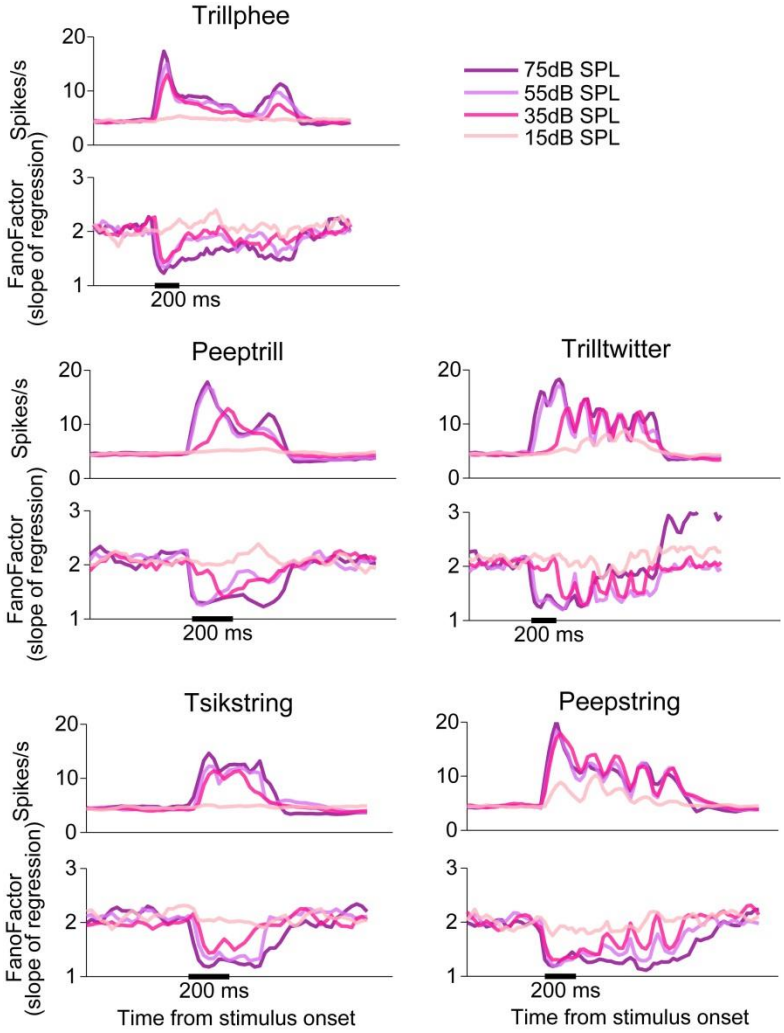

Figure 5.1 Population-averaged responses to five vocalizations at multiple intensities and the corresponding population activity variability with respect to time.

The spectrotemporal acoustic patterns of marmoset vocalizations can induce correlated discharge patterns in a subpopulation of individual neurons in A1 of marmosets (Wang et al., 1995). We wondered whether the same correlated patterns exist in the variability of population neural responses**,** in comparison with averaged population response at multiple intensities. **Figure 5.1** shows that population responses to all five vocalizations at the loudest 75 dB SPL

exhibited discharge patterns that follow the acoustic envelope belonging to each vocalization (see Figure 2.1). As intensity decreased, the population responses diminished to near the spontaneous activity range at the softest level. Though the intensity decreased over an equal interval, population responses were not scaled linearly. In addition to the general trend, vocalization-dependent changes were also observed. For instance, spiking rates over time at 55 dB SPL were very similar to those at 75 dB SPL, while spiking rates at 35 dB SPL not only shrank in scale but were also deformed, leading to delay in onset responses for some vocalizations. It is most noticeable for the vocalizations Peeptrill and Trilltwitter, to which the onset responses were delayed about 100 ms and 200 ms, respectively, at 35 dB SPL. The probable reason is that 35 dB SPL is around the hearing threshold and serves as a transition point at which some acoustic features in the vocalizations were not well perceived. For vocalizations Trillphee, Tsikstring, and Peepstring, neural responses largely maintained their structures even at 35 dB SPL.

Table 5.1 Pearson correlation between spiking rate and variability for vocalizations at multiple intensities

Vocalization	75 dB SPL	55 dB SPL	35 dB SPL	15 dB SPL
Trillphee	***r* = -0.886** *p* = 2.60e-17	***r* = -0.801** *p* = 4.99e-12	***r* = -0.582** *p* = 1.12e-05	***r* = 0.331** *p* = 0.0200
Peeptrill	***r* = -0.737** *p* = 1.39e-04	***r* = -0.930** *p* = 1.04e-09	***r* = -0.962** *p* = 3.39e-12	***r* = 0.719** *p* = 2.38e-04
Trilltwitter	***r* = -0.785** *p* = 2.95e-10	***r* = -0.853** *p* = 1.83e-13	***r* = -0.937** *p* = 9.06e-21	***r* = -0.571** *p* = 5.09e-05
Tsikstring	***r* = -0.938** *p* = 3.54e-10	***r* = -0.962** *p* = 3.79e-12	***r* = -0.959** *p* = 7.44e-12	***r* = -0.531** *p* = 0.0133
Peepstring	***r* = -0.5573** *p* = 1.88e-04	***r* = -0.856** *p* = 1.96e-12	***r* = -0.878** *p* = 9.45e-14	***r* = -0.625** *p* = 5.16e-06

The variability of population responses was a mirror image of the spiking rate of population responses over time. The Pearson correlation coefficients between these two metrics were computed and as displayed in **Table 5.1**. Variability was significantly correlated in a negative way with spiking rate for all vocalizations above 15 dB SPL. The strongest correlation, however, was not necessarily associated with the loudest intensity. For example, Peepstring's correlation value at 75 dB SPL was relatively much lower than those at 55 and 35 dB SPL, and the variability over time did not show a profile that tracked individual phrases in the vocalization. The relationship between variability and spiking rate at the softest 15 dB SPL was less consistent across vocalizations, and either positive or negative correlation was possible. The population response variability of Trilltwitter, Tsikstring, and Peepstring at that level still had a negative correlation with spiking rate, indicating that A1 neurons are probably more sensitive to the three vocalizations than Trillphee and Peeptrill.

Therefore, the information of the vocalization envelope is also represented in the population activity variability. Furthermore, a nonlinear relationship between vocalization intensity and population response variability was revealed.

5.3.2 Population Response Trajectory of Vocalizations at Multiple Intensities in 3D Space

An intuitive understanding of population neural responses can be obtained by visualizing their spatiotemporal structure. A powerful tool to achieve this is to project the high-dimensional response vectors onto a lower dimensional space, in which enough variance in the high-dimensional dataset is captured by three principal components (i.e., virtual neurons).

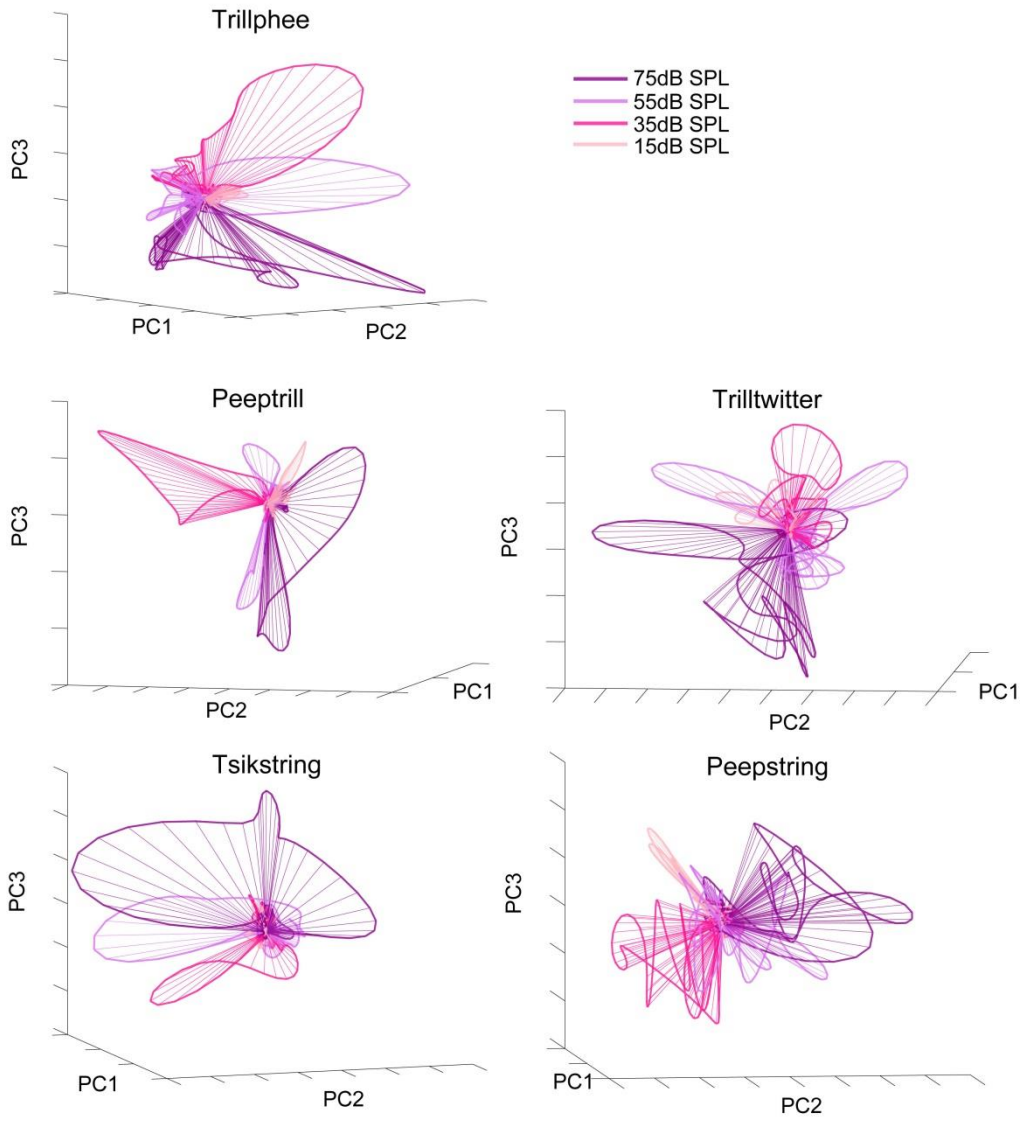

Figure 5.2 Trajectories of population responses to vocalizations at multiple intensities in 3D space.

For each vocalization at multiple intensities, population responses were reduced to the same 3D space, and the resulting response trajectories are displayed in **Figure 5.2**. Trajectories were formed by connecting the response points from three time stages: pre-stimulus, during-stimulus, and post-stimulus (not explicitly marked in Figure 5.2). Skeletons, which link the first point on the trajectory with the remaining points, were plotted to visualize the response

hyperplane. Hyperplanes belonging to different vocalizations all have very distinct shapes. Some are relatively smooth and simple, such as Trillphee, while some are more tangled and twisted, such as Peepstring.

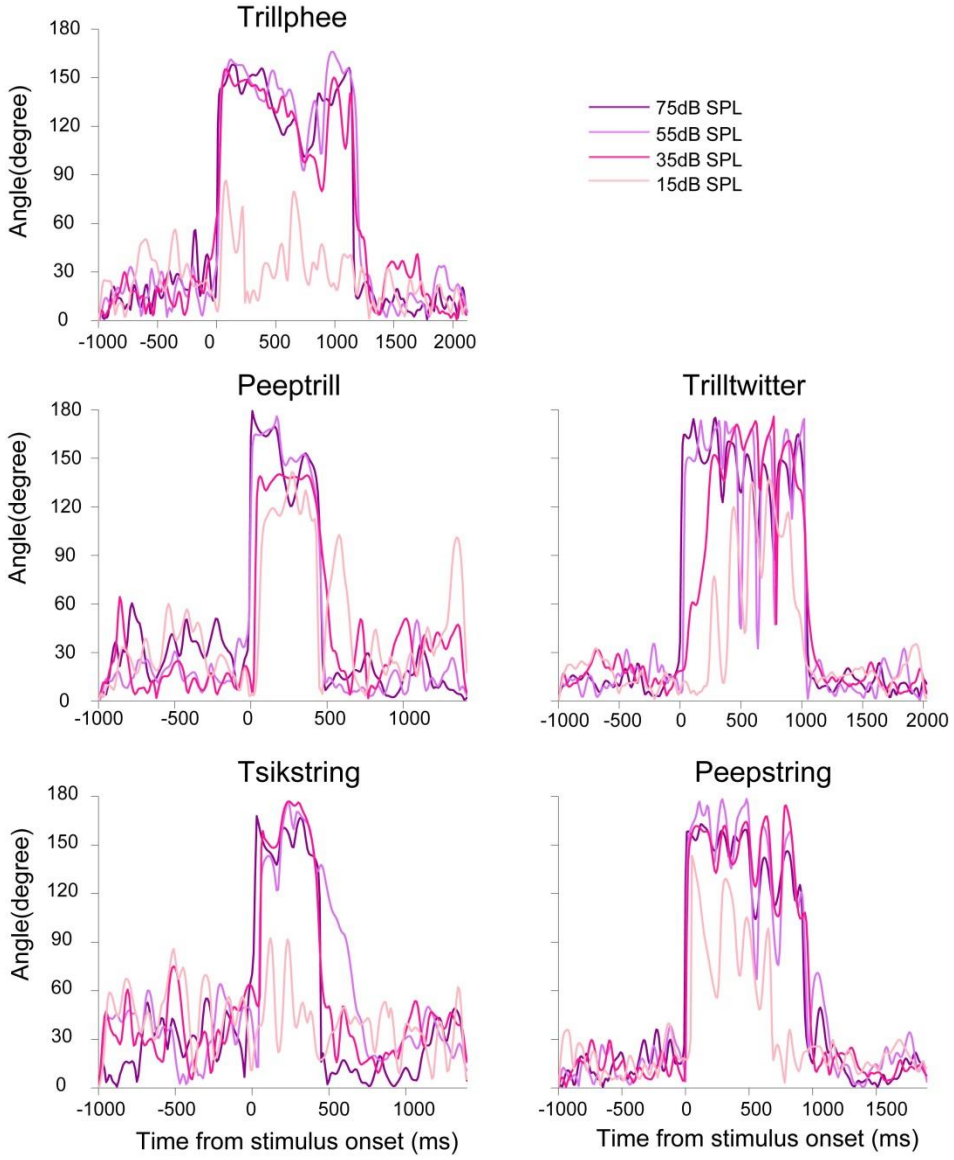

Figure 5.3 Evolution of rotation angles relative to the first time point (in silence) of the population response at multiple intensities in 3D space.

How does the population hyperplane change in response to a decrease in intensity? Here we consider the hyperplane at 75 dB SPL as the reference hyperplane. If neuronal populations linearly scaled their responses' amplitudes, we would expect to see the response hyperplane shrink without changing its position in 3D space. Alternatively, the hyperplane could change in a way that only rotates its position relative to the reference hyperplane. As a matter of fact, the hyperplane seems to both resize and rotate. It is worth noting that the more the intensity decreases, the further the hyperplane deviates from the reference, in a consistent direction.

To quantify the response hyperplane and the changed induced by intensity, we calculated the angle between response vectors in two ways. First, we quantitatively described the spatiotemporal structure of a hyperplane by computing the angle between the first response vector on the hyperplane and the remaining response vectors over time in **Figure 5.3**. Clearly, across vocalizations and intensities (except for 15 dB SPL), the intra-trajectory angle fluctuated between 0 and 60 degrees at the pre-stimulus stage. To process the upcoming stimulus, an acute increase in the angle immediately followed the stimulus onset and further evolved during the stimulus presentation. As the end of the stimulus presentations approached, the angles acutely declined back to the pre-stimulus level. Therefore, the angles of response vectors during the stimulus presentations occupied a distinctively different range from the pre/post presentation. Comparing the angles over time across different intensities, we noticed that at measured intensities above 15 dB SPL, the angles over time were very similar without the scaling shown in the spiking rate and response variability. This similarity indicates that population responses may represent a vocalization identity in an intensity-invariant manner by encoding the information in the angle evolution of a trajectory.

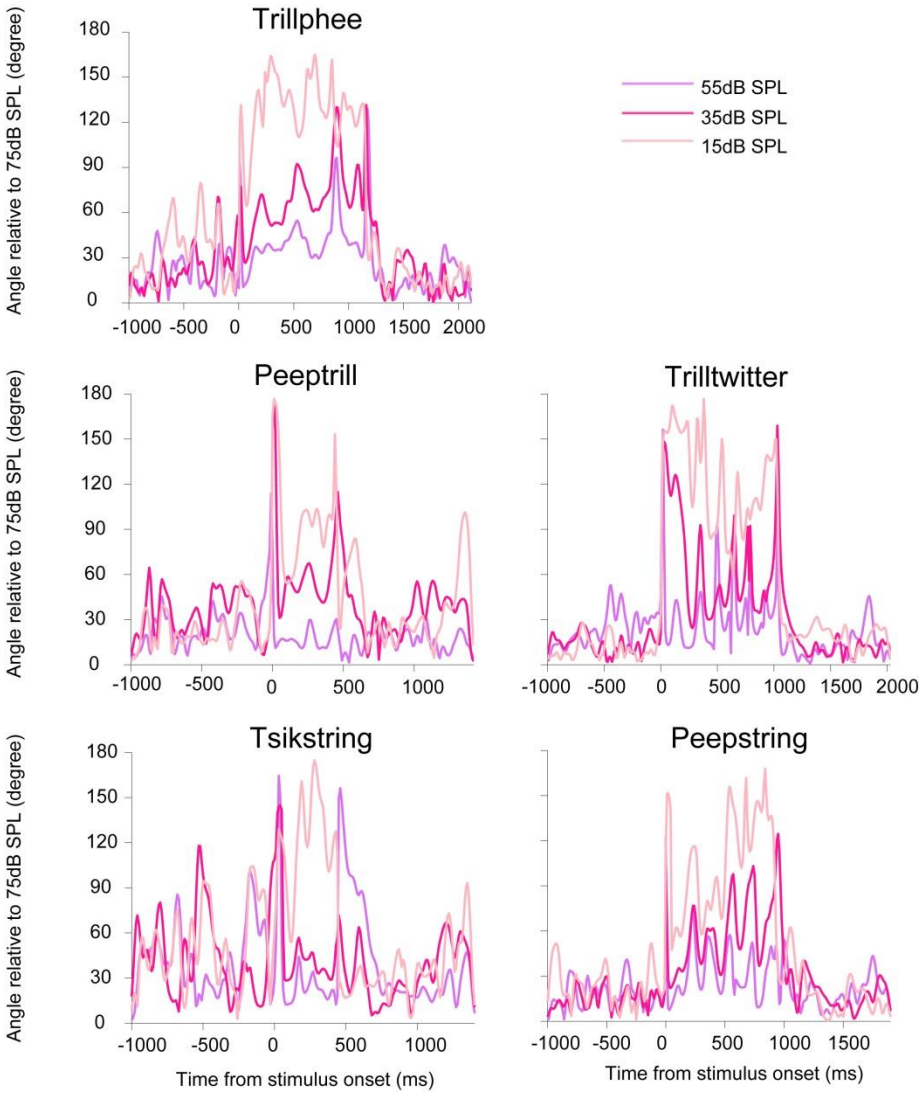

Figure 5.4 Evolution of the rotation angles of population responses at multiple intensities relative to the population response at 75dB SPL in 3D space.

Next, we quantified the influence of intensity on the deviation of response trajectories by computing the inter-trajectory angles between response trajectories of decreasing intensities relative to the reference trajectory at 75 dB SPL over time, as shown in **Figure 5.4**. The less intense the vocalization, the further away the corresponding population trajectory was from the reference trajectory in terms of angle rotations, which is consistent with a qualitative visual inspection in Figure 5.2. The rotation angles, however, are not equal over time. The pre-stimulus

and post-stimulus periods have rotation angles that fluctuated in the same range as that in Figure 5.3. For the stimulus-driven angle evolution, finer structures potentially related to the acoustic features of vocalizations can be observed. With regard to the angles between two neighboring intensities, for instance, 55 dB SPL vs 35 dB SPL, their difference at each time point is generally smaller than the angle difference between different points belonging to the same intensity. Rotation of population responses may serve as an indicator to encode the information of intensity.

To summarize, population responses to the same vocalization largely retain their intrinsic structures within trajectories in 3D space at multiple intensities. By contrast, the relationship between hyperplanes at different intensities is more complicated than just an equal angle shift.

5.3.3 Population Response Discrimination of Vocalizations across Intensities

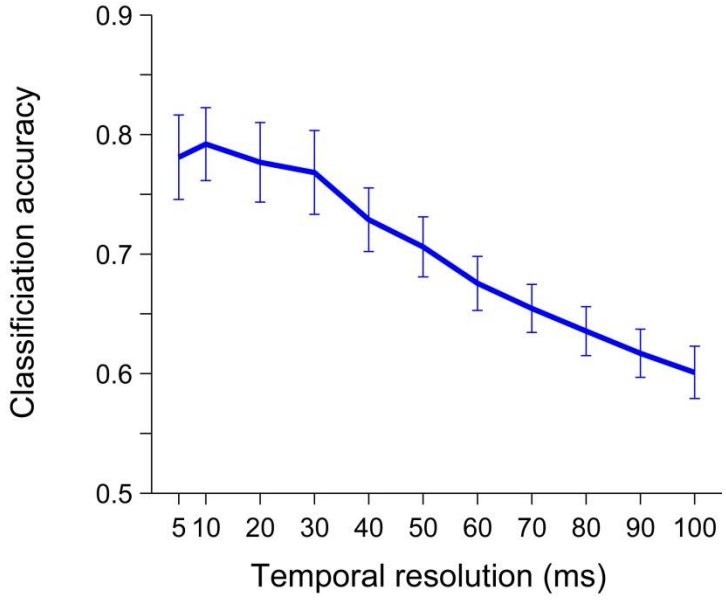

Figure 5.5 Population response discrimination across multiple intensities as a function of temporal resolution.

In previous analyses, we studied the variability and spatiotemporal structures of population responses to vocalizations at multiple intensities. Population responses averaged over

five to ten trials exhibited rich temporal dynamics in terms of rotation angles. Marmoset vocalizations have features spanning a wide range of time scales (Agamaite et al., 2015). Here, we further ask how the trial-by-trial population response discrimination between vocalizations at multiple intensities depends on the temporal resolution. How does the discrimination evolve over time? And how does the number of neurons in the population affect the discrimination?

We quantified the discriminability of population spike trains by building predictive models to decode vocalization types based upon a series of different temporal resolutions, which were used to bin spike trains into response vectors (spike trains of all vocalizations were truncated to the length of the shortest vocalization). As shown in **Figure 5.5**, the discrimination accuracy reaches an optimal level at ~ 10 ms, and degrades substantially with widened temporal resolutions. The minimum temporal resolution we tested here is 5 ms, and performance at that level also shows a decreasing trend.

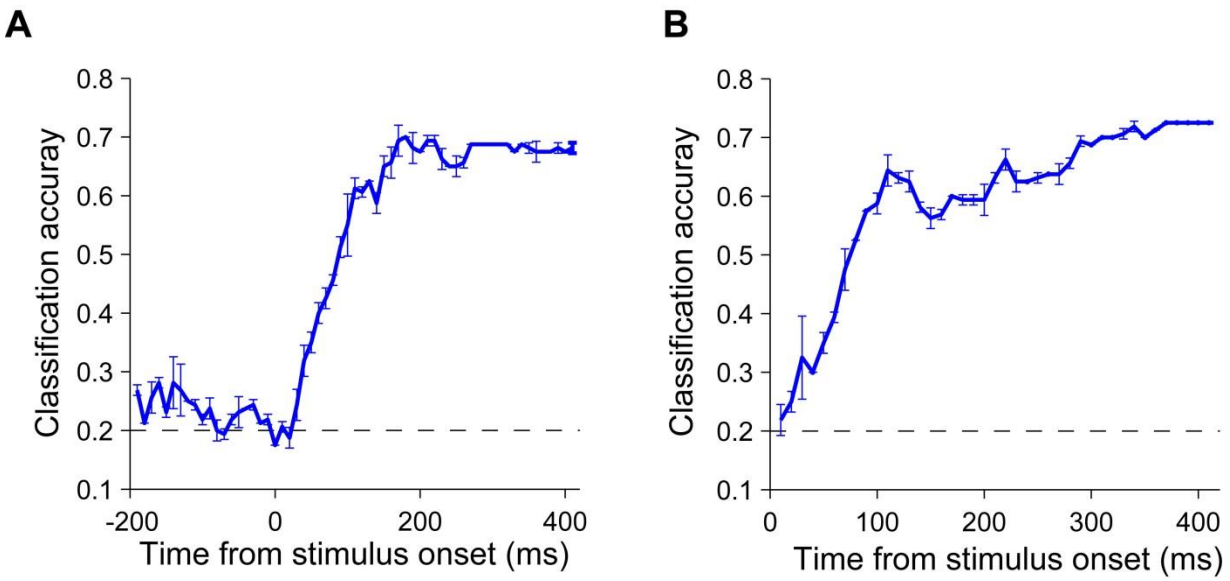

Figure 5.6 Time course of population response discriminations across multiple intensities (mean ± s.d.). (A) Discrimination with a single time bin. (B) Discrimination with cumulatively increasing numbers of time bins. Dashed lines denote performance at chance level.

To describe the discrimination dynamics over time, we built predictive models with both a single time bin and with increasing numbers of time bins, and obtained the results in **Figure 5.6**. Performance of predictive models based upon spontaneous activities is displayed in **Figure 5.6A**, along with that of models based upon stimulus–driven activities, as a control. Discrimination with a single time bin begins at the chance level, gradually increases following the onset of vocalizations, and achieves a steady state within 200 ms after stimulus onset. Discrimination with increasing lengths of spike trains is shown in **Figure 5.6B**, demonstrating a similar but slightly different trend. It also begins at the chance level, and steadily increases at a relatively fast speed within the first 100 ms after the onset of vocalizations. Later, it enters an oscillating and slowly increasing mode for about 300 ms, and finally reaches a plateau not long before the whole spike train is included.

Lastly, to evaluate the influence of the number of neurons on discriminability, we randomly sampled various numbers of neurons to build predictive models classifying 20 stimulus labels, until all neurons were included. The resulting discrimination result for each vocalization intensity condition is displayed in **Figure 5.7A**. Generally speaking, as more and more neurons are included, discrimination improves from the chance level to a plateau when the neuron numbers are between 200 and 300. Neural responses to all vocalizations at 75 dB SPL can be 100% classified when enough neurons are included. The responses at other intensities, however, are not all well classified. In addition, a relative higher intensity does not necessarily guarantee a better performance, as seen by comparing the 55 dB SPL performance of Trillphee and Peeptrill

comparing with their 35 dB SPL performance. A better evaluation of the classification performance can be obtained by the confusion matrix in **Figure 5.7B**. The confusion matrix provides a clearer view of how likely each neural response to a particular stimulus is to be mistakenly classified as another label. It shows that neural responses to the same vocalization but at different intensities above 15 dB SPL are less likely to be classified as other vocalization types. The matrix demonstrates that the vocalization type is relatively more robustly encoded than the intensity. At 15 dB SPL, misclassified labels seem to be equally distributed among different vocalizations, which makes sense given that the vocalizations are hardly audible at that level.

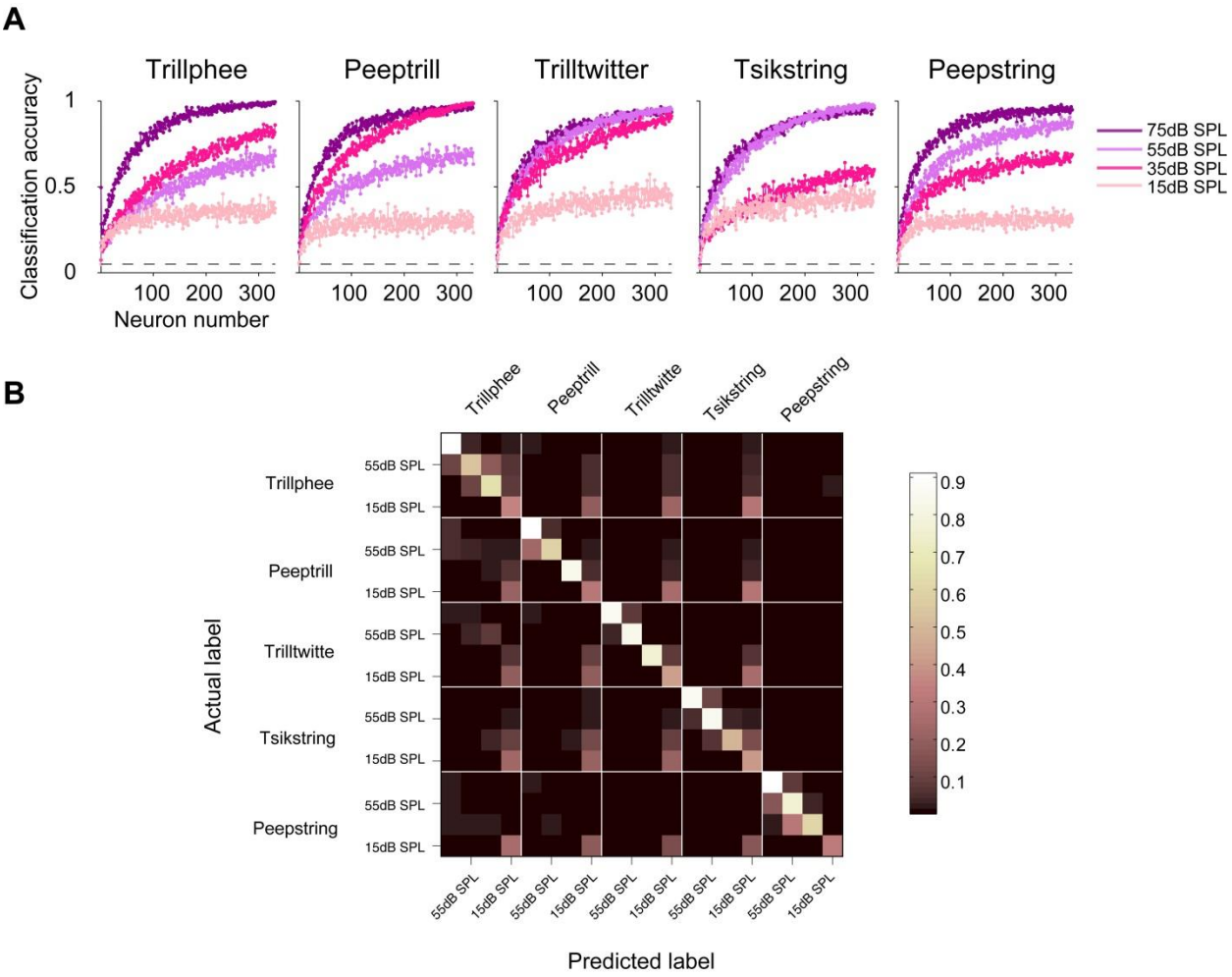

Figure 5.7 Discrimination of population response as a function of number of neurons in population (mean). (A) Discrimination accuracy for each vocalization intensity condition as a function of number of neurons. (B) Confusion matrix of discrimination performance averaged over numbers of neurons.

In the second half of this chapter, the influences of two noises on the population neural responses were studied in a similar way as that of intensity.

5.3.4 Population Response Variability of Vocalizations at Multiple SNRs

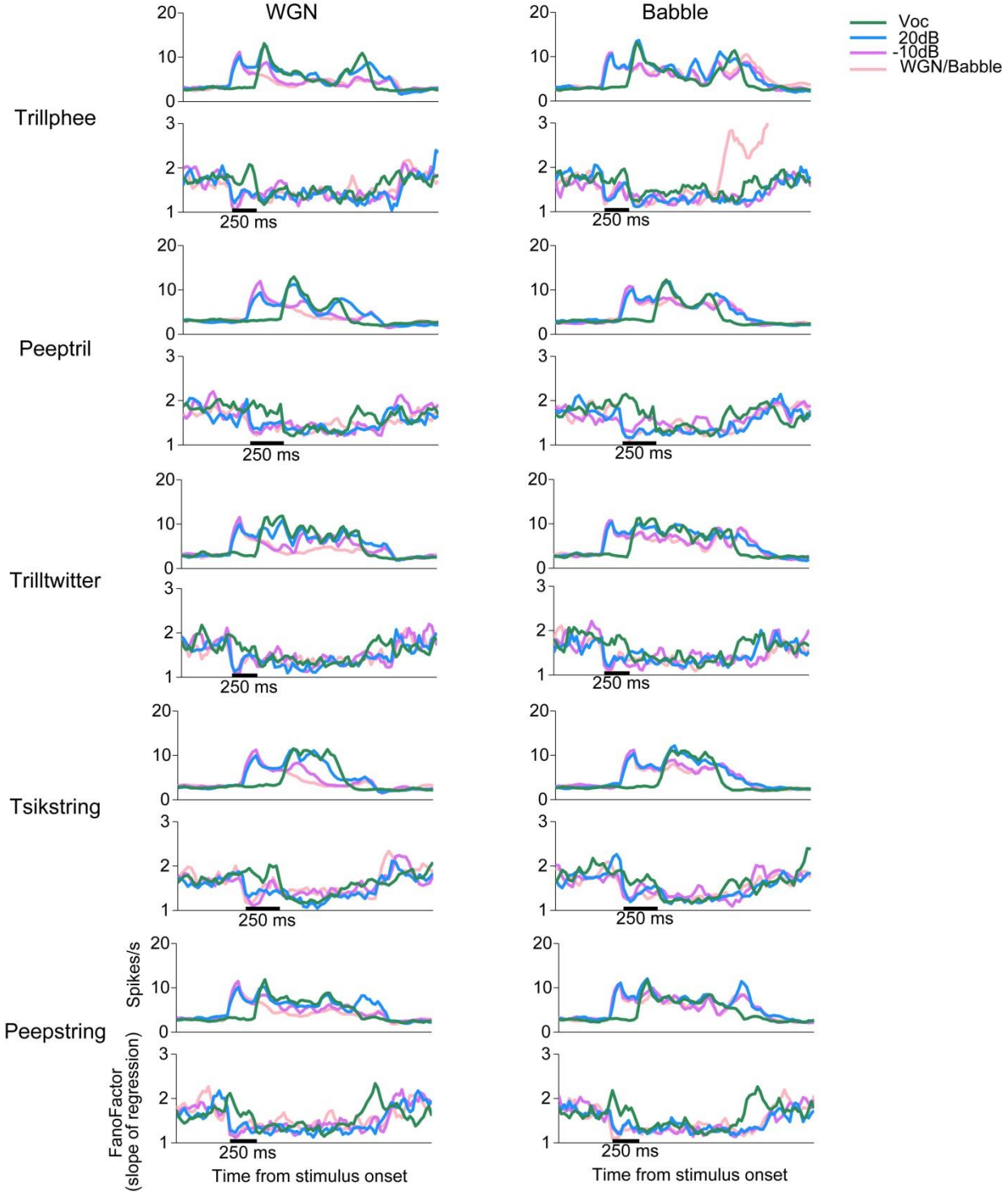

Figure 5.8 Population-averaged responses to vocalizations at multiple SNRs in WGN/Babble condition and the corresponding population activity variability with respect to time. The WGN condition is shown in the left column, and the Babble condition is shown in the right column.

How do the two different noises affect the population responses in terms of response variability? We computed the population-averaged spiking rate and the corresponding variability with the results shown in **Figure 5.8**. For visualization, only four SNR conditions out of ten were plotted. The onset responses under both noise conditions are very distinct across SNR levels, keeping in mind that the onset of the pure vocalization stimuli is 250 ms later than the vocalizations with noise, which have a preceding pure noise component. For vocalizations with 20 dB SNR, the neural responses seem to have a second onset once the vocalization component is introduced, while for -10 dB SNR, the second onset response is hardly discernable. The population response variability declines consistently at the stimulus onset for all stimuli, and the later introduction of vocalization in the auditory scene is somewhat captured by a second slight decline, but the degree of change compared to the preceding variability is much smaller than the first onset decline. The variability dynamics during noisy vocalization presentation fluctuate greatly, and do not seem to mirror the spiking rate.

We further computed the correlation between spiking rate and variability under both noise conditions in **Table 5.2** and **Table 5.3**. Comparing with pure vocalizations, which have significant negative correlation between these two metrics, population response variability at 20 dB SNR is less strong correlated with spiking rate in a negative way. As noise becomes the dominant component in the stimulus, correlations belonging to different vocalization types do not follow a consistent trend, with a wide of range values from negative to positive. Population neural responses to WGN generally do not exhibit significant correlation between spiking rate and variability, while for Babble, both significant positive and negative correlations are possible. Therefore, population neural responses to auditory scenes have a degrading negative correlation

between the time course of the spiking rate and variability as SNRs become lower, and the increase of noises yield a less simple and consistent trend .

Table 5.2 Pearson correlation between spiking rate and variability for vocalizations at multiple SNRs in the WGN condition

Vocalization	Voc	20 dB	-10 dB	WGN
Trillphee	***r* = -0.725** *p* = 3.77e-09	***r* = -0.339** *p* = 0.0173	***r* = 0.195** *p* = 0.181	***r* = -0.329** *p* = 0.0211
Peeptrill	***r* = -0.6124** *p* = 0.00320	***r* = -0.256** *p* = 0.264	***r* = 0.727** *p* = 1.91e-04	***r* = -0.0618** *p* = 0.790
Trilltwitter	***r* = -0.306** *p* = 0.0435	***r* = -0.246** *p* = 0.108	***r* = -0.754** *p* = 3.52e-09	***r* = -0.123** *p* = 0.426
Tsikstring	***r* = -0.851** *p* = 1.00e-06	***r* = -0.495** *p* = 0.022	***r* = -0.197** *p* = 0.391	***r* = 0.00880** *p* = 0.961
Peepstring	***r* = -0.664** *p* = 2.99e-06	***r* = -0.207** *p* = 0.201	***r* = 0.0812** *p* = 0.618	***r* = 0.084** *p* = 0.608

Table 5.3 Pearson correlation between spiking rate and variability for vocalizations at multiple SNRs in the Babble condition

Vocalization	Voc	20 dB	-10 dB	Babble
Trillphee	***r* = -0.725** *p* = 3.77e-09	***r* = -0.533** *p* = 8.19e-05	***r* = -0.326** *p* = 0.0225	***r* = 0.321** *p* = 0.0245
Peeptrill	***r* = -0.6124** *p* = 0.00320	***r* = -0.206** *p* = 0.370	***r* = 0.154** *p* = 0.504	***r* = -0.0341** *p* = 0.883
Trilltwitter	***r* = -0.306** *p* = 0.0435	***r* = -0.0386** *p* = 0.803	***r* = -0.361** *p* = 0.016	***r* = -0.145** *p* = 0.348
Tsikstring	***r* = -0.851** *p* = 1.00e-06	***r* = -0.572** *p* = 0.00670	***r* = -0.5051** *p* = 0.0195	***r* = 0.43** *p* = 0.0535
Peepstring	***r* = -0.664** *p* = 2.99e-06	***r* = -0.198** *p* = 0.221	***r* = -0.476** *p* = 0.00190	***r* = -0.647** *p* = 6.455e-06

5.3.5 Population Response Trajectory of Vocalizations at Multiple SNRs in 3D Space

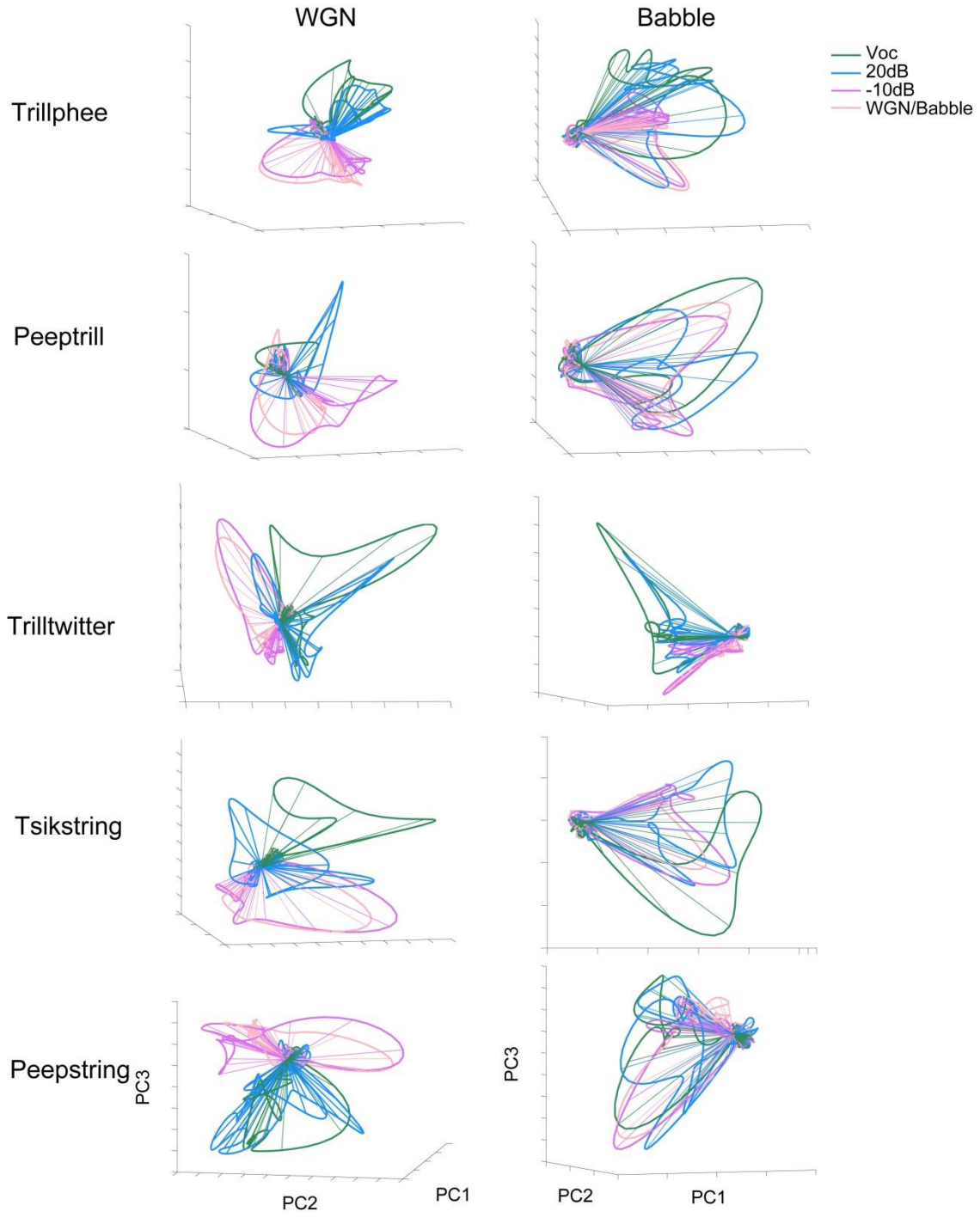

Figure 5.9 Trajectories of population responses to vocalizations at multiple SNRs with WGN/Babble in 3D space. The WGN condition is shown in the left column, and the Babble condition is shown in the right column.

To characterize the spatiotemporal structures of population responses to vocalizations with increasing amounts of noise, the associated population response trajectories based upon three principal components are displayed in **Figure 5.9**. Responses were projected to different 3D spaces under two noise conditions, but a salient differences can detected between how increasing the amount of different types of noise affects the population response structures. Under the WGN condition, two groups can be identified. Trajectories to vocalization and 20 dB SNR are clustered together, while trajectories to -10 dB SNR and pure WGN noise share a similar subspace. In contrast, trajectories to vocalizations masked with Babble noise do not form individual clusters, with a large portion overlapped across SNR levels.

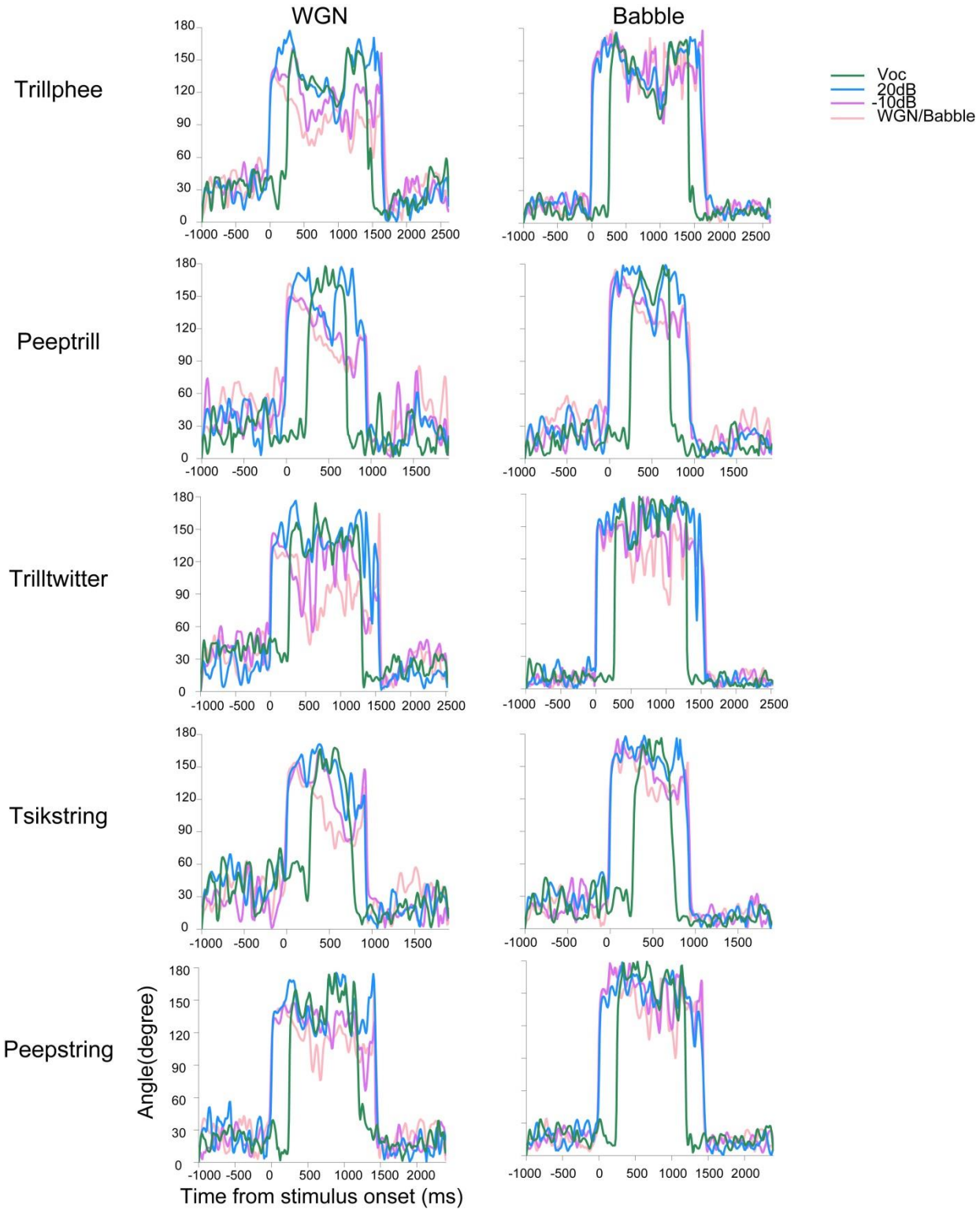

Figure 5.10 Evolution of rotation angles of the population response at multiple SNRs, relative to the first time point (in silence) with WGN/Babble in 3D space.

Rotation angles of each trajectory are quantified in **Figure 5.10**. Trajectories start with a pre-stimulus portion fluctuating below 60 degrees, greatly increase rotation angles to over 150 degrees following the stimulus onset, and further evolve during stimulus presentation, with particular structures associated with each vocalization. Trajectories at 20 dB SNR share a majority of features with those of clean vocalizations, and trajectories at -10 dB SNR are more noise-like. Again, angle evolution within trajectories of pure vocalizations and pure noise are more separated from each other in the WGN condition than in the Babble condition.

To quantify the distance between trajectories, we computed the rotation angles of trajectories of vocalizations at multiple SNRs levels relative to the trajectories of pure vocalizations, with the results shown in **Figure 5.11**. Two big peaks indicate the onset and offset responses induced by the two 250-ms noise segments. Time courses between these two peaks show that trajectories at 20 dB SNR have the smallest angular difference from that of pure vocalizations, below 30 degrees. Trajectories of -10 dB SNR and pure noise are further away. Figure 5.11 also quantitatively shows that WGN leads to more separated response trajectories than Babble.

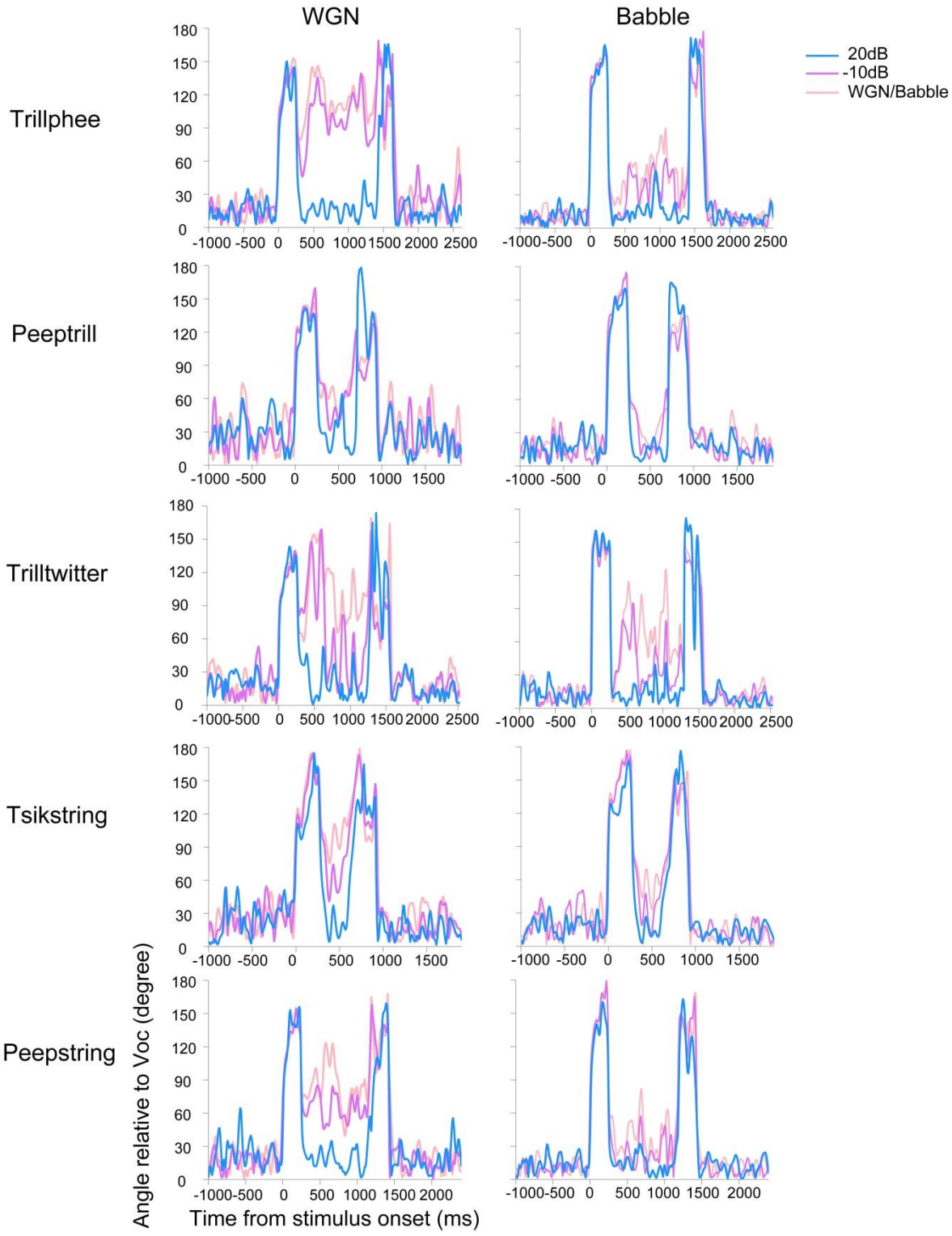

Figure 5.11 Evolution of rotation angles of population responses at multiple SNR, relative to clean vocalizations in WGN/Babble in 3D space.

5.3.6 Population Response Discrimination of Vocalizations across SNRs

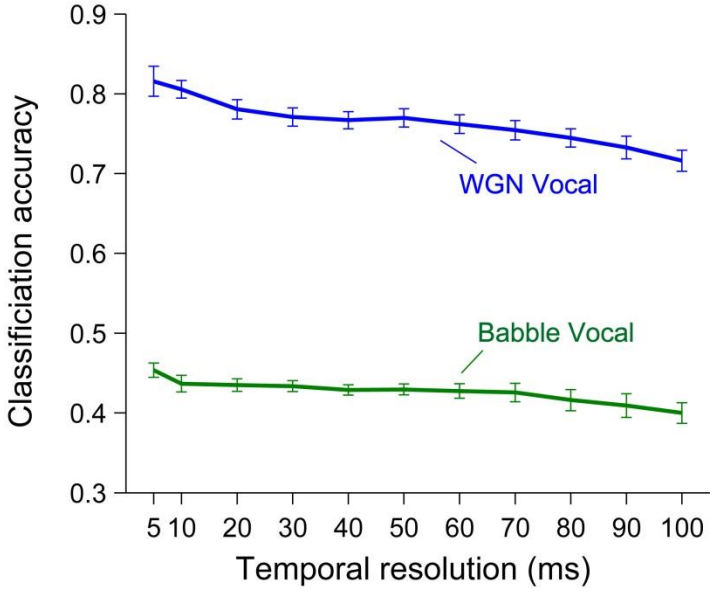

Figure 5.12 Population response discrimination across multiple SNRs in WGN/Babble condition as a function of temporal resolutions.

To evaluate the dependence of discriminability of population responses for vocalizations across multiple SNRs on the temporal resolution, we built predictive models based upon spiking trains binned by time windows of different lengths. Models were built to classify single trial population response to one of the five vocalizations or pure noise (c = 6), and evaluated by the percentage of correctly classified labels. Here, the number of time bins possessed by the shortest vocalization was used. The temporal resolution, in **Figure 5.12**, appears to be negatively associated with the classification accuracy. Based upon the range of time bins we investigated (5ms ~ 100 ms), a finer temporal resolution appears to provide a better discriminability. In addition, discrimination performance under WGN is about twice that under Babble.

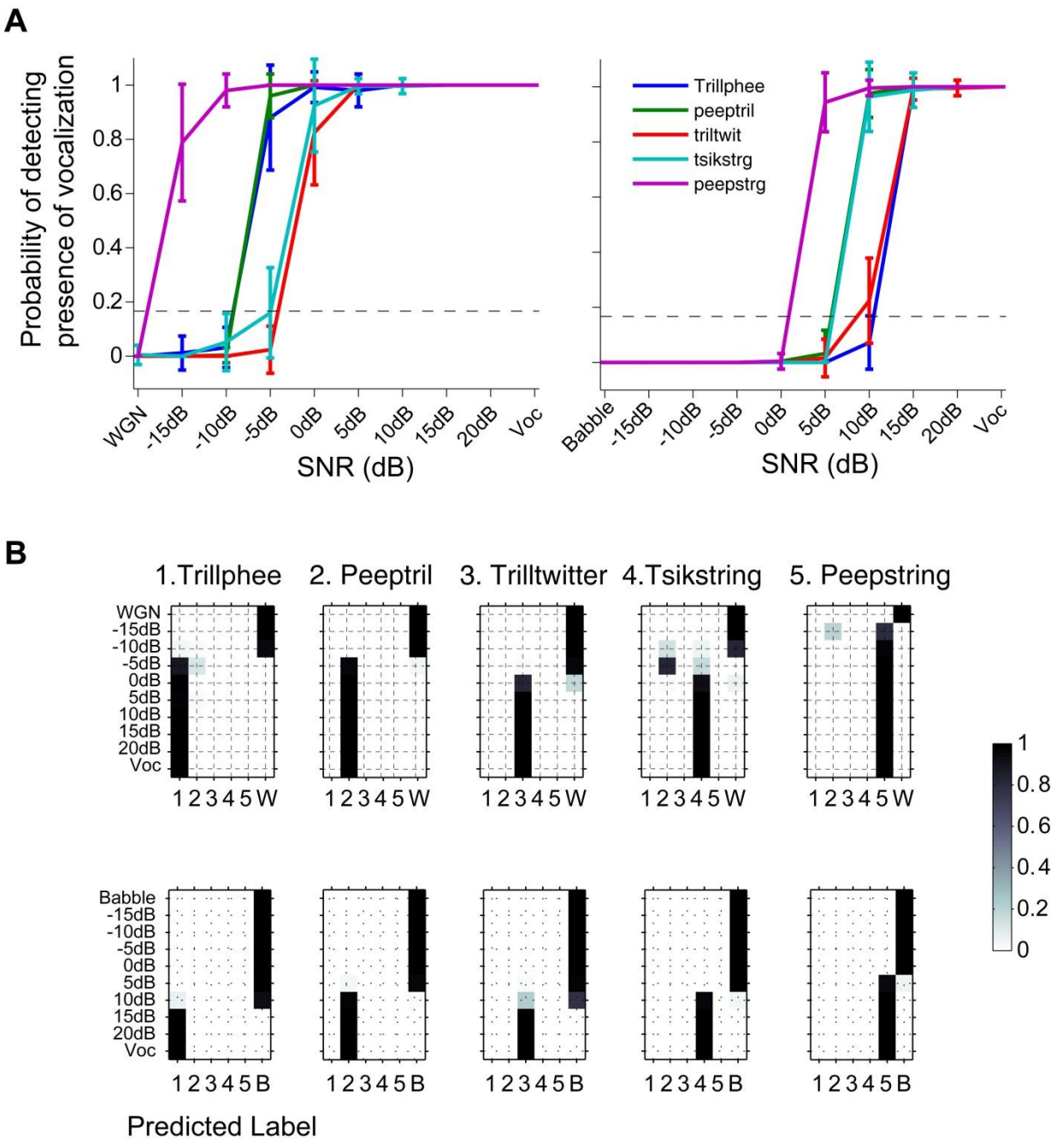

Figure 5.13 Classification performance of population neural responses. (A) Probability of detecting the presence of vocalizations as a function of SNR level under WGN/ Babble condition. (B) Confusion matrix under WGN/Babble condition. Labels of vocalizations are indicated by numbers from 1 to 5.

More details of the classification performance can be obtained by segregating the accuracy for each vocalization and SNR level as in **Figure 5.13**. The performance of each

vocalization is displayed as a function of SNR in **Figure 5.13A**. For the WGN condition, neural responses to vocalizations delivered with SNR above 5 dB can largely be identified as driven by the correct vocalization type, and neural responses delivered with SNR under -10 dB SNR are most likely to be classified as purely noise-induced, with -5dB and 0dB as the transition points. Neural responses to vocalizations under Babble noise tend to have higher detection thresholds between 0 dB SNR and 10 dB SNR. Under both noise conditions, Peepstring vocalization had the best discrimination over lower SNR levels than other four vocalizations. Whether those wrongly classified neural responses were classified as other types of vocalization of pure noise can be further inferred from **Figure 5.13B**. The confusion matrices clearly show that neuron responses driven by a particular vocalization are rarely wrongly classified as other types of vocalizations, except for Tsikstring at -5dB under WGN condition. Noises, instead, exert more interference on the neural responses.

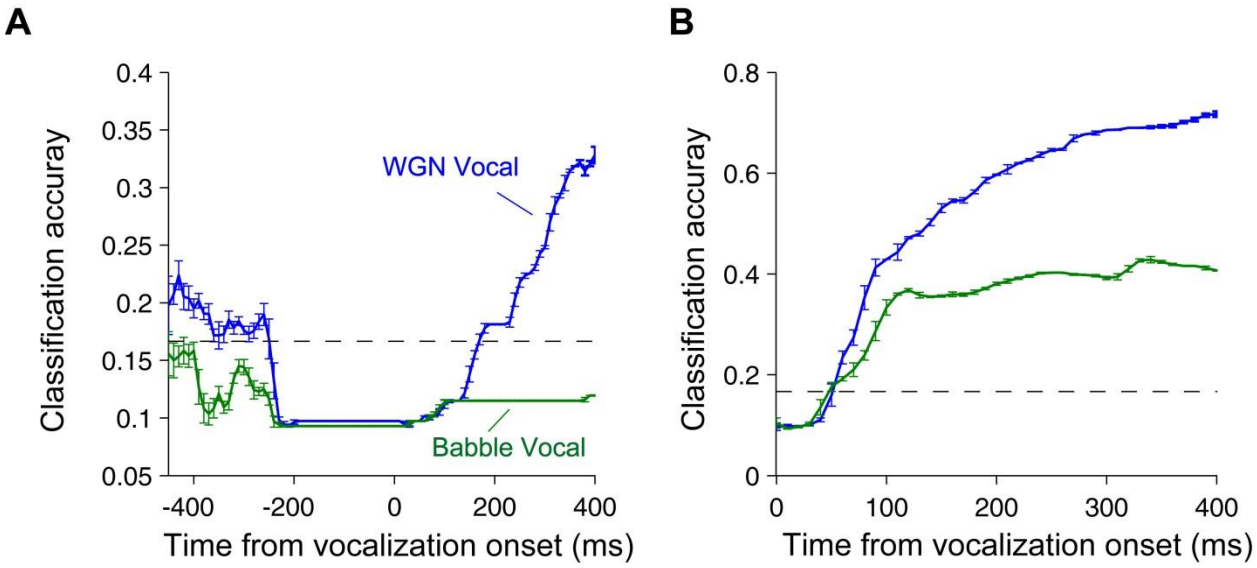

Figure 5.14 Time course of population response discrimination across multiple SNRs in the WGN/Babble condition (mean ± s.d.). (A) Discrimination with a single time bin. (B) Discrimination with cumulatively increasing numbers of time bins. Dashed lines denote performance at chance level.

We next studied the evolution of population discrimination over time by building predictive models using a single time bin and increased numbers of time bins, as shown in **Figure 5.14**. Discrimination based upon a single bin begins at the chance level, stabilizes at 0.1 for 250 ms of noise preceding the vocalization, and steadily increases following the onset of vocalization in the auditory scene (Figure 5.15A). Babble has a rather low performance based upon single bin response, even below the chance level. When information was integrated over more and more time bins, the discrimination of population neural responses improved with a steep slope for the first 100 ms following the vocalization onset, and were further boosted under the WGN condition, but reached a plateau under the Babble condition.

5.3.7 Discrimination Generalization over Multiple SNRs

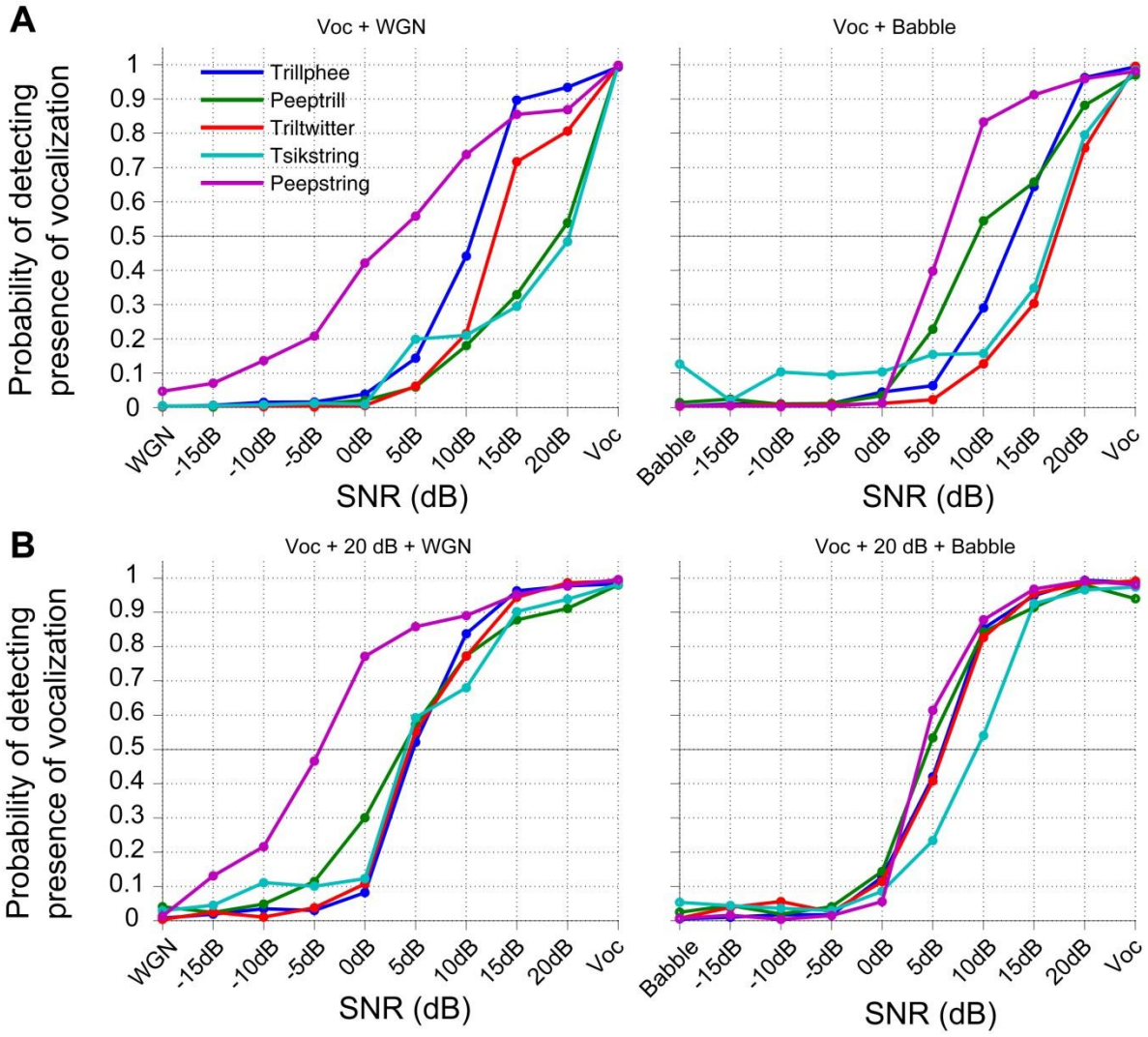

Figure 5.15 Discrimination of population responses using different training datasets (mean). (A) Performance of classifiers using neural responses to pure vocalization and pure noise as training samples (left, WGN; right, Babble). (B) Performance of classifiers using neural responses to pure vocalization, 20 dB SNR, and pure noise as training samples (left, WGN; right, Babble).

Performances of classifiers heavily depend on the quality of the training dataset. In the machine learning field, it is well known that adding an extra small amount of noise to the training dataset can improve the classifiers' generalization and obtain better performance (Bishop, 1995). Here, we explored the generalization of neural response classifiers by using different training datasets.

For each vocalization, we built separate binary SVM classifiers by using different numbers of time bins, ranging from a single time bin to all the time bins available to that vocalization. Given a trial of population response, the task of the classifiers was to predict whether the response was induced by pure noise or not. The performances of classifiers were averaged over different numbers of time bins. Two groups of training datasets were studied. The first group includes only neural responses to pure noise (labeled as noise) and pure vocalization (labeled as vocalization). The second group includes neural responses to 20 dB SNR as extra training samples labeled as vocalization. The resulting classifier performances are displayed in **Figure 5.15**. When only responses to pure noise and pure vocalization are used as training samples, the performances under both noise conditions are not ideal, and greatly degrade around 15 dB SNR. The performance of classifiers trained by the second group of neural responses shows an overall improvement, however. All the lines shift towards the left, with smaller differences between vocalizations, and lead to a lower detection threshold of around 5 dB SNR regardless of noise type. Therefore, by training on neural responses contaminated by a small amount of noise in the stimuli, we can obtain classifiers with more generalized performance over multiple SNR levels.

5.3.8 Subpopulation Response Discrimination of Vocalizations across SNRs

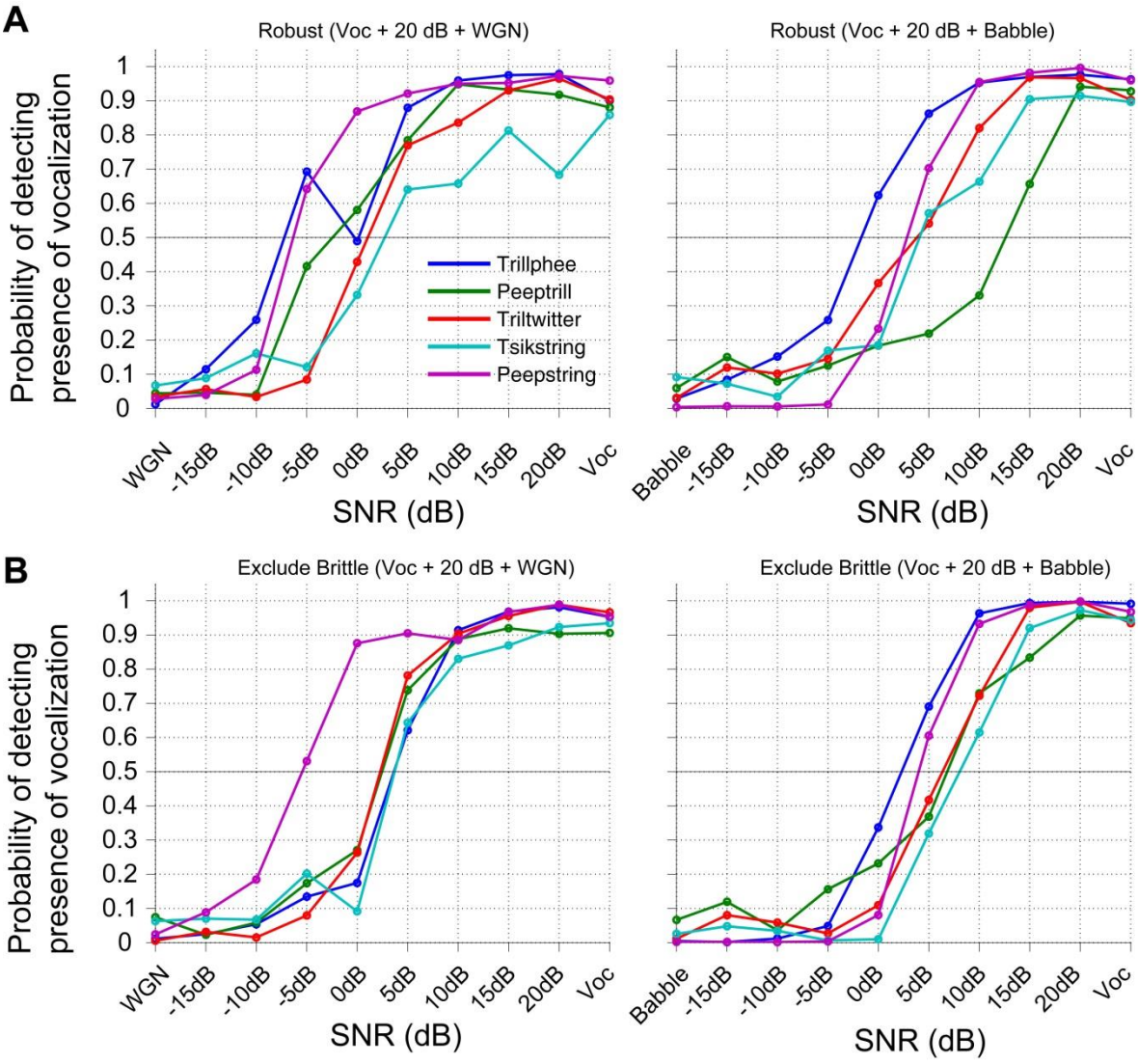

Figure 5.16 Discrimination of subpopulations of neurons using predictive models trained by neural response to pure vocalization, 20dB SNR, and pure noise (mean). (left, WGN; right,: Babble) (A) Performance of classifiers built upon only the robust group of neurons. (B) Performance of classifiers built upon a population of neurons, excluding the brittle group.

In Chapter 4, we showed that responses of individual neurons to noisy vocalizations can be categorized into four different groups: *robust*, *balanced*, *insensitive*, and *brittle*. Here, we investigate the discrimination of subpopulations of neurons by using pure vocalization, 20 dB SNR and pure noise collectively to train the classifiers. Two subpopulations of neurons are

shown in **Figure 5.16**: the robust group of neurons and population of neurons excluding the brittle group. Compared with Figure 5.16B, robust groups of neurons generally produce classifiers with a slightly lower detection threshold, but their performance curves are less smoothed. Considering all neurons except the brittle group adds smoothness and consistency between vocalizations.

5.4 Discussion

We examined how the responses of a population of A1 neurons encodes vocalizations at multiple intensities and SNR levels by studying the time course of population response variability and the spatiotemporal structures of reduced population responses. We also investigated how well the combined responses of populations of neurons could be used to discriminate among vocalizations under different conditions.

Stimulus-driven decline in the variability of neural states has been demonstrated to be a widespread feature of cortical responses, from the occipital to frontal cortex (Churchland et al., 2010). In our datasets of neural responses in the auditory cortex, the same trends, declining in the across-trial variability of the underlying firing rate following the onset of stimulus were overserved for both different intensity and SNR levels. The result is consistent with previous studies (Monier et al., 2003; Finn et al., 2007; Monier et al., 2008; Churchland et al., 2010). Moreover, we further probed the dynamics of variability during the stimulus presentation, and found that the firing-rate variability was significantly negatively correlated with the firing rate for vocalizations at multiple intensities. Under a more difficult perception condition, in auditory scenes, the negative correlation was not clearly exhibited. One potential explanation for the inconsistent findings between intensity and SNR is that we have nearly twice as many individual neurons in the dataset of intensity than noisy vocalizations. The relationship between the firing-

rate variability and the firing rate might not be well captured by a relatively small population of neurons. Alternatively, such a relationship might exist in subpopulations of neurons in response to noisy vocalizations, for instance, the robust group of neurons revealed in Chapter 4. If the second explanation holds, the change in variability might serve as a coding channel for vocalizations.

Temporally unstructured stimuli presentations have been demonstrated to induce neural codes of dynamic evolution (Sugase et al., 1999; Friedrich and Laurent, 2001; Stopfer et al., 2003; Hegdé and Van Essen, 2004; Bartho et al., 2009). By projecting high-dimensional neural responses into a lower dimensional space, we visualized the spatiotemporal structures of population responses induced by complex stimuli: vocalizations at multiple intensities and multiple SNR levels. Even though vocalizations delivered at different intensities are perceptually similar, population responses were progressively differentiated over time, and produced finer discrimination. Different vocalizations can be easily identified by their unique trajectories in space. Some trajectories are relatively smooth and simple, while others are more convoluted, and this variety is associated with the acoustic features of the vocalizations. Differentiation of population response trajectories over time in an auditory scene was dependent on the noise type. WGN noise led to more separable trajectories across SNR levels than Babble, and demonstrated spatiotemporal analysis as a useful indicator of the difficulty of vocalization perception. Consistent with the population coding of tone stimuli (Bartho et al., 2009), population response vectors had the largest rotation during the initial hundreds of milliseconds in response to vocalizations under different conditions, which is probably a common feature shared by population responses to acoustic stimuli regardless of the complexity of stimuli. In addition, we also implemented the same angle evolution analysis using raw population PSTH (data not

shown), and revealed a much weaker relationship between the angle evolution and vocalization temporal envelope. This finding indicates that the information of vocalizations is well encoded in a subset of neurons, as the reduced population responses are actually representations of partial covariance in the whole population.

Building neural response classifiers allowed us to investigate the optimal temporal resolution and temporal dynamics of cortical detection and discrimination. Cortical discrimination has been extensively studied for single units, and the optimal temporal resolution was demonstrated to be 10 ms (Rieke, 1999; Machens et al., 2003; Narayan et al., 2006; Schneider and Woolley, 2010). For our population of neurons, we also found that the temporal resolution for cortical discrimination between vocalizations at multiple intensities on the population level was optimized around 10 ms. This time scale is small enough to capture temporal structures of vocalization, and wide enough to allow integration of information over time, thereby reducing noise. The temporal resolution for cortical discrimination between noisy vocalizations, however, was smaller than 10 ms. In our analysis, the best value was at 5 ms, which is the finest temporal resolution studied here. It is possible that the optimal temporal resolution for noisy vocalization discrimination is below 5 ms. A finer temporal resolution might reduce the interference of the noise component in the auditory scene on vocalization recognition, because a longer time window potentially introduces more noise information, thus confounding the vocalization discrimination. The analysis of the temporal dynamics of discrimination revealed a range for the time scale of integration on the order of hundreds of milliseconds, with ~100 ms for vocalizations at multiple intensities and ~300 ms for noisy vocalizations. The time scale of integration provides information about the speed of accumulation of discrimination accuracy at the population level, and is in similar range to that of single units.

Whether information about sensory stimuli is best represented by the whole population of neurons or a subpopulation of neurons was debated. The discrimination by subpopulations of neurons was particularly studied for noisy vocalizations. We found that the subpopulation of the robust group of neurons and the subpopulations of neurons excluding the brittle group both yielded slightly lower detection thresholds than the whole population, thus a better discrimination performance. This result indicates that the brittle group of neurons contributes as a neural distractor for noisy vocalization discrimination, and that information about vocalization is better encoded by a subpopulation of neurons instead. We also built classifiers to demonstrate that we can generalize the discrimination over lower SNR levels by using neural responses contaminated by a little acoustic noise as training samples. While we are not saying that the brain actually decodes vocalization information in the same way that our classifiers do, the results are consistent with a previous psychoacoustic study that demonstrated that introducing weak noises in perception improved the detection thresholds of target signals (Zeng et al., 2000).

In summary, we investigated our data with population analytic techniques and revealed population response dynamics that cannot be fully evaluated by single-unit analysis alone.

Chapter 6: Conclusions and Recommendations for Future Work

6.1 Conclusions

A long-standing puzzle in auditory neuroscience is the neural mechanism of robust perception of behaviorally relevant acoustic signals. Auditory scene analysis was proposed as a model to explain this phenomenon, which refers to the ability of integrating segregated acoustic elements and form auditory streams. In this dissertation, individual neurons in the primary auditory cortex of awake marmoset monkeys were collected, and their responses to marmoset conspecific vocalizations masked with WGN/Babble noise were investigated. Datasets were analyzed using both single-unit analysis and population analysis. The results generally showed there were subgroups of neurons to correspondingly encode the vocalization and noise streams in the stimuli.

Several conclusions were drawn from the single-unit analysis. First, response averaged over individual neurons demonstrated that Babble had a greater degradation on the vocalization encoding than WGN regarding spiking rate and response reliability. Second, four consistent response types (*robust*, *balanced*, *sensitive*, and *brittle*) in terms of individual neurons' ability to resist noise were found regardless of noise types. However, which response type an individual neuron belongs to was noise-dependent. Third, the ability of individual neurons to discriminate vocalizations at multiple intensities was not significantly correlated with each neuron's ability to resist noise interference. Last, a subset of neurons in A1 was found to have feature-aligned responses to WGN, and these neurons encoded both vocalizations and WGN with high information rates.

In this dissertation, only neurons in A1 were investigated. It is possible that neurons in other core areas, such as the rostral (R) field or the rostrotemporal (RL) field, may have shown greater resistance to noise degradation of vocalization-induced activities given those areas' different response properties from A1. There is also evidence that noise degradation would be less pronounced in the belt areas, such as anterolateral belt (AL), in old-world monkeys. For the neurons recorded in A1, their responses to noisy vocalizations were categorized into four different groups based upon EI profiles. However, this does not mean individual neurons' responses are strictly discrete. As the distribution of mean EI values is unimodal instead of multimodal, it is more likely that individual neurons' noise resistance lies on a continuum between vocalization sensitive and noise sensitive, while the location of the individual neurons' noise resistance on the continuum is still noise-dependent.

One limitation of the study is the fundamental response properties of each group of neurons were not explicitly studied. Given the response properties of higher-order auditory neurons and the previous literature, neurons can be highly responsive to vocalizations without a clearly meaningful response field, as indicated by our spike-triggered analysis on neurons with feature-aligned responses to both vocalizations and WGN. It is likely that neural response properties such as integration time and input-output function slope do correlate with vocalization-in-noise response classes described in this dissertation, but the limitation on holding neurons long enough to acquire their responses to many similar yet slightly different features prohibited this additional analysis in this case. Our finding that robust A1 vocalization responses were often generated by different neurons in different contexts likely means that the important acoustic features for this phenomenon vary, and/or the robust extraction of vocalization is incomplete at the level of A1. A productive future study may be to repeat this analysis in

downstream areas to determine if robust vocalization encoding appears there, as it seems to in songbird forebrain.

Complementing the single-unit analysis, the population analysis revealed more details about the dynamics of a population of neurons. First, population response variability was found to mirror the population spiking rate in response to vocalizations at multiple intensities, but the trend was much less significant for vocalizations in noise at multiple SNR levels. Second, population responses to vocalizations across intensities exhibited distinct spatiotemporal structures and differences in masking effects of WGN and Babble could be visualized by the angle distance between population response trajectories to noisy vocalizations and pure vocalizations. Third, discrimination of population responses to noisy vocalizations had a finer optimal temporal resolution and a longer time scale for integration than those of discrimination of population response to pure vocalizations. Finally, we demonstrated that subpopulations of neurons had slightly better discrimination than the whole population when the brittle group of neurons was not considered.

Consistent with studies in the olfactory system, our population analysis showed that population response trajectories are able to systematically track the alternation induced in neural responses by modulation of complex acoustic stimuli, which were quantified by intra-trajectory and inter-trajectory angle evolutions. There are two limitations in the current trajectory analysis, which can be further addressed in future study. First of all, the current trajectory analysis was conducted on trial-averaged population responses. It would be helpful to visualize the variance of trial-by-trial population response trajectories as a supplement to the trial-by-trial discriminability analysis. Furthermore, dimensionality reduction was implemented for each vocalization separately, thus population responses to different vocalizations were projected to different

reduced spaces. Although it is shown that each vocalization is represented by a population response with unique spatiotemporal structures, we cannot directly compare those trajectories belonging to different vocalizations. To determine whether similar acoustic features also lead to similar spatiotemporal structures, population responses to different vocalizations should be reduced to the same 3D space. With regard to the discriminability analysis, it was very interesting to show that adding 20 dB SNR level neural response as training dataset achieved more robust classifiers. A more systematic investigation could be done to sequentially test the effect of neural response of each individual SNR level as training dataset, and finding the highest SNR level where overall performance begins to decline.

6.2 Recommendation for Future Work

Both the single-unit and population analysis enhanced our understanding of how neurons in A1 cope with the noise distractions. Extensions based upon this dissertation are recommendation for future work. Contextual effects of individual neurons were demonstrated using two types of noises. To further test the degree of generalizability of this effect, more types of noise could be investigated, such as natural environment sounds. Further, the auditory scene studies in this dissertation only have two sources: vocalization and one distracting noise. It is worth well to add a third stream into the auditory scene to test whether the neuron response types would vary as the number of auditory streams in the stimuli. Though there is evidence to show that sequential recording and simultaneous recording yield roughly the same results, simultaneous recording is of great value for future work. Simultaneous recording is more efficient to collecting a large number of neurons in a relatively short time frame, and allows us to form and test new hypothesis more conveniently. Last but not least, experimental design can

include behavioral tasks so that both neural representation and perception can be probed, and also allows for studying the effect of attention on the auditory scene analysis.

References

Abolafia JM, Martinez-Garcia M, Deco G, Sanchez-Vives MV (2013) Variability and information content in auditory cortex spike trains during an interval-discrimination task. J Neurophysiol 110:2163-2174.

Aertsen AM, Johannesma PI (1981) The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol Cybern 42:133-143.

Agamaite JA, Chang CJ, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). J Acoust Soc Am 138:2906-2928.

Aggelopoulos NC, Franco L, Rolls ET (2005) Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. J Neurophysiol 93:1342-1357.

Aitkin LM, Merzenich MM, Irvine DR, Clarey JC, Nelson JE (1986) Frequency representation in auditory cortex of the common marmoset (Callithrix jacchus jacchus). J Comp Neurol 252:175-185.

Anderson B, Sanderson MI, Sheinberg DL (2007) Joint decoding of visual stimuli by IT neurons' spike counts is not improved by simultaneous recording. Experimental brain research 176:1-11.

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

Baeg E, Kim Y, Huh K, Mook-Jung I, Kim H, Jung M (2003) Dynamics of population code for working memory in the prefrontal cortex. Neuron 40:177-188.

Bar-Yosef O, Nelken I (2007) The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. Front Comput Neurosci 1:3.

Bar-Yosef O, Rotman Y, Nelken I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J Neurosci 22:8619-8632.

Barbour DL (2011) Intensity-invariant coding in the auditory system. Neurosci Biobehav Rev 35:2064-2072.

Barbour DL, Wang X (2003a) Auditory cortical responses elicited in awake primates by random spectrum stimuli. J Neurosci 23:7194-7206.

Barbour DL, Wang X (2003b) Contrast tuning in auditory cortex. Science 299:1073-1075.

Bartho P, Curto C, Luczak A, Marguet SL, Harris KD (2009) Population coding of tone stimuli in auditory cortex: dynamic rate vector analysis. Eur J Neurosci 30:1767-1778.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161-1165.

Bendor D, Wang XQ (2008) Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. J Neurophysiol 100:888-906.

Bezerra BM, Souto A (2008) Structure and usage of the vocal repertoire of Callithrix jacchus. International Journal of Primatology 29:671-701.

Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K (2008) Invariance and sensitivity to intensity in neural discrimination of natural sounds. J Neurosci 28:6304-6308.

Billings CJ, McMillan GP, Penman TM, Gille SM (2013) Predicting perception in noise using cortical auditory evoked potentials. Journal of the Association for Research in Otolaryngology 14:891-903.

Bishop CM (1995) Training with noise is equivalent to Tikhonov regularization. Neural Comput 7:108-116.

Boulton AA, Baker GB, Vanderwolf CH (1990) Neurophysiological techniques: applications to neural systems: Humana Press.
Bregman AS (1994) Auditory scene analysis: The perceptual organization of sound: MIT press.
Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat Neurosci 7:456-461.
Brungart DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am 109:1101-1109.
Buzsáki G (2004) Large-scale recording of neuronal ensembles. Nat Neurosci 7:446-451.
Carhart R, Tillman TW, Greetis ES (1969) Perceptual masking in multiple sound backgrounds. J Acoust Soc Am 45:694-703.
Carney LH, Geisler CD (1986) A temporal analysis of auditory‐nerve fiber responses to spoken stop consonant–vowel syllables. J Acoust Soc Am 79:1896-1914.
Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I (2006) Reduction of information redundancy in the ascending auditory pathway. Neuron 51:359-368.
Cherry EC (1953) Some Experiments on the Recognition of Speech, with One and with 2 Ears. Journal of the Acoustical Society of America 25:975-979.
Churchland MM et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat Neurosci 13:369-378.
Churchland P, Sejnowsky T (1991) Computational neuroscience: MIT Press: Cambridge, Mass.
de Boer R, Kuyper P (1968) Triggered correlation. IEEE Trans Biomed Eng 15:169-179.
de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. Science 275:1805-1808.
DeCasper AJ, Fifer WP (1980) Of human bonding: newborns prefer their mothers' voices. Science 208:1174-1176.
Delgutte B, Kiang NY (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. J Acoust Soc Am 75:866-878.
Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85:1220-1234.
DiMattina C, Wang X (2006) Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. J Neurophysiol 95:1244-1262.
Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. P Natl Acad Sci USA 109:11854-11859.
Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728-5735.
Dirks DD, Bower DR (1969) Masking effects of speech competing messages. Journal of Speech, Language, and Hearing Research 12:229-245.
Egan JP, Carterette EC, Thwing EJ (1954) Some Factors Affecting Multi‐Channel Listening. J Acoust Soc Am 26:774-782.
Eggermont JJ, Johannesma PM, Aertsen AM (1983) Reverse-correlation methods in auditory research. Q Rev Biophys 16:341-414.
Elhilali M, Fritz JB, Klein DJ, Simon JZ, Shamma SA (2004) Dynamics of precise spike timing in primary auditory cortex. J Neurosci 24:1159-1172.
Feng AS, Hall JC, Gooler DM (1990) Neural basis of sound pattern recognition in anurans. Prog Neurobiol 34:313-329.

Finn IM, Priebe NJ, Ferster D (2007) The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. Neuron 54:137-152.
Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116:1656-1670.
Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151:167-187.
Fletcher H, Galt RH (1950) The perception of speech and its relation to telephony. J Acoust Soc Am 22:89-151.
French N, Steinberg J (1947) Factors governing the intelligibility of speech sounds. J Acoust Soc Am 19:90-119.
Friedrich RW, Laurent G (2001) Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. Science 291:889-894.
Gai Y, Carney LH (2008) Influence of inhibitory inputs on rate and timing of responses in the anteroventral cochlear nucleus. J Neurophysiol 99:1077-1095.
Gehr DD, Komiya H, Eggermont JJ (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. Hear Res 150:27-42.
Gochin PM, Colombo M, Dorfman GA, Gerstein GL, Gross CG (1994) Neural ensemble coding in inferior temporal cortex. J Neurophysiol 71:2325-2337.
Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. J Neurophysiol 89:472-487.
Grimsley JM, Shanbhag SJ, Palmer AR, Wallace MN (2012) Processing of communication calls in Guinea pig auditory cortex. PLoS One 7:e51646.
Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. Nature 452:220-224.
Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol 441:197-222.
Hashikawa T, Nakatomi R, Iriki A (2015) Current models of the marmoset brain. Neurosci Res 93:116-127.
Hegdé J, Van Essen DC (2004) Temporal dynamics of shape analysis in macaque visual area V2. J Neurophysiol 92:3030-3042.
Imig TJ, Irons WA, Samson FR (1990) Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. J Neurophysiol 63:1448-1466.
Jancke L, Mirzazade S, Shah NJ (1999) Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neurosci Lett 266:125-128.
Jolliffe I (2002) Principal component analysis: Wiley Online Library.
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.
Libersat F, Murray JA, Hoy RR (1994) Frequency as a releaser in the courtship song of two crickets, Gryllus bimaculatus (de Geer) and Teleogryllus oceanicus: a neuroethological analysis. J Comp Physiol A 174:485-494.

Machens CK, Schütze H, Franz A, Kolesnikova O, Stemmler MB, Ronacher B, Herz AV (2003) Single auditory neurons rapidly discriminate conspecific communication signals. Nat Neurosci 6:341-342.
Marmosetcare.com (2011) http://www.marmosetcare.com/understanding-behaviour/calls.html. In.
McIlwain JT (2001) Population coding: a historical sketch. Prog Brain Res 130:3-7.
Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol 100:1407-1419.
Miller LM, Schreiner CE (2000) Stimulus-based state control in the thalamocortical system. J Neurosci 20:7011-7016.
Monier C, Fournier J, Frégnac Y (2008) In vitro and in vivo measures of evoked excitatory and inhibitory conductance dynamics in sensory cortices. Journal of neuroscience methods 169:323-365.
Monier C, Chavane F, Baudot P, Graham LJ, Frégnac Y (2003) Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. Neuron 37:663-680.
Moore RC, Lee T, Theunissen FE (2013) Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. PLoS Comput Biol 9:e1002942.
Morel A, Kaas JH (1992) Subdivisions and connections of auditory cortex in owl monkeys. J Comp Neurol 318:27-63.
Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol 87:1723-1737.
Narayan R, Grana G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. J Neurophysiol 96:252-258.
Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K (2007) Cortical interference effects in the cocktail party problem. Nat Neurosci 10:1601-1607.
Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14:474-480.
Newman JD, Wollberg Z (1973) Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. Brain research 54:287-304.
Nikolić D, Haeusler S, Singer W, Maass W (2006) Temporal dynamics of information content carried by neurons in the primary visual cortex. In: Advances in neural information processing systems, pp 1041-1048.
Oram MW, Wiener MC, Lestienne R, Richmond BJ (1999) Stochastic nature of precisely timed spike patterns in visual system neuronal responses. J Neurophysiol 81:3021-3033.
Panzeri S, Pola G, Petersen RS (2003) Coding of sensory signals by neuronal populations: the role of correlated activity. The Neuroscientist 9:175-180.
Paxinos G, Watson C, Petrides M, Rosa M, Tokuno H (2012) The marmoset brain in stereotaxic coordinates: Elsevier.
Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biol 4:e215.
Phillips DP, Cynader MS (1985) Some neural mechanisms in the cat's auditory cortex underlying sensitivity to combined tone and wide-spectrum noise stimuli. Hear Res 18:87-102.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Current Biology 18:1124-1128.
Rabinowitz NC, Willmore BD, King AJ, Schnupp JW (2013) Constructing noise-invariant representations of sound in the auditory pathway. PLoS Biol 11:e1001710.
Rieke F (1999) Spikes: exploring the neural code: MIT press.
Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc Biol Sci 262:259-265.
Ruggero MA (1973) Response to noise of auditory nerve fibers in the squirrel monkey. J Neurophysiol 36:569-587.
Sadagopan S, Wang X (2008) Level invariant representation of sounds by populations of neurons in primary auditory cortex. J Neurosci 28:3415-3426.
Saha D, Leong K, Li C, Peterson S, Siegel G, Raman B (2013) A spatiotemporal coding mechanism for background-invariant odor recognition. Nat Neurosci 16:1830-1839.
Schneider DM, Woolley SM (2010) Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. J Neurophysiol 103:3248-3265.
Schneider DM, Woolley SM (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. Neuron 79:141-152.
Schreiber S, Fellous JM, Whitmer D, Tiesinga P, Sejnowski TJ (2003) A new correlation-based measure of spike timing reliability. Neurocomputing 52-54:925-931.
Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. J Acoust Soc Am 115:813-821.
Shamma SA, Micheyl C (2010) Behind the scenes of auditory perception. Curr Opin Neurobiol 20:361-366.
Stephan H, Baron G, Schwerdtfeger WK (2012) The brain of the common marmoset (Callithrix jacchus): a stereotaxic atlas: Springer Science & Business Media.
Stopfer M, Jayaraman V, Laurent G (2003) Intensity versus identity coding in an olfactory system. Neuron 39:991-1004.
Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. Nature 400:869-873.
Tervaniemi M, Kruck S, De Baene W, Schroger E, Alter K, Friederici AD (2009) Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. Eur J Neurosci 30:1636-1642.
Theunissen FE, Elie JE (2014) Neural processing of natural sounds. Nat Rev Neurosci 15:355-366.
Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315-2331.
Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63:411-423.
Valentine PA, Eggermont JJ (2004) Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. Hear Res 196:119-133.
Van Gestel T, Suykens JA, Lanckriet G, Lambrechts A, De Moor B, Vandewalle J (2002) Bayesian framework for least-squares support vector machine classifiers, gaussian processes, and kernel Fisher discriminant analysis. Neural Comput 14:1115-1147.

Vinje WE, Gallant JL (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. J Neurosci 22:2904-2915.
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.
Wang X (2007) Neural coding strategies in auditory cortex. Hear Res 229:81-93.
Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685-2706.
Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341-346.
Watkins PV, Barbour DL (2011) Rate-level responses in awake marmoset auditory cortex. Hear Res 275:30-42.
Willmore BD, Cooke JE, King AJ (2014) Hearing in noisy environments: noise invariance and contrast gain control. J Physiol 592:3371-3381.
Woolley SM, Gill PR, Theunissen FE (2006) Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. J Neurosci 26:2499-2512.
Zeng F-G, Fu Q-J, Morse R (2000) Human hearing enhanced by noise. Brain research 869:251-255.