THE EFFECTS OF PASSIVE FAMILIARIZATION ON NEURAL AND BEHAVIORAL DISCRIMINATION OF ACOUSTIC SIGNALS

By

EFE SOYMAN

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

David S. Vicario, PhD

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January 2018

ABSTRACT OF THE DISSERTATION

The Effects of Passive Familiarization on Neural and Behavioral Discrimination of Acoustic Signals

by EFE SOYMAN

Dissertation Director:

David S. Vicario, PhD

Organisms encounter numerous novel sensory signals throughout life. Thus, sensory representations in the adult brain, set up during ontogeny through the interaction of early experience with innate organizational principles, must undergo dynamic changes to adapt to the complexity of the external world. This thesis investigates how passive exposure to novel sounds modifies neural representations to facilitate recognition and discrimination, using the zebra finch model organism. Zebra finches use complex, learned acoustic signals for social communication with many parallels to human speech. Furthermore, the neural responses in an auditory structure in the zebra finch brain, Caudal Medial Nidopallium (NCM), undergo a long-term form of adaptation with repeated stimulus presentation, providing an excellent substrate to probe the neural underpinnings of adaptive sensory representations. In Experiment 1, electrophysiological activity in NCM was recorded under passive listening conditions as novel natural vocalizations were familiarized through playback. Neural decoding of stimuli using the temporal profiles of both multi-unit and single-unit neural responses improved dramatically during the first

few stimulus presentations. During subsequent encounters, these signals were successfully recognized after hearing fewer initial acoustic features. Remarkably, the accuracy of neural decoding was higher when different stimuli were heard in separate blocks compared to when they were presented randomly in a shuffled sequence. Experiment 2 supported and extended these findings by showing that the rapid gains in neural decoding of natural vocalizations with passive familiarization were long-lasting, maintained for 20 hours after the initial encounter. Experiment 3 investigated how the degree of acoustic similarity between sounds related to these rapid dynamic changes in stimulus representations, using synthesized vocalizations that vary parametrically along a single dimension. Single-unit responses demonstrated that rapid differentiation of the temporal profiles of neural responses to different signals were more pronounced for stimulus pairs that are acoustically less similar to each other, although these results were mixed for multi-unit responses. Finally, in Experiment 4, the effects of passive familiarization on subsequent behavioral discrimination of two acoustically similar synthesized vocalizations were investigated. Surprisingly, this experiment did not indicate an effect of pre-exposure on behavioral responses. Taken together, the experiments in this thesis provide valuable insights into the mechanisms by which the nervous system dynamically modulates sensory representations to improve discrimination of novel complex signals over short and long time-scales. Similar mechanisms may also be engaged during processing of human speech signals, and thus may have potential translational relevance to elucidate the neural underpinnings of speech perception and comprehension difficulties.

ACKNOWLEDGMENT

First, I would like to thank my supervisor David S. Vicario for his inspiring guidance and endless support throughout the years. I also thank my committee members Lori L. Holt, Kasia M. Bieszczad, and John P. McGann for their contributions. I am thankful to the members of the Vicario Lab, especially Mimi L. Phan, Brittany A. Klawiter, and Mingwen Dong for their great support and friendship. Special thanks go to Dhruti Trivedi, who has helped me with histological analysis for this thesis. I would also like to thank my family and my friends for their unconditional love and support. Last but not least, I am truly grateful to Gokce Terzioglu, who has provided me with all the love and support that I could only dream of from a significant other during all the difficulties I have been through over the years.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION

One of the major challenges the nervous system faces is rapid recognition, classification, and discrimination of novel sensory signals. Innate organizational principles and early sensory experiences interact during ontogeny to adapt sensory brain systems to efficiently process the recurring physical stimulus parameters in the environment (White & Fitzpatrick, 2007; Sanes & Bao, 2009). However, throughout life, organisms encounter novel sensory signals, whether they are unfamiliar patterns drawn from the distribution of familiar stimulus statistics or instances of completely novel stimulus features. To cope with this problem, sensory systems retain a considerable degree of plasticity in the adult brain (Kourtzi & DiCarlo, 2006; Dahmen & King, 2007). This plasticity works at multiple time-scales, from rapid, transient dynamics to signals embedded in a context or sequence (Dahmen, Keating, Nodal, Schulz, & King, 2010; Adibi, McDonald, Clifford, & Arabzadeh, 2013) to long-lasting changes in neural representations (Weinberger, 2004). The neural mechanisms by which sensory experiences induce these changes remain largely unknown. This thesis approaches this problem by examining how passive exposure to novel sounds modifies neural

representations to facilitate recognition and discrimination, using the songbird model organism.

Complex Acoustic Signals in Songbird Vocal Communication

Songbirds (order *Passeriformes*, suborder *Passeri*) are among the few taxa known to use a set of complex learned acoustic signals in social communication, which makes them excellent model organisms for studying the neurobiological basis of auditory perception and vocal production. They learn their vocalizations from conspecific tutors during development through a process of vocal imitation with many parallels to human speech acquisition (Doupe & Kuhl, 1999). Initially, a young bird cannot produce complex vocalizations; however, during development, the bird forms the auditory memory of an adult tutor vocalization and gradually learns to imitate that signal using vocal practice and auditory feedback (for a review, see Mooney, Prather, & Roberts, 2008). When the bird reaches adulthood, its vocalizations become highly structured and similar to that of the tutor, with subtle differences that are unique to each particular bird. Some songbirds sing a fixed song for life, while others show greater adult plasticity (Beecher & Brenowitz, 2005). These vocal signals primarily serve territorial defense and/or mate attraction in adulthood (Catchpole & Slater, 1995). In addition to vocal learning and production, behavioral studies documented that songbirds use individually-specific complex vocalizations to form auditory memories and recognize familiar conspecifics such as tutors (Miller, 1979; Clayton, 1988) and mates (Vignal, Mathevon, & Mottin, 2004, 2008).

This thesis studies the zebra finch (*Taeniopygia guttata*), which has been the most widely investigated songbird species for vocal communication in the laboratory setting

due to its relative ease of breeding in captivity and stereotyped song output. Zebra finches are not territorial, but live in big colonies in the wild (Zann, 1996). Thus, male zebra finches produce learned songs primarily to attract potential mates, rather than defend territories, whereas female zebra finches do not sing. In addition, both sexes produce long calls for long-distance communication, however only male calls contain learned features, while female calls are innately determined (Zann, 1985). Male zebra finches are closed learners, that is, their learned vocalizations become crystallized when they reach adulthood and remain unchanged thereafter under normal conditions (Immelmann, 1969). The male song consists of the sequential repetition of essentially identical motifs, which can be further divided into stereotyped syllables comprised of a few notes that show complex frequency and temporal modulations (**Fig. 1A**). A single motif usually lasts about 0.5 to 1 second. The long calls, on the other hand, consist of a single syllable and usually last about 0.1 to 0.3 seconds (**Fig. 1B**). Most zebra finch vocalizations have a fundamental frequency of 400 to 2000 Hz and contain several harmonics with considerable power up to 10 kHz. Furthermore, the acoustic features of these vocalizations are modulated by social context (Sossinka & Böhner, 1980). Thus, zebra finch vocalizations provide a rich and complex repertoire that encodes several acoustic and ethological cues. Thanks to years of anatomical and functional work, we now know the detailed brain circuitry that serves the perception of these complex auditory signals.

Dynamic Responses in the Songbird Auditory Forebrain

In the songbird brain, the primary ascending auditory pathway carries auditory information from the cochlea to the midbrain nucleus mesencephalicus lateralis pars dorsalis (MLd, homolog of the inferior colliculus) via the hindbrain relay centers

cochlear nucleus, lateral lemniscus, and superior olive (Boord, 1968; **Fig. 2**). MLd then projects to the songbird homolog of the mammalian medial geniculate nucleus, ovoidalis (Ov; Karten, 1967). Auditory information is conveyed from Ov to the thalamo-recipient field L2 (Vates, Broome, Mello, & Nottebohm, 1996), which is analogous to the layer IV of the mammalian primary auditory cortex (Wang, Brzozowska-Prechtl, & Karten, 2010). One of the major targets of field L2 is the caudal medial nidopallium (NCM), which receives auditory input via field L3. NCM is believed to be analogous to the secondary auditory cortex in the mammalian brain (Calabrese & Woolley, 2015). Another high-order auditory region is the caudal mesopallium (CM), which also receives projections from field L2 via field L1 (Vates, Broome, Mello, & Nottebohm, 1996). NCM and the medial portion of CM are also reciprocally connected.

NCM has attracted much attention due to its interesting auditory response characteristics. Electrophysiological responses to pure tones in NCM reveal broader tuning functions as compared to field L2 responses (Terleph, Mello, & Vicario, 2007). Furthermore, field L2 neurons are highly responsive to simple tone stimuli, whereas neurons in NCM are selective towards more complex signals such as noise bands and amplitude modulations (Muller & Leppelsack, 1985). Song, but not pure tone, presentation elicits an increase in mRNA levels of the immediate early gene ZENK (also known as zif-268, egr-1, NGFI-A, or Krox-24) in NCM, as well as in other auditory forebrain structures, but not in field L2 (Mello, Vicario, & Clayton, 1992; Mello & Clayton, 1994). Most importantly, NCM responses are stronger for conspecific sounds than for the vocalizations of other species as indicated by both ZENK (Mello et al., 1992) and electrophysiological studies (Chew, Vicario, & Nottebohm, 1996a). We recently

showed that electrophysiological responses of different sites in NCM show low levels of coherence with each other, but high levels of mutual information for conspecific songs at rapid time-scales (Soyman & Vicario, 2017). Taken together, these studies suggest that NCM consists of functionally heterogeneous neurons and/or neural subpopulations that reliably represent different features of the individually-specific complex vocalizations of other conspecifics at fine temporal resolutions. In addition to the physical characteristics of acoustic signals, neural responses in NCM are also sensitive to social context (Vignal, Andru, & Mathevon, 2005) and acquired predictive value (Thompson & Gentner, 2010; Bell, Phan, & Vicario, 2015).

Auditory responses in NCM are also marked by a long-term form of adaptation. Electrophysiological studies documented that neural responses to initial presentations of a novel stimulus are robust, but gradually decrease with repeated presentation (Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995; Chew, Vicario, & Nottebohm, 1996a, 1996b; Stripling, Volman, & Clayton, 1997; Stripling, Kruse, & Clayton, 2001; for a review, see Dong & Clayton, 2009; **Fig. 3**). When another novel sound is then presented, the initial responses are again robust and adapt independently from the first stimulus. This is not a general habituation effect, but is stimulus-specific, because when the first stimulus is presented again after several presentations of other sounds, neural responses do not start at initial high magnitudes, but remain at adapted levels (Chew et al., 1996a). In this sense, the adaptation process forms neuronal memories that can be detected hours to days after the initial induction. Molecular studies also confirmed the process of adaptation in NCM and showed that ZENK mRNA induction first increases, then rapidly decreases in response to repeated stimulus presentation and this stimulus-specific reduction lasts at

least for 1 day (Mello, Nottebohm, & Clayton, 1995). Importantly, this adaptation phenomenon is not observed in field L2 (Terleph, Mello, & Vicario, 2006).

Although different complex acoustic signals show similar immediate adaptation profiles, for any given stimulus presentation regime, responses to conspecific vocalizations remain at adapted levels for longer durations than responses to other complex sounds (Chew et al., 1996b). The adaptation phenomenon can be so pervasive for special acoustic signals such that the responses to tutor songs in NCM remain at adapted levels even months after the last time those signals were heard (Phan, Pytte, & Vicario, 2006). The consolidation of immediate reductions in firing rates requires both RNA and new protein synthesis as injections of cycloheximide or actinomycin-D after the initial stimulus presentation reinstates the original response levels at later time points (Chew et al., 1996b). Furthermore, enabling gene transcription by inhibition of the histone deacetylase-3 activity via RGFP966 injections turns subthreshold stimulus presentations, which are not consolidated under normal conditions, into long-term memories in NCM (Phan et al., 2017). Several other molecular mechanisms such as the extracellular receptor kinase (Cheng & Clayton, 2004) and the caspase-3 (Huesmann & Clayton, 2006) activity are also implicated in the induction and consolidation of neuronal memory for familiar sounds.

In addition to encoding stimulus familiarity, the process of adaptation in NCM may also incorporate contextual information. ZENK induction in response to previously familiarized stimuli is reinstated to original levels when presented with flashing colored lights (Kruse, Stripling, & Clayton, 2004). Furthermore, presenting those same stimuli at a reduced sound level or from a speaker at a different location in the same environment

can also reset ZENK adaptation. Chronic electrophysiological recordings from NCM corroborated these findings, indicating that the process of adaptation is sensitive to the location of the source of the sound in addition to its identity (Smulders & Jarvis, 2013). The acoustic context in which a stimulus is embedded also affects the profile of immediate adaptation. Electrophysiological responses in NCM to a target stimulus follows the predicted adaptation profile in the context of other conspecific signals, whereas, in the context of heterospecific sounds, the responses are enhanced, although adaptation continues latently and is revealed when the contrasting context is removed (Lu & Vicario, 2017). However, in a recent study, we showed that the magnitude of neural responses in NCM is not sensitive to the sequence in which the stimuli are presented (Soyman & Vicario, 2017).

Previous studies have successfully documented the dynamics governing the process of adaptation in NCM. What remains still unknown is how adaptation affects stimulus encoding and decoding. The main reason for this lack of knowledge is because the majority of studies that investigated adaptation in NCM based their analyses exclusively on spike-count rates or similar measures. With these measures, neural discrimination can only be analyzed between novel and familiar, but not between two novel or two familiar stimuli. However, going beyond the total amount of activity and looking at the temporal profiles of neural responses reveals interesting observations. The process of adaptation does not affect neural responses uniformly, but has a temporal structure (**Fig. 4**). Although there is strong adaptation at certain time points along the stimulus duration, there is no change in neural response magnitudes at other time points. More importantly, these patterns differ between different stimuli. Thus, adaptation in

NCM may represent, or reflect, a mechanism by which the nervous system increases the contrasts between different signals for rapid and efficient discrimination. This working hypothesis represents the core of this thesis and is investigated in detail in the experiments presented in the following chapters. However, to provide a foundation for these studies, I would like to review the current state of knowledge of the relationship between adaptation and neural discrimination in other species.

Different Forms of Adaptation in the Mammalian Brain

Stimulus-specific adaptation (SSA) and deviance detection have been extensively studied in the mammalian auditory system using the classical oddball paradigm, in which a deviant stimulus is presented with a low probability in a series that also includes a standard stimulus with a high probability of occurrence. Neural responses to the same stimulus are weaker when that stimulus is standard compared to when it is deviant, suggesting that extensive repeated presentation leads to SSA (Ulanovsky, Las, & Nelken, 2003). This phenomenon is observed in the inferior colliculus (Malmierca, Cristaudo, Perez-Gonzalez, & Covey, 2009) and medial geniculate nucleus of the thalamus (Anderson, Christianson, & Linden, 2009), as well as in A1 (Ulanovsky et al., 2003), although functional anatomy studies suggest that SSA occurs independently in subcortical and cortical areas (Nelken, 2014; Pérez-González & Malmierca, 2014). That is, the lemniscal pathway that provides the main auditory input to A1 shows little or no SSA (Antunes, Nelken, Covey, & Malmierca, 2010; Duque, Perez-Gonzalez, Ayala, Palmer, & Malmierca, 2012), while the strong SSA observed in the non-lemniscal regions of the inferior colliculus and the medial geniculate nucleus is not abolished by

deactivation of A1 (Antunes & Malmierca, 2011; Anderson & Malmierca, 2013; for a review, see Malmierca, Anderson, & Antunes, 2015).

Although they both can be described as repetition suppression, the forms of adaptation typically studied in the mammalian and the songbird auditory system differ from each other in certain aspects. SSA in mammals has been studied almost exclusively using artificial pure tone stimuli that differ only in the frequency domain, although it is also observed in processing of spectro-temporally complex sounds (Nelken, Yaron, Polterovich, & Hershenhoren, 2013). Adaptation in the songbird brain, on the other hand, is always studied using ecologically relevant, complex conspecific vocalizations. In fact, adaptation to pure tone stimuli in NCM is markedly low as compared to complex signals (Chew et al., 1996a). Furthermore, SSA in the mammalian brain is typically studied with rapid stimulations with interstimulus intervals (ISI) in the order of hundreds of milliseconds. In the songbird brain, much longer ISIs, such as 6 or 11 seconds, are used and it is shown that longer ISIs lead to more long-lasting adaptation (Chew et al., 1996b). Most importantly, the mammalian SSA is transient and only reflects stimulus statistics in a relatively local stimulation regime, whereas adaptation in the NCM of the songbird brain can last several days (Chew et al., 1996b). There are also anatomical differences between the types of adaptation observed in the mammalian and the songbird brain. Although SSA in the mammalian A1 is calculated *de novo* from non-adapting auditory inputs, the same SSA as a phenomenon is observed at the level of the auditory midbrain and thalamus (Anderson et al., 2009; Malmierca et al., 2009). Nevertheless, the form of adaptation described in the NCM of the songbird brain is not seen in field L2, which provides the main source of auditory input to the forebrain (Terleph et al., 2006).

Adaptation has not yet been tested in the auditory midbrain or thalamus of the songbird brain. Taken together, whether the listed differences between the forms of adaptation seen in the mammalian and the songbird brain point to completely separate processes or reflect only quantitative differences stemming from methodological dissimilarities cannot be dissected due to a lack of comparative studies. Regardless, similar to the previous studies on adaptation in the songbird brain, neural discrimination has only been assessed between a common and a rare stimulus using the experimental and analytical methods described in studies of the mammalian SSA. This is significantly different from the neural discrimination investigated in this thesis, which compares the temporal profiles of neural responses to stimuli with equal probabilities of occurrence as they become familiar.

Other studies have investigated the effects of neural adaptation to recurrent stimulus statistics on receptive fields or after-effects in various regions of the mammalian brain. In the guinea pig inferior colliculus, neural responses adapt to the mean, variance, and bimodality of sound intensity level distributions to improve the encoding of sounds with most probable intensities (Dean, Harper, & McAlpine, 2005). Similarly, adaptation to the variance or the higher-order statistics such as the kurtosis of dynamic sounds that vary in their modulation depth leads to changes in the receptive fields of the cat inferior colliculus neurons (Kvale & Schreiner, 2003). Furthermore, adaptation to incoming stimulus statics by rapidly presenting sounds from different locations, but with a fixed mean or standard deviation angle along the azimuth, enhances the contrast between those recurrent signals and upcoming target stimuli in the ferret inferior colliculus (Dahmen, Keating, Nodal, Schulz, & King, 2010). Parallel effects have been documented in other

sensory modalities such as for whisker stimulation frequency in the rat barrel cortex (Adibi, McDonald, Clifford, & Arabzadeh, 2013) and for orientation (Ghisovan, Nemri, Shumikhina, & Molotchnikoff, 2009) and spatial frequency (Bouchard, Gillet, Shumikhina, & Molotchnikoff, 2008) of visual signals in the cat primary visual cortex. This form of adaptation may serve to adjust the limited dynamic range of neurons to the relevant stimulus statistics in the environment and/or to enhance the detection of changes in the incoming stimulus stream. However, the adaptation and neural discrimination studied in the present thesis deviate significantly from this form of adaptation since it represents long-lasting changes that serve encoding of individual complex auditory objects, rather than, or in addition to, transient changes to adjust neural responses to incoming stimulus statistics.

One concept that is highly relevant to this thesis is sparse coding in sensory systems, which refers to a coding scheme that relies on a small fraction of neurons (population sparseness) and/or a small amount of neural firing (life-time sparseness) to represent any given stimulus (Willmore & Tolhurst, 2001). Sparse coding provides several advantages such as expenditure of less metabolic energy and increased storage capacity for distinct representations. However, there has been no report up to date as to whether sparsening of representations within a structure as multiple novel sensory objects are passively familiarized also serves to improve neural discriminations among those signals, which is the central question addressed in this thesis. Another line of research that is relevant to the approach described here is the established literature on the effects of passive exposure on behavioral discrimination of sensory signals in humans.

Short-term Adaptive Plasticity in the Human Auditory System

The classical oddball paradigm described above has also been widely used to study SSA and deviance detection in human auditory processing. The most well-known event-related potential marker of this phenomenon is the mismatch negativity (MMN), which represents an activity difference between stimuli with high- and low-probability of occurrence dominantly in frontal-temporal cortical regions 100-250 ms after stimulus onset (Näätänen, Gaillard, & Mäntysalo, 1978). MMN occurs for both simple acoustic feature deviations and violations of complex pattern regularities (for a review, see Garrido, Kilner, Stephan, & Friston, 2009). However, the latency of MMN is much longer than deviance detection responses in the mammalian auditory system reported in animal studies using similar paradigms (Grimm, Escera, & Nelken, 2016). Studies looking for early markers of SSA in event-related potentials discovered that the mid-latency response, which peaks around 10-50 ms after stimulus onset, shows the oddball effect for acoustic feature repetitions, but not for more complex pattern regularities (Cornella, Leung, Grimm, & Escera, 2012; Althen, Grimm, & Escera, 2013). Thus, these findings suggest that there might be a hierarchical organization where acoustic feature repetitions induce short-term adaptation in early sensory regions, whereas more abstract pattern regularities lead to prediction-based suppression of activity in temporal-frontal networks to detect complex rule violations. Nevertheless, in terms of the aims of this thesis, all of these studies suffer from the same limitation as the studies of SSA in other animal studies: while neural discrimination of common and rare stimuli or more complex

regularities can be studied, it is not possible to investigate how neural discrimination of multiple novel signals are affected while they become equally familiar.

There is compelling evidence that passive auditory exposure to distorted speech signals improves behavioral measures of recognition and comprehension (for a review, see Guediche, Blumstein, Fiez, & Holt, 2014). For instance, passive familiarization with foreign-accented speech for few minutes greatly improves comprehension of subsequent vocalizations with similar acoustic characteristics (Clarke & Garrett, 2004; Bradlow & Bent, 2008). Similar gains in speech recognition are also observed for other distorted signals such as time-compressed (Pallier, Sebastian-Gallés, Dupoux, Christophe, & Mehler, 1998), dysarthric (Liss, Spitzer, Caviness, & Adler, 2002), and noise-vocoded, spectrally shifted speech (Guediche, Fiez, & Holt, 2016). In addition, presenting listeners with an artificial accent, which reverses the correlation between two acoustic dimensions normally found in English, induces short-term changes in the weight placed on specific dimensions in judging ambiguous speech signals (Idemaru & Holt, 2011; Liu & Holt, 2015). Taken together, these findings are in line with the idea that dynamic sensory representations are updated accordingly during passive exposure to improve the mapping between the incoming sensory signals and learned linguistic categories, such as phonemes.

The small number of studies investigating the neural basis of these adaptive changes reveal a rather complicated picture (for a review, see Guediche, Blumstein, Fiez, & Holt, 2014). Rapid adaptation to time-compressed speech for sentence verification induces greater activation in auditory association cortices and the left ventral premotor cortex (Adank & Devlin, 2010). Another fMRI study using degraded speech signals

indicated that individual differences in adaptation is associated with inferior frontal gyrus activity, whereas gains in comprehension are related to activation in inferior parietal regions (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010). In addition, adaptation to vocoded speech signals for purposes of vocal repetition is associated with upregulation of activity in posterior cingulate and left ventral premotor cortex and downregulation of activity in the anteroventral thalamic nucleus, caudate nucleus, frontal, occipital, and cerebellar regions (Erb, Henry, Eisner, & Obleser, 2013). The crus I region of the cerebellum is also related to adaptive plasticity to distorted signals in a word recognition task (Guediche, Holt, Laurent, Lim, & Fiez, 2015).

In sum, several auditory, motor, and sensorimotor brain regions seem involved in underlying the improvements in comprehension of distorted speech signals following passive familiarization. However, these studies do not provide information as to how neural representations of these novel signals change as they become familiar. This problem is due to the fact that the noninvasive brain imaging technologies that are readily available in humans lack spatial and/or temporal granularity sufficient to address this problem. Methodological tools available for animal studies, on the other hand, provide a window into the millisecond range dynamics of individual and populations of neurons that is required for investigating adaptive sensory representations in real time. Thus, this thesis uses electrophysiological techniques in a brain area specialized for processing communication signals in an established animal model.

Overview of This Thesis

This thesis investigates the effects of passive familiarization on neural and behavioral discrimination of acoustic signals using the zebra finch as a model system.

Specifically, Experiment 1 examines how changes in the representation of natural vocalizations with passive familiarization affect immediate neural discrimination. This experiment also tests whether these rapid changes reflect exposure not only to the target stimulus itself, but also to other signals embedded in a sequence, in such a way as to improve the neural contrast between signals. Experiment 2 investigates whether immediate changes in neural representations caused by passive exposure also produce long-term changes that alter processing of sounds at later time-points. Experiment 3 assesses the relationship between improvements in neural discrimination with passive familiarization and the degree of acoustic similarity between stimuli, using synthesized vocalizations that vary along a single dimension. Finally, Experiment 4 tests the contribution of passive exposure to subsequent behavioral discrimination of signals that are acoustically highly similar to each other. A general discussion of all empirical findings is presented at the end. The findings of these studies shed light on the mechanisms by which experience dynamically modulates sensory representations to improve recognition and discrimination of external signals over multiple different time-scales.

EXPERIMENT 1: NEURAL DISCRIMINATION OF NOVEL VOCAL SIGNALS AS THEY BECOME FAMILIAR

The primary purpose of this experiment was to test whether passive familiarization improves neural discrimination of novel acoustic signals. The process of adaptation in NCM represents a reduction in total firing rates in response to the repeated presentation of an auditory signal (Chew et al., 1995, 1996a, 1996b). However, a careful examination of neural response profiles reveals that, not only the total firing rates, but also the temporal patterns of neural responses undergo considerable changes (**Fig. 4**). Most importantly, these temporal profile changes appear to accentuate response differences between stimuli. These observations suggest that adaptation in NCM may represent a mechanism to increase the contrast between neural responses to different signals. Thus, it is hypothesized that stimulus discrimination and decoding improve with ongoing adaptation as novel signals become familiar, as reflected in the dynamic temporal profiles of neural responses.

A second major goal of this experiment was to assess whether the rapid changes in temporal profiles of neural responses with stimulus repetition reflect exposure not only to the target stimulus itself, but also to other signals embedded in a sequence, in such a way as to improve the neural contrast between them. In addition to stimulus familiarity, neural responses in NCM are known to reflect various kinds of relationships among signals in a sequence such as transition probabilities (Lu & Vicario, 2014) and acoustic category contrasts (Lu & Vicario, 2017). To probe the effects of stimulus context on neural discrimination, this experiment used blocked and shuffled presentation sequences. The shuffled sequence provides opportunities for pairwise stimulus contrasts as neural responses undergo adaptation, whereas, in the blocked sequence, stimuli undergo adaptation one by one. Hence, it is hypothesized that neural discrimination and decoding of acoustic signals are greater in a shuffled than in a blocked stimulus presentation sequence.

To test these hypotheses, electrophysiological responses to novel zebra finch songs and calls in NCM were recorded under passive listening conditions. One group of birds was presented with a sequence where each stimulus was repeated in a block, whereas the other group heard a sequence containing all the stimuli in shuffled order. Both multi-unit and single-unit responses were quantified by applying several neural discrimination metrics that represented the temporal profiles of neural activity, but that were explicitly designed to be insensitive to the absolute firing rates. Previous studies showed that neurons in the songbird auditory forebrain can be classified based on spike waveforms into two clusters, the narrow and wide spike neurons, which differ in their response properties (Meliza & Margoliash, 2012; Jeanne, Sharpee, & Gentner, 2013).

Thus, these two neuron types were assessed separately to probe their respective roles in neural discrimination.

Methods

Subjects

Sixteen naïve adult (>120 days) male zebra finches purchased from a commercial supplier (Magnolia Bird Farm, Anaheim, CA) were used in this experiment. All birds lived in same-sex cages in a general aviary (LD 12:12, 21-25 C^{o}) with *ad libitum* food and water throughout the experiment. All procedures were approved by Rutgers University Institutional Animal Care and Use Committee (Protocol Number 02-217).

Stimuli

Two different stimulus sets from the natural vocal repertoire of the zebra finch were used in this experiment. One set consisted of 8 male songs and the other set consisted of 8 male long calls. All experimental stimuli were selected from a corpus of recorded vocalizations that the experimental birds had never heard before. Within each stimulus set, the acoustic similarities between all pairs of stimuli were calculated using Sound Analysis Pro software as described in Tchernichovski, Nottebohm, Ho, Pesaran, and Mitra (2000). Briefly, four different acoustic features - pitch, frequency modulation, spectral continuity, and Wiener entropy - were calculated for each stimulus pair, a probability-based goodness of the match measure was estimated for each of these features between the two sounds, and these estimations were finally integrated into a global percent similarity score. The experimental stimuli were selected from the corpus such that all pairwise acoustic similarity scores within a stimulus set captured the natural range for that stimulus type with no outliers. This resulted in pairwise percent similarity scores

between 39% and 63% (Mean ± SEM = 51 ± 1%) for the song set and between 23% and 75% (Mean ± SEM = 53 ± 3%) for the call set. The durations of the experimental stimuli ranged from 658 to 825 ms (Mean ± SEM = 745 ± 19 ms) and from 148 to 211 ms (Mean ± SEM = 176 ± 7 ms) for songs and calls, respectively. All stimuli in each set were equated for root-mean-square-amplitude. All birds were tested with the same two stimulus sets.

Surgery

One to 2 days before electrophysiological recordings, birds underwent a surgery in which they were anesthetized with isoflurane (2-3% in oxygen; Henry Schein Animal Health, Dublin, OH) and placed in a stereotaxic apparatus. A craniotomy was performed over the region of interest and a metal pin was attached anterior to this opening with dental cement (Dentsply Caulk, Milford, DE) to be used to fix the bird's head during subsequent electrophysiological recordings. All birds were injected with meloxicam (0.01 ml of 5 mg/ml; Boehringer Ingelheim, Ingelheim am Rhein, Germany) at the end of the surgery and recovered within an hour.

Electrophysiology

Electrophysiological recordings in awake, restrained birds were conducted in a walk-in sound attenuation chamber (Industrial Acoustics Company, Bronx, NY). Two silicon probes (NeuroNexus, Ann Arbor, MI), one for each hemisphere, were used for recordings. Each probe had 16 recording sites (0.4-1 MΩ impedance at 1 kHz) in a 4-by-4 grid layout. The probes were implanted in a para-sagittal plane such that the 4-by-4 grid layout extended in anterior-posterior and dorsal-ventral axes. Each probe was used for the right hemisphere for half of the birds and for the left hemisphere for the other half. Prior

to insertion, the probes were dipped into a DiI solution (10% in ethanol; Sigma Aldrich, St. Louis, MO) and allowed to dry to label probe insertion tracks for later histological analyses. The dura was opened and the probes were placed on the surface of the brain, one for each hemisphere, above NCM according to stereotaxic coordinates. Then, the probes were lowered by means of hydraulic microdrives, while playing zebra finch songs that were different from those to be used as experimental stimuli. When firing patterns characteristic of NCM neurons were observed at the majority of the recording sites, the experiment started. Multi-unit neural recordings were high- and low-pass filtered (0.3 and 5 kHz), amplified (10,000x), digitized (25 kHz), and saved to disk using Spike2 software (Cambridge Electronic Design, Cambridge, UK).

Stimulus Presentation

Birds were divided into two groups based on the type of stimulus presentation sequence they would be administered during electrophysiological recordings. For 8 birds, the experimental stimuli were presented in a blocked sequence, while a shuffled sequence was used for the other 8 birds. Each stimulus was played 25 times at an onset-onset ISI of 6 s from a speaker located 30 cm in front of the bird at an amplitude of 55 dB SPL (A scale) and a sampling frequency of 44.444 kHz. The song and call stimulus sets were always presented separately such that half of the birds in each stimulus presentation sequence group were presented with songs first and the other half were presented with calls first. For a given bird, the two stimulus sets were both presented either in the blocked or in the shuffled sequence. In the blocked sequence, the order of stimulus blocks was pseudorandomly counterbalanced across birds such that each stimulus within a set occurred in each position once.

Histology

After the electrophysiological recordings, birds were deeply anesthetized with an overdose of pentobarbital (0.15 ml of 39 mg/ml; Vortech Pharmaceutical, Dearborn, MI), transcardially perfused with saline (0.9%, 40 ml) and paraformaldehyde (4%, 40 ml), and decapitated. The brains were extracted and post-fixed with paraformaldehyde for at least 4 days, after which 50-µm sagittal sections were cut on a vibratome. Unstained sections were visualized under a fluorescence microscope and grayscale digital images of the same sections were collected under 450-490/515-cut-on and 510-560/590-cut-on nm excitation/emission filters for anatomical markers and DiI, respectively. Two images from the same sections were superimposed to create composite images, and scaled drawings of the silicon probes were used to validate the recording sites that fell within the boundaries of NCM (**Fig. 5**). Only recording sites that were at least 200 µm posterior of field L, which can be clearly identified by its cytoarchitecture, were included in data analyses.

Data Analysis

Raw neural recordings were visually assessed by a human operator and trials with movement artifacts were excluded. Then, multi-unit spiking activity at each recording site was thresholded at 2 standard deviations from the mean amplitude (calculated from the whole recording) and the peaks of positive threshold-crossings were marked with time-stamps, each representing a spike, with a time window of 1.24 ms. In addition to these multi-unit spike trains, single-unit spike trains were also extracted by spike-sorting

the raw neural recordings via an unsupervised technique as described in Quiroga, Nadasdy, and Ben-Shaul (2004). From the resulting single-unit clusters, only units with < 2% of their spikes within a 2 ms refractory period and with more than 2000 spikes throughout the entire recording were included in the final data set. All further analyses were carried out using custom scripts in MATLAB (The Mathworks, Natick, MA) and Statistica (Statsoft, Tulsa, OK) software.

Response magnitude. To quantify the magnitude of stimulus-driven neural responses, the firing rate during each baseline period (500 ms window preceding each stimulus presentation, FR_{base}) and each stimulus period (stimulus duration period plus 100 ms, FR_{stim}) was calculated as spikes/second. The response magnitude for each presentation of each stimulus was calculated as

$$Respone\ Magnitude = FR_{stim} - \overline{FR_{base}}$$

where $\overline{FR_{base}}$ is the average of the baseline firing rates across all repetitions of that particular stimulus. To control for between-stimulus and between-unit variability in the analysis of adaptation profiles, the response magnitudes to all presentations of a given stimulus were calculated as a percent of the response magnitude to the first presentation of that particular stimulus.

Adaptation rate. To quantify the rate of adaptation of neural responses, stimulus presentations 1 to 6 and 6 to 25 were analyzed separately as in previous studies (Phan, Pytte, & Vicario, 2006; Bell, Phan, & Vicario, 2015). For each of these two sets of presentations, a linear regression analysis between stimulus presentation numbers and response magnitudes was conducted for each stimulus separately. The adaptation rate for each stimulus was calculated as

$$Adaptation\ Rate = 100\ (b\ /\ \overline{RM}\ )$$

where b is the slope of the linear regression and $\overline{RM}$ is the average of the response magnitudes on the set of stimulus presentations that was used for that analysis. This adaptation rate metric provides a normalized measure of adaptation, enabling comparisons across stimuli and units with varying average response magnitudes.

Neural dissimilarity. To calculate the dissimilarities between the temporal profiles of different neural responses, the spike counts during the stimulus-evoked response period were first grouped into 10-ms bins since peak mutual information estimations in NCM are seen at 5 to 10-ms temporal resolutions (Soyman & Vicario, 2017). The duration of the response period for each stimulus set was equal to the minimum stimulus duration for that particular set plus 100 ms, which gave response durations of 750 and 240 ms for songs and calls, respectively. To develop a dissimilarity metric that is only sensitive to the temporal profiles, but not to the total firing rates, neural responses were standardized via taking the z-score of each bin by normalizing it with the average and the standard deviation across all bins within the same trial. Then, neural dissimilarity was quantified by calculating the Euclidean distance between these z-scored response profiles of the same unit to different pairs of stimulus presentations as

$$Neural\ Dissimilarity = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2}$$

where A and B are the binned response profiles in the two stimulus presentations and n is the number of bins. Similar Euclidean distance-based metrics have been widely used as measures of spike train dissimilarity (van Rossum, 2001). Within each stimulus set, the pairwise dissimilarities between each presentation of a particular stimulus and all

presentations of all of the other stimuli were averaged to calculate between-stimulus neural dissimilarities. Similarly, the pairwise dissimilarities between each presentation of a particular stimulus and all of the other presentations of the same stimulus were averaged to calculate within-stimulus neural dissimilarities.

Neural decoding accuracy. Neural dissimilarity calculations described above were used to decode stimulus identities from the temporal profiles of neural responses. The dissimilarities of a particular response to the responses on all presentations of each stimulus were averaged, which produced 8 average neural dissimilarities, one for each of the 8 stimuli. The response was assigned to the stimulus with the minimal average neural dissimilarity. To assess the decoding performance from stimulus onset to any given point along the stimulus duration, this decoding procedure was conducted by progressively increasing the number of bins that went into the calculation starting from the stimulus onset. That is, for the first bin, the decoding was based solely on the neural responses in the first bin; for the second bin, the neural responses in the first and second bins were used for decoding; responses in bins 1 through 3 were used for decoding in the third bin, and so on. For decoding at each bin, the probability of correct decoding was calculated by counting how many of the 8 stimuli were correctly classified for a given stimulus presentation. The chance level for correct decoding probability was $1/8 = 0.125$. In addition to the correct decoding probabilities at selected bins, the latencies to reach specific probability levels along the stimulus duration were analyzed in detail.

Mutual information. The decoding procedure described above was further used to calculate the mutual information between stimulus identities and temporal profiles of neural responses. For decoding at each bin, the true and the neurally decoded stimulus

identities were used to construct a confusion matrix, from which mutual information was calculated using Shannon's formula as

$$Mutual\ Information = \sum_{s,r} p(s,r) log_2 \left(\frac{p(s,r)}{p(s)p(r)} \right)$$

where s is the true and r is the neurally decoded stimulus identity. The multiplication *0*log2(0)* was equated to 0 and the prior probability $p(r)$ for each stimulus was taken as 1/8 = 0.125, since all 8 stimuli were presented an equal number of times. The maximum possible mutual information was $\log_2 8 = 3$ bits. To correct for the bias in mutual information estimations, for every unit, the calculation above was repeated 5 times while randomly shuffling the stimulus-response relationships across trials for each iteration. The mean across these 5 mutual information calculations was taken as the bias and subtracted from the real estimates so that bias-corrected mutual information estimates were used for all statistical analyses.

Slopes of the neural discrimination metrics. The directions and rates of changes in neural discrimination metrics with repeated stimulus presentation were quantified using linear regression and normalization methods as described for adaptation rates. These calculations were conducted for between-stimulus and within-stimulus neural dissimilarities, as well as correct decoding probabilities and latencies, separately. Trends for presentations 1 to 6 and 6 to 25 were analyzed separately in conjunction with the analysis of adaptation rates.

Spike waveform clustering. Following previous studies, single-units were divided into two clusters based on their spike waveforms. To do this, first, the average waveform of each single-unit was normalized by its peak amplitude. Then, the waveforms of single-units from all experiments in this thesis were aligned by their peaks

and processed via a principal components analysis. The first two components were used in an affinity propagation clustering algorithm (Frey & Dueck, 2007) to classify waveforms into two clusters. Similar to previous reports (Meliza & Margoliash, 2012; Jeanne, Sharpee, & Gentner, 2013), this method nicely separated single-units into narrow and wide spike neurons (**Fig. 6**).

Statistical analyses. Parametric statistical tests were preferred for the analysis of multi-unit response measures, because they did not deviate significantly from normality. Factorial models with interactions were assessed via ANOVAs with designs specified for each particular analysis below. Post-hoc comparisons were conducted via Tukey's HSD tests. For testing the difference of a sample from a single value, a one-sample t-test was used and, for testing the strength of a pairwise relationship, Pearson's correlation was calculated. Single-unit response measures, on the other hand, deviated severely from normality, thus nonparametric statistical tests were used for single-unit analysis. Due to a lack of nonparametric ANOVA for assessing factorial designs, Mann-Whitney U and Wilcoxon signed-rank tests were used for between-subjects and within-subjects comparisons, respectively, in combination to analyze main effects and all pairwise comparisons for interactions. When this was done, the significance level was Bonferroni-corrected for the number of comparisons conducted on the same data set. For the single-unit response measures, a one-sample Wilcoxon signed-rank test was used for testing the difference of a sample from a single value and Spearman's correlation was used for testing the strength of a pairwise relationship.

Results

Multi-unit Responses

For song playbacks, 184 multi-unit sites in the blocked sequence and 152 multi-unit sites in the shuffled sequence were histologically verified to be in NCM. For call playbacks, the numbers of multi-unit recording sites in NCM were 186 and 157 in blocked and shuffled sequences, respectively. Preliminary analyses did not show systematic hemispheric differences in any of the basic response properties or the neural discrimination metrics in either the blocked or the shuffled sequence. Thus, the two hemispheres were combined for all subsequent analyses.

Multi-unit Adaptation Rates

Songs. First, to compare the adaptation profiles of song responses between the two sequences, a mixed ANOVA with sequence (Blocked, Shuffled) as a between-subjects variable and presentation (2 through 25) as a within-subjects variable was conducted on percent response magnitudes. Overall, there was no significant difference between the two sequences ($F(1,334) = 0.050$, $p = 0.821$). However, the effect of presentation was highly significant ($F(23,7682) = 571.65$, $p < 0.001$, **Fig. 7A**). Post-hoc comparisons revealed that there was a significant decrease in percent response magnitudes from the 1st until the 20th stimulus presentation, after which there was no consistent change. There was also a significant interaction between sequence and presentation ($F(23,7682) = 12.68$, $p < 0.001$, **Fig. 7A**). However, post-hoc analyses did not reveal any significant difference between the two sequences in corresponding presentations and indicated that the interaction was driven by differences in the rates of adaptation in the two sequences.

Following previous studies, these adaptation rates were analyzed separately for stimulus presentations 1 through 6 and 6 through 25 (hereafter referred to as 1-6 and 6-

25, respectively). A mixed ANOVA using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable revealed that, overall, the adaptation rates in the blocked sequence were significantly more negative than those in the shuffled sequence ($F(1,334) = 5.82$, $p = 0.016$, **Fig. 7B**). There was also a significant effect of presentation ($F(1,334) = 428.09$, $p < 0.001$, **Fig. 7B**), indicating more negative adaptation rates for presentations 1-6 than for presentations 6-25. Most importantly, there was a significant interaction between sequence and presentation ($F(1,334) = 32.22$, $p < 0.001$, **Fig. 7B**). Post-hoc comparisons showed that, for presentations 1-6, adaptation rates were significantly more negative in the blocked than in the shuffled sequence ($p < 0.001$), whereas there was no such difference for presentations 6-25 ($p = 0.595$). In both sequences, adaptation rates were more negative for presentations 1-6 than for presentations 6-25 (both $p < 0.001$). Adaptation rates for presentations 1-6 in both sequences separately, as well as for presentations 6-25 together, were significantly less than 0 (all $p < 0.001$).

Summary: These findings indicate that the responses to songs adapted more steeply in the blocked than in the shuffled sequence during the first 6 presentations and then continued to adapt at similar rates during subsequent stimulus repetitions.

Calls. The adaptation profiles and rates of responses to calls were also compared between the two sequences using similar ANOVAs. The analysis of percent response magnitudes revealed similar results, indicating no overall effect of sequence ($F(1,341) < 0.01$, $p = 0.985$). The effect of presentation was significant ($F(23,7843) = 317.71$, $p < 0.001$, **Fig. 7C**), such that there was a significant decrease in percent response magnitudes from the 1st until the 19th presentation, after which there was no consistent

change. There was also a significant interaction between sequence and presentation ($F(23,7843) = 12.38$, $p < 0.001$, **Fig. 7C**). However, once again, post-hoc comparisons indicated no significant difference between the two sequences in corresponding presentations and that the interaction was due to differences in the adaptation profiles of the two sequences.

The analysis of adaptation rates did not show an effect of sequence ($F(1,341) = 0.79$, $p = 0.374$). Nevertheless, there was a significant effect of presentation ($F(1,341) = 39.21$, $p < 0.001$, **Fig. 7D**), such that the adaptation rates were in general more negative in the blocked than in the shuffled sequence. Similar to the findings in song responses, there was also an interaction between sequence and presentation ($F(1,341) = 34.00$, $p < 0.001$, **Fig. 7D**). Post-hoc tests demonstrated that the adaptation rates for presentations 1-6 were significantly more negative in the blocked than in the shuffled sequence ($p < 0.001$), while there was no significant difference for presentations 6-25 ($p = 0.059$). Looking at the interaction from the other perspective, the adaptation rates were significantly more negative for presentations 1-6 than for presentations 6-25 only in the blocked sequence ($p < 0.001$), whereas no such difference was observed in the shuffled sequence ($p = 0.991$). Adaptation rates for presentations 1-6 in both sequences separately, as well as for presentations 6-25 together, were significantly less than 0 (all $p < 0.001$).

Summary: These analyses showed patterns almost identical to the findings in song recordings, indicating that the responses to calls underwent stronger adaptation in the blocked than in the shuffled sequence during the first 6 presentations and then continued to adapt at similar rates during consequent stimulus repetitions.

Validation of the Neural Discrimination Metrics

Before the analysis of actual neural recordings, the neural discrimination metrics developed in this study were first tested to validate that they were only sensitive to the temporal profile, but not to the total magnitude, of neural responses. To do this, a randomized data set was simulated using the song responses in the shuffled sequence. This was done by randomizing the timings of spikes, while keeping the total number of spikes for each stimulus presentation constant, so that any change in the neural discrimination metrics as a function of stimulus presentation would be solely due to the changes in the total number of spike counts across repetitions. As predicted, these analyses did not reveal any consistent trend in the neural discrimination metrics with repeated stimulus presentation (**Fig. 8A-C**). That is, none of the slopes of between- and within-stimulus dissimilarities, as well those of the correct decoding probabilities and latencies, for either presentations 1-6 or 6-25 were significantly different from 0 (all $p > 0.130$). In addition, the bias-corrected mutual information estimations in this randomized data set were not significantly different from 0 ($t(151) = 0.75$, $p = 0.457$, **Fig. 8D**), suggesting that when the temporal profiles of neural responses were disrupted, the neural decoding method used throughout this thesis did not yield accurate results. Taken together, all of these findings strongly validate that the neural discrimination metrics used in this study were exclusively sensitive to the temporal profiles of neural responses and were not affected by variations in total firing rates.

Multi-unit Between-stimulus Neural Dissimilarities

Songs. First, the dissimilarities between the temporal profiles of neural responses to different songs were analyzed via a mixed ANOVA on between-stimulus neural dissimilarities using sequence (Blocked, Shuffled) as a between-subjects variable and

presentation (1 through 25) as a within-subjects variable. Overall, there was no significant difference between the two sequences ($F(1,334) = 1.50$, $p = 0.223$). However, the effect of presentation was highly significant ($F(24,8016) = 102.12$, $p < 0.001$, **Fig. 9A**). Post-hoc analyses revealed that the between-stimulus neural dissimilarities significantly increased from the 1st until the 15th presentation, after which no consistent change was observed. The interaction between sequence and presentation was also significant ($F(24,8016) = 6.08$, $p < 0.001$, **Fig. 9A**). However, post-hoc comparisons did not show a significant difference between the two sequences in corresponding presentations and indicated that the interaction was driven by differences in the rates between-stimulus neural dissimilarities changed with stimulus presentation in the two sequences.

These differences were further analyzed in detail via a mixed ANOVA on the slopes of between-stimulus neural dissimilarities using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable. There was neither an effect of sequence ($F1,334) = 2.19$, $p = 0.140$) nor an interaction between sequence and presentation ($F(1,334) = 0.194$, $p = 0.660$). However, the effect of presentation was significant ($F(1,334) = 109.44$, $p < 0.001$, **Fig. 9B**), indicating greater slopes for between-stimulus neural dissimilarities for presentations 1-6 than for presentations 6-25. These slopes were significantly greater than 0 for both presentations 1-6 and 6-25 (both $p < 0.001$).

Summary: Taken together, these analyses revealed that the dissimilarities between the temporal profiles of neural responses to different songs increased markedly during the first 6 presentations and then continued to increase more gradually during the following

repetitions. There was no difference between blocked and shuffled sequences either in the levels of these neural dissimilarities or in the rates at which they changed with stimulus repetition.

Calls. Between-stimulus neural dissimilarities, as wells as their linear trends as a function of stimulus repetition, for call responses were also analyzed via similar ANOVAs. Overall, between-stimulus neural dissimilarities were significantly greater in the shuffled than in blocked sequence ($F(1,341) = 11.48$, $p < 0.001$, **Fig. 9C**). The effect of presentation was also highly significant ($F(24,8184) = 55.46$, $p < 0.001$, **Fig. 9C**). Post-hoc tests revealed that between-stimulus neural dissimilarities increased with stimulus presentation. Although this increase was less pronounced after the 11^{th} presentation, no clear asymptotic level was reached with 25 stimulus repetitions. There was also a significant interaction between sequence and presentation ($F(24,8184) = 8.34$, $p < 0.001$, **Fig. 9C**). Nevertheless, once again, post-hoc comparisons did not show a significant difference between the two sequences in corresponding presentations and indicated that the interaction was due to differences in the way between-stimulus neural dissimilarities changed with presentation in the two sequences.

The analysis of these differences in detail showed that the slopes of between-stimulus neural dissimilarities were significantly greater in the shuffled than in the blocked sequence ($F(1,341) = 4.75$, $p = 0.030$, **Fig. 9D**). There was also a significant effect of presentation ($F(1,341) = 6.97$, $p = 0.009$, **Fig. 9D**), such that the between-stimulus neural dissimilarity slopes were significantly greater for presentations 1-6 than for 6-25. There was no significant interaction between sequence and presentation ($F(1,341) = 0.10$, $p = 0.748$). Finally, the slopes of between-stimulus neural

dissimilarities for both sequences and presentations were significantly greater than 0 (all $p < 0.001$).

Summary: These findings were mostly in line with what was observed in song responses and showed that the temporal profiles of neural responses to different calls become more and more dissimilar from each other with repeated stimulus presentation. These changes were more robust during the first 6 presentations than the more gradual changes during the subsequent stimulus repetitions. Furthermore, neural response profiles were generally more dissimilar from each other and increased with stimulus repetition more robustly in the shuffled than in the blocked sequence.

Multi-unit Within-stimulus Neural Dissimilarities

Songs. Next, the dissimilarities between the temporal profiles of neural responses to different repetitions of the same songs were analyzed via a mixed ANOVA on within-stimulus neural dissimilarities using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. The effect of sequence was highly significant ($F(1,334) = 32.60$, $p < 0.001$, **Fig. 10A**), indicating dramatically greater within-stimulus neural dissimilarities in the shuffled than in the blocked sequence. The effect of presentation was also significant ($F(24,8016) = 19.86$, $p < 0.001$, **Fig. 10A**). Post-hoc tests demonstrated that within-stimulus neural dissimilarities significantly decreased from the 1st until the 6th stimulus presentation, after which there was no consistent change. There was also a significant interaction between sequence and presentation ($F(24,8016) = 8.68$, $p < 0.001$, **Fig. 10A**); however post-hoc comparisons indicated no significant difference between the two sequences in corresponding presentations and the interaction was driven by differences in the rates

within-stimulus neural dissimilarities changed with stimulus presentation in the two sequences.

These differences were further examined via a mixed ANOVA on the slopes of within-stimulus neural dissimilarities using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable. Overall, the slopes of within-stimulus neural dissimilarities were significantly more negative in the blocked than in the shuffled sequence ($F(1,334) = 25.11$, $p < 0.001$, **Fig. 10B**). Moreover, the slopes of within-stimulus neural dissimilarities were significantly smaller for presentations 1-6 than for presentations 6-25 ($F(1,334) = 152.92$, $p < 0.001$, **Fig. 10B**). Most importantly, there was a significant interaction between sequence and presentation ($F(1,334) = 16.64$, $p < 0.001$, **Fig. 10B**). Post-hoc comparisons revealed that, for presentations 1-6, within-stimulus neural dissimilarity slopes were significantly more negative in the blocked than in the shuffled sequence ($p < 0.001$), whereas no such difference was observed for presentations 6-25 ($p = 0.888$). In both sequences, the slopes of within-stimulus neural dissimilarities were significantly more negative for presentations 1-6 than for presentations 6-25 (both $p < 0.001$). Finally, the slopes for presentations 1-6 in both sequences were significantly less than 0 (both $p < 0.001$). Conversely, the two sequences together indicated that the slopes of within-stimulus neural dissimilarities for presentations 6-25 were slightly, but significantly, greater than 0 ($p = 0.025$).

Summary: To summarize, these analyses revealed that the dissimilarities between the temporal profiles of neural responses to different repetitions of the same songs underwent an initial decrease during the first 6 presentations, then slightly increase

during the following stimulus repetitions. The initial reduction in these dissimilarities was more pronounced in the blocked than in the shuffled sequence. However, the most crucial finding was that, across stimulus presentations, neural response profiles in different repetitions of the same songs were much more dissimilar to each other in the shuffled than in the blocked sequence.

Calls. Within-stimulus neural dissimilarities and their linear trends with repeated stimulus presentation for call responses were also an analyzed with similar ANOVAs. Similar to results in song responses, within-stimulus neural dissimilarities were significantly greater in the shuffled than in the blocked sequence ($F(1,341) = 15.28$, $p < 0.001$, **Fig. 10C**). The effect of presentation was also significant ($F(24,8184) = 23.53$, $p = 0.001$, **Fig. 10C**), however the post-hoc comparisons did not show a consistent pattern, except that the within-stimulus neural dissimilarities in the second half of stimulus presentations were generally greater than those in the first half. There was also a significant interaction between sequence and presentation ($F(24,8184) = 8.65$, $p < 0.001$, **Fig. 10C**); however, once again, there was no difference between the two sequences in corresponding presentations and the interaction was due to differences in the rates within-stimulus neural dissimilarities changed as a function of stimulus presentation.

The analysis of these differences revealed that the slopes of within-stimulus neural dissimilarities were greater in the shuffled than in the blocked sequence ($F(1,341) = 11.65$, $p < 0.001$, **Fig. 10D**). However, there was no effect of presentation ($F(1,341) = 0.98$, $p = 0.323$) or an interaction between sequence and presentation ($F(1,341) = 2.67$, $p = 0.103$). The slopes of within-stimulus neural dissimilarities were significantly greater

than 0 in the shuffled sequence ($p < 0.001$), whereas no such difference was found in the blocked sequence ($p = 0.746$).

Summary: Taken together, these analyses revealed patterns that were partially in line with the findings in songs responses. Most importantly, the dissimilarities between the temporal profiles of neural responses to different presentations of the same calls were much higher in the shuffled than in the blocked sequence. These dissimilarities gradually increased with stimulus repetition in the shuffled sequence, but did not change in the blocked sequence.

Multi-unit Correct Decoding Probabilities

Songs. Having examined how repeated stimulus presentation changed the dissimilarities between the temporal profiles of neural responses, the effects of these changes on neural decoding of songs were assessed next. **Figures 11A and 11B** show the correct neural decoding probabilities across time points along the stimulus duration and across stimulus presentations in the blocked and the shuffled sequence, respectively. Visual examination of these figures clearly indicates that the probability of correct neural decoding increases with time along the stimulus duration and also with stimulus presentation. To analyze these changes in detail, first, several different time points were selected and the probabilities across stimulus presentations at each of these time points were analyzed separately. These analyses all produced similar findings, thus only the results for the 500 ms time point are presented here. The probabilities of correct neural decoding were analyzed via a mixed ANOVA using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. The effect of sequence was highly significant ($F(1,334) = 35.39$, $p < 0.001$, **Fig. 11C**),

such that the probabilities were greater in the blocked than in the shuffled sequence. The effect of presentation was also significant ($F(24,8016) = 31.88$, $p < 0.001$, **Fig. 11C**). Post-hoc tests showed that there was a significant increase in probabilities during the first 6 presentations, after which there was no consistent change. There was also a significant interaction between sequence and presentation ($F(24,8016) = 3.06$, $p < 0.001$, **Fig. 11C**), however post-hoc comparisons did not indicate a significant difference between the two sequences in corresponding presentations. The interaction was driven by differences in the rates at which probabilities changed with stimulus presentation in the two sequences.

These differences were further investigated via a mixed ANOVA on the slopes of probabilities using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable. Overall, the slopes of probabilities did not significantly differ between the blocked and the shuffled sequence ($F(1,334) = 1.45$, $p = 0.229$). Furthermore, there was also no significant interaction between sequence and presentation ($F(1,334) = 1.38$, $p = 0.242$). However, the effect of presentation was highly significant ($F(1,334) = 97.72$, $p < 0.001$, **Fig. 11D**), such that the slopes of probabilities for presentations 1-6 were significantly greater than those for presentations 6-25. Finally, the slopes of probabilities were significantly greater than 0 for both presentations 1-6 and 6-25 in both sequences (all $p < 0.032$).

Summary: Taken together, these analyses showed that neural decoding of song identities using the temporal profiles of neural responses improved with repeated stimulus exposure. These improvements were stronger during the first 6 presentations than the more gradual improvements during the subsequent stimulus repetitions.

Furthermore, neural decoding of songs across presentations was much more accurate in the blocked than in the shuffled sequence.

Calls. Correct neural decoding probabilities across time points along the stimulus duration and across stimulus presentations for call responses are shown in the blocked and the shuffled sequence are shown in **Figures 12A and 12B**, respectively. Similar to the findings in song responses, correct decoding probabilities increased as a function of time along the stimulus duration and stimulus presentation. The analysis of probabilities across stimulus presentations at different time points revealed similar results, thus only the results of 160 ms time point are reported here. These probabilities, as well as their slopes as a function of stimulus repetition, in call responses were analyzed via ANOVAs similar to the ones described for song responses. Once again, probabilities were significantly greater in the blocked than in the shuffled sequence ($F(1,341) = 6.03$, $p = 0.015$, **Fig. 12C**). The effect of presentation was also significant ($F(24,8184) = 7.21$, $p < 0.001$, **Fig. 12C**). Post-hoc analysis showed that probabilities gradually increased with stimulus repetition. There was no significant interaction between sequence and presentation ($F(24,8184) = 1.28$, $p = 0.162$)

The analysis of the slopes of probabilities for call responses did not indicate a significant difference between the two stimulus presentation sequences ($F(1,341) = 0.09$, $p = 0.768$, **Fig. 12D**) or between the two presentation parts ($F(1,341) = 2.85$, $p = 0.092$). There was also no significant interaction between sequence and presentation ($F(1,341) = 0.95$, $p = 0.331$). The slopes of probabilities were significantly greater than 0 for both presentations 1-6 and 6-25 in the two sequences together (both $p < 0.005$).

Summary: These findings were exactly in line with those seen in song responses. The accuracy of neural decoding of calls using the temporal profiles of neural responses improved strongly during the first 6 and more gradually during the following stimulus presentations. Again, the neural decoding performance was much better in the blocked than in the shuffled sequence.

Multi-unit Correct Decoding Latencies

Songs. Another way of dissecting the neural decoding probabilities shown in **Figures 11A and 11B**, rather than fixing time and allowing probabilities to vary, is to keep a probability level constant and focus on the changes in latencies to reach that probability level across stimulus presentations. Visual examinations indicate that the neural decoding process reached any given probability level sooner and sooner with repeated stimulus presentation. The analysis of several different probability levels in song responses revealed similar patterns, thus only the 0.75 probability level is presented here. A mixed ANOVA on neural decoding latencies using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable revealed that latencies were significantly greater in the shuffled than in the blocked sequence ($F(1,334) = 12.90$, $p < 0.001$, **Fig. 11E**). The effect of presentation was also highly significant ($F(24,8016) = 14.63$, $p < 0.001$, **Fig. 11E**). Post-hoc comparisons showed that there was a sharp decrease in latencies in the first 3 presentations, after which no consistent change was observed. There was also a significant interaction between sequence and presentation ($F(24,8016) = 2.53$, $p < 0.001$, **Fig. 11E**), however post-hoc tests did not indicate a significant difference between the two sequences in

corresponding presentations and the interaction seemed driven by differences in the way latencies changed with stimulus presentation in the two sequences.

These differences were further analyzed in detail via a mixed ANOVA on the slopes of latencies using sequence (Blocked, Shuffled) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable. There was a significant effect of sequence ($F(1,334) = 6.23$, $p = 0.013$, **Fig. 11F**), indicating more negative slopes for latencies in the blocked than in the shuffled sequence. Furthermore, the slopes of latencies for presentations 1-6 were significantly more negative than those for presentations 6-25 ($F(1,334) = 55.29$, $p < 0.001$, **Fig. 11F**). Most importantly, there was an interaction between sequence and presentation ($F(1,334) = 10.38$, $p = 0.001$, **Fig. 11F**). Although the slopes of latencies for presentations 1-6 were significantly more negative in the blocked than in the shuffled sequence ($p < 0.001$), no difference was found for presentations 6-25 ($p = 0.942$). Looking at the interaction from the other perspective, the slopes of latencies for presentations 1-6 was significantly more negative than those for presentations 6-25 in both the blocked ($p < 0.001$) and the shuffled sequence ($p = 0.023$). Finally, the slopes of latencies for presentations 1-6 in both sequences separately, as well as for presentations 6-25 together, were significantly less than 0 (all $p < 0.002$).

Summary: Overall, these findings showed that neural decoding of songs using the temporal profiles of neural responses reached a certain confidence level sooner and sooner with repeated stimulus exposure. These changes were strong during the first 6 presentations and continued gradually thereafter. In addition, the same confidence level

for neural decoding was reached faster along the stimulus duration in the blocked than in the shuffled sequence.

Calls. Latencies to reach selected correct neural decoding probabilities in call responses were analyzed via similar ANOVAs. The analysis of different probability levels revealed similar results, thus only the 0.5 probability level is presented here. There was no significant difference between the two sequences ($F(1,341) = 0.15$, $p = 0.697$, **Fig. 12E**) or across presentations ($F(24,8184) = 1.05$, $p = 0.394$). There was also no significant interaction between sequence and presentation ($F(24,8184) = 0.74$, $p = 0.820$). The analysis of the slopes of latencies also showed no effect of sequence ($F(1,341) < 0.01$, $p = 0.952$, **Fig. 12F**), presentation ($F(1,341) = 1.44$, $p = 0.231$), or an interaction between sequence and presentation ($F(1,341) = 0.16$, $p = 0.687$). The slopes of latencies were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p < 0.173$).

Summary: In contrast to the findings in song responses, the analysis of correct neural decoding latencies in call responses did not reveal a change with stimulus presentation or a difference between the two sequences.

Multi-unit Mutual Information

Songs. Finally, the mutual information between songs and the temporal profiles of neural responses across time points along the stimulus duration was analyzed via a mixed ANOVA with the between-subjects variable sequence (Blocked, Shuffled) and the within-subjects variable bin (1 to 75). Across bins, mutual information was significantly greater in the blocked than in the shuffled sequence ($F(1,334) = 29.53$, $p < 0.001$, **Fig. 13A**). The effect of bin was also significant ($F(74,24716) = 2597.95$, $p < 0.001$, **Fig.**

13A), such that mutual information significantly increased from the 1st until the 68th bin (680 ms), after which there was no change. Most importantly, there was a significant interaction between sequence and bin ($F(74,24716) = 27.26$, $p < 0.001$, **Fig. 13A**). Planned comparisons indicated that mutual information was significantly greater in the blocked than in the shuffled sequence between the 2nd and 7th bins (20-70 ms) and again from the 28th bin until the end of the stimulus period (280-750 ms).

Summary: This analysis provided further support for the differences between the two sequences in informativeness of the temporal profiles of neural responses for decoding songs. Mutual information was greater in the blocked than in the shuffled sequence as early as 20 ms after the stimulus onset.

Calls. Mutual information between neural responses and calls were also analyzed via a similar ANOVA. Again, significantly higher mutual information estimations were found in the blocked than in the shuffled sequence ($F(1,341) = 5.77$, $p = 0.017$, **Fig. 13B**). There was also a significant effect of bin ($F(23,7843) = 414.72$, $p < 0.001$, **Fig. 13B**), indicating significant increases in mutual information estimations from the 1st until the 21st bin (210 ms) and no change thereafter. The interaction between sequence and bin was also significant ($F(23,7843) = 10.12$, $p < 0.001$, **Fig. 13B**). Planned comparisons revealed that mutual information estimations were significantly greater in the blocked than in the shuffled sequence between the 2nd and 4th bins (20-40 ms) and again from the 22nd bin until the end of the stimulus period (220-260 ms).

Summary: The analysis of mutual information in call responses revealed exactly the same findings seen in song responses. Mutual information was greater in the blocked than in the shuffled sequence as early as 20 ms after the stimulus onset.

Multi-unit Adaptation and Neural Discrimination Relationships

Songs. To assess whether the adaptation of neural responses was responsible for changes in the neural discrimination with repeated stimulus presentation, the correlations between the adaptation rates and the slopes of each on the neural discrimination metrics were analyzed separately for presentations 1-6 and 6-25. Preliminary analyses did not reveal systematic differences between the two sequences, so the data was combined for subsequent analyses. The analysis of song responses revealed that, for presentations 1-6, adaptation rates were negatively correlated with the slopes of between-stimulus neural dissimilarities ($r(334) = -0.24$, $p < 0.001$, **Fig. 14A**) and positively correlated with the slopes of within-stimulus neural dissimilarities ($r(334) = 0.26$, $p < 0.001$, **Fig. 14B**). Furthermore, adaptation rates were negatively correlated with the slopes of probabilities ($r(334) = -0.14$, $p = 0.013$, **Fig. 14C**) and positively correlated with the slopes of latencies for presentations 1-6 ($r(334) = 0.13$, $p = 0.018$, **Fig. 14D**). For presentations 6-25, adaptation to songs were again negatively correlated with the slopes of between-stimulus neural dissimilarities ($r(334) = -0.26$, $p < 0.001$, **Fig. 15A**). However, unlike for presentations 1-6, there was a negative correlation between adaptation rates and the slopes of within-stimulus neural dissimilarities for presentations 6-25 ($r(334) = -0.25$, $p < 0.001$, **Fig. 15B**). Adaptation rates were not found to be related either to the slopes of probabilities ($r(334) = 0.03$, $p = 0.636$) or to the slopes of latencies for presentations 6-25 ($r(334) = 0.05$, $p = 0.323$).

Summary: Taken together, these analyses showed that the changes in neural discrimination of songs across stimulus repetitions were associated with adaptation of responses. During the first 6 presentations, multi-unit sites that adapted more steeply also

displayed stronger increases in the dissimilarities between the temporal profiles of neural responses to different songs and decreases in the dissimilarities between the temporal profiles of neural responses to the same songs. Thus, multi-units that underwent stronger adaptation during the first 6 presentations also showed greater improvements in correct decoding probabilities and greater reductions in correct decoding latencies. These patterns in neural decoding performance were not observed during the following stimulus presentations, however, surprisingly, the links between the adaptation rates and changes in within-stimulus neural dissimilarities were reversed.

Calls. The same analyses were also conducted on responses to calls. For presentations 1-6, adaptation rates were not related to the slopes of between-stimulus neural dissimilarities ($r(341) = 0.05$, $p = 0.373$), but were positively related to the slopes of within-stimulus neural dissimilarities ($r(341) = 0.19$, $p < 0.001$, **Fig. 16A**). Although the slopes of probabilities were negatively related to adaptation rates for presentations 1-6 ($r(341) = -0.14$, $p = 0.012$, **Fig. 16B**), the slopes of latencies were not related to adaptation rates for the same presentations ($r(341) = -0.10$, $p = 0.073$). For presentations 6-25, adaptation rates were negatively related to the slopes of both the between-stimulus ($r(341) = -0.22$, $p < 0.001$, **Fig. 17A**) and the within-stimulus neural dissimilarities ($r(341) = -0.21$, $p < 0.001$, **Fig. 17B**). There was also a negative relationship between adaptation rates and the slopes of probabilities for presentations 6-25 ($r(341) = -0.15$, $p = 0.005$, **Fig. 17C**), whereas there was no relationship between the slopes of latencies and adaptation rates for the same presentations ($r(341) = -0.08$, $p = 0.161$).

Summary: These results were partially in line with the findings in song responses. Once again, stronger adaptation was related to stronger decreases in within-stimulus

neural dissimilarities during the first 6 presentations; however, during the following presentations, the reverse was true. Most importantly, multi-units that underwent stronger adaptation during both the first 6 and the subsequent 20 stimulus presentations showed greater improvements in correct decoding probabilities.

Single-unit Responses

For song playbacks, 106 neurons (61 narrow and 45 wide spike) were recorded in the blocked sequence; 113 neurons (53 narrow and 60 wide spike) were recorded in the shuffled sequence. For call playbacks, 111 neurons (58 narrow and 53 wide spike) were recorded in the blocked sequence; 123 neurons (64 narrow and 59 wide spike) were recorded in the shuffled sequence. Chi-square analyses indicated that there was no significant difference among the frequencies either in song ($X^2 = 2.48$, $p = 0.115$) or call recordings ($X^2 = 0.02$, $p = 0.887$).

Single-unit Response Properties

Songs. Firing rates of narrow spike neurons during the silent baseline condition in song recordings were significantly greater than those of wide spike neurons ($z = 2.92$, $p = 0.004$, **Fig. 18A**). There was no overall difference between the blocked and shuffled sequences in baseline firing rates ($z = 0.06$, $p = 0.952$). The more detailed pairwise comparisons of the two neuron types and the two sequences revealed that, in the blocked sequence, there was no difference between the baseline firing rates of narrow and wide spike neurons ($z = 0.046$, $p = 0.648$). However, in the shuffled sequence, narrow spike neurons had significantly higher baseline firing rates than did wide spike neurons ($z = 3.75$, $p < 0.001$, **Fig. 18A**). Looking at the interaction from the other perspective, baseline

firing rates did not differ between the two sequences either for narrow ($z = 1.67$, $p = 0.095$) or wide spike neurons ($z = 1.31$, $p = 0.191$).

The analysis of the magnitude of stimulus-driven responses revealed a marked difference between the two neuron types such that, on average, narrow spike neurons had twice the response magnitude of wide spike neurons ($z = 7.40$, $p < 0.001$, **Fig. 18B**). There was also a significant difference between the two sequences, indicating higher response magnitudes in the blocked than in the shuffled sequence ($z = 3.29$, $p < 0.001$, **Fig. 18B**). Neuron type and sequence also seemed to interact with each other as no significant difference was found between the two sequences in the response magnitudes of narrow spike neurons ($z = 0.94$, $p = 0.350$), whereas wide spike neurons had significantly greater response magnitudes in the blocked than in the shuffled sequence ($z = 3.09$, $p = 0.002$, **Fig. 18B**). The response magnitudes of narrow spike neurons were significantly greater than those of wide spike neurons in both the blocked and the shuffled sequence (both $p < 0.001$, **Fig. 18B**).

The analysis of adaptation rates for presentations 1-6 did not show a significant difference between the two neuron types ($z = 2.33$, $p = 0.020$) or the two sequences ($z = 2.11$, $p = 0.035$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and sequence groups (all $p > 0.015$). The adaptation rates of both neuron types together for presentations 1-6 were significantly less than 0 ($z = 7.26$, $p < 0.001$, **Fig. 18C**).

Adaptation rates for presentations 6-25 significantly differed between the two neuron types, indicating more negative adaptation rates for wide than for narrow spike neurons ($z = 4.72$, $p < 0.001$, **Fig. 18D**). There was no significant difference between the

two sequences ($z = 0.79$, $p = 0.427$). Neuron type and sequence also did not interact with each other as indicated by similarly more negative adaptation rates for wide than for wide spike neurons both in the blocked ($z = 3.32$, $p < 0.001$) and the shuffled sequence ($z = 3.26$, $p = 0.001$, **Fig. 18D**). There was no difference between the two sequences in the adaptation rates of either narrow ($z = 0.38$, $p = 0.701$) or wide spike neurons ($z = 1.00$, $p = 0.317$). Finally, adaptation rates for presentations 6-25 were significantly less than 0 for both narrow ($z = 3.48$, $p < 0.001$) and wide spike neurons ($z = 6.53$, $p < 0.001$, **Fig. 18D**).

Summary: Taken together, these analyses indicated that narrow spike neurons had higher baseline firing rates and stronger responses to songs than did wide spike neurons. Although there was no difference between the two neuron types in rates of adaptation during the first 6 stimulus presentations, wide spike neurons adapted more strongly than did narrow spike neurons over subsequent stimulus repetitions. In addition, wide, but not narrow, spike neurons had stronger responses to songs in the blocked than in the shuffled sequence.

Calls. Basic properties of responses to calls in the two neuron types and in the two sequences were analyzed similarly. Again, narrow spike neurons had significantly greater firing rates during baseline conditions as compared to wide spike neurons ($z = 6.48$, $p < 0.001$, **Fig. 19A**). The baseline firing rates did not differ between the blocked and shuffled sequences ($z = 0.54$, $p = 0.588$). There was also no indication of an interaction between neuron type and sequence as revealed by significantly higher baseline firing rates for narrow than for wide spike neurons in both the blocked ($z = 4.92$, $p < 0.001$) and the shuffled sequence ($z = 4.26$, $p < 0.001$, **Fig. 19A**). The baseline firing

rates did not differ between the two sequences either for narrow ($z = 0.75$, $p = 0.451$) or wide spike neurons ($z = 0.51$, $p = 0.61$).

The stimulus-driven response magnitudes of narrow spike neurons were markedly greater as compared those of wide spike neurons ($z = 6.52$, $p < 0.001$, **Fig. 19B**). There was a trend toward higher response magnitudes in the blocked than in the shuffled sequence, however it was not statistically significant at the corrected alpha level ($z = 2.39$, $p = 0.017$). There was also no indication of an interaction between neuron type and sequence as shown by similarly greater response magnitudes for narrow than for wide spike neurons in both the blocked ($z = 4.72$, $p < 0.001$) and the shuffled sequence ($z = 4.57$, $p < 0.001$, **Fig. 19B**). There was no significant difference between the two sequences in the response magnitudes of either narrow ($z = 2.26$, $p = 0.024$) or wide spike neurons ($z = 1.34$, $p = 0.179$).

The analysis of adaptation rates for presentations 1-6 did not reveal a significant difference between narrow and wide spike neurons ($z = 2.24$, $p = 0.025$). There was also no significant difference between the two sequences ($z = 0.25$, $p = 0.802$). However, there seemed to be an interaction between neuron type and sequence such that, in the blocked sequence, wide spike neurons had significantly more negative adaptation rates than did narrow spike neurons ($z = 2.70$, $p = 0.007$, **Fig. 19C**), whereas there was no such difference in the shuffled sequence ($z = 0.65$, $p = 0.518$). The adaptation rates in the two sequences did not differ from each other either for narrow ($z = 0.87$, $p = 0.384$) or for wide spike neurons ($z = 0.89$, $p = 0.371$). The adaptation rates for presentations 1-6 were significantly less than 0 for both narrow ($z = 2.65$, $p = 0.008$) and wide spike neurons ($z =$

3.97, $p < 0.001$) in the blocked sequence, as well as for the two neuron types together in the shuffled sequence ($z = 3.75$, $p < 0.001$, **Fig. 19C**).

The adaptation rates for presentations 6-25 were significantly more negative for wide than for narrow spike neurons ($z = 3.08$, $p = 0.002$, **Fig. 19D**). There was no significant difference between the adaptation rates in the two sequences ($z = 2.36$, $p = 0.018$). Neuron type and sequence seemed to interact with each other as shown by significantly more negative adaptation rates for wide than for narrow spike neurons in the blocked ($z = 3.68$, $p < 0.001$, **Fig. 19D**), but not in the shuffled sequence ($z = 1.25$, $p = 0.212$). Looking at the interaction from the other perspective, the adaptation rates in the two sequences did not differ from each other either for narrow ($z = 2.25$, $p = 0.024$) or for wide spike neurons ($z = 1.24$, $p = 0.214$). Finally, the adaptation rates for presentations 6-25 were significantly less than 0 for both the narrow ($z = 4.68$, $p < 0.001$) and wide spike neurons ($z = 4.99$, $p < 0.001$) in the blocked sequence, as well as for the two neuron types together in the shuffled sequence ($z = 5.89$, $p < 0.001$, **Fig. 19D**).

Summary: Analysis of basic properties of responses to calls revealed patterns similar to the findings in song responses. Narrow spike neurons had higher baseline firing rates and stronger responses than did wide spike neurons. On the other hand, responses to calls adapted more strongly for wide than for narrow spike neurons, however this difference was only observed in the blocked sequence.

Single-unit Between-stimulus Neural Dissimilarities

Songs. Following the same approach as for multi-unit responses, the dissimilarities between the temporal profiles of neural responses to different songs were compared between the two sequences in single-units. Across presentations, between-

stimulus neural dissimilarities were significantly greater for wide than for narrow spike neurons ($z = 4.44$, $p < 0.001$, **Fig. 20A**). In addition, there were significantly higher between-stimulus neural dissimilarities in the shuffled than in the blocked sequence ($z = 3.24$, $p = 0.001$, **Fig. 20A**). Neuron type and sequence did not seem to interact with each other as indicated by similarly greater between-stimulus neural dissimilarities for wide than for narrow spike neurons in both the blocked ($z = 2.98$, $p = 0.003$) and the shuffled sequence ($z = 3.14$, $p = 0.003$, **Fig. 20A**). There was no significant difference between the between-stimulus neural dissimilarities in the two sequences either for narrow ($z = 2.37$, $p = 0.018$) or for wide spike neurons ($z = 1.96$, $p = 0.051$).

The changes in between-stimulus neural dissimilarities as a function of stimulus repetition were analyzed via the slopes for presentations 1-6 and 6-25 as in multi-unit responses. The analysis of the slopes of between-stimulus neural dissimilarities for presentations 1-6 did not show any difference between the two neuron types ($z = 0.57$, $p = 0.567$) or the two sequences ($z = 0.44$, $p = 0.660$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and sequence conditions (all $p > 0.331$). Similarly, the slopes of between-stimulus neural dissimilarities for presentations 6-25 were also not different between the two neuron types ($z = 0.43$, $p = 0.669$) or the two sequences ($z = 1.43$, $p = 0.152$). There was also no interaction (all $p > 0.284$). For presentations 1-6, the slopes of between-stimulus neural dissimilarities were not significantly different from 0 ($z = 0.61$, $p = 0.540$), whereas, for presentations 6-25, they were significantly greater than 0 ($z = 3.17$, $p = 0.002$, **Fig. 20B**).

Summary: These findings in single-units were partially in line with those seen in multi-unit responses. Most importantly, the temporal profiles of single-unit responses to

different songs were more dissimilar from each other in the shuffled than in the blocked sequence. Furthermore, wide spike neurons responded more differently to different songs than did narrow spike neurons. Neural dissimilarities for different songs increased after the first 6 stimulus presentations, but there was no difference between the two neuron types in the two sequences in terms of the rates of these improvements.

Calls. The between-stimulus neural dissimilarities and their slopes were also analyzed for single-unit responses to calls. Wide spike neurons had significantly greater between-stimulus neural dissimilarities than did narrow spike neurons ($z = 4.89$, $p < 0.001$, **Fig. 21A**). There was no overall difference between the two sequences ($z = 1.84$, $p = 0.066$). However, there seemed to be an interaction between neuron type and sequence such that the between-stimulus neural dissimilarities of wide spike neurons were significantly greater than those of narrow spike neurons only in the blocked sequence ($z = 4.77$, $p < 0.001$, **Fig. 21A**), whereas no such difference was observed in the shuffled sequence ($z = 2.29$, $p = 0.022$). In parallel, narrow spike neurons showed significantly higher between-stimulus neural dissimilarities in the shuffled than in the blocked sequence ($z = 2.67$, $p = 0.008$, **Fig. 21A**), however there was no such difference for wide spike neurons ($z = 0.044$, $p = 0.965$).

The slopes of between-stimulus neural dissimilarities for presentations 1-6 were not different between the two neuron types ($z = 0.81$, $p = 0.417$) or the two sequences ($z = 0.20$, $p = 0.840$). There was also no interaction between neuron type and sequence (all $p > 0.404$). The analysis of the slopes of between-stimulus neural dissimilarities also did not reveal a neuron type ($z = 1.43$, $p = 0.153$), a sequence ($z = 1.93$, $p = 0.054$), or an interaction effect (all $p > 0.011$). The slopes of between-stimulus neural dissimilarities

were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.067$).

Summary: These findings were generally similar to the ones seen in single-unit responses to songs, but were more mixed. Wide spike neurons responded more dissimilarly to different calls than did narrow spike neurons, except that this was only seen in the blocked sequence. In addition, between-stimulus neural dissimilarities were greater in the shuffled than in the blocked sequence, but only for narrow spike neurons. Unlike the multi-unit responses to songs and calls and single-unit responses to songs, the temporal profiles of single-unit responses to different calls did not become more and more dissimilar from each other with repeated stimulus exposure.

Single-unit Within-stimulus Neural Dissimilarities

Songs. Next, the dissimilarities between the temporal profiles of neural responses to different presentations of the same songs were compared between the two sequences in single-units. Overall, within-stimulus neural dissimilarities were significantly greater for wide than for narrow spike neurons ($z = 5.99$, $p < 0.001$, **Fig. 20C**) and in the shuffled than in the blocked sequence ($z = 5.45$, $p < 0.001$, **Fig. 20C**). There was no indication of an interaction between neuron type and sequence as demonstrated by similarly greater within-stimulus neural dissimilarities for wide than for narrow spike neurons in both the blocked ($z = 2.98$, $p = 0.003$) and the shuffled sequences ($z = 3.14$, $p = 0.002$, **Fig. 20C**). Moreover, within-stimulus neural dissimilarities were similarly greater in the shuffled than in the blocked sequence both for narrow ($z = 3.50$, $p < 0.001$) and wide spike neurons ($z = 4.05$, $p < 0.001$, **Fig. 20C**).

The analysis of the slopes of within-stimulus neural dissimilarities for presentations 1-6 did not show any difference between the two neuron types ($z = 0.01$, $p = 0.992$) or the two sequences ($z = 1.29$, $p = 0.198$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and sequence conditions (all $p > 0.179$). The slopes of within-stimulus neural dissimilarities for presentations 6-25 also did not reveal a significant difference between the two neuron types ($z = 1.67$, $p = 0.095$). However, there was a significant difference between the two sequences such that the slopes of within-stimulus neural dissimilarities for presentations 6-25 were greater in the shuffled than in the blocked sequence ($z = 2.93$, $p = 0.003$, **Fig. 20D**). There was no interaction between neuron type and sequence (all $p > 0.021$). The slopes of within-stimulus neural dissimilarities for presentations 1-6 were significantly greater than 0 ($z = 6.63$, $p < 0.001$, **Fig. 20D**). For presentations 6-25, within-stimulus neural dissimilarity slopes were significantly greater than 0 in the blocked ($z = 2.60$, $p = 0.009$, **Fig. 20D**), but not in the shuffled sequence ($z = 1.45$, $p = 0.148$).

Summary: Taken together, these analyses showed that the temporal profiles of neural responses to different presentations of the same songs were more dissimilar from each other in the shuffled than in blocked sequence, similar to the findings in multi-unit responses. In addition, wide spike neurons responded more dissimilarly to different presentations of the same songs than did narrow spike neurons. Within-stimulus neural dissimilarities increased with stimulus presentation during the first 6 presentations in both sequences and during the subsequent repetitions only in the blocked sequence.

Calls. Within-stimulus neural dissimilarities, as well as their slopes, were analyzed similarly for single-unit responses to calls. Within-stimulus neural

dissimilarities were significantly greater for wide than for narrow spike neurons ($z = 4.23$, $p < 0.001$, **Fig. 21B**) and in the shuffled than in the blocked sequence ($z = 2.99$, $p = 0.003$, **Fig. 21B**). There was no indication of an interaction between neuron type and sequence as indicated by similarly greater within-stimulus neural dissimilarities for wide than for narrow spike neurons both in the blocked ($z = 3.15$, $p = 0.002$) and the shuffled sequence ($z = 3.14$, $p = 0.002$, **Fig. 21B**). There was no significant difference between the two sequences either for narrow ($z = 2.46$, $p = 0.015$) or for wide spike neurons ($z = 2.01$, $p = 0.044$).

The slopes of within-stimulus neural dissimilarities for presentations 1-6 were not different between the two neuron types ($z = 0.41$, $p = 0.685$) or the two sequences ($z = 0.40$, $p = 0.690$). There was also no interaction between neuron type and sequence as indicated by lack of differences in pairwise comparisons of neuron type and sequence conditions (all $p > 0.727$). The analysis of the slopes of within-stimulus neural dissimilarities for presentations 6-25 also did not reveal a neuron type ($z = 0.09$, $p = 0.931$), a sequence ($z = 1.30$, $p = 0.193$), or an interaction effect (all $p > 0.257$). The slopes of within-stimulus neural dissimilarities were significantly less than 0 for presentations 1-6 ($z = 3.39$, $p < 0.001$, **Fig. 21C**), but not for presentations 6-25 ($z = 1.11$, $p = 0.268$).

Summary: These findings were similar to results in single-unit responses to songs in that the temporal profiles of neural responses to different presentations of the same calls were more dissimilar from each other in the shuffled than in blocked sequence and for wide than for narrow spike neurons. In addition, within-stimulus neural dissimilarities decreased only during the first 6 presentations, but not afterwards.

Single-unit Correct Decoding Probabilities

Songs. Correct neural decoding probabilities across time points along the stimulus duration and stimulus presentations in single-unit responses to songs were also analyzed similar to the multi-unit responses. There was a marked difference between the neural decoding probabilities of the two neuron types such that narrow spike neurons had significantly higher probabilities than did wide spike neurons ($z = 7.01$, $p < 0.001$, **Fig. 22A**). In addition, correct decoding probabilities were significantly greater in the blocked than in the shuffled sequence ($z = 4.67$, $p < 0.001$, **Fig. 22A**). There was no indication of an interaction between neuron type and sequence as shown by similarly higher probabilities for narrow than for wide spike neurons in both the blocked ($z = 3.94$, $p < 0.001$) and the shuffled sequence ($z = 5.87$, $p < 0.001$, **Fig. 22A**). Significantly greater probabilities in the blocked than in the shuffled sequence was also observed for both narrow ($z = 2.55$, $p = 0.011$) and wide spike neurons ($z = 3.85$, $p < 0.001$, **Fig. 22A**).

The analysis of the slopes of correct decoding probabilities for presentations 1-6 did not show any difference between the two neuron types ($z = 0.54$, $p = 0.587$) or the two sequences ($z = 0.23$, $p = 0.819$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and sequence conditions (all $p > 0.466$). Similarly, the slopes of probabilities for presentations 6-25 were also not different between the two neuron types ($z = 0.46$, $p = 0.649$) or the two sequences ($z = 0.69$, $p = 0.490$). There was also no interaction (all $p > 0.207$). The slopes of correct decoding probabilities were significantly greater than 0 for presentations 1-6 ($z = 2.55$, $p = 0.001$, **Fig. 22B**), whereas, for presentations 6-25, they were not significantly different from 0 ($z = 1.63$, $p = 0.103$).

Summary: Taken together, these analyses in single-unit responses to songs strongly supported the findings observed in multi-unit responses. Decoding of songs using the temporal profiles of neural responses yielded much more accurate results in the blocked than in shuffled sequence. Furthermore, a completely novel finding was that decoding accuracy was greater for narrow than for wide spike neurons. The two neuron types both showed improvements in neural decoding accuracy in the blocked compared to the shuffled sequence, so the differences between the two sequences cannot be explained by the differential involvement of the two neuron types. In addition, neural decoding accuracy improved at the single-unit level during the first 6 stimulus presentations, but did not change during the following repetitions.

Calls. Correct neural decoding probabilities of narrow spike neurons were markedly greater than those of wide spike neurons in single-unit responses to calls ($z = 6.39$, $p < 0.001$, **Fig. 23A**). Furthermore, probabilities were significantly higher in the blocked than in the shuffled sequence ($z = 2.92$, $p = 0.004$, **Fig. 23A**). There was no indication of an interaction between neuron type and sequence as correct decoding probabilities of narrow spike neurons were similarly greater than those of wide spike neurons in both the blocked ($z = 4.73$, $p < 0.001$) and the shuffled sequence ($z = 4.53$, $p < 0.001$, **Fig. 23A**). The probabilities in the blocked sequence were also significantly greater than those in the shuffled sequence both for narrow ($z = 2.47$, $p = 0.014$) and wide spike neurons ($z = 2.06$, $p = 0.040$, **Fig. 23A**).

The slopes of correct decoding probabilities for presentations 1-6 were not different between the two neuron types ($z = 1.00$, $p = 0.921$) or the two sequences ($z = 0.07$, $p = 0.946$). There was also no interaction between neuron type and sequence (all p

> 0.218). The analysis of the slopes of probabilities for presentations 6-25 also did not reveal a neuron type ($z = 0.16$, $p = 0.875$), a sequence ($z = 0.57$, $p = 0.571$), or an interaction effect (all $p > 0.154$). Furthermore, the slopes of correct decoding probabilities were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.108$).

Summary: These findings in single-unit responses to calls were exactly in line with those observed for songs. Decoding of calls using the temporal profiles of neural responses yielded more accurate results for narrow than for wide spike neurons and in the blocked than in the shuffled sequence. However, unlike in multi-unit responses to songs and calls, and single-unit responses to songs, neural decoding accuracy in single-unit responses to calls did not improve with repeated stimulus presentation.

Single-unit Correct Decoding Latencies

Songs. Latencies along the stimulus duration to reach a selected probability level in single-unit responses to songs were analyzed as for multi-unit responses. Overall, correct decoding latencies were not significantly different between the two neuron types ($z = 0.94$, $p = 0.925$) and the two sequences ($z = 0.34$, $p = 0.734$). There was also no indication of an interaction as none of the pairwise comparisons between the two neuron types in the two sequences revealed significant differences (all $p > 0.038$). Furthermore, the analysis of the slopes of correct decoding latencies for presentations 1-6 did not show any difference between the two neuron types ($z = 0.25$, $p = 0.806$) or the two sequences ($z = 0.64$, $p = 0.523$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and sequence conditions (all $p > 0.488$). Similarly, the slopes of latencies for presentations 6-25 were also not different

between the two neuron types ($z = 0.84$, $p = 0.404$) or the two sequences ($z = 0.48$, $p = 0.634$). There was also no interaction (all $p > 0.398$). The slopes of correct decoding latencies were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.615$).

Summary: In contrast to the findings in multi-unit responses to songs, these analyses showed that correct neural decoding latencies did not change with repeated stimulus presentation and were not different between the two sequences or between the two neuron types in single-unit responses to songs.

Calls. Correct neural decoding latencies in single-unit responses to calls were significantly shorter for wide than for narrow spike neurons ($z = 3.39$, $p < 0.001$, **Fig. 23B**). Latencies in the shuffled sequence were not significantly different from those in the blocked sequence ($z = 2.57$, $p = 0.010$). The more detailed pairwise comparisons indicated an interaction between neuron type and sequence, such that the correct decoding latencies of wide spike neurons were significantly shorter than those of narrow spike neurons in the shuffled ($z = 2.68$, $p = 0.007$, **Fig. 23B**), but not in the blocked sequence ($z = 2.16$, $p = 0.031$). There was no significant difference between the two sequences for either narrow or wide spike neurons (both $p > 0.563$).

The slopes of correct decoding latencies for presentations 1-6 were not different between the two neuron types ($z = 0.83$, $p = 0.407$) or the two sequences ($z = 2.17$, $p = 0.030$). There was also no interaction between neuron type and sequence (all $p > 0.012$). The analysis of the latency slopes for presentation 6-25 also did not reveal a neuron type ($z = 0.96$, $p = 0.336$), a sequence ($z = 1.25$, $p = 0.212$), or an interaction effect (all $p >$

0.108). The slopes of correct decoding latencies did not significantly differ from 0 for either presentations 1-6 or 6-25 (both $p > 0.219$).

Summary: Similar to the findings in single-unit responses to songs, these analyses did not show a decrease in neural decoding latencies with stimulus repetition. In addition, latencies did not differ between the two sequences. The only difference was that the correct neural decoding latencies were shorter for wide than for narrow spike neurons only in the shuffled sequence.

Single-unit Mutual Information

Songs. Finally, mutual information between songs and the temporal profiles of single-unit responses to songs were analyzed. Overall, the responses of narrow spike neurons were markedly more informative about song stimulus identities compared to the responses of wide spike neurons ($z = 7.34$, $p < 0.001$, **Fig. 24A**). Furthermore, mutual information was significantly higher in the blocked than in the shuffled sequence ($z = 4.94$, $p < 0.001$, **Fig. 24A**). There was no indication of an interaction between neuron type and sequence as mutual information estimations of narrow spike neurons were similarly greater than those of wide spike neurons in both the blocked ($z = 4.80$, $p < 0.001$) and the shuffled sequence ($z = 5.59$, $p < 0.001$, **Fig. 24A**). Mutual information in the blocked sequence was also significantly greater than those in the shuffled sequence both for narrow ($z = 3.15$, $p = 0.002$) and wide spike neurons ($z = 3.80$, $p < 0.001$, **Fig. 24A**).

Summary: Similar to the findings in correct neural decoding probabilities, temporal profiles of single-unit responses to songs for narrow spike neurons were more informative about external signals compared to wide spike neurons. Moreover, in parallel

with multi-unit findings, mutual information was higher in the blocked than in the shuffled sequence at the single-unit level.

Calls. The analysis of the mutual information between neural responses and calls revealed exactly the same results. Mutual information was significantly greater for narrow than for wide spike neurons ($z = 3.77$, $p < 0.001$, **Fig. 24B**) and in the blocked than in the shuffled sequence ($z = 3.36$, $p < 0.001$, **Fig. 24B**). However, pairwise comparisons indicated that, in shuffled sequence, narrow spike neurons had significantly higher mutual information than did wide spike neurons ($z = 3.28$, $p = 0.001$, **Fig. 24B**), whereas, in the blocked sequence, the difference between the mutual information estimations of narrow and wide spike neurons did not reach significance at the corrected alpha level ($z = 2.19$, $p = 0.029$). In addition, mutual information was significantly greater in the blocked sequence than in the shuffled sequence for wide spike neurons ($z = 2.81$, $p = 0.005$, **Fig. 24B**), but not for narrow spike neurons ($z = 2.15$, $p = 0.032$).

Summary: These findings were in line with the patterns observed in single-unit responses to songs. That is, temporal profiles of single-unit responses to calls were more informative about external signals for narrow spike neurons compared to wide spike neurons and in the blocked sequence than in the shuffled sequence. Even though there seemed to be interactions between the two neuron types and the two sequences, trends in each group were in line with the main effects.

Single-unit Adaptation and Neural Discrimination Relationships

Songs. The analysis of the relationship between adaptation rates and the slopes of each one of the neural discrimination metrics in single-unit song responses did not revealed any significant correlations for presentations 1-6 (all $p > 0.064$). However, for

presentations 6-25, adaptation rates were negative correlated with the slopes of both the between-stimulus ($r(205) = -0.16$, $p = 0.023$, **Fig. 25A**) and the within-stimulus neural dissimilarities ($r(205) = -0.25$, $p < 0.001$, **Fig. 25B**). In contrast, adaptation rates were positively correlated with both the slopes of correct decoding probabilities ($r(205) = 0.16$, $p = 0.002$, **Fig. 25C**) and the latencies ($r(205) = 0.09$, $p = 0.188$, **Fig. 25D**) for presentations 6-25.

Summary: These analyses revealed mixed patterns. Unlike in multi-unit responses to songs, none of the neural discrimination metrics were related to adaptation rates during the first 6 presentations in single-units. The neurons that showed stronger adaptation during the following 20 presentations also showed stronger improvements in both between-stimulus and within-stimulus neural dissimilarities. Surprisingly, in contrast to the findings in multi-units, neurons with steeper adaptation profiles also displayed stronger reductions in their neural decoding accuracies.

Calls. The same correlation analyses were also conducted in single-unit responses to calls. Again, for presentations 1-6, adaptation rates were not significantly correlated to the slopes of any of the neural discrimination metrics (all $p > 0.295$). The adaptation rates for presentations 6-25 were again positively correlated with the slopes of correct decoding probabilities ($r(217) = 0.17$, $p = 0.012$, **Fig. 26**), similar to what was observed on song responses. However, the slopes of none of the other neural discrimination metrics significantly correlated with adaptation rates for presentation 6-25 (all $p > 0.071$).

Summary: These findings were mainly in line with the results described for single-unit responses to songs in that none of the neural discrimination metrics were

related to adaptation rates during the first 6 presentations. In addition, neurons that underwent stronger adaptation also decreased their neural decoding accuracies more.

Neuron Type Differences Controlled for Response Magnitude Variations

Songs. An ANCOVA with spike width (narrow, wide) as a between-subjects variable and response magnitude as a covariate was conducted on correct decoding probabilities for single-unit responses to songs. Even after controlling for variation in response magnitudes, narrow spike neurons still had significantly greater correct decoding probabilities than did wide spike neurons ($F(1,216) = 30.15$, $p < 0.001$). The analysis of mutual information estimates via a similar ANCOVA also showed that the significantly higher mutual information for narrow than for wide spike neurons was not due to differing levels of response magnitudes between the two neuron types ($F(1,216) = 27.57$, $p < 0.001$).

Summary: These analyses demonstrated that the higher neural decoding accuracies observed for narrow as compared to wide spike neurons for songs could not be explained by differential absolute firing rates between the two neuron types.

Calls. The correct decoding probabilities and mutual information estimations for single-unit responses to calls were also subjected to the same analyses. Similar to the results for songs, correct decoding probabilities were still significantly greater for narrow than for wide spike neurons, even when the variations in response magnitudes were controlled ($F(1,231) = 8.31$, $p = 0.004$). However, the analysis of mutual information estimations showed that, when response magnitudes were partialled out, there was no longer a significant difference between the two neuron types ($F(1,231) = 1.14$, $p = 0.287$).

Summary: These findings suggest that the higher neural decoding accuracies observed for narrow as compared to wide spike neurons for calls could not be accounted for by the differences in total firing rates between the two neuron types. However, mutual information difference between the two neuron types became absent when the absolute levels of responding were controlled.

Discussion

The results of this experiment provided strong evidence that passive familiarization with novel natural vocalizations changes the temporal profiles of neural responses in the zebra finch NCM to improve neural decoding. Both multi-unit responses to songs and calls and single-unit responses to songs showed this effect, which suggests that the changes in the temporal profiles of responses observed in the neural population most likely results from modifications in spike timings of individual neurons with repeated stimulus presentation. It is thus not simply due to changes in the multi-unit population across presentations. In multi-unit responses, the enhancements in neural decoding for both songs and calls were robust during the first 6 stimulus presentations and then continued at more modest rates during the following repetitions. In single-unit responses, neural decoding improved for songs during only the first 6 presentations and did not change with further exposure. Interestingly, no consistent change in neural decoding accuracies with repeated stimulus exposure were seen in single-unit responses to calls. In addition, neural decoding of songs using the temporal profiles of multi-unit responses reached a given confidence level sooner and sooner along the stimulus duration with repeated stimulus presentation. Nevertheless, this effect was not observed in multi-unit responses to calls and in single-unit responses to both songs and calls.

The changes in the dissimilarities between the temporal profiles of neural responses between different presentations of the same stimuli, as well as between presentations of different stimuli, were assessed to dissect the dynamics that produced the improvements in neural decoding accuracies with repeated stimulus exposure. However, these analyses did not reveal clear results. The clearest patterns were observed in the multi-unit responses to songs, which indicated that between-stimulus neural dissimilarities increased, while within-stimulus neural dissimilarities decreased during the first 6 stimulus presentations. This means that the temporal profiles of neural responses to any given stimulus became more consistent for that stimulus and more differentiated from the temporal profiles of neural responses to other stimuli. These dynamics together resulted in the enhancement of neural decoding during these stimulus presentations. During the following 20 stimulus presentations, responses to songs indicated that both the between-stimulus and the within-stimulus neural dissimilarities increased. As the net effect of these changes, gains in neural decoding accuracy continued at a much more modest rate. Although the analysis of these two opposing dynamics helped to explain the improvements in neural decoding accuracy observed in the multi-unit responses to songs, they did not always produce conclusive results. This was demonstrated by the multi-unit responses to calls and single-unit responses to both songs and calls. In these responses, between-stimulus neural dissimilarities either increased or did not change, while within-stimulus neural dissimilarities also tended to increase or not change with stimulus repetition. The net effects of these trends were increases in neural decoding accuracy in some cases and no changes in others. Thus, examination of the directions of changes in between-stimulus and within-stimulus neural

dissimilarities separately were not sufficient to understand the more complex interactions among neural representations governing the improvements in neural decoding accuracy with repeated stimulus presentation.

Contrary to predictions, neural decoding accuracy and mutual information between the temporal profiles of neural responses and natural vocalizations were markedly higher in the blocked than in the shuffled stimulus presentation sequence. This effect was clear in both the multi-unit and single-unit responses to both songs and calls suggesting that the stimulus presentation sequence modulated temporal profiles of neural activity at the individual neuron level. This difference between the two stimulus presentation sequences was due to the differences in between- and within-stimulus neural dissimilarities. In the shuffled sequence, dissimilarities between the temporal profiles of neural responses to different stimuli were generally higher than those in the blocked sequence, which, under normal conditions, would lead to more accurate neural decoding in the shuffled sequence. However, neural dissimilarities between different presentations of the same stimulus were strikingly higher in the shuffled than in the blocked sequence, which shows that the temporal profiles of neural responses to repeated presentations of a particular stimulus were much less consistent (more variable) in the shuffled sequence. The net effect of these two opposing factors was that the ability to identify songs or calls using the temporal profiles of neural responses was diminished in the shuffled compared to the blocked sequence. Taken together, these findings suggest that the more predictable context produced by the blocked stimulus presentation sequence, in contrast to the unpredictable random stimulus transitions in the shuffled sequence, led to more reliable

neural response profiles across repetitions of any given stimulus, which in turn enhanced the neural differentiation among different stimuli.

The direction and the degree of changes in neural discrimination metrics in multi-unit responses were generally found to be related to rates of adaptation. Neural decoding accuracies were negatively correlated with adaptation rates for both stimulus sets and during both presentation parts, except for songs after the first 6 presentations. These negative correlations mean that multi-units which showed stronger adaptation (more negative slope) also showed stronger enhancement in neural decoding accuracy. Between-stimulus neural dissimilarities were also all negatively correlated with adaptation rates, except during the first 6 presentations of calls. Thus, stronger adaptation was associated with stronger increase in the dissimilarity between the temporal profiles of neural responses to different stimuli. Within-stimulus neural dissimilarities, on the other hand, revealed an interesting effect. For both songs and calls, within-stimulus neural dissimilarities were positively correlated with the adaptation rates for the first 6 presentations and negatively correlated with the following presentations. This suggests that stronger adaptation meant more change towards reliability for the neural response profiles of any given stimulus during the initial presentations; however, after one point, it was related to more reductions in neural response reliability. Despite the abundant evidence that the changes in neural discrimination metrics were related to the rates of adaptation, even in the strongest case, adaptation rates explained only less than 7% of the total variation in the slopes of neural discrimination metrics. Thus, adaptation as measured (across the whole response) does not seem to be the main driving factor in

improving neural decoding accuracies; there are likely to be other yet unknown mechanisms in effect.

Following previous studies, single-units were clustered based on spike waveforms into narrow and wide spike neurons. These two neuron types differed significantly from each other in various basic response properties. Narrow spike neurons generally fired at much higher rates in silent baseline conditions and responded to auditory stimulation much more robustly compared to wide spike neurons. Conversely, wide spike neurons generally underwent stronger adaptation with repeated stimulus presentation than did narrow spike neurons. Most remarkably, the two neuron types also differed in neural discrimination of auditory signals, such that the temporal profiles of neural responses for narrow spike neurons more accurately decoded songs and calls and thus had higher mutual information compared to those of wide spike neurons. These differences were generally not due to the differences in the total response magnitudes between the two neuron types. Despite these differences in stimulus representations, the two neuron types were not differentially involved in the blocked and the shuffled stimulus presentation sequences. That is, both narrow and wide spike neurons had similarly increased neural decoding accuracy and mutual information levels in the blocked compared to in the shuffled sequence.

Narrow and wide spike neurons also did not systematically differ in their contributions to the changes in neural discrimination with repeated stimulus presentation. The neural decoding accuracy of both neuron types together increased with stimulus repetition for songs, but not for calls. Despite the overall similarities to multi-unit responses, the analysis of the correlations between the improvements in neural decoding

accuracy and adaptation rates indicated puzzling results. For both songs and calls, adaptation rates were positively correlated with the changes in neural decoding accuracy after the first 6 stimulus presentations, which suggests that the single-units that adapted more strongly, also showed bigger reductions in neural decoding performance. There was no relationship between adaptation rates and the neural decoding changes during the first 6 stimulus presentations.

In summary, passive familiarization of novel natural vocalizations rapidly updated the temporal profiles of neural responses in such a way as to improve discrimination and recognition of those signals. Whether these immediate changes in neural representations were retained over the long-term to affect the processing of the same auditory stimuli at later times was investigated in the following experiment.

EXPERIMENT 2: LONG-TERM EFFECTS OF FAMILIARIZATION ON NEURAL DISCRIMINATION OF VOCAL SIGNALS

The goal of this experiment was to test whether rapid improvements in neural discrimination seen with passive familiarization were long-lasting and could affect processing of the same sounds at a later time point. It is well-established that neural responses to conspecific vocalizations remain adapted at least 20 hours after the initial stimulus presentation (Chew et al., 1996a, 1996b). Accordingly, familiar and novel signals can be differentiated by comparing their profiles of adaptation following the initial familiarization (Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995; Phan, Pytte, & Vicario, 2006). Experiment 1 revealed a relationship between adaptation rates and the improvements in neural discrimination. Thus, it is hypothesized that neural discrimination, induced during initial exposure to a set of songs, is maintained and can be detected during testing 20 hours after that exposure.

To assess this hypothesis, electrophysiological responses to zebra finch songs in NCM were recorded under passive listening conditions. One group of birds were pre-exposed to test songs 20 hours before the electrophysiological recordings; the other group

was presented with other songs that were unrelated to the test songs. Only shuffled presentation sequences were used for both the pre-exposure and the test phase. Neural discrimination metrics were calculated and analyzed as in Experiment 1.

Methods

Subjects

All subjects were as in Experiment 1 (16 naïve adult male zebra finches), except they lived in the general aviary only until the beginning of the experiment, after which they were housed individually in auditory isolation as described below.

Stimuli

Two different stimulus sets were used in this experiment. The experimental stimulus set consisted of the eight songs described in Experiment 1. The control stimulus set included eight other zebra finch songs, also selected from a corpus of unfamiliar songs such that the percent acoustic similarity scores between the control and experimental stimuli ranged from 36% to 65% (Mean ± SEM = 52 ± 1%). The durations of the songs in the control set were between 652 and 840 ms (Mean ± SEM = 781 ± 20 ms). All birds were naïve to the stimuli in both sets in the beginning of the experiment.

Surgery

All surgical procedures were as in Experiment 1, except surgeries were conducted one to three days before the passive auditory exposure phase.

Passive auditory exposure

Birds were divided into two groups of 8 birds based on the stimulus sets to which they would be passively exposed prior to electrophysiological recordings. The eight birds in the pre-exposed group were presented with the experimental stimulus set that was also

used during electrophysiological recordings, whereas the eight birds in the control group were presented with the control stimulus set. Note that the control group was also tested with the experimental stimulus set during electrophysiological recordings. Passive auditory exposures were conducted individually in the walk-in sound attenuation chamber also used for electrophysiological recordings. During auditory exposure, the birds were not head-fixed, but were housed in a cage. Each stimulus was played 200 times in a shuffled sequence at an onset-onset ISI of 6 s from a speaker located 70 cm in front of the cage at an amplitude of 60 dB SPL (A scale) and a sampling frequency of 44.444 kHz. At the end of the auditory exposure, birds were left isolated in the sound attenuation chamber until the beginning of electrophysiological recordings.

Electrophysiology

All electrophysiological procedures were as in Experiment 1.

Stimulus presentation

Electrophysiological recordings of neural responses to experimental stimuli started 20 h ± 15 min from the beginning of the passive auditory exposure. Since the passive auditory exposure lasted for 2 h 40 min, there was ~17 h 20 min from the last stimulus pre-exposure to the beginning of electrophysiological recordings. All birds were tested with the experimental stimulus set with which the pre-exposed group, but not the control group, was presented 20 h earlier. All other stimulus presentation parameters were as in Experiment 1 (25 stimulus repetitions, 6 s ISI), except that only shuffled stimulus presentation sequences were used.

Histology

All histological procedures were as in Experiment 1.

Data analysis

All data analysis procedures were as in Experiment 1.

Results

Multi-unit Responses

There were 145 multi-unit sites in the pre-exposed group and 149 multi-unit sites in the control group that were histologically verified to be in NCM. Preliminary analyses did not show systematic lateral differences, thus the two hemispheres were combined for all subsequent analyses.

Multi-unit Adaptation Rates

To compare the profiles of adaptation between the two exposure conditions in song responses, a mixed ANOVA on percent response magnitudes using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (2 through 25) as a within-subjects variable was conducted. Overall, percent response magnitudes were significantly greater in the control than in the pre-exposed condition ($F(1,292) = 16.68$, $p < 0.001$, **Fig. 27A**). The effect of presentation was also significant ($F(23,6716) = 225.35$, $p < 0.001$, **Fig. 27A**). Post-hoc comparisons revealed that there was a significant decrease in percent response magnitudes from the 1st until the 17th presentation, after which there was no further change. Most importantly, there was a significant interaction between exposure and presentation ($F(23,6716) = 13.81$, $p < 0.001$, **Fig. 27A**). Post-hoc tests indicated no difference between the two exposure groups in corresponding presentations, however the way percent response magnitudes decreased with stimulus presentation differed markedly between the two exposure condition. In the control exposure condition, there was a significant decrease in percent response magnitudes from the 1st until the 18th

presentation, whereas the pre-exposed group showed significant decreases only until the 7^{th} presentation and no consistent change thereafter.

To further examine these differences in detail, a mixed ANOVA was conducted on adaptation rates using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1-6, 6-25) as a within-subjects variable. There was a significant exposure effect ($F(1,292) = 10.41$, $p = 0.001$, **Fig. 27B**), such that the adaptation rates in the pre-exposed condition were more negative than those in the control exposure condition. There was also a significant effect of presentation ($F(1,292) = 223.28$, $p < 0.001$, **Fig. 27B**), indicating more negative adaptation rates for presentations 1-6 than for presentations 6-25. Most importantly, the interaction between exposure and presentation was highly significant ($F(1,292) = 64.90$, $p < 0.001$, **Fig. 27B**). Post-hoc analyses demonstrated that, for presentations 1-6, adaptation rates were significantly more negative in the pre-exposed than in the control condition ($p < 0.001$), whereas this pattern was reversed for presentations 6-25, indicating significantly more negative adaptation rates in the control than in the pre-exposed condition ($p = 0.031$). In both exposure conditions, adaptation rates for presentations 1-6 were significantly more negative compared to those for presentations 6-25 (both $p < 0.001$). Finally, adaptation rates in both exposure conditions for both presentations 1-6 and 6-25 were significantly less than 0 (all $p < 0.001$).

Summary: Taken together, these analyses clearly demonstrated a neural memory for the test songs in birds that were passively exposed to those stimuli 20 hours earlier. Neural responses to songs adapted more rapidly and remained at asymptotic levels in pre-

exposed birds, whereas the typical gradual adaptation profile for novel signals was observed in control birds that were hearing the test songs for the first time.

Multi-unit Between-stimulus Neural Dissimilarities

Similar to the analyses in Experiment 1, the dissimilarities between the temporal profiles of neural responses to different songs were analyzed. A mixed ANOVA on between-stimulus neural dissimilarities using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable revealed a significant effect of exposure ($F(1,292) = 12.35$, $p < 0.001$, **Fig. 28A**), indicating greater between-stimulus neural dissimilarities in the control than in the pre-exposed condition. There was also a significant effect of presentation ($F(24,7008) = 47.50$, $p < 0.001$, **Fig. 28A**). Post-hoc comparisons showed that there was a sharp increase in between-stimulus neural dissimilarities in the first 4 presentations, followed by a more gradual increase until the 15^{th} presentation, and no consistent change afterwards. The interaction between exposure and presentation was also significant ($F(24,7008) = 4.41$, $p < 0.001$, **Fig. 28A**), however post-hoc comparisons did not show a significant difference between the two exposure groups in corresponding presentations and the interaction was driven by unsystematic differences in the way between-stimulus neural dissimilarities changed with presentation in the two exposure conditions.

These differences were further probed via a mixed ANOVA on the slopes of between-stimulus neural dissimilarities with exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. There was neither an effect of exposure ($F(1,292) = 1.52$, $p = 0.218$) nor an interaction between exposure and presentation ($F(1,292) = 1.06$, $p = 0.304$). However, the effect of

presentation was significant ($F(1,292) = 29.27$, $p < 0.001$, **Fig. 28B**), showing that the slopes of between-stimulus neural dissimilarities for presentations 1-6 were greater than those for presentations 6-25. These slopes were significantly greater than 0 for both presentations 1-6 and 6-25 (both $p < 0.001$).

Summary: These findings corroborate results from Experiment 1 showing an increase in the dissimilarities between the temporal profiles of neural responses to different songs as a function of repeated stimulus exposure. These improvements were stronger during the first 6 presentations compared to the more gradual changes during the following stimulus repetitions. Interestingly, across presentation, the temporal profiles of neural responses to different songs were more dissimilar from each other when these stimuli were completely novel compared to when they were tested 20 hours after familiarization.

Multi-unit Within-stimulus Neural Dissimilarities

Next, the dissimilarities between the temporal profiles of neural responses to different presentations of the same songs were analyzed via a mixed ANOVA on within-stimulus neural dissimilarities using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. The within-stimulus neural dissimilarities in the control exposure condition were markedly greater than those in the pre-exposed condition ($F(1,292) = 48.94$, $p < 0.001$, **Fig. 28C**). The effect of presentation was also highly significant ($F(24,7008) = 41.11$, $p < 0.001$, **Fig. 28C**) and post-hoc analyses indicated a significant decrease in within-stimulus neural dissimilarities during the first 6 presentations and no systematic change thereafter. There was also a significant interaction between exposure and presentation ($F(24,7008) =$

2.91, $p < 0.001$, **Fig. 28C**). Post-hoc tests indicated no significant difference between the two exposure conditions in corresponding presentations and the interaction was driven by a sharp decrease in within-stimulus neural dissimilarities in the first 5 presentations in the pre-exposed condition as compared to a more gradual decrease in the first 4 presentations in the control condition.

The slopes of within-stimulus neural dissimilarities were further probed in detail to examine these differences. A mixed ANOVA on the within-stimulus neural dissimilarity slopes using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable revealed a significant effect of exposure ($F(1,292) = 13.35$, $p < 0.001$, **Fig. 28D**), such that the within-stimulus neural dissimilarities in the pre-exposed condition had more negative slopes than did those in the control condition. The effect of presentation was also highly significant ($F(1,292) = 187.81$, $p < 0.001$, **Fig. 28D**), showing more negative slopes for within-stimulus neural dissimilarities for presentations 1-6 than for presentations 6-25. Most importantly, there was a significant interaction between exposure and presentation ($F(1,292) = 7.57$, $p = 0.006$, **Fig. 28D**). Post-hoc analyses demonstrated that, for presentations 1-6, the slopes of within-stimulus neural dissimilarities were significantly more negative in the pre-exposed than in the control conditions ($p < 0.001$), whereas no such difference was observed for presentations 6-25 ($p = 0.958$). In both exposure conditions, within-stimulus neural dissimilarity slopes for presentations 1-6 were significantly more negative compared to those for presentations 6-25 (both $p < 0.001$). For presentations 1-6, the slopes of within-stimulus neural dissimilarities in both exposure conditions were

significantly less than 0 (both $p < 0.001$), however, for presentations 6-25, they were together not different from 0 ($p = 0.323$).

Summary: Taken together, these analyses revealed patterns mostly in line with the results found in Experiment 1 for song responses. The temporal profiles of neural responses to different presentations of the same songs became less and less dissimilar from each other during the first 6 presentations and did not change afterwards. These initial changes happened more rapidly in birds that had been passively familiarized with these same songs compared to birds that were naïve to them. In addition, across presentations, the temporal profiles of neural responses to the same songs were more dissimilar from each other when these stimuli were completely novel compared to when they were tested 20 hours after familiarization.

Multi-unit Correct Decoding Probabilities

The correct neural decoding probabilities across time points along the stimulus duration and stimulus presentations for songs are shown in **Figures 29A** and **29B** for the pre-exposed and the control exposure condition, respectively. Correct neural decoding probabilities at the 500 ms time point across stimulus presentations were analyzed as in Experiment 1 via a mixed ANOVA with exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. Overall, probabilities were significantly greater in the pre-exposed than in the control condition ($F(1,292) = 35.20$, $p < 0.001$, **Fig. 29C**). The effect of presentation was also highly significant ($F(24,7008) = 14.77$, $p < 0.001$, **Fig. 29C**), such that there was a sharp increase in probabilities during the first 6 presentations and no systematic change afterwards. There was also a significant interaction between exposure and presentation

($F(24,7008) = 4.28$, $p < 0.001$, **Fig. 29C**). Post-hoc comparisons revealed that, in the pre-exposed condition, probabilities in the first presentation was significantly lower than those in most of the other presentations, whereas, in the control exposure condition, probabilities increased more gradually in the first 5 presentations.

These difference were further examined via a mixed ANOVA on the slopes of probabilities using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. There was a significant effect of exposure ($F(1,292) = 23.24$, $p < 0.001$, **Fig. 29D**), indicating grater slopes for probabilities in the control than in the pre-exposed group. The effect of presentation was also highly significant ($F(1,292) = 70.15$, $p < 0.001$, **Fig. 29D**), such that the slopes of probabilities for presentations 1-6 were greater than those for presentations 6-25. Most importantly, there was a significant interaction between exposure and presentation ($F(1,292) = 19.16$, $p < 0.001$, **Fig. 29D**). For presentations 1-6, the probabilities in the control condition had significant greater slopes than did those in the pre-exposed condition ($p < 0.001$), whereas there was no difference between the two exposure conditions for presentations 6-25 ($p = 0.989$). Looking at the interaction from the other perspective, the slopes of probabilities were significantly greater for presentations 1-6 than for presentations 6-25 in both the pre-exposed ($p = 0.026$) and the control condition ($p < 0.001$). The slopes of probabilities for presentations 1-6 were significantly greater than 0 in both the pre-exposed and the control condition (both $p < 0.001$). On the other hand, the slopes of probabilities for presentations 6-25 in the two exposure groups together were not significantly different from 0 ($p = 0.197$).

Summary: These analyses strongly supported the findings in Experiment 1 for song responses and showed that the accuracy of neural decoding using the temporal profiles of neural responses improved with repeated stimulus presentations. These improvements occurred only during the first 6 presentations and did not continue afterwards. Of utmost importance, the rapid gains in neural decoding of songs with repeated stimulus exposure lasted for at least 20 hours as indicated by better neural decoding performance for the same songs when they were previously familiarized as compared to when they were heard for the first time. As a result of already high neural decoding accuracy in the pre-exposed condition, rapid improvements were more pronounced when the songs were novel as compared to when they were familiar.

Multi-unit Correct Decoding Latencies

Latencies along the stimulus duration to reach the correct neural decoding probability level of 0.75 in song responses were analyzed as in Experiment 1 via a mixed ANOVA with exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. The effect of exposure was highly significant ($F(1,292) = 33.55$, $p < 0.001$, **Fig. 29E**), showing shorter latencies in the pre-exposed than in the control condition. The effect of presentation was also significant ($F(24,7008) = 15.53$, $p < 0.001$, **Fig. 29E**) and post-hoc tests revealed that there was a significant decrease in latencies during the first 6 presentations and no systematic trend afterwards. There was also a significant interaction between exposure and presentation ($F(24,7008) = 2.87$, $p < 0.001$, **Fig. 29E**). Post-hoc comparisons indicated that there was a clear decrease in latencies during the first 6 presentations in the

control condition, whereas, in the pre-exposed condition, latencies decreased more gradually in the first 5 presentations.

These difference were further probed via a mixed ANOVA on the slopes of latencies using exposure (Pre-exposed, Control) as a between-subjects variable and presentation (1 through 25) as a within-subjects variable. The slopes of latencies were significantly more negative in the control than in pre-exposed condition ($F(1,292) = 8.58$, $p = 0.004$, **Fig. 29F**) and also for presentations 1-6 than for presentations 6-25 ($F(1,292) = 58.03$, $p < 0.001$, **Fig. 29F**). Most importantly, there was a significant interaction between exposure and presentation ($F(1,292) = 9.14$, $p = 0.003$, **Fig. 29F**). Post-hoc analyses showed that, for presentations 1-6, the slopes of latencies were significantly more negative in the control than in the pre-exposed condition ($p < 0.001$), whereas no difference was observed between the two exposure conditions for presentations 6-25 ($p = 0.999$). The slopes of latencies were significantly more negative for presentations 1-6 than for presentations 6-25 in both the pre-exposed ($p < 0.001$) and the control condition ($p = 0.007$). Finally, the slopes of latencies for presentations 1-6 were significantly less than 0 in both the pre-exposed and the control condition (both $p < 0.001$). On the other hand, the slopes of latencies for presentations 6-25 in the two exposure groups together were not significantly different from 0 ($p = 0.116$).

Summary: These findings were in line with the results obtained in Experiment 1 for song responses. Decoding of songs using the temporal profiles of neural responses reached a given confidence level sooner and sooner with stimulus exposure during the first 6 presentations but did not change during subsequence repetitions. The long-term effects of passive familiarization on neural decoding accuracies were paralleled in

latencies, showing earlier correct neural decoding latencies for the same songs when they were previously familiarized as compared to when they were heard for the first time. Again, due to already low levels of neural decoding latencies in the pre-exposed condition, rapid changes in latencies were more pronounced when the songs were novel as compared to when they were familiar.

Multi-unit Mutual Information

Finally, mutual information between the temporal profiles of neural responses and songs were analyzed across time points along the stimulus duration via a mixed ANOVA with the between-subjects variable exposure (Pre-exposed, Control) and the within-subjects variable bin (1 to 75). Across bins, mutual information was significantly greater in the pre-exposed than in the control exposure condition ($F(1,292) = 50.00$, $p < 0.001$, **Fig. 30**). The effect of bin was also significant ($F(74,21608) = 3993.78$, $p < 0.001$, **Fig. 30**), such that mutual information significantly increased from the 1st until the 66th bin (660 ms), after which there was no change. Most importantly, there was a significant interaction between sequence and bin ($F(74,21608) = 22.03$, $p < 0.001$, **Fig. 30**). Planned comparisons indicated that mutual information was significantly greater in the pre-exposed than in the control exposure condition starting from the 4th bin until the end of the stimulus period (40-750 ms).

Summary: Analysis of mutual information corroborated the differences in neural decoding probabilities of the pre-exposed and control exposure conditions and showed that the temporal profiles of neural responses were more informative about songs when those songs were passively familiarized 20 hours earlier as compared to when they were

heard for the first time. These differences emerged as early as 40 ms after the stimulus onset.

Single-unit Responses

A total of 52 neurons (27 narrow and 25 wide spike) were recorded in the pre-exposed condition; a total of 63 neurons (32 narrow and 31 wide spike) were recorded in the control condition. A chi-square analysis indicated that there was no significant difference between the frequencies of the two neuron types in the two exposure conditions ($X^2 = 0.08$, $p = 0.772$).

Single-unit Response Properties

Firing rates of narrow spike neurons during silent baseline conditions were significantly greater than those of wide spike neurons ($z = 2.65$, $p = 0.008$, **Fig. 31A**). There was no difference between the pre-exposed and the control condition in baseline firing rates ($z = 2.23$, $p = 0.026$). The analysis of the more detailed pairwise comparisons between the two neuron types and the two exposure conditions revealed that, in the control condition, narrow spike neurons had significantly higher baseline firing rates than did wide spike neurons ($z = 3.33$, $p < 0.001$, **Fig. 31A**), whereas, in the pre-exposed condition, there was no difference between the baseline firing rates of the two neuron types ($z = 0.39$, $p = 0.694$). In addition, the baseline firing rates of wide spike neurons were significantly higher in the pre-exposed than in the control condition ($z = 3.40$, $p < 0.001$, **Fig. 31A**), however, no such difference was found in the baseline firing rates of narrow spike neurons ($z = 0.26$, $p = 0.796$).

The magnitude of stimulus-driven responses differed markedly between the two neuron types such that narrow spike neurons had significantly greater response

magnitudes than did wide spike neurons ($z = 5.53$, $p < 0.001$, **Fig. 31B**). There was no difference between the pre-exposed and the control condition in response magnitudes ($z = 0.13$, $p = 0.897$). Neuron type and exposure did not seem to interact as revealed by similarly greater response magnitude for narrow than for wide spike neurons in both the pre-exposed ($z = 3.82$, $p < 0.001$) and the control condition ($z = 3.91$, $p < 0.001$, **Fig. 31B**). The response magnitudes in the two exposure conditions did not differ from each other either for narrow ($z = 0.81$, $p = 0.420$) or for wide spike neurons ($z = 1.46$, $p = 0.145$).

The analysis of adaptation rates for presentations 1-6 revealed that wide spike neurons had significantly more negative adaptation rates than did narrow spike neurons ($z = 4.45$, $p < 0.001$, **Fig. 31C**). No difference between the two exposure conditions was found ($z = 2.13$, $p = 0.033$). The analysis of the interaction between neuron type and exposure indicated that the adaptation rates of wide spike neurons were significantly more negative than those of narrow spike neurons in the pre-exposed ($z = 4.23$, $p < 0.001$, **Fig. 31C**), but not in the control condition ($z = 2.42$, $p = 0.016$). There was no significant difference between the two exposure conditions either for narrow ($z = 1.05$, $p = 0.294$) or for wide spike neurons ($z = 2.55$, $p = 0.011$). The adaptation rates of wide spike neurons were significantly less than 0 ($z = 5.08$, $p < 0.001$, **Fig. 31C**), whereas those of narrow spike neurons were not different from 0 ($z = 1.79$, $p = 0.074$).

The difference between the adaptation rates of the two neuron types for presentations 6-25 was highly significant indicating more negative adaptation rates for wide than for narrow spike neurons ($z = 3.62$, $p < 0.001$, **Fig. 31D**). The adaptation rates for presentations 6-25 did not differ between the two sequences ($z = 0.06$, $p = 0.951$). The

analysis of the interaction between neuron type and exposure showed that wide spike neurons had significantly more negative adaptation rates than did narrow spike neurons in the pre-exposed ($z = 2.93$, $p = 0.004$, **Fig. 31D**), but not in the control condition ($z = 2.20$, $p = 0.028$). The adaptation rates in the two exposure conditions did not differ from each other either for narrow ($z = 0.20$, $p = 0.839$) or for wide spike neurons ($z = 0.22$, $p = 0.827$). Finally, the adaptation rates of wide spike neurons were significantly less than 0 for presentations 6-25 ($z = 4.55$, $p < 0.001$, **Fig. 31D**), whereas those of narrow spike neurons were not different from 0 ($z = 1.58$, $p = 0.113$).

Summary: Taken together, these analyses supported the results found in Experiment 1 in that narrow spike neurons fired more strongly both during silent conditions and in response to songs than did wide spike neurons. With repeated stimulus presentation, wide spike neurons underwent gradual adaptation, but narrow spike neurons did not adapt. Surprisingly, there was no difference in adaptation rates between the pre-exposed and the control condition. This was not in line with the findings in multi-unit responses showing faster adaptation in birds that were passively familiarized with the same songs compared to birds that heard those songs for the first time.

Single-unit Between-stimulus Neural Dissimilarities

The dissimilarities between the temporal profiles of neural responses to different songs in single-unit responses were analyzed similar to the multi-units. Between-stimulus neural dissimilarities were not significantly different between the two neuron types ($z = 1.67$, $p = 0.095$) or the two exposure conditions ($z = 0.68$, $p = 0.497$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and exposure conditions (all $p > 0.084$).

The analysis of the slopes of between-stimulus neural dissimilarities for presentations 1-6 did not show a neuron type ($z = 1.48$, $p = 0.138$), an exposure ($z = 0.56$, $p = 0.578$), or an interaction effect (all $p > 0.113$). Similarly, the slopes of between-stimulus neural dissimilarities for presentations 6-25 were not significantly different between the two neuron types ($z = 0.39$, $p = 0.695$) or the two exposure conditions ($z = 1.74$, $p = 0.083$). There was also no interaction between neuron type and exposure (all $p > 0.013$). The slopes of between-stimulus neural dissimilarities were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.086$).

Summary: Unlike for multi-unit responses, the dissimilarities between the temporal profiles of neural responses to different songs in single-unit responses did not change with stimulus repetition and were not different between the two neuron types or the two exposure conditions.

Single-unit Within-stimulus Neural Dissimilarities

The dissimilarities between the temporal profiles of neural responses to different presentations of the same songs in single-unit responses were analyzed similarly. Within-stimulus neural dissimilarities did not significantly differ between the two neuron types ($z = 2.47$, $p = 0.013$) or the two exposure conditions ($z = 1.06$, $p = 0.288$). However, there was an indication of an interaction between neuron type and exposure such that, in the pre-exposed condition, wide spike neurons had significantly greater within-stimulus neural dissimilarities compared to narrow spike neurons ($z = 2.90$, $p = 0.004$, **Fig. 32A**), whereas, in the control condition, there was no difference between the two neuron types ($z = 0.65$, $p = 0.518$). There was no significant difference between the two exposure

conditions either for narrow ($z = 1.86$, $p = 0.063$) or for wide spike neurons ($z = 0.55$, $p = 0.581$).

The slopes of within-stimulus neural dissimilarities for presentations 1-6 were also not different between the two neuron types ($z = 0.21$, $p = 0.832$) or the two exposure conditions ($z = 1.20$, $p = 0.231$). There was also no indication of an interaction between neuron type and exposure (all $p > 0.224$). The analysis of the slopes of within-stimulus neural dissimilarities for presentations 6-25 also did not reveal a neuron type ($z = 1.49$, $p = 0.137$), an exposure ($z = 0.07$, $p = 0.942$), or an interaction effect (all $p > 0.184$). The slopes of within-stimulus neural dissimilarities were significantly less than 0 for presentations 1-6 ($z = 5.50$, $p < 0.001$, **Fig. 32B**), but not for presentations 6-25 ($z = 1.33$, $p = 0.182$).

Summary: These findings were mostly in line with the patterns observed in between-stimulus neural dissimilarities, showing no robust differences between the temporal profiles of neural responses to different presentations of the same songs between the two neuron types or the two exposure conditions in single-unit responses. Neural responses to the same songs became more similar to each other during the first 6 presentations, but did not change thereafter.

Single-unit Correct Decoding Probabilities

Next, the correct neural decoding probabilities across stimulus presentations in single-unit responses to songs were analyzed as in multi-unit responses There was a marked difference between the neural decoding accuracies of the two neuron types such that narrow spike neurons had significantly higher probabilities than did wide spike neurons ($z = 5.18$, $p < 0.001$, **Fig. 33A**). However, correct decoding probabilities were

not significantly different between the two exposure conditions ($z = 0.89$, $p = 0.375$). There was no interaction between neuron type and exposure as indicated by similarly higher probabilities for narrow than for wide spike neurons in both the pre-exposed ($z = 4.29$, $p < 0.001$) and the control condition ($z = 3.03$, $p = 0.002$, **Fig. 33A**). Correct decoding probabilities were not significantly different between the two exposure conditions either for narrow or for wide spike neurons (both $p > 0.097$).

The analysis of the slopes of correct decoding probabilities for presentations 1-6 did not show any difference between the two neuron types ($z = 0.74$, $p = 0.459$) or the two exposure conditions ($z = 0.71$, $p = 0.475$). There was also no interaction as indicated by lack of significant differences in pairwise comparisons of neuron type and exposure conditions (all $p > 0.118$). Similarly, the slopes of probabilities for presentations 6-25 were also not different between the two neuron types ($z = 0.39$, $p = 0.697$) or the two exposure conditions ($z = 1.33$, $p = 0.184$). There was also no interaction (all $p > 0.323$). The slopes of correct decoding probabilities were significantly greater than 0 for presentations 1-6 ($z = 3.31$, $p < 0.001$, **Fig. 33B**), whereas, for presentations 6-25, they were not significantly different from 0 ($z = 0.50$, $p = 0.615$).

Summary: These analyses revealed that, similar to the findings in Experiment 1, neural decoding of songs was more accurate using the temporal profiles of responses of narrow as compared to wide spike neurons. Furthermore, in the two neuron types together, neural decoding accuracy increased during the first 6 stimulus presentations, but did not change during the following repetitions. However, unlike in multi-unit responses, neural decoding accuracies were not different in single-units when the same songs were passively familiarized 20 hours earlier or when they were completely novel. Thus, in

parallel to the lack of evidence of neural memory in single-units for familiar songs, there was no evidence of long-lasting improvements in neural decoding performance for previously familiarized songs in these single-units.

Single-unit Correct Decoding Latencies

Correct neural decoding latencies were also analyzed in single-unit responses to songs as in multi-unit responses. Latencies were significantly longer for wide than for narrow spike neurons ($z = 3.19$, $p = 0.001$, **Fig. 33C**). However, there was no difference between the two exposure conditions ($z = 0.91$, $p = 0.364$). The detailed pairwise comparisons indicated an interaction between neuron type and exposure such that the correct decoding latencies were significantly longer for wide than for narrow spike neurons in the pre-exposed ($z = 2.80$, $p = 0.005$, **Fig. 33C**), but not in the control exposure condition ($z = 1.71$, $p = 0.087$). Latencies were not different between the two exposure conditions either for narrow or wide spike neurons (both $p > 0.121$).

The analysis of the slopes of correct decoding latencies for presentations 1-6 did not show a neuron type ($z = 1.48$, $p = 0.138$), an exposure ($z = 0.05$, $p = 0.960$), or an interaction effect (all $p > 0.031$). Similarly, the slopes of latencies for presentations 6-25 were also not significantly different between the two neuron types ($z = 1.37$, $p = 0.170$) or the two exposure conditions ($z = 1.86$, $p = 0.063$). There was also no indication of an interaction between neuron type and exposure (all $p > 0.069$). The slopes of correct decoding latencies were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.482$).

Summary: The latencies along the stimulus duration to reach a given confidence level for neural decoding of songs did not change with change with stimulus repetition

and were not different between the two exposure conditions. The only difference in latencies was that correct neural decoding was attained sooner for narrow than for wide spike neurons only in the pre-exposed condition.

Single-unit Mutual Information

Finally, the mutual information between songs and neural response profiles were analyzed in single-unit responses to songs. Overall, the responses of narrow spike neurons were markedly more informative about songs compared to the responses of wide spike neurons ($z = 5.17$, $p < 0.001$, **Fig. 34**). However, there was no significant difference between mutual information in the pre-exposed and the control exposure condition ($z = 0.99$, $p = 0.323$). The more detailed analysis of pairwise differences also revealed similar results. Narrow spike neurons had significantly higher mutual information in both the pre-exposed ($z = 4.39$, $p < 0.001$) and the control condition ($z = 2.87$, $p = 0.004$, **Fig. 34**). There was no significant difference between the mutual information estimations in the two exposure conditions either for narrow ($z = 1.66$, $p = 0.097$) or for wide spike neurons ($z = 0.72$, $p = 0.473$).

Summary: Once again, these analyses were in line with findings in Experiment 1 in that the temporal profiles of narrow spike neuron responses were more informative about songs than those of wide spike neurons. However, unlike in multi-unit responses, mutual information in single-unit responses did not differ whether the same songs were completely novel or familiarized 20 hours earlier.

Discussion

The results of this experiment clearly demonstrate that passive familiarization with novel natural vocalizations leads to long-lasting changes in the temporal profiles of

multi-unit neural responses in the zebra finch NCM that serve improved stimulus discrimination and recognition at later time-points. Both the correct decoding probabilities and mutual information were greater in the group that was pre-exposed with the test stimuli 20 hours earlier as compared to the group that were completely naïve to those stimuli. Furthermore, a given level of neural decoding accuracy was reached sooner along the stimulus duration in the pre-exposed than in the control condition. The analysis of neural dissimilarities for further investigation of the dynamics that brought about the improvements in neural decoding with pre-exposure revealed that, although the between-stimulus neural dissimilarities were slightly greater in the control condition, the within-stimulus neural dissimilarities were dramatically reduced in the pre-exposed compared to in the control exposure condition. This suggests that previous passive familiarization increased neural decoding accuracy and mutual information mainly by increasing the consistency of the temporal profiles of neural responses to repetitions of any given stimulus. Multi-unit responses also showed a neural memory for the test songs as shown by an initial more rapid reduction in firing rates during the first 6 presentations followed by a relatively flatter adaptation profile during the subsequent stimulus repetitions in the pre-exposed than in the control condition. These differences in in the adaptation profiles of familiar and novel sounds, especially after the first 6 presentations, are typical markers of neural memory as also shown in previous studies (Chew et al., 1995; Phan et al., 2006).

Contrary to findings in multi-unit responses, single-unit responses did not show these effects. To begin with, there was no difference between the adaptation rates of single-units in the pre-exposed and the control conditions, suggesting no indication of

neural memory for previously heard stimuli. All of the studies that investigated long-term neural memory in NCM in the past based their analyses exclusively on multi-unit, but not on single-unit responses. Thus, it is not clear whether the lack of neural memory effects in single-unit responses found in the present experiment is due to methodological shortcomings or represents an actual difference between the multi-unit and single-unit responses. The former possibility appears more likely because these findings were based on a limited number of single-units (n = 52 for pre-exposed; n = 63 for control). The yield of single-units was substantially lower in this experiment than in Experiment 1, for reasons that are not understood. None of the neural decoding metrics in single-unit responses differed between the pre-exposed and control conditions. However, this matches the lack of neural memory effects between the two exposure conditions. Taken together, more data is necessary to derive conclusive results regarding neural memory and long-term effects of passive exposure on neural discrimination in single-unit responses.

The rapid effects of passive exposure on neural discrimination metrics observed in Experiment 1 were largely replicated in the multi-unit response profiles of the control condition, for which the test songs were completely novel in the beginning of neural recordings. Decoding of songs using the temporal profiles of neural responses improved in accuracy during the first 6 stimulus presentations. These improvements were smaller in the pre-exposed condition because of already high levels of neural decoding accuracy in the beginning of stimulus presentations due to the long-term effect of prior familiarization. However, unlike the findings in Experiment 1, neural decoding accuracy in both exposure conditions did not increase after the first 6 stimulus presentations.

Between-stimulus neural dissimilarities in multi-unit responses generally increased with stimulus repetition, while within-stimulus neural dissimilarities decreased only during the first 6 presentations and did not change thereafter. These findings suggest that the enhancement in neural decoding accuracy during the first 6 presentations in multi-unit responses can be explained by the simultaneous convergence of the temporal profiles of responses to the same songs and differentiation of neural response profiles to different songs with repeated exposure. In addition to these effects, the latencies along the stimulus duration to reach a given level of confidence for neural decoding accuracy also decreased with stimulus presentation in multi-unit responses.

Similar to what was observed in Experiment 1, wide spike neurons had lower baseline firing rates and stimulus-driven response magnitudes, but underwent stronger adaptation as compared to narrow spike neurons. Moreover, the accuracy of neural decoding and mutual information were higher for narrow than for wide spike neurons. None of the other neural discrimination metrics or their trends as a function of repeated stimulus presentation was consistently different between the two neuron types. However, the accuracy of decoding of songs using the temporal profiles of the responses of the two neuron types increased during the first 6 stimulus presentations. This, similar to the findings in Experiment 1, indicates that the narrow and wide spike neurons in NCM contribute similarly to the improvements in neural decoding accuracy with passive exposure.

In summary, the findings in multi-unit responses clearly showed that rapid improvements in neural discrimination with passive familiarization were long-lasting and could affect the processing of the same signals 20 hours after the initial encounter. Thus,

Experiments 1 and 2 established that repeated stimulus presentation leads to more accurate neural decoding of complex natural vocalizations that could be useful in ongoing sensory processing. One crucial detail that was left unexamined in these experiments was the relationship between the acoustic similarity of the stimuli and the degree of neural differentiation with repeated stimulus exposure. Thus, this factor was investigated in the next experiment using synthesized vocalizations that parametrically varied along a single dimension.

EXPERIMENT 3: THE RELATIONSHIP BETWEEN ACOUSTIC SIMILARITY AND NEURAL DISCRIMINATION OF VOCAL SIGNALS

The purpose of this experiment was to examine the relationship between acoustic similarity and the direction and degree of change in neural discrimination with passive familiarization. Although natural songs and calls used in Experiments 1 and 2 allow us to investigate the full complexity of neural responses to ecologically-relevant, complex sounds, they make it impossible to control the acoustic similarities between different stimuli. Synthesized zebra finch calls, generated using natural stimulus parameters, have been used widely and produce similar behavioral responses to their natural counterparts (Vicario, Naqvi, & Raksin, 2001; Vignal, Andru, & Mathevon, 2005). In addition, zebra finches are sensitive in their behavioral responses to variations along different stimulus parameters, which validates the use of synthesized vocalizations – whose dissimilarity can be accurately quantified - in assessing the effects of acoustic similarity on neural discrimination. Experiment 1 demonstrated that passive exposure improves neural discrimination and decoding of natural calls. Thus, the hypothesis in this experiment was

that passive familiarization improves neural discrimination of synthesized calls and this improvement is greater for stimuli that are highly dissimilar from each other.

To test this hypothesis, a varied set of male long calls was first analyzed to extract the natural distribution of time, frequency, and amplitude parameters that could then be used to synthesize calls using mathematical functions. From these analyses, the duration of the initial frequency modulation (FM) was selected to create a set of synthesized calls that varied only in this parameter (**Fig. 35**). Electrophysiological responses to these stimuli were recorded in NCM under passive listening conditions. All birds received a shuffled stimulus presentation sequence. Neural discrimination metrics were calculated and analyzed as in Experiment 1.

Methods

Subjects

All subjects were as in Experiment 1, except that this experiment used 8 naïve adult male zebra finches.

Stimuli

From a large corpus of male long calls, sixteen calls (each from a different bird) consisting of an initial downward frequency modulation (FM) followed by a relatively unmodulated section were selected (**Fig. 35A**). The fundamental frequency contour of each of these calls was detected from the spectrogram and was broken down into the downward FM and unmodulated sections. Some calls also had a brief upward modulation between these two sections. The initial and end frequencies, as well as the duration of each of these sections were extracted. To determine an ethologically-relevant dimension for synthesis of calls using mathematical functions, the relationships between the derived

parameters were examined. These analyses revealed that, the rate of the initial FM was not fixed, that is, the duration of the FM was not correlated with the amount of decrease in frequencies ($r(14) = 0.19$, $p = 0.490$). Thus, the start and the end frequencies were fixed and the duration of the initial FM was varied independently as the critical dimension. The durations of the FM ranged between 18 and 59 ms in natural calls. To keep the number of stimuli the same across experiments, 8 durations separated by 6-ms intervals were used to synthesize 8 different calls. Thus, the FM durations of the 8 synthesized calls were 18, 24, 30, 36, 42, 48, 54, and 60 ms (**Fig. 35B**). The same average start and end frequencies of 1500 and 600 Hz, respectively, were used for synthesis of all calls. For the downward FM section, the start and end frequencies and a given duration were fitted by an exponential function. For the relatively stable, unmodulated section, the frequency contours of these sections in natural calls were each fitted by a power function and the fit that produced the parameters closest to the average values across calls was used for synthesis of all calls. To keep the entire duration of all stimuli equal to the average natural call duration of 150 ms, the stable section was trimmed from its beginning according to the duration of the specific initial FM used to synthesize that call. The two sections were connected with a 10-ms linear interpolation of an upward FM. These steps produced a fundamental frequency contour, from which 25 harmonics were generated. For harmonic emphasis, the power spectrum of the stable sections of all natural calls were calculated. Harmonic peak amplitudes were detected, linearly interpolated, and averaged across all natural calls. The resulting average power spectrum was used to calculate the frequency-dependent amplitude function for each harmonic separately. Finally, the amplitude envelopes of all natural calls were extracted, smoothed,

and averaged to generate an average amplitude envelope. All frequency contours and amplitude functions were brought together to generate the complex sound and the average of the amplitude envelopes of all natural calls was applied to produce the final synthesized call. This general approach to synthesizing calls was similar to that described in Vicario, Naqvi, and Raksin (2001).

Surgery

All surgical procedures were as in Experiment 1.

Electrophysiology

All electrophysiological procedures were as in Experiment 1.

Stimulus presentation

All stimulus presentation parameters were as in Experiment 1 (25 stimulus repetitions, 6 s ISI), except only shuffled stimulus presentation sequences were used.

Histology

All histological procedures were as in Experiment 1.

Data analysis

All data analysis procedures were as in Experiment 1, except both neural dissimilarities and neural classifications were analyzed in more detail with respect to pairwise acoustic similarities among different signals.

Neural dissimilarity. Neural dissimilarities were calculated as in Experiment 1. Then, for each multi-unit or single-unit, a neural dissimilarity matrix was constructed to analyze the neural dissimilarities for all possible synthesized call pairs. To assess how acoustic similarity was related to neural dissimilarity, the trends of the diagonals of this matrix across FM durations were quantified using linear regression and normalization

methods as described for adaptation rates in Experiment 1. To analyze the relationship between acoustic similarity and the changes in neural dissimilarities with repeated stimulus presentation, two more such matrices were constructed and analyzed as described. However, instead of the average neural dissimilarities across stimulus repetitions, one of these matrices contained the slopes of neural dissimilarities for presentations 1-6 and the other contained the slopes of neural dissimilarities for presentations 6-25.

Correct classification ratio. Neural classifications resulting from the decoding algorithm were used to construct confusion matrices, which show the frequencies for all true and neurally classified stimulus identity pairs. For each unit, these frequencies were used to calculate the ratio of correct classifications between a given stimulus pair as

$$Correct\ Classification\ Ratio = \frac{fq(s_A, r_A) + fq(s_B, r_B)}{fq(s_A, r_A) + fq(s_A, r_B) + fq(s_B, r_A) + fq(s_B, r_B)}$$

where fq is the frequency, s and r are the true and the neurally classified stimulus identities, respectively, and A and B are the particular stimuli used for the calculation. These ratios were calculated separately for all stimulus pairs that differed by 6 ms and 12 ms in their FM durations. If two stimuli are completely confused, then this calculation is at 0.5 chance level. However, if this ratio is significantly greater than 0.5, then the two stimuli are correctly classified and not confused.

Results

Multi-unit Responses

A total of 174 multi-unit sites were histologically verified to be in NCM and were included in statistical analyses. Due to the lack of systematic lateral differences, the two hemispheres were combined for all subsequent analyses.

Multi-unit Adaptation Rates

To test whether there was a difference in the rates of adaptation among the synthesized calls, a repeated-measures ANOVA with stimulus (1 through 8) and presentation (1-6, 6-25) as within-subjects variables was conducted on adaptation rates. As expected, adaptation rates were significantly more negative for presentations 1-6 than for presentations 6-25 ($F(1,173) = 74.68$, $p < 0.001$). There were also significant differences among stimuli ($F(7,1211) = 7.61$, $p < 0.001$, **Fig. 36**). Post-hoc comparisons indicated that the adaptation rates of the synthesized calls with the shortest two FMs (18 and 24 ms) were significantly higher (less negative) than the adaptation rates of all the other synthesized calls, except for the one with the 30 ms FM duration, the adaptation rate of which was not different from any of the other synthesized calls. There was also a significant interaction between stimulus and presentation ($F(7,1211) = 8.92$, $p < 0.001$, **Fig. 36**). Post-hoc analyses revealed that the difference between the adaptation rates among the synthesized calls described above was only true for presentations 1-6, whereas no difference in adaptation rates was observed for presentations 6-25.

Summary: This analysis showed that, although multi-unit responses to all synthesized calls underwent adaptation, the rates of these changes were stronger for the calls with longer FMs.

Multi-unit Neural Dissimilarities

Average neural dissimilarities between pairs of synthesized calls are shown in **Figure 37A**. Each entry in the main diagonal of this matrix shows the average neural dissimilarity across different repetitions of a particular synthesized call. That is, the main diagonal shows within-stimulus neural dissimilarities. The entries directly above the main

diagonal, which is hereafter referred to as diagonal 1, indicate the average neural dissimilarities between synthesized call pairs that are separated by 6 ms in their FM durations. Similarly, higher-order diagonals, from diagonals 2 through 6, contain the average neural dissimilarities between synthesized call pairs that are acoustically more and more dissimilar from each other. Thus, off-main-diagonal entries indicate between-stimulus neural dissimilarities between pairs of synthesized calls with different degrees of acoustic similarity. The entries in the anti-diagonal of this matrix contains the average neural dissimilarities for all synthesized call pairs that share the same acoustic similarity levels. The slopes of all of these diagonals as functions of FM durations were analyzed to examine the relationships between acoustic parameters and the dissimilarities between the temporal profiles of neural responses. Across multi-units, the slopes of the main diagonals of these matrices were significantly lower than 0 ($t(173) = 8.45$, $p < 0.001$, **Fig. 37B**), suggesting that within-stimulus neural dissimilarities decreased as the FM of synthesized calls increased. Furthermore, the slopes of all the other higher-order diagonals, from diagonals 1 through 6, were also significantly lower than zero (all $p < 0.002$, **Fig. 37B**). This means that there was an overall asymmetry in between-stimulus neural dissimilarities such that, for any given acoustic similarity level, neural dissimilarities tended to be greater for synthesized calls with shorter FMs. The slopes of the anti-diagonals of these matrices were significantly greater than 0 ($t(173) = 12.13$, $p < 0.001$, **Fig. 37C**), indicating that neural dissimilarities increased as the acoustic dissimilarity increased.

Next, the relationship of the changes in neural dissimilarities during presentations 1-6 to the FMs of synthesized calls was analyzed. The average of the slopes of neural

dissimilarities for presentations 1-6 between each pair of synthesized calls are shown in **Figure 37D**. Across multi-units, none of the slopes of diagonals 1 through 6 of these matrices were significantly different from 0 (all $p > 0.288$), indicating that, for any given acoustic similarity level, the changes in neural dissimilarities during presentations 1-6 were not related to the FM duration of synthesized calls. Moreover, the slopes of the anti-diagonals of these matrices also did not significantly differ from 0 ($t(173) = 1.11$, $p = 0.269$), which shows that the rates of changes in neural dissimilarities during presentations 1-6 were not related to the acoustic similarities between the synthesized calls. Surprisingly, for presentations 1-6, the slopes of neural dissimilarities for all acoustic similarity levels along the anti-diagonal were significantly less than 0 (all $p < 0.030$), except for the acoustically most dissimilar stimulus pair ($p = 0.403$), indicating that the between-stimulus neural dissimilarities generally decreased during presentations 1-6.

Similar to the above analysis, the relationship of the changes in neural dissimilarities to the FMs of synthesized calls were analyzed for presentations 6-25 (**Fig. 37E**). These analyses also revealed that, across multi-units, none of the slopes of diagonals 1 through 6 of these matrices significantly differed from 0 (all $p > 0.052$), indicating that the changes in neural dissimilarities during presentations 6-25 were not related to the FM duration of the synthesized calls for any given acoustic similarity level. The slopes of the anti-diagonals of these matrices also did not significantly differ from 0 ($t(173) = 0.17$, $p = 0.869$), suggesting that the magnitude of changes in neural dissimilarities were not related to the acoustic similarities between the synthesized calls during presentations 6-25. Unlike for presentations 1-6, the slopes of neural

dissimilarities for all acoustic similarity levels along the anti-diagonal were significantly greater than 0 (all $p < 0.034$), indicating that the between-stimulus neural dissimilarities generally increased during presentations 6-25.

Summary: Taken together, these analyses revealed interesting results. The temporal profiles of neural responses were more dissimilar from each other for synthesized call pairs with shorter FM durations. In addition, as expected, neural responses were more dissimilar for synthesized call pairs that were acoustically more dissimilar from each other. Surprisingly, the temporal profiles of neural responses to different synthesized calls became less dissimilar from each other during the first 6 stimulus presentations, but then reversed and started to become more dissimilar from each other with further stimulus repetition. In contrast to predictions, the rates of these changes did not differ as a function of the acoustic similarities between synthesized call pairs.

Multi-unit Correct Decoding Probabilities

The correct neural decoding probabilities across time points along the stimulus duration and stimulus presentations for synthesized calls are shown in **Figure 38A**. A repeated-measures ANOVA on correct neural decoding probabilities using presentation (1 through 25) as a within-subjects variable revealed a significant effect of presentation ($F(24,4152) = 2.10$, $p = 0.001$, **Fig. 38B**). However, post-hoc comparisons did not reveal any significant difference except that probabilities in the first presentation was greater than those for presentations 11, 21, 22, 24, and 25. To examine these differences further in detail, the slopes of correct decoding probabilities for presentations 1-6 and 6-25 were compared. The slopes of probabilities were significantly greater for presentations 1-6

than for presentations 6-25 ($t(173) = 3.33$, $p = 0.001$, **Fig. 38C**). Furthermore, the slopes of between-stimulus neural dissimilarities were significantly greater than 0 for presentations 1-6 ($t(173) = 3.65$, $p < 0.001$), whereas, for presentations 6-25, they did not significantly differ from 0 ($t(173) = 1.32$, $p = 0.187$).

Summary: These analyses provided further support for the results found for natural songs and calls in Experiments 1 and 2 and showed that decoding of synthesized calls using the temporal profiles of multi-unit responses increased during the first 6 stimulus presentations, but did not change with subsequent exposure.

Multi-unit Correct Decoding Latencies

Correct neural decoding latencies in synthesized call responses were also analyzed via a repeated-measures ANOVA using presentation (1 through 25) as a within-subjects variable. This analysis did not show a significant effect of presentation ($F(24,4152) = 1.44$, $p = 0.077$, **Fig. 38D**). The comparison of the slopes of correct decoding latencies for presentations 1-6 and 6-25 also did not reveal a significant difference ($t(173) = 0.43$, $p = 0.665$, **Fig. 38E**). Moreover, the slopes of latencies were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.617$).

Summary: These analyses revealed that, similar to the findings for natural call responses in Experiment 1, latencies along the stimulus duration to reach a correct neural decoding probability did not change with repeated stimulus presentation.

Multi-unit Correct Classification Ratios

Classifications resulting from the neural decoding process were used to construct a confusion matrix (**Fig. 39A**), from which correct neural classification ratios were calculated for pairs of synthesized calls with different acoustic dissimilarities. The ratios

of correct classifications between pairs of stimuli that were separated by only 6 ms in their FM durations were not significantly greater than 0.5 in any of the stimulus pairs analyzed (all $p > 0.052$, **Fig. 39B**). When the same analysis was conducted on pairs of stimuli that are separated by 12 ms in their FM durations, the correct classification ratios were all clearly greater than 0.5 for all tested stimulus pairs (all $p < 0.008$, **Fig. 39C**).

Summary: These analyses indicated that the temporal profiles of multi-unit responses were not able to discriminate synthesized calls that only differed by 6 ms in their FM durations. However, they successfully discriminated synthesized call pairs with 12 ms FM duration differences.

Multi-unit Mutual Information

Finally, the changes in mutual information between synthesized calls and the temporal response profiles were analyzed as a function of time along the stimulus duration. A repeated-measures ANOVA on mutual information using bin (1 through 26) as a within-subjects variable revealed a significant effect ($F(25,4325) = 62.24$, $p < 0.001$, **Fig. 40**). Post-hoc comparisons indicated that there was a significant increase in mutual information until the 8^{th} bin (80 ms), after which no significant change was observed. However, even the peak mutual information levels of ~0.13 bits for the 8 synthesized calls were drastically low as compared to the high mutual information documented for natural songs (~2.5 bits) and calls (~0.6 bits).

Summary: These findings indicate that the informativeness of the temporal profiles of multi-unit responses about synthesized calls improved until 80 ms after stimulus onset. However, the overall informativeness of these responses was lower than those observed for natural songs and calls, presumably because the high degree of

acoustic similarity among synthesized calls made them very difficult for neurons to classify accurately.

Single-unit Responses

The spike waveform clustering algorithm revealed a total of 90 neurons (43 narrow and 47 wide spike) for synthesized call playbacks. A chi-square analysis indicated that there was no significant difference between the frequencies of the two neuron types ($X^2 = 0.18$, $p = 0.673$).

Single-unit Response Properties

The firing rates of narrow spike neurons during baseline conditions were significantly greater than those of wide spike neurons ($z = 2.99$, $p = 0.003$, **Fig. 41A**). Similarly, narrow spike neurons also had significantly greater stimulus-driven response magnitudes than did wide spike neurons ($z = 2.61$, $p = 0.009$, **Fig. 41B**). The adaptation rates of the two neuron types for presentations 1-6 did not differ from each other ($z = 1.11$, $p = 0.266$) and were together not significantly different from 0 ($z = 1.05$, $p = 0.295$). However, for presentations 6-25, wide spike neurons had significantly more negative adaptation rates as compared to narrow spike neurons ($z = 1.99$, $p = 0.047$, **Fig. 41C**). The adaptation rates of wide spike neurons for presentations 6-25 were significantly less than 0 ($z = 2.98$, $p = 0.003$, **Fig. 41C**), whereas those of narrow spike neurons were not significantly different from 0 ($z = 0.01$, $p = 0.992$).

Summary: These analyses supported the differences between the two neuron types found in Experiments 1 and 2 and showed stronger firing for narrow spike neurons and stronger adaptation for wide spike neurons. Similar to the findings in Experiment 2, narrow spike neurons did not undergo adaptation at all with repeated stimulus exposure.

Single-unit Neural Dissimilarities

The relationship between acoustic parameters and the dissimilarities between the temporal profiles of single-unit responses were analyzed as for multi-unit responses. Average neural dissimilarities for pairs of synthesized calls are shown in **Figure 42A**. Similar to the neural dissimilarity matrix described for multi-units, the main-diagonal of this matrix shows within-stimulus neural dissimilarities, whereas off-main-diagonal entries indicate between-stimulus neural dissimilarities between pairs of synthesized calls with different degrees of acoustic similarity. Across single-units, the slopes of the main diagonals of these matrices did not significantly differ between the two neuron types ($z = 0.54$, $p = 0.586$). These slopes in the two neuron types together were also not significantly different from 0 ($z = 1.05$, $p = 0.293$), suggesting that single-unit within-stimulus neural dissimilarities were not related to the FM durations among synthesized calls. No difference was observed between the two neuron types in the slopes of any of the other higher-order diagonals (all $p > 0.286$). Furthermore, none of these slopes were significantly different from 0 (all $p > 0.188$), indicating that, for any given acoustic similarity level, neural dissimilarities were not related to the FM duration of the synthesized calls. Across single-units, there was also no significant difference between the two neuron types in the slopes of the anti-diagonals of these matrices ($z = 1.42$, $p = 0.155$). However, these slopes in two neuron types together were significantly greater than 0 ($z = 7.65$, $p < 0.001$, **Fig. 42B**), showing that neural dissimilarities increased as the acoustic dissimilarity between synthesized calls increased.

Next, the relationship of the changes in neural dissimilarities during presentations 1-6 to the FM durations of synthesized calls were analyzed in single-unit responses. The

average of the slopes of neural dissimilarities for presentations 1-6 between each pair of synthesized calls are shown in **Figure 42C**. Across single-units, none of the slopes of diagonals 1 through 6 of these matrices differed between the two neuron types (all $p > 0.069$) or differed from 0 ($p > 0.123$), which suggest that there was no relationship between the FM durations of synthesized calls and the rates of changes in neural dissimilarities during presentations 1-6. However, the slopes of the anti-diagonals of these matrices were significantly greater for wide than for narrow spike neurons ($z = 2.07$, $p = 0.039$, **Fig. 42D**). These slopes were significantly greater than 0 for wide spike neurons ($z = 3.69$, $p < 0.001$, **Fig. 42D**), whereas no such difference was found for narrow spike neurons ($z = 0.84$, $p = 0.400$). This suggests that, for wide spike neurons, neural dissimilarities increased during presentations 1-6 more strongly for synthesized call pairs with more acoustic dissimilarity.

Similar to the above analysis, the relationship of the changes in neural dissimilarities to the FMs of synthesized calls were also analyzed for presentations 6-25 (**Fig. 42E**). Across single-units, none of the slopes of diagonals 1 through 6 of these matrices differed between the two neuron types (all $p > 0.204$) or differed from 0 (all $p > 0.084$). There was also no significant difference between the two neuron types in the slopes of the anti-diagonals of these matrices ($z = 0.59$, $p = 0.559$). However, these slopes in two neuron types together were significantly greater than 0 ($z = 3.16$, $p = 0.002$, **Fig. 42F**) showing that, during presentations 6-25, the increase in neural dissimilarities were greater for stimulus pairs that were acoustically more dissimilar from each other.

Summary: Taken together, these findings indicated that there were no strong differences between the neural dissimilarities of the two neuron types, similar to the

patterns observed in Experiments 1 and 2. In sharp contrast to the findings in multi-unit responses, after the first 6 stimulus presentations, the improvements in the dissimilarities between the temporal profiles of neural responses were stronger for acoustically more dissimilar synthesized calls in single-unit responses. There was also a similar effect during the first 6 presentations, but only for wide spike neurons.

Single-unit Correct Decoding Probabilities

The correct neural decoding probabilities did not differ between the two neuron types ($z = 0.22$, $p = 0.825$). Similarly, there was no significant difference between the slopes of correct decoding probabilities of the two neuron types for either presentations 1-6 ($z = 0.11$, $p = 0.916$) or 6-25 ($z = 1.76$, $p = 0.079$). However, the correct decoding probability slopes of the two neuron types together were significantly greater than 0 for presentations 1-6 ($z = 1.97$, $p = 0.049$, **Fig. 43**). For presentations 6-25, the slopes of probabilities did not significantly differ from 0 ($z = 1.36$, $p = 0.174$).

Summary: Unlike what was observed in Experiments 1 and 2, narrow spike neurons did not yield more accurate neural decoding compared to wide spike neurons. However, similar to the findings for natural vocalizations in Experiments 1 and 2, decoding of synthesized calls using the temporal profiles of single-unit responses improved during the first 6 presentations and did not change with further repetition.

Single-unit Correct Decoding Latencies

There was no significant difference between the two neuron types either for correct decoding latencies ($z = 0.71$, $p = 0.475$) or for the slopes of latencies for presentations 1-6 ($z = 1.20$, $p = 0.229$) or 6-25 ($z = 1.37$, $p = 0.170$). Moreover, the slopes

of correct decoding latencies in the two neurons types together were not significantly different from 0 for either presentations 1-6 or 6-25 (both $p > 0.649$).

Summary: These analyses revealed that, similar to the findings for natural call responses in Experiment 1, latencies along the stimulus duration to reach a correct neural decoding probability did not change with repeated stimulus presentation.

Single-unit Correct Classification Ratios

Similar to multi-units, classifications resulting from the neural decoding process were used to construct a confusion matrix for single-unit responses (**Fig. 44A**). From these, correct neural classification ratios were calculated for pairs of synthesized calls with different degrees of acoustic dissimilarity. The ratios of correct classifications between pairs of stimuli that are separated by only 6 ms in their FM durations did not significantly differ between the two neuron types in any of the stimulus pairs (all $p > 0.501$), except one. The only significant difference was that wide spike neurons had greater correct classification ratios than did narrow spike neurons for the comparison between the stimuli with 24 and 30 ms FM durations ($z = 2.27$, $p = 0.023$). Similarly, the analysis of stimulus pairs that are separated by 12 ms in their FM durations did not reveal any significant difference between the two neuron types (all $p > 0.070$). The correct neural classification ratios of the two neuron types together were not significantly different from the 0.5 chance level for any of the comparisons between stimulus pairs with FMs that are 6 ms apart from each other (all $p > 0.107$, **Fig. 44B**). However, when the same analysis was conducted on pairs of stimuli that are separated by 12 ms in their FM durations, the correct classification ratios were significantly greater than 0.5 for all

tested stimulus pairs (all $p < 0.049$, **Fig. 44C**), except only the stimulus pair with 42 and 54 ms FM durations ($z = 1.18$, $p = 0.240$).

Summary: These analyses indicated that, similar to the multi-unit responses, the temporal profiles of single-unit responses were not able to discriminate synthesized calls that only differed by 6 ms in their FM durations. However, they successfully discriminated synthesized call pairs with 12 ms FM duration differences. There were no consistent differences between the confusions for the two neuron types.

Single-unit Mutual Information

Finally, mutual information between synthesized calls and the temporal profiles of neural responses did not significantly differ between narrow and wide spike neurons ($z = 1.90$, $p = 0.058$).

Summary: Unlike in Experiments 1 and 2, the temporal profiles of neural responses of the two neuron types were not found to be differentially informative about synthesized calls.

Discussion

The results of this experiment revealed some clear effects and some mixed patterns regarding the relationship between acoustic similarity and the magnitudes of changes in neural discrimination with repeated stimulus exposure. In multi-unit responses, the slopes of neural dissimilarities as a function of stimulus repetition were not found to be related to the acoustic similarities between different pairs of synthesized calls. This suggests that, regardless of how similar a given synthesized call pair was, the dissimilarities between the temporal profiles of neural responses to those signals changed with passive exposure at similar rates. Nevertheless, single-unit responses showed a

completely different pattern. The wide spike neurons during the first 6 stimulus presentations and both neuron types during the subsequent repetitions indicated a negative relationship between acoustic similarity and the changes in neural dissimilarity. That is, neural dissimilarities between acoustically less similar synthesized call pairs increased at faster rates with repeated stimulus exposure. The reasons behind these conflicting findings between multi-unit and single-unit responses were not clear.

Despite these conflicting findings, the average neural dissimilarities across all stimulus presentations were positively related with acoustic dissimilarity in both multi-unit and single-unit responses. This suggests that, irrespective of the changes with repeated stimulus presentation, acoustically more dissimilar synthesized call pairs were neurally discriminated to a greater degree. In addition, the analysis of confusion matrices showed that the temporal profiles of either the multi-unit or the single-unit responses could not discriminate synthesized call pairs that differed by only 6 ms in their FM durations. Successful neural discriminations in both multi-unit and single-unit responses were observed at the next acoustic dissimilarity level, which was a 12 ms FM duration difference. The similar findings in multi-unit and single-unit responses suggest that considering multiple neurons together, compared to analyzing single neurons in isolation, does not improve the sensitivity of stimulus discrimination. These findings can be seen as reflecting an innate just noticeable difference threshold for discrimination of FM duration differences in NCM neurons. However, this interpretation should be made with caution, because the neural decoding algorithm used in this experiment digitized neural responses at a 10 ms temporal resolution. Thus, complete confusion of synthesized calls with 6 ms

FM duration differences might be due to the specific temporal resolution used to decode responses and not necessarily to the temporal precision of neural responses themselves.

Interestingly, for any given acoustic similarity level, the temporal profiles of multi-unit responses were more dissimilar for stimulus pairs with shorter FM durations. This might reflect a relatively more superior sensitivity to detect subtle differences between calls with faster FMs, which might serve individual recognition and discrimination of conspecific males. Furthermore, male zebra finches are shown to be extremely sensitive in their call back responses to whether a brief FM is present in the beginning of the calls they hear, which signals whether a call is produced by a male or a female (Vicario et al., 2001). Thus, the increased sensitivity to discriminate calls with faster FMs might be a reflection of this functional ability that enables birds to rapidly determine the sex of conspecifics based on calls.

When all stimuli were analyzed together, the results mostly confirmed the main findings in Experiments 1 and 2. In both multi-unit and single-unit responses, the accuracy of neural decoding improved with passive exposure during the first 6 presentations and did not show a consistent trend afterwards. The latencies to reach the highest levels of neural decoding accuracy, on the other hand, did not systematically change with repeated stimulus presentation. In addition, narrow spike neurons had higher baseline firing rates and stimulus-driven response magnitude as compared to wide spike neurons. The two neuron types also differed in their adaptation rates during stimulus repetitions following the first 6 presentations, such that wide spike neurons underwent strong adaptation, whereas narrow spike neurons as a group did not show adaptation. Unlike in Experiments 1 and 2, there was no difference between the two neuron types in

neural decoding accuracies or mutual information. This lack of difference might be due to overall poor neural decoding accuracy and mutual information levels for synthesized calls, which were more difficult to discriminate than natural songs and calls.

In summary, these findings suggest that the neural discrimination of synthesized calls that vary along a single dimension - the initial FM duration - improved with passive repeated exposure, as was seen for natural calls and songs. In single neurons, these improvements in the dissimilarities between the temporal profiles of neural responses were more pronounced for synthesized call pairs that were acoustically less similar from each other. Whether these neural changes influence the behavioral discrimination of the same signals was examined in the following experiment.

EXPERIMENT 4: BEHAVIORAL DISCRIMINATION OF PASSIVELY FAMILIARIZED VOCAL SIGNALS

The primary goal of this experiment was to test whether prior passive familiarization with two auditory stimuli that are highly similar to each other improves the subjects' ability to discriminate them in a subsequent behavioral task. Experiment 3 showed that the neural discrimination of synthesized calls improves with passive familiarization. Furthermore, in humans, a brief period of exposure to distorted speech signals enhances behavioral measures of comprehension in later time points (Liss, Spitzer, Caviness, & Adler, 2002; Clarke & Garrett, 2004; Bradlow & Bent, 2008). Thus, it was hypothesized that behavioral discrimination of two similar synthesized calls is greater when passively pre-exposed to those signals as compared to encountering them for the first time.

To assess this hypothesis, birds were first trained in a Go/No-Go paradigm to discriminate the stimuli at the two ends of the synthesized call continuum (**Fig. 35B**). Then, one group was pre-exposed with two calls that fell between the training stimuli in the synthesized call continuum, whereas the other group heard two other stimuli that were

acoustically unrelated to synthesized calls. The stimuli with which the pre-exposed group was familiarized were used for both groups, together with other stimuli from the same synthesized call set, in a subsequent probe test. Behavioral discrimination between different pairs of stimuli was quantified using normalized measures. In humans, responses to ambiguous signals along a stimulus dimension in similar paradigms are marked by a contrast effect, which represents a shift in the perceptual boundary away from the most recently heard signals (Diehl, Elman, & McCusker, 1978; for a review, see Kleinschmidt & Jaeger, 2016). Thus, responses to ambiguous probe signals were also analyzed in detail as a function of the preceding stimulus.

Methods

Subjects

All subjects were as in Experiment 1, except that 10 naïve adult male zebra finches were used in this experiment. The birds lived in the general aviary until the beginning of the experiment, after which they were housed individually in sound isolation boxes with *ad libitum* food. Access to water was restricted throughout the experiment, except on non-training days, in which the birds were given 1.5 ml water.

Stimuli

From the parametrically varying synthesized call set described in Experiment 3, the stimuli at the two ends of the continuum, the ones with 18 and 60 ms FM durations, were used in the initial stage of auditory discrimination training. In the passive auditory exposure stage, the stimuli with 30 and 48 ms FM durations from the same set were used for the experimental group, whereas two stimuli that were unrelated to the synthesized calls were used for the control group. These two control stimuli were synthesized using

similar procedures; however, they differed from the parametrically varying synthesized call set in two important aspects: they lack any FM and differed from each other only in their fundamental frequencies. The fundamental frequencies of 437 and 493 Hz were used so that they were within the natural range of fundamental frequencies of zebra finch vocalizations and they did not share any harmonics either with each other or with the experimental stimuli. The durations of the control stimuli were same as the experimental stimuli (150 ms). In the final probe test stage of the experiment, the stimuli with 30, 36, 42, and 48 ms FM durations from the parametrically varying synthesized call set were used for all birds.

Apparatus

The behavioral training and the probe test were conducted in soundproof booths. On one wall of the training cage, a 25 Gauge needle (the tip blunted) was mounted 2 cm from the floor to deliver 5-10 µl water droplet rewards. To detect pecking behavior, an infrared system was placed 3 cm above the water delivery unit. On the panel behind the infrared system, a yellow LED was placed right between the receiver and the sensor to signal the bird availability of trial initialization. Stimuli were played back at an amplitude of 65 dB SPL (A scale) and a sampling frequency of 44.100 Hz from a speaker located 7 cm behind the infrared system.

Behavioral training

Birds were placed in individual sound isolation boxes and water-deprived 24 hr before the beginning of the behavioral training. The training consisted of three consecutive stages. In the first stage, water droplets were delivered at random intervals between 30 and 60 s with no auditory stimulus or response requirement in order to

acclimate the birds to the training apparatus. All birds successfully learned to obtain the water droplets within one day and proceeded to the second stage. In this stage, shaping of the pecking behavior was accomplished by delivering the water droplets only after the bird pecked to the panel between the sensors to successfully break the infrared beam. Initially, pecking was facilitated by taping seeds to the panel between the infrared sensors. An intertrial interval (ITI) of 3 s was used to prevent continuous pecking. Birds proceeded to the next stage if they completed at least 100 trials in two consecutive days. In the third stage, a Go/No-Go paradigm was employed for auditory discrimination training. Before the initialization of a trial by the bird, the yellow LED stayed turned on. A peck to initialize a trial turned the LED off and led to the presentation of either the Go or the No-Go stimulus. Go and No-Go stimuli were always the synthesized calls with 60 and 18 ms FM durations, respectively. There was a 1-s period from the onset of the stimulus presentation, in which the birds could not respond, followed by a 3-s response window. A successful peck within the response window to the Go stimulus was rewarded with a water droplet, while a peck in response to the No-Go stimulus led to a 30-s lights-out period. Failure to respond to either stimulus within the response window did not deliver any outcome. The ITI was 1 s. The LED stayed off throughout the trial and turned on at the end of the ITI to signal the bird that the next trial was ready to be initialized. Each training day lasted for 6 h. The success criterion for the discrimination was an overall accuracy of correct responses on at least 80% of the trials for two consecutive days.

Passive auditory exposure

After successful acquisition of auditory discrimination, the birds were assigned to one of two groups. The pre-exposed group was presented with the stimuli with 30 and 48 ms FM durations from the synthesized call set, whereas the control group was presented with the two control stimuli that varied in fundamental frequency. For both groups, the passive auditory exposure was administered in home cages the day following the second criterion day of behavioral training. Stimuli were played back from a speaker located 5 cm in front of the cage at an amplitude of 65 dB SPL (A scale) and a sampling frequency of 44.100 kHz for 200 times each with a 6 s ISI in a shuffled order.

Probe test

Ten minutes after the end of the passive auditory exposure, the birds were placed in the training cage for the probe test. The dynamics and the parameters of this test was exactly as in the initial auditory discrimination task, except the original Go and No-Go stimuli were presented only in 80% of the trials. In the remaining 20% of the trials, the stimuli with 30, 36, 42, and 48 ms FM durations from the parametrically varying synthesized call set were presented with equal probability. Note that the pre-exposed group was passively presented with two of these probe stimuli prior to the test phase, while the control group was completely naïve to all of the probe stimuli. The responses to the probe stimuli were neither rewarded nor punished. The probe test was conducted in 8-h sessions for 3 consecutive days.

Data analysis

Percent response. The responses to the Go stimulus were quantified by taking the ratio of Go trials with a pecking response to the total number of Go trials as a

percentage. Similarly, to quantify the responses to the No-Go stimulus, the percentage of No-Go trials with a pecking response in all No-Go trials was calculated.

Percent accuracy. To quantify the overall discrimination of the training stimuli, the fraction of the sum of the Go trials with pecking responses and the No-Go trials with no pecking responses to the total number of trials was calculated as a percentage.

Discrimination index. The magnitude of the discrimination between a particular probe stimulus pair was quantified via a metric that normalizes the difference between the responses to those signals by the response difference between the Go and No-Go stimuli as

$$Discrimination\ Index = 100 * \frac{\%R_A - \%R_B}{\%R_{Go} - \%R_{No-Go}}$$

where $\%R$ is percent response and A and B are the particular probe stimuli used for the calculation. Stimulus A was always the stimulus that was acoustically more similar to the Go stimulus so that the calculation yields positive results. This discrimination index was computed for two probe stimulus pairs separately, one between the two stimuli that were passively presented to the pre-exposed group before the test sessions (30 and 48 ms FM durations) and one between the two probe stimuli that fell between the two pre-exposure signals along the FM duration dimension (36 and 42 ms FM durations).

Results

Responses in the Auditory Discrimination Training

First, the behavioral performances in the initial Go/No-Go auditory discrimination task were analyzed. On average across training days, birds completed 244.12 to 680.80 (Mean ± SEM = 431.91 ± 47.03) trials per training day, which did not significantly differ between the pre-exposed and the control exposure condition ($t(8) = 0.06$, $p = 0.956$).

Latencies to reach the criterion for successful acquisition ranged from 9 to 36 days (Mean ± SEM = 21.80 ± 2.91, **Fig. 45A-B**). Despite this considerable individual variability, acquisition latencies were also not significantly different between the two exposure conditions ($t(8) = 0.26$, $p = 0.801$).

The auditory discrimination performances in the beginning and at the end of training were compared between the two groups by calculating the average percent accuracies in the first and the last 2 training days separately. A mixed ANOVA on overall percent accuracies using exposure (Pre-exposed, Control) as a between-subjects variable and training days (First 2, Last 2) as a within-subject variable did not indicate an effect of exposure ($F(1,8) = 1.47$, $p = 0.260$) or an interaction between exposure and training days ($F(1,8) = 0.06$, $p = 0.819$). However, the effect of training days was highly significant ($F(1,8) = 1666.86$, $p < 0.001$), indicating greater percent accuracies in the last 2 than in the first 2 days of training. The percent accuracies in the first 2 training days were not significantly different from 50% chance level ($t(9) = 0.12$, $p = 0.905$). However, in the last 2 days of training, percent accuracies were significantly greater than 80% criterion level ($t(9) = 8.14$, $p < 0.001$).

Next, the responses to Go and No-Go stimuli were analyzed separately to examine the nature of auditory discrimination acquisition more in detail. A mixed ANOVA on percent responses to Go stimulus using exposure (Pre-exposed, Control) as a between-subjects variable and training days (First 2, Last 2) as a within-subjects variable did not show an effect of exposure ($F(1,8) = 0.15$, $p = 0.710$) or an interaction between exposure and training days ($F(1,8) = 0.01$, $p = 0.983$). There was a small but significant increase in percent responses to Go stimulus from the first 2 to last 2 training days ($F(1,8)$

= 9.90, p = 0.014, **Fig. 45C**), however the percent responses to Go stimulus in the first 2 days of training were already at very high levels (Mean ± SEM = 92.46 ± 1.89 %). A similar ANOVA was used to examine the responses to No-Go stimulus and also revealed no significant effect of exposure ($F(1,8) = 0.19$, $p = 0.678$) or an interaction between exposure and training days ($F(1,8) = 0.01$, $p = 0.946$). However, the effect of training days was highly significant ($F(1,8) = 736.34$, $p < 0.001$, **Fig. 45D**). The percent responses to No-Go stimulus in the first 2 days of training were at very high levels (Mean ± SEM = 92.30 ± 1.59 %), but dramatically decreased at the end of the training (Mean ± SEM = 26.43 ± 1.34 %).

Summary: Taken together, these analyses indicated that the birds in the two exposure groups had comparable numbers of trials per day, latencies to successfully learn the task, and behavioral responses to both the Go and the No-Go stimulus prior to passive exposure and probe test. In addition, acquisition of the initial auditory discrimination task was mainly driven by learning to inhibit responding to the presentations of the No-Go stimulus.

Responses to Probe Stimuli

Having established comparable responses between the two exposure groups in the initial discrimination training, responses in the probe tests were analyzed next. The percent responses to each stimulus during the three days of probe test for the pre-exposed and control birds are shown in **Figure 46**. There was no significant difference between the pre-exposed and the control condition in percent responses to either the Go or the No-Go stimulus in any of the 3 days of probe tests (all $p > 0.122$). Thus, the discrimination of the initial training stimuli in probe tests were comparable between the two exposure

conditions. The discrimination between the probe stimuli with 30 and 48 ms FM durations, which were used in passive familiarization for the pre-exposed birds, were compared between the two exposure conditions via a mixed ANOVA on discrimination indices using exposure (Pre-exposed, Control) as a between-subjects variable and test day (1, 2, 3) as a within-subjects variable. This analysis did not reveal an overall difference between the two exposure conditions ($F(1,8) = 0.65$, $p = 0.443$, **Fig. 47A**) or an interaction between exposure and test day ($F(2,16) = 1.06$, $p = 0.370$). There was also no main effect of test day ($F(2,16) = 1.49$, $p = 0.254$). The discrimination between the probe stimuli with 36 and 42 ms FM durations was also assessed via a similar ANOVA. Again, this analyses did not reveal an effect of exposure ($F(1,8) < 0.01$, $p = 0.995$, **Fig. 47B**), an effect of test day ($F(2,16) = 0.96$, $p = 0.404$), or an interaction between exposure and test day ($F(2,16) = 0.17$, $p = 0.846$).

The number of trials completed throughout the 3 probe test days varied between 951 and 3914 (Mean ± SEM = 2289.40 ± 342.77) across birds, although there was no significant difference between the two exposure conditions ($t(8) = 0.80$, $p = 0.449$). Due to this strong individual variability, probe test performances were further examined by focusing on blocks of trials rather than days. Block size was determined to be 200 trials so that there were 10 presentations of each probe stimulus in each block. This analysis was conducted on 4 blocks since the bird with the minimum number of trials completed 951 trials. There was no significant difference between the pre-exposed and the control condition in percent responses to either the Go or the No-Go stimulus in any of the 4 test blocks (all $p > 0.066$). Thus, the original training stimuli were discriminated comparably between the two exposure conditions in each test block. The analysis of the

discrimination between the probe stimuli with 30 and 48 ms FM durations via a mixed ANOVA on discrimination indices using exposure (Pre-exposed, Control) as a between-subjects variable and test block (1, 2, 3, 4) as a within-subjects variable did not reveal an effect of exposure ($F(1,8) = 0.01$, $p = 0.911$, **Fig. 47C**) or an interaction between exposure and test block ($F(3,24) = 0.60$, $p = 0.622$). There was also no significant difference among the test blocks ($F(3,24) = 3.00$, $p = 0.051$). The discrimination between the probe stimuli with 36 and 42 ms FM durations was also assessed via a similar ANOVA. Once again, there was no effect of exposure ($F(1,8) = 1.99$, $p = 0.196$, **Fig. 47D**), effect of test block ($F(3,24) = 0.91$, $p = 0.451$), or an interaction between exposure and test block ($F(3,24) = 0.61$, $p = 0.615$).

Summary: Contrary to predictions, these analyses did not show an effect of prior familiarization on subsequent behavioral discrimination of the probe signals. There was no difference in the discrimination of synthesized calls whether they were familiarized right before testing or they were heard for the first time.

Relationships Between Training and Test Performances

Since pre-exposure did not affect the behavioral discrimination of probe stimuli, relationships between training and test performances were assessed to examine whether consistent individual differences in the initial training could account for variations in responses to probe stimuli. For discrimination of the probe stimuli with 30 and 48 ms FM durations, there was a significant negative correlation between latencies to acquire the initial discrimination and magnitude of discrimination indices in the control condition ($r(3) = -0.95$, $p = 0.013$). The correlation between acquisition latencies and the same type of discrimination indices was also strongly negative in the pre-exposed condition;

however, it did reach statistical significance ($r(3) = -0.74$, $p = 0.154$). These correlations were not significantly different between the two exposure conditions ($z = 0.88$, $p = 0.379$). Thus, when analyzed together, acquisition latencies and discrimination indices were highly negatively correlated ($r(8) = -0.75$, $p = 0.013$, **Fig. 48A**). The analysis of the correlation between acquisition latencies and the discrimination indices between the probe stimuli with 36 and 42 ms FM durations, on the other hand, did not reveal a significant correlation in either the pre-exposed ($r(3) = 0.15$, $p = 0.806$) or the control condition ($r(3) = -0.80$, $p = 0.108$). These two correlation coefficients also did not significantly differ from each other ($z = 1.25$, $p = 0.211$).

The relationship between the initial discrimination acquisition latency and the average level of responding to all probe stimuli, whether they were correct or not, was also analyzed. There was a significant positive correlation in the pre-exposed condition ($r(3) = 0.95$, $p = 0.013$). The same analysis also indicated a positive correlation coefficient in the control condition, but this correlation was not statistically significant ($r(3) = 0.62$, $p = 0.269$). However, these correlation coefficients were not significantly different between the two exposure conditions ($z = 1.11$, $p = 0.267$). Hence, the two exposure conditions were analyzed together, which revealed a significant positive correlation between the acquisition latencies and the overall levels of percent responses to probe stimuli ($r(8) = 0.77$, $p = 0.009$, **Fig. 48B**).

Summary: These analyses indicated that a significant portion of the variations in probe test behaviors could be accounted for by the variations in latencies to acquire the initial discrimination task. Birds that were quicker to learn the first auditory discrimination, also differentiated synthesized calls with 30 and 48 ms FM durations to a

greater degree. However, this effect was not observed for discrimination of acoustically more similar synthesized calls. In addition, birds that needed longer training to acquire the initial discrimination, also tended to respond more frequently to all probe stimuli, whether these responses were correct or not.

Contrast Effects

Last, responses to probe stimuli were analyzed separately when they were presented after the Go or the No-Go stimulus to investigate whether the acoustic contrast to the preceding stimulus affected responses in the following trial. A mixed ANOVA was conducted on percent responses using exposure (Pre-exposed, Control) as a between-subjects variable and target stimulus (30, 36, 42, 48 ms FM durations) and previous stimulus (Go, No-Go) as within-subjects variables. There was no effect of exposure ($F(1,8) < 0.01$, $p = 0.972$), or an interaction between exposure and target stimulus ($F(3,24) = 0.13$, $p = 0.940$), or an interaction between exposure and previous stimulus ($F(1,8) < 0.01$, $p = 0.992$), or a three-way interaction ($F(3,24) = 2.20$, $p = 0.114$). The effect of target stimulus was highly significant ($F(3,24) = 56.12$, $p < 0.001$, **Fig. 49**). Post-hoc analyses indicated that all pairwise comparisons were significantly different (all $p < 0.025$), such that the highest percent responses were seen for the probe stimulus with the 48 ms FM duration, followed by the one with the 42, 36, and 32 ms FM durations. Most importantly, there was a significant effect of previous stimulus ($F(1,8) = 23.55$, $p = 0.001$, **Fig. 49**), indicating greater percent responses for target stimuli when they were preceded by the No-Go as compared to the Go stimulus. There was also a significant interaction between the target and the previous stimulus ($F(3,24) = 3.20$, $p = 0.041$, **Fig. 49**). Post-hoc comparisons revealed that, for target stimuli with 30 and 42 ms FM

durations, percent responses were significantly greater when they were preceded by the No-Go as compared to the Go stimulus (both $p < 0.009$), whereas no such difference was observed for target stimuli with 36 and 48 ms FM durations (both $p > 0.148$).

Summary: This analyses revealed a clear contrast effect in behavioral responses to ambiguous probe stimuli. When preceded by the No-Go stimulus, which was a synthesized call with an 18 ms FM duration, the perception of probe stimuli were apparently "pushed away" from these short FM durations, resulting in more Go stimulus-like, stronger behavioral responses, as compared to when they were preceded by the Go stimulus (a synthesized call with a 60 ms FM duration). When analyzed further in detail, this contrast effect was only significant for synthesized calls with 30 and 42 ms FM durations, however, the trends were also in the same direction for the other two probe stimuli.

Discussion

The results of this experiment indicated no effect of prior passive familiarization with two acoustically similar synthesized calls on their behavioral discrimination in a subsequent task. There was no difference between the pre-exposed and the control exposure condition in the discrimination of either the synthesized calls that were used in passive exposure or the synthesized calls that fell between those signals along the FM duration dimension of the synthesized call set. These findings were in direct contrast with the expectations from the results in Experiment 3, which showed that the neural discrimination of the same synthesized calls that were used for pre-exposure in this experiment improved with passive repeated presentation. One potential reason for lack of an effect in this experiment might be that the passive exposure was administered in

housing cages, which provide a context different from that of the behavioral discrimination task. The pre-exposure was not conducted in the behavioral test apparatus to avoid any extinction-like interference effects. However, if the gains in neural discrimination with passive exposure are context-specific, then this might explain why no effect of pre-exposure was found in subsequent behavioral discrimination. Partial evidence for this idea comes from previous studies that investigated the context-dependence of adaptation in NCM and reported that adaptation in one context, such as the specific location of the speaker or additional visual cues that were presented together with the auditory signal, is partially or completely reversed when the same signals are presented in a different context (Kruse, Stripling, & Clayton, 2004; Smulders & Jarvis, 2013). As shown in Experiment 1, there is a relationship between adaptation and improvements in neural discrimination, thus it is possible that the gains in neural discrimination with passive exposure in the housing cage did not transfer to the behavioral discrimination context in this experiment. Another alternative explanation for the lack of differences between the pre-exposed and the control condition might be the relatively small sample size (n = 5 per group) used in this experiment. The effect of passive exposure on behavioral discrimination might not be very large, which would require more birds to reveal a significant difference between the two exposure conditions.

More than half of the variation in the behavioral discrimination of the pre-exposure stimuli was explained by the variation in the latencies to acquire the initial auditory discrimination. This strong relationship indicates that the birds that learned the Go/No-Go discrimination quicker, also tended to discriminate the synthesized calls with 30 and 42 ms FM durations better. This correlation might potentially reflect individual

differences in common general auditory capabilities that similarly affect detection of subtle acoustic differences between pairs of stimuli in the training and the test stage of the experiment. However, it is also possible that the relationship between the initial discrimination acquisition speed and the subsequent discrimination of probe signals might be due to individual differences in general task-related behavioral differences that are unrelated to auditory sensitivities. The analysis of the Go and the No-Go responses in the initial discrimination task indicated that birds improved their performance mainly by learning to inhibit responding to the No-Go stimulus. Furthermore, birds that took longer to acquire the initial discrimination, also showed higher levels of responding to the probe stimuli, whether these responses were correct or not. Taken together, these findings indicate that individual differences in executive control of responses may potentially contribute to the correlations between training and test performances. The two described possibilities, general auditory sensitivity and executive response control, are not mutually exclusive and may together underlie the individual differences seen in behavioral discrimination of synthesized calls. Nevertheless, it is important to note that the strong relationship between the initial discrimination acquisition speed and the subsequent discrimination magnitude for test signals was not observed in the discrimination of the two probe stimuli that were only differentiated by a 6 ms difference in their FM durations.

There were clear effects of the preceding training stimuli on behavioral responses to subsequent probe signals. If a probe synthesized call was presented after the No-Go stimulus (a synthesized call with an 18 ms FM duration) birds were more likely to respond as if the ambiguous probe signal was the Go stimulus (a synthesized call with a

60 ms FM duration). Thus, hearing a synthesized call with a short FM duration led to perceiving the FM duration of an ambiguous synthesized call as longer compared to hearing it after a synthesized call with a long FM duration. This finding suggests that responses to probe stimuli were not solely based on their acoustic properties, but also were influenced by their contrasts with recent signals. Similar phenomena are well-known in humans as contrast effects, which represent a shift in perceptual boundaries away from the particular properties of the most recently encountered signals along a stimulus dimension (Diehl et al., 1978; for a review, see Kleinschmidt & Jaeger, 2016).

In summary, contrary to rapid and long-term improvements in neural discrimination of novel acoustic signals with passive familiarization, this experiment did not provide any evidence that these improvements affected behavioral discrimination.

GENERAL DISCUSSION

In a series of experiments, this thesis investigated the effects of passive familiarization on neural and behavioral discrimination of novel acoustic signals. Experiment 1 provided strong evidence that the temporal profiles of neural responses to different novel natural vocalizations rapidly become more dissimilar from each other with repeated exposure, which improves the decoding of these stimuli from neural responses. In addition, the results of Experiment 1 indicated that the temporal profiles of neural responses are sensitive to the sequence in which the signals are presented such that neural responses are more informative about acoustic stimuli when they are presented in a blocked than in a shuffled sequence. Experiment 2 supported and extended these findings by showing that the rapid gains in neural discrimination and decoding of natural vocalizations with passive familiarization remained in effect 20 hours after the initial encounter. Experiment 3 investigated how the degree of acoustic similarity between sounds related to these rapid dynamic changes in stimulus representations using synthesized vocalizations and revealed mixed effects. Finally, contrary to findings in

neural responses, Experiment 4 did not yield any effect of passive familiarization on subsequent behavioral discrimination of two acoustically similar synthesized vocalizations.

Rapid and Long-lasting Improvements in Neural Discrimination

The main finding from the first three experiments was that the decoding of novel stimuli using the temporal profiles of neural responses improves with passive familiarization. This was true for two different types of natural, as well as synthesized, vocalizations that varied in length, complexity, and ecological significance. These gains in neural recognition and discrimination were very rapid, exhibiting a sharp increase during initial stimulus presentations and either little or no change with further exposure. In most cases, these improvements were paralleled by a reduction in the latency along the stimulus duration to accurately decode stimulus identities with passive exposure. This means that, as they become more familiar, novel signals can be recognized by hearing shorter sections from the beginning of the vocalization. Although for songs, the improvements in neural decoding performance could be clearly explained by neural responses to the same stimuli becoming more and more similar to each other and neural responses to different stimuli becoming more and more dissimilar from each other, the neural decoding of natural and synthesized calls showed various other patterns. Thus, the analysis of the between-stimulus and within-stimulus neural dissimilarities separately was not always helpful in understanding the gains in neural decoding with repeated stimulus presentation.

The main working hypothesis explored in this thesis was that the rapid improvements in neural discrimination and recognition of acoustic signals with passive

familiarization were related to the process of adaptation. The analysis of the correlations between the rates of adaptation and the linear trends of the neural discrimination metrics as a function of stimulus repetition indeed showed that the sites that adapted more steeply, also tended to increase in neural decoding more strongly. However, although significant, this relationship was not very strong, accounting for only about 7% of the total variation in the changes in neural decoding. Thus, adaptation, as measured in the zebra finch brain in previous studies (Chew et al., 1995, 1996a, 1996b), did not correlate strongly with the improvements in neural decoding with repeated stimulus presentation. However, in past studies and throughout this thesis, adaptation was quantified using the total magnitude of neural activity during the whole response, disregarding its temporal profile. This was in direct contrast to the neural discrimination metrics used in this thesis, which rely on the neural response profiles, not the absolute magnitudes. The contribution of non-stimulus-specific reductions in firing rates to the overall measure of adaptation rate might have overshadowed the effects of stimulus-specific adaptations to local features that maximally differentiate the temporal profiles of neural responses with stimulus repetition. Thus, if adaptation rates were examined at specific time points along the stimulus duration, points of maximum adaptation could potentially account for much more of the variation in neural discrimination improvements. However, although theoretically fruitful, this idea represents a methodological challenge, as the time points of maximum local adaptation can differ between different stimuli. In depth exploratory analytical techniques in future studies will potentially reveal the true strength of the relationship between the rates of adaptation and the improvements in neural discrimination.

The phenomenon of adaptation to features or complex statistics of external signals studied in the mammalian brain does not necessarily imply that the neural response magnitudes decrease with repeated exposure. Rather, the receptive fields of neurons change in complex ways to adjust the dynamic range of neurons to the relevant stimulus statistics in the environment and/or to enhance the detection of changes in the incoming stimulus stream. In a similar vein, recent studies in our lab demonstrated that the spectro-temporal receptive fields of neurons in NCM undergo rapid changes with repeated stimulus exposure (Yang & Vicario, unpublished results). In fact, adaptation may just be a measure of dynamic changes in receptive field selectivity, with more selective receptive fields showing smaller responses with repeated stimulation. Whether total firing rates decrease, increase, or remain unchanged with stimulus repetition for any given neuron in NCM, it is possible that a significant portion of the variation in the changes in the temporal profiles of neural responses can potentially be explained by modulations in the receptive fields that change the time along the stimulus duration at which the neuron fires, irrespective of the changes in the total amount of firing. This might explain the reported low level of correlation between adaptation and neural discrimination improvements. Having discussed this possibility, it is important to reemphasize here that all of the neural discrimination metrics used in this thesis were exclusively based on the temporal profiles of neural responses and completely insensitive to changes in the total firing rates with adaptation. Thus, other neural decoding algorithms that also take into account the total amount of activity could potentially provide a stronger relationship between the rates of overall adaptation and the changes in neural decoding.

Another important piece of evidence in support of the relationship between adaptation and neural discrimination gains was provided by the long-term effects. As shown in previous studies, the responses to novel songs gradually adapted with repeated exposure, whereas the responses to songs that were passively familiarized 20 hours ago decreased dramatically with few repetitions and remained at adapted levels thereafter. This marked difference between the adaptation profiles of novel and familiar signals is taken as an indication of neural auditory memory in NCM (Chew et al., 1995; Phan, Pytte, & Vicario, 2006). These differences were paralleled by higher levels of neural decoding accuracy and mutual information for exactly the same songs when they had been previously familiarized as compared to when they were heard for the first time. Thus, there seems to be a link between conditions that produce long-term adaptation and neural decoding improvements. However, a closer examination of the trial-by-trial dynamics reveals interesting details. For instance, it is well-established that presentation of a novel song for 20 times does not produce a long-term neural memory that can be detected 20 hours later in NCM (Chew et al., 1996a, 1996b). In Experiment 2, 200 stimulus repetitions were used for passive familiarization, which successfully induced long-term memory. However, the results of Experiments 1 and 2 indicate that there is little or no gain in neural decoding accuracy after about the first 6 stimulus presentations. This suggests that adaptation during the following stimulus presentations does not contribute to rapid improvements in neural decoding, but is needed for consolidation of those improvements for later processing. The direct relationships between rapid and long-term effects are hard to decipher from the experiments in this thesis, because electrophysiological responses from given sites and neurons were recorded either only

during the test phase or only during the initial induction phase of memories. That is, there was no longitudinal recording. Having now established the first evidence for long-lasting gains in neural decoding with passive exposure, future studies utilizing chronic recordings are needed to address the dynamics governing the amount of information retained or lost at particular sites after the initial familiarization of novel acoustic signals.

Acoustic Features and Neural Discrimination

As expected, physical stimulus features were tightly related to the overall levels of the dissimilarity between the temporal profiles of neural responses to different auditory signals. The analysis of synthesized calls that only vary along the initial FM duration dimension indicated that acoustically more similar signals elicited more similar neural response profiles. The reason for choosing FM duration as the critical variable was that it was intrinsically a temporal feature, which was more likely to have its effects on the temporal profiles of neural responses as compared to other features, such as the fundamental frequency, that might lead to variations primarily in total firing rates. Despite the clear effects in multi-unit responses, the analysis of single neuron temporal response profiles did not reveal a relationship between acoustic and neural dissimilarity. One potential reason for this can be that neurons in NCM, due to their complex spectro-temporal receptive fields (Kozlov & Gentner, 2016), might have shown differences for the highly similar short synthesized calls in their total firing rates, but not in temporal response patterns. If this was true, the neural dissimilarity metric that was utilized throughout this thesis would not be able to reflect those differences. Thus, to gain a more complete understanding of the relationship between physical stimulus characteristics and neural discrimination, one valuable future direction would be to assess the relative

contributions of total firing rates and the temporal activity patterns by employing different neural decoding techniques.

Contrary to overall neural dissimilarity measures, the rates of changes in the dissimilarities between the neural response profiles with repeated stimulus presentation was not related to acoustic similarity in multi-unit responses. This means that, regardless of the degree of acoustic similarity between stimulus pairs, dissimilarities between the temporal profiles of neural responses to those signals changed at similar rates with passive exposure. In contrast, single neuron responses indicated that the neural dissimilarities between acoustically less similar synthesized call pairs increased more with repeated stimulus exposure compared to acoustically more similar stimulus pairs. This finding indicates that, at least in responses of single neurons, acoustic contrasts between auditory signals affect the improvements in neural discrimination with repeated stimulus exposure.

Stimulus Sequence Effects on Neural Discrimination

In addition to physical acoustic stimulus features, contextual factors also had a huge impact on neural discrimination and recognition of acoustic signals. Contrary to what was hypothesized, the neural decoding of the same songs and calls was faster and more accurate when those stimuli were presented in a blocked compared to in a shuffled sequence. Mutual information between the temporal profiles of neural responses and stimulus identities also showed the same effect. Similar findings were reported in studies that investigated the effects of talker variability on speech comprehension. Recognition of speech signals is poorer and takes more time when stimuli are spoken by different talkers in a shuffled setting as compared to when utterances from different speakers are

presented one by one (Creelman, 1957; Mullennix, Pisoni, & Martin, 1989). These findings have been interpreted as an indication of a talker normalization process that more successfully improves the mapping of sounds to phonetic categories by accumulating talker-specific evidence under the stable conditions of a blocked setup as compared to a shuffled sequence, which triggers the normalization process every time the talker changes and leads to discontinuities in incoming talker-specific vocalization characteristics. Although in the case of zebra finch vocalizations, it is impossible to talk about a mapping mechanism between sounds and phoneme-like abstract processing units, either due to the absence of, or our ignorance of, such a system, the difference between blocked and shuffled sequences seem to be explained by similar principles. The main drive behind the neural decoding accuracy and mutual information differences was that the within-stimulus neural dissimilarities in the blocked sequence were markedly lower than those in the shuffled sequence. This suggests that the temporal profiles of neural responses to different presentations of any particular signal were reliably more consistent in the blocked compared to those in the shuffled presentation. Thus, similar to the proposed talker normalization process in humans, the discontinuities in the presentations of any particular stimulus in the shuffled sequence may induce variability and thus hinder the neural response profiles to reliably represent acoustic signals in NCM.

Neuron Type Differences in Neural Discrimination

Neurons in NCM separated nicely into narrow and wide spike neurons based on their spike waveforms. In all experiments, single-unit populations consisted of comparable numbers of the two neuron types, whereas a previous study reported a ~1:2 narrow to wide spike neuron ratio in NCM (Meliza & Margoliash, 2012). These

conflicting findings might be due to the differences in the recording techniques or the spike-sorting algorithms used in this and other studies. In the mammalian cortex, it is shown that narrow spike waveforms indicate inhibitory neurons and wide spike waveforms indicate excitatory neurons (Bartho et al., 2004). If this formulation is correct in the songbird forebrain, then the equal proportions for the two neuron types found in this study would be a more accurate depiction of the underlying circuitry since both histological (Pinaud & Mello, 2007) and functional (Pinaud et al., 2008) analyses showed that roughly half of the neurons in NCM are GABAergic inhibitory neurons.

Narrow spike neurons had higher baseline firing rates and stronger stimulus-driven responses compared to wide spike neurons. In parallel, intracellular recordings in brain slices from NCM indicated that half of the neurons fire phasically, whereas the other half fires tonically or transiently in response to electrical stimulation (Dagostin, Lovell, Hilscher, Mello, & Leão, 2015). Importantly, the spike waveforms of the phasic neurons are wider than those of the tonic and transient neuron types. Unfortunately, the authors did not examine whether these differences in firing properties and spike waveforms reflected an excitatory/inhibitory neuron type differentiation. The results of this thesis also revealed a marked difference between the adaptation profiles of the two neuron types. Wide spike neurons underwent strong adaptation with repeated stimulus presentation, whereas there was little or no adaptation for narrow spike neuron responses. In the mammalian primary auditory cortex, two types of inhibitory interneurons are shown to contribute differently to the stimulus-specific adaptation (SSA) of excitatory neuron responses in the oddball paradigm (Natan et al., 2015). Parvalbumin-positive neurons provide a global inhibition, whereas somatostatin-positive neurons inhibit

excitatory responses to only the frequently repeated stimuli. The net effect of these two inhibitory processes is to adapt the excitatory neuron activity to repeatedly presented stimuli. If the narrow and wide spike neurons in NCM reflect inhibitory and excitatory neurons, respectively, then the strong adaptation observed for wide spike neurons might potentially be explained by similar circuitry interactions. However, to date, the representation of parvalbumin- and somatostatin-positive interneurons has not been well-characterized in songbird NCM. Future studies utilizing in vivo imaging, together with cell type-specific optogenetic manipulations, are needed to address whether the long-term form of adaptation in the songbird NCM is an emergent property of such complex interactions between excitatory and different types of inhibitory neurons.

Narrow spike neurons yielded higher neural decoding accuracy and mutual information estimations than did wide spike neurons. A previous study reported that wide spike neurons in the starling auditory lobule display higher levels of stimulus selectivity compared to narrow spike neurons (Meliza & Margoliash, 2012). This might seem in conflict with the present findings, however the selectivity measure used in the mentioned study was completely based on total firing rates as opposed to the neural decoding method based exclusively on the temporal profiles of responses used throughout this thesis. Even when the variations in total response magnitudes were controlled, the responses of narrow spike neurons were generally more informative about stimulus identities. This suggests that the differences between the two neuron types in terms of neural decoding accuracy cannot be simply explained by the absolute magnitudes of responses and thus might potentially reflect real differences in stimulus representations in different neuron types.

Despite the differences between narrow and wide spike neurons in overall mutual information, the two neuron types showed similar improvements in neural discrimination with repeated stimulus presentation. This suggests that the improvement in neural decoding performance observed in multi-unit responses with stimulus familiarization does not stem from differential contributions of the two neuron types, but rather occurs similarly at the single neuron level for both kinds of cells. Still, there might be other currently unknown parameters, based on which neurons in NCM can be classified and differ in their degrees of contribution to neural discrimination gains with passive familiarization.

Neural Discrimination Gains in the Absence of Behavioral Improvements

Contrary to strong evidence for neural discrimination gains with repeated stimulus exposure, there was no indication of an improvement in behavioral discrimination of acoustic signals with passive familiarization. As discussed, this absence of behavioral effects might be due to methodological constraints, such as the contextual differences between the pre-exposure and the probe test environments and small sample size. However, if there were indeed no direct gains in behavioral discrimination with passive exposure, how could the conflicting findings at the neural and behavioral levels be interpreted? It has been argued that one of the main advantages of sparse coding is increased storage capacity for distinct representations (Willmore & Tolhurst, 2001). Thus, the improvements in the dissimilarities of neural representations with the long-term form of adaptation in the songbird NCM may represent a process to maximize this capacity for encoding of unique sensory memories. Adaptations of this sort can benefit the organism in its ecology without necessarily having direct behavioral outcomes. It is

also possible that the effects of neural discrimination improvements with passive exposure are much subtler than previously imagined. Future studies using chronic electrophysiological recordings that span both passive exposure and subsequent behavioral testing are needed to examine the details of the relationship between neural and behavioral discrimination of novel signals as they become familiar.

Related Observations in a Vocal Motor Area of the Songbird Brain

This thesis focused on processing of novel signals in NCM, but it is only one area that responds to sounds in the songbird forebrain. In addition to other areas that receive ascending auditory input, such as field L and CM, neurons in a vocal motor area, HVC (**Fig. 2**), also have auditory responses, and these responses are selective for species-specific vocalizations. The role or roles of these auditory responses in a motor area remains somewhat unclear, but lesions of HVC do disrupt auditory-driven behaviors (Brenowitz, 1991; Del Negro, Gahr, Leboucher, & Kreutzer, 1998). In a recent study, changes in neural responses to passive exposure under blocked and shuffled conditions were compared between HVC and NCM (Soyman & Vicario, 2017). Analysis of data from simultaneous recordings in both structures revealed quantitative and qualitative differences in the way responses adapted with repeated presentation. In NCM, the response decrease with repetition of each stimulus was gradual and long-lasting, and the absolute magnitudes of the responses did not differ between the two sequences. In contrast, HVC responses to songs decreased much more rapidly in the blocked than in the shuffled sequence. Furthermore, contrary to long-lasting adaptation in NCM, adaptation in HVC was much more transient: there was a decrease in responding from the first to the second presentation only after zero or one intervening stimulus presentation (within 12

seconds); 2 or more intervening stimuli (separated by >12 seconds) resulted in a response to the second presentation as if the stimulus was completely novel. In addition, auditory responses in NCM were more informative and less internally coherent than in HVC. Thus, these findings suggest that NCM processes the individually-specific complex vocalizations of others based on prior familiarity, while HVC responses appear to be modulated strongly by transitions and/or timing in the ongoing sequence of sounds. Thus, the kinds of changes documented in this thesis for NCM are only one of a set of parallel processes that discriminate sounds in the service of learning and individual recognition.

Conclusion

The experiments in this thesis provide valuable insights into the mechanisms by which the nervous system dynamically modulates sensory representations to improve discrimination of external signals at short and long time-scales. When faced with auditory signals that have never been heard before, neural representations are rapidly modulated with just a few exposures to dramatically improve recognition, discrimination, and classification. During subsequent exposures, these signals are then successfully recognized after hearing fewer initial acoustic features. Furthermore, the nervous system can better recognize and discriminate these sounds when they are encountered in blocks of the same stimulus as compared to in an unpredictable sequence. The rapid plasticity in neural representations of novel auditory signals not only affects immediate processing, but is also long-lasting. That is, the discrimination of previously familiarized sounds is improved and occurs faster as compared to discrimination of completely novel signals. How these changes in neural representations affect behavioral outcomes remains unclear. Taken together, these findings shed light on how the adult sensory systems retain

neuroplasticity to adapt the organism to rapidly encode and classify sensory signals in an everchanging world. Similar mechanisms may also be engaged during processing of human speech signals, and thus have a significant potential translational relevance to understand the neural underpinnings of speech perception and comprehension difficulties.

REFERENCES

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage, 49*(1), 1124–1132.

Adibi, M., McDonald, J. S., Clifford, C. W. G., & Arabzadeh, E. (2013). Adaptation improves neural coding efficiency despite increasing correlations in variability. *Journal of Neuroscience, 33*(5), 2108–2120.

Althen, H., Grimm, S., & Escera, C. (2013). Simple and complex acoustic regularities are encoded at different levels of the auditory hierarchy. *European Journal of Neuroscience, 38*(10), 3448–3455.

Anderson, L. A., Christianson, G. B., & Linden, J. F. (2009). Stimulus-specific adaptation occurs in the auditory thalamus. *Journal of Neuroscience, 29*(22), 7359–7363.

Anderson, L. A., & Malmierca, M. S. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *European Journal of Neuroscience, 37*(1), 52–62.

Antunes, F. M., & Malmierca, M. S. (2011). Effect of auditory cortex deactivation on stimulus-specific adaptation in the medial geniculate body. *Journal of Neuroscience, 31*(47), 17306–17316.

Antunes, F. M., Nelken, I., Covey, E., & Malmierca, M. S. (2010). Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. *PLoS ONE, 5*(11).

Bartho, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., & Buzsaki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology, 92*(1), 600–608.

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution, 20*(3), 143–149.

Bell, B. A., Phan, M. L., & Vicario, D. S. (2015). Neural responses in songbird forebrain reflect learning rates, acquired salience, and stimulus novelty after auditory discrimination training. *Journal of Neurophysiology, 113*(5), 1480–1492.

Boord, R. L. (1968). Ascending projections of the primary cochlear nuclei and nucleus laminaris in the pigeon. *Journal of Comparative Neurology, 133*(4), 523–541.

Bouchard, M., Gillet, P.-C., Shumikhina, S., & Molotchnikoff, S. (2008). Adaptation changes the spatial frequency tuning of adult cat visual cortex neurons. *Experimental Brain Research, 188*(2), 289–303.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

Brenowitz, E. A. (1991). Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science, 251*(4991), 303–5.

Calabrese, A., & Woolley, S. M. N. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences, 112*(11), 3517–3522.

Catchpole, C., & Slater, P. (1995). *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press.

Cheng, H. Y., & Clayton, D. F. (2004). Activation and habituation of extracellular signal-regulated kinase phosphorylation in zebra finch auditory forebrain during song presentation. *Journal of Neuroscience*, *24*(34), 7503–7513.

Chew, S. J., Mello, C., Nottebohm, F., Jarvis, E., & Vicario, D. S. (1995). Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(8), 3406–10.

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996a). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(5), 1950–5.

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996b). Quantal duration of auditory memories. *Science*, *274*(5294), 1909–14.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Clayton, N. S. (1988). Song discrimination learning in zebra finches. *Animal Behaviour*, *36*(4), 1016–1024.

Cornella, M., Leung, S., Grimm, S., & Escera, C. (2012). Detection of simple and pattern regularity violations occurs at different levels of the auditory hierarchy. *PLoS ONE*, *7*(8).

Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, *29*(5), 655.

Dagostin, A. A., Lovell, P. V, Hilscher, M. M., Mello, C. V, & Leão, R. M. (2015). Control of phasic firing by a background leak current in avian forebrain auditory neurons. *Frontiers in Cellular Neuroscience*, *9*, 471.

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*, *66*(6), 937–948.

Dahmen, J. C., & King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Current Opinion in Neurobiology*, *17*(4), 456–464.

Dean, I., Harper, N. S., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, *8*(12), 1684–1689.

Del Negro, C., Gahr, M., Leboucher, G., & Kreutzer, M. (1998). The selectivity of sexual responses to song displays: effects of partial chemical lesion of the HVC in female canaries. *Behavioural Brain Research*, *96*(1–2), 151–9.

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 599–609.

Dong, S., & Clayton, D. F. (2009). Habituation in songbirds. *Neurobiology of Learning*

and Memory, *92*(2), 183–188.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631.

Duque, D., Perez-Gonzalez, D., Ayala, Y. A., Palmer, A. R., & Malmierca, M. S. (2012). Topographic distribution, frequency, and intensity dependence of stimulus-specific adaptation in the inferior colliculus of the rat. *Journal of Neuroscience*, *32*(49), 17762–17774.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*(21), 7179–7186.

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, *33*(26), 10688–10697.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*(5814), 972–976.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463.

Ghisovan, N., Nemri, A., Shumikhina, S., & Molotchnikoff, S. (2009). Long adaptation reveals mostly attractive shifts of orientation tuning in cat primary visual cortex. *Neuroscience*, *164*(3), 1274–1283.

Grimm, S., Escera, C., & Nelken, I. (2016). Early indices of deviance detection in humans and animal models. *Biological Psychology*, *116*, 23–27.

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, *7*(January), 1–16.

Guediche, S., Fiez, J. A., & Holt, L. L. (2016). Adaptive plasticity in speech perception: Effects of external information and internal predictions. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 1048–1059.

Guediche, S., Holt, L. L., Laurent, P., Lim, S. J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, *25*(7), 1867–1877.

Huesmann, G. R., & Clayton, D. F. (2006). Dynamic role of postsynaptic caspase-3 and BIRC4 in zebra finch song-response habituation. *Neuron*, *52*(6), 1061–1072.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956.

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R. A. Hinde (Ed.), *Bird Vocalizations* (pp. 61–74). London: Cambridge UP.

Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2013). Associative learning enhances

population coding by inverting interneuronal correlation patterns. *Neuron*, *78*(2), 352–363.

Karten, H. J. (1967). The organization of the ascending auditory pathway in the pigeon (Columba livia) I. Diencephalic projections of the inferior colliculus (nucleus mesencephali lateralis, pars dorsalis). *Brain Research*, *6*(3), 409–427.

Kleinschmidt, D. F., & Jaeger, T. F. (2016). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*, *23*(3), 678–691.

Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology*, *16*(2), 152–158.

Kozlov, A. S., & Gentner, T. Q. (2016). Central auditory neurons have composite receptive fields. *Proceedings of the National Academy of Sciences*, *113*(5), 1441–1446.

Kruse, A. A., Stripling, R., & Clayton, D. F. (2004). Context-specific habituation of the zenk gene response to song in adult zebra finches. *Neurobiology of Learning and Memory*, *82*(2), 99–108.

Kvale, M. N., & Schreiner, C. E. (2003). Short-term adaptation of auditory receptive fields to dynamic stimuli. *Journal of Neurophysiology*, *91*(2), 604–612.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America*, *112*(6), 3022–3030.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783–1798.

Lu, K., & Vicario, D. S. (2014). Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences*, *111*(40), 14553–14558.

Lu, K., & Vicario, D. S. (2017). Familiar but unexpected: Effects of sound context statistics on auditory responses in the songbird forebrain. *Journal of Neuroscience*, *5722*(12).

Malmierca, M. S., Anderson, L. A., & Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. *Frontiers in Systems Neuroscience*, *9*, 1–14.

Malmierca, M. S., Cristaudo, S., Perez-Gonzalez, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *Journal of Neuroscience*, *29*(17), 5483–5493.

Meliza, C. D., & Margoliash, D. (2012). Emergence of selectivity and tolerance in the avian auditory cortex. *Journal of Neuroscience*, *32*(43), 15158–15168.

Mello, C., Nottebohm, F., & Clayton, D. (1995). Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in

zebra finch telencephalon. *Journal of Neuroscience*, *15*(10), 6919–25.

Mello, C. V, & Clayton, D. F. (1994). Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *Journal of Neuroscience*, *14*(11), 6652–66.

Mello, C. V, Vicario, D. S., & Clayton, D. F. (1992). Song presentation induces gene expression in the songbird forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(15), 6818–22.

Miller, D. B. (1979). Long-term recognition of father's song by female zebra finches. *Nature*, *280*, 389–391.

Mooney, R., Prather, J., & Roberts, T. (2008). Neurophysiology of birdsong learning. In H. Eichenbaum (Ed.), *Memory Systems* (pp. 441–474). Oxford, UK: Elsevier.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Muller, C. M., & Leppelsack, H. J. (1985). Feature extraction and tonotopic organization in the avian auditory forebrain. *Experimental Brain Research*, *59*(3).

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329.

Natan, R. G., Briguglio, J. J., Mwilambwe-Tshilobo, L., Jones, S. I., Aizenberg, M., Goldberg, E. M., & Geffen, M. N. (2015). Complementary control of sensory adaptation by two types of cortical interneurons. *eLife*, *4*, 1–27.

Nelken, I. (2014). Stimulus-specific adaptation and deviance detection in the auditory system: Experiments and models. *Biological Cybernetics*, *108*(5), 655–663.

Nelken, I., Yaron, A., Polterovich, A., & Hershenhoren, I. (2013). Stimulus-specific adaptation beyond pure tones. In B. C. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing* (pp. 411–418). New York, NY: Springer.

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, *26*(4), 844–851.

Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: an overview. *Frontiers in Integrative Neuroscience*, *8*, 1–10.

Phan, M. L., Gergues, M. M., Mahidadia, S., Jimenez-Castillo, J., Vicario, D. S., & Bieszczad, K. M. (2017). HDAC3 inhibitor RGFP966 modulates neuronal memory for vocal communication signals in a songbird model. *Frontiers in Systems Neuroscience*, *11*, 1–12.

Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1088–93.

Pinaud, R., & Mello, C. V. (2007). GABA immunoreactivity in auditory and song control brain areas of zebra finches. *Journal of Chemical Neuroanatomy*, *34*(1–2), 1–21.

Pinaud, R., Terleph, T. a, Tremere, L. a, Phan, M. L., Dagostin, A. a, Leão, R. M., … Vicario, D. S. (2008). Inhibitory network interactions shape the auditory processing of natural communication signals in the songbird auditory forebrain. *Journal of Neurophysiology*, *100*(1), 441–55.

Quiroga, R. Q., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, *16*(8), 1661–1687.

Sanes, D. H., & Bao, S. (2009). Tuning up the developing auditory CNS. *Current Opinion in Neurobiology*, *19*(2), 188–199.

Smulders, T. V., & Jarvis, E. D. (2013). Different mechanisms are responsible for dishabituation of electrophysiological auditory responses to a change in acoustic identity than to a change in stimulus location. *Neurobiology of Learning and Memory*, *106*, 163–176.

Sossinka, R., & Böhner, J. (1980). Song types in the zebra finch. *Z. Tierpsychol.*, *53*, 123–132.

Soyman, E., & Vicario, D. S. (2017). Principles of auditory processing differ between sensory and premotor structures of the songbird forebrain. *Journal of Neurophysiology*, *117*(3), 1266–1280.

Stripling, R., Kruse, A. A., & Clayton, D. F. (2001). Development of song responses in the zebra finch caudomedial neostriatum: Role of genomic and electrophysiological activities. *Journal of Neurobiology*, *48*(3), 163–180.

Stripling, R., Volman, S. F., & Clayton, D. F. (1997). Response modulation in the zebra finch neostriatum: Relationship to nuclear gene regulation. *Journal of Neuroscience*, *17*(10), 3883–3893.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, *59*(6), 1167–1176.

Terleph, T. A., Mello, C. V., & Vicario, D. S. (2007). Species differences in auditory processing dynamics in songbird auditory telencephalon. *Developmental Neurobiology*, *67*(11), 1498–1510.

Terleph, T. A., Mello, C. V, & Vicario, D. S. (2006). Auditory topography and temporal response dynamics of canary caudal telencephalon. *Journal of Neurobiology*, *66*(3), 281–92.

Thompson, J. V., & Gentner, T. Q. (2010). Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *Journal of Neurophysiology*, *103*(4), 1785–1797.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398.

van Rossum, M. C. W. (2001). A novel spike distance. *Neural Computation*, *13*(4), 751–763.

Vates, G. E., Broome, B. M., Mello, C. V., & Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (Taenopygia guttata). *Journal of Comparative Neurology*, *366*(4), 613–642.

Vicario, D. S., Naqvi, N. H., & Raksin, J. N. (2001). Sex differences in discrimination of vocal communication signals in a songbird. *Animal Behaviour*, *61*(4), 805–817.

Vignal, C., Andru, J., & Mathevon, N. (2005). Social context modulates behavioural and brain immediate early gene responses to sound in male songbird. *European Journal of Neuroscience*, *22*(4), 949–955.

Vignal, C., Mathevon, N., & Mottin, S. (2004). Audience drives male songbird response to partner's voice. *Nature*, *430*(6998), 448–451.

Vignal, C., Mathevon, N., & Mottin, S. (2008). Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*, *77*(2), 191–198.

Wang, Y., Brzozowska-Prechtl, A., & Karten, H. J. (2010). Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences*, *107*(28), 12676–12681.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, *5*(4), 279–290.

White, L. E., & Fitzpatrick, D. (2007). Vision and cortical map development. *Neuron*, *56*(2), 327–338.

Willmore, B., & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, *12*(3), 255–270.

Zann, R. (1985). Ontogeny of the zebra finch distance call: I. Effects of cross-fostering to bengalese finches. *Zeitschrift Für Tierpsychologie*, *68*(1), 1–23.

Zann, R. (1996). *Zebra finch: A synthesis of field and laboratory studies*. Oxfprd, UK: Oxford University Press.

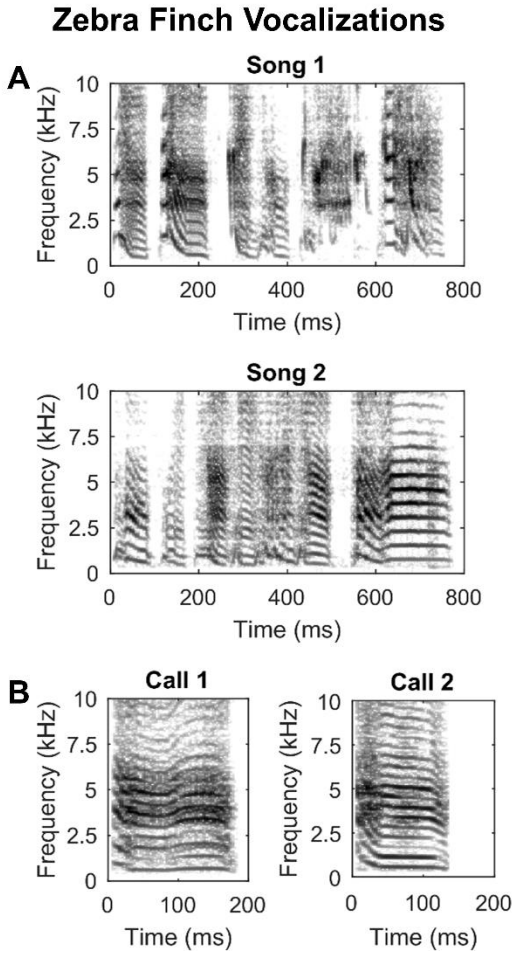

Figure 1. Male zebra finch vocalizations. **A** shows the spectrograms of two songs. **B** shows the spectrograms of two long calls.

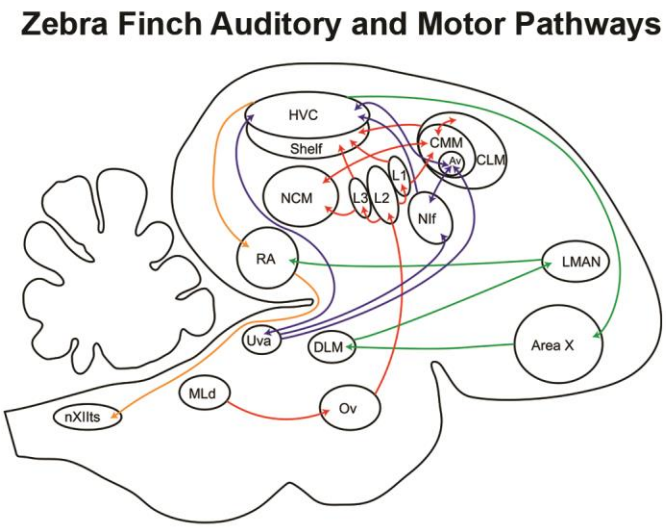

Figure 2. Overview of the zebra finch auditory and motor pathways. Red connections denote the conventional auditory thalamus-driven ascending auditory pathway, blue projections show the ascending auditory and motor feedback pathway, orange connections indicate the song production pathway, and green projections show the corticostriatal anterior forebrain pathway. Abbreviations: HVC (used as a proper name); Shelf: HVC Shelf; NCM: Caudal Medial Nidopallium; CMM: Caudal Medial Mesopallium; CLM: Caudal Lateral Mesopallium; Av: Avalanche; L1, L2, and L3: Subregions 1, 2, and 3 of the general Field L region; Nif: Nucleus Interfacialis of the Nidopallium; RA: Robust Nucleus of the Archopallium; LMAN: Lateral Magnocellular Nucleus of the Anterior Neostriatum; Area X (used as a proper name); Uva: Uveaformis; DLM: Medial Nucleus of the Dorsomedial Thalamus; Ov: Nucleus Ovoidalis; MLd: Mesencephalicus Lateralis Dorsalis; nXIIts: Tracheosyringeal Motor Nucleus.

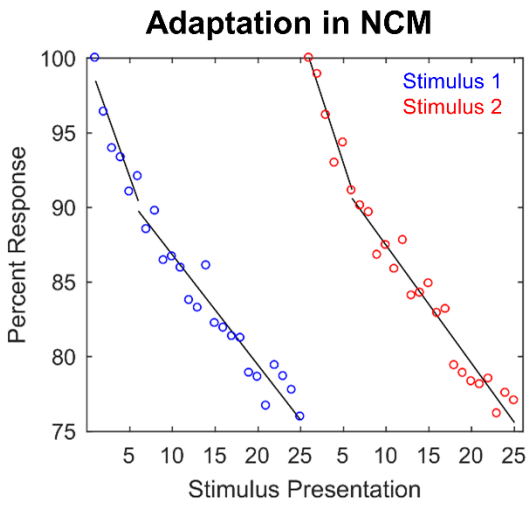

Figure 3. Adaptation of neural responses to repeated stimulus presentation in NCM. Adaptation rates are typically quantified via linear regression methods for the first 6 and the remaining 20 presentations separately.

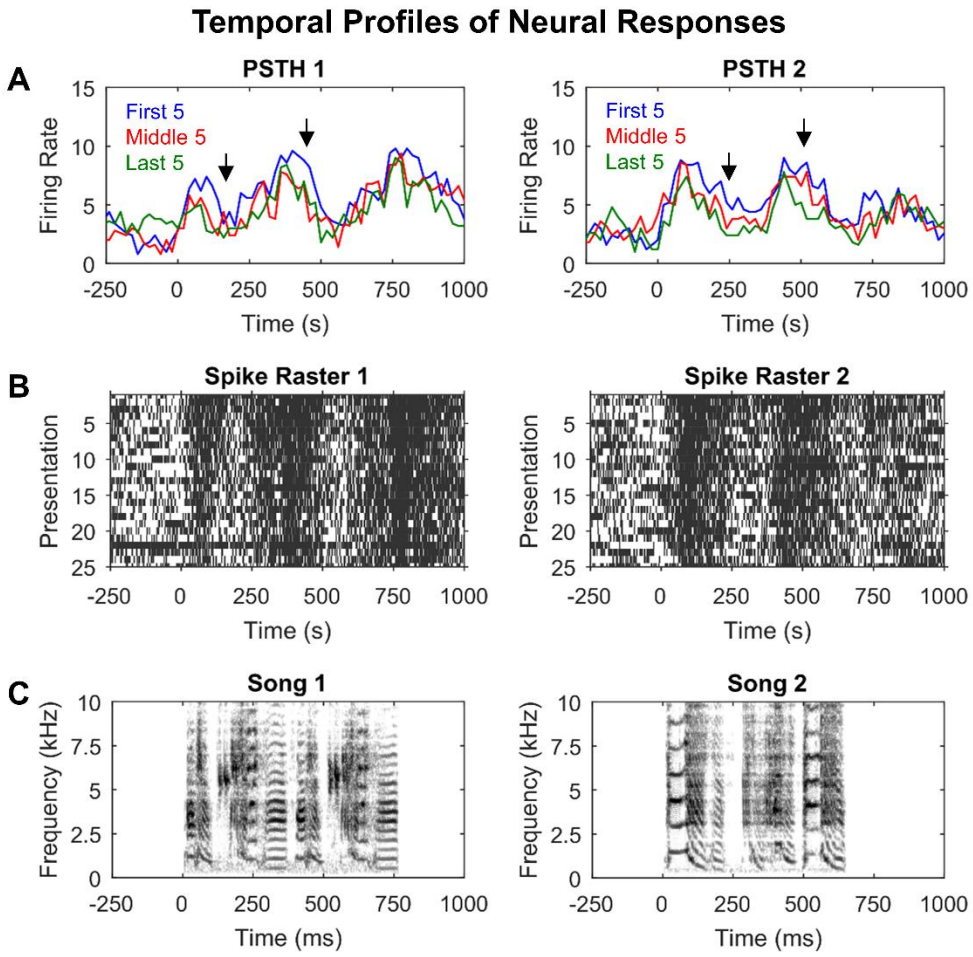

Figure 4. Temporal profiles of multi-unit neural responses to two different songs in NCM. **A** shows the peristimulus time histograms for the first 5, middle 5, and last 5 stimulus presentations. Arrows indicate example time points that show strong adaptation. **B** shows the spike rasters for the same recordings. **C** shows the spectrograms of the two songs.

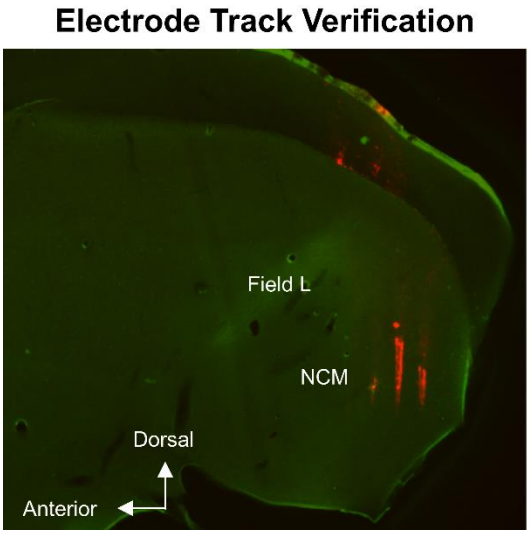

Figure 5. Histological verification of electrode tracks. Field L is identified by its densely packed neurons. NCM is the region posterior to field L. Red marks show the shanks of the silicon probe.

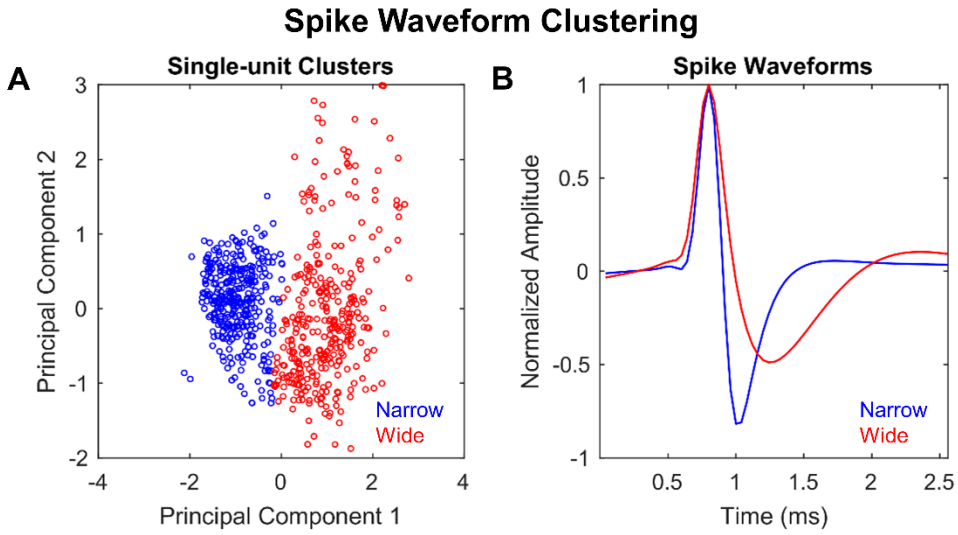

Figure 6. Clustering of single-units based on spike waveforms. **A** shows the single-unit clusters in the first 2 principal component space. **B** shows the average waveforms of narrow and wide spike neurons.

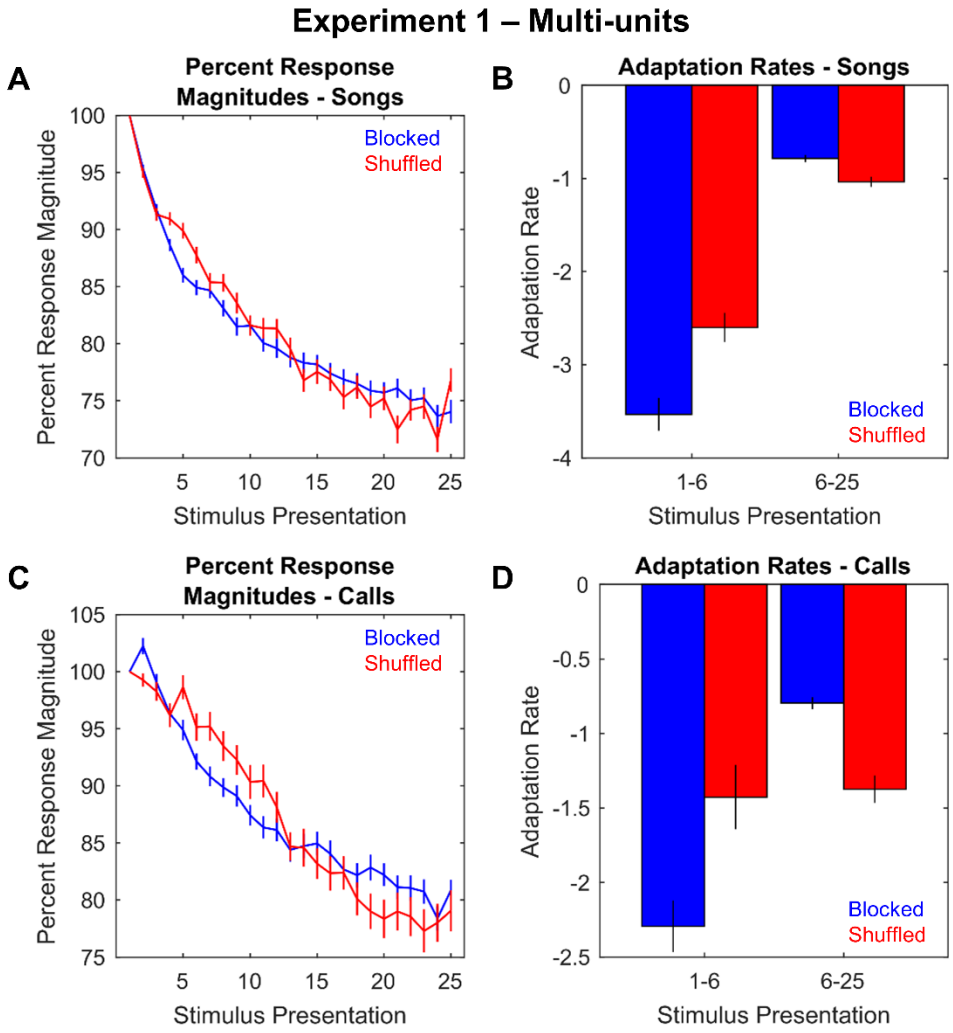

Figure 7. Multi-unit adaptation profiles for songs and calls in blocked and shuffled sequences. **A** shows the percent response magnitudes as a function of stimulus presentation for songs in the two sequences. **B** shows the adaptation rates for presentations 1-6 and 6-25 for songs in the two sequences. **C** and **D** are similar to A and B, respectively, but they show the same metrics for calls. Plotted values are means ± SEMs across multi-units.

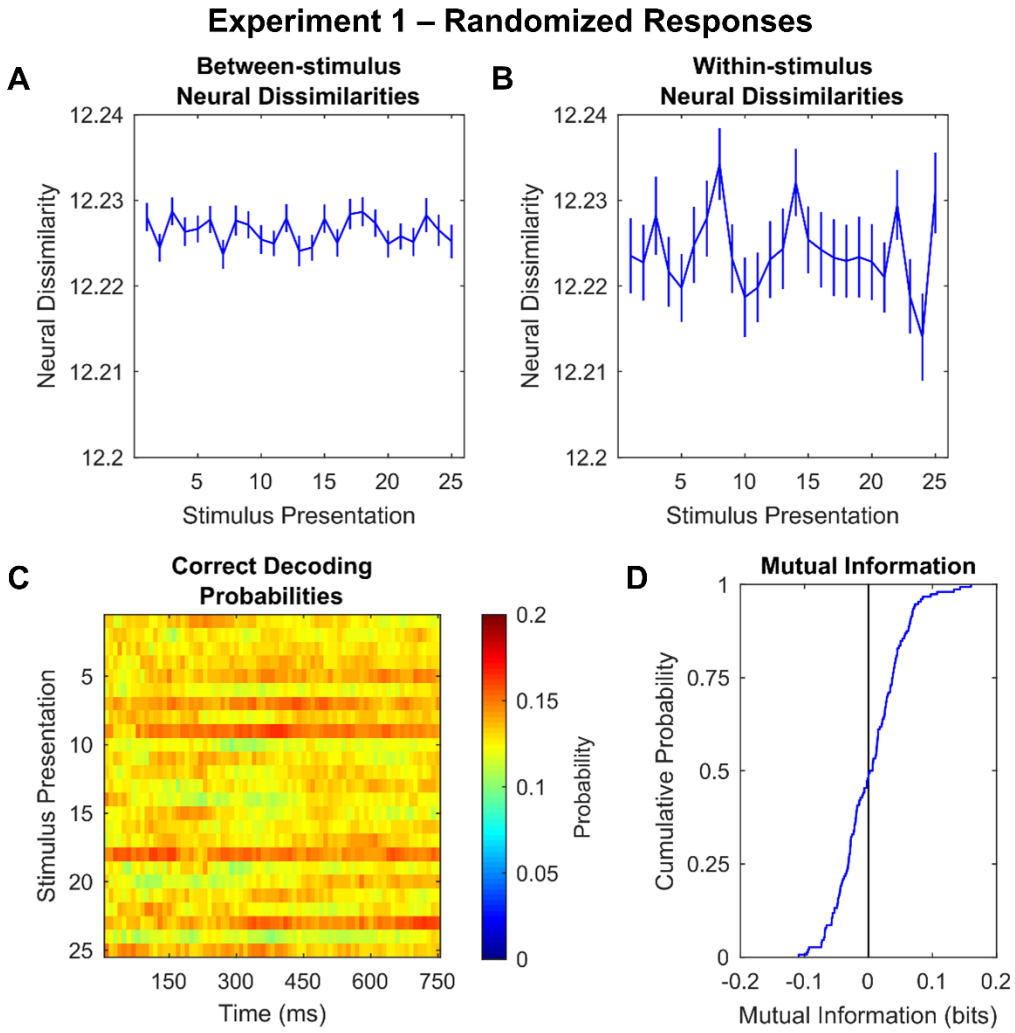

Figure 8. Validation of the neural discrimination metrics via the analysis of simulated data that randomizes the timings of spikes, while keeping the absolute firing rates constant. **A** shows the between-stimulus neural dissimilarities as a function of stimulus presentation. **B** shows the within-stimulus neural dissimilarities in a similar way. **C** shows the correct decoding probabilities across time points along the stimulus duration and across stimulus presentations. **D** shows the bias-corrected mutual information estimations.

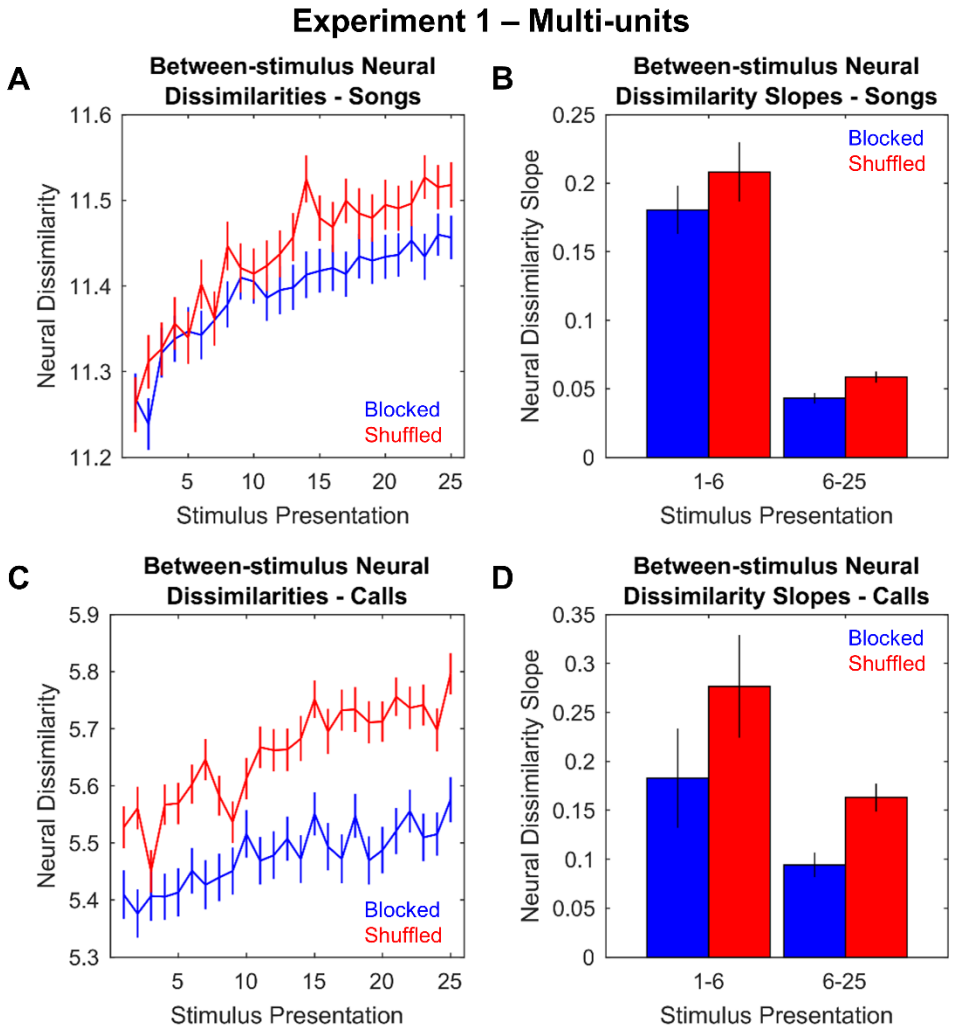

Figure 9. Multi-unit between-stimulus neural dissimilarities for songs and calls in blocked and shuffled sequences. **A** shows the between-stimulus neural dissimilarities as a function of stimulus presentation for songs in the two sequences. **B** shows the slopes of between-stimulus neural dissimilarities for presentations 1-6 and 6-25 for songs in the two sequences. **C** and **D** are similar to A and B, respectively, but they show the same metrics for calls.

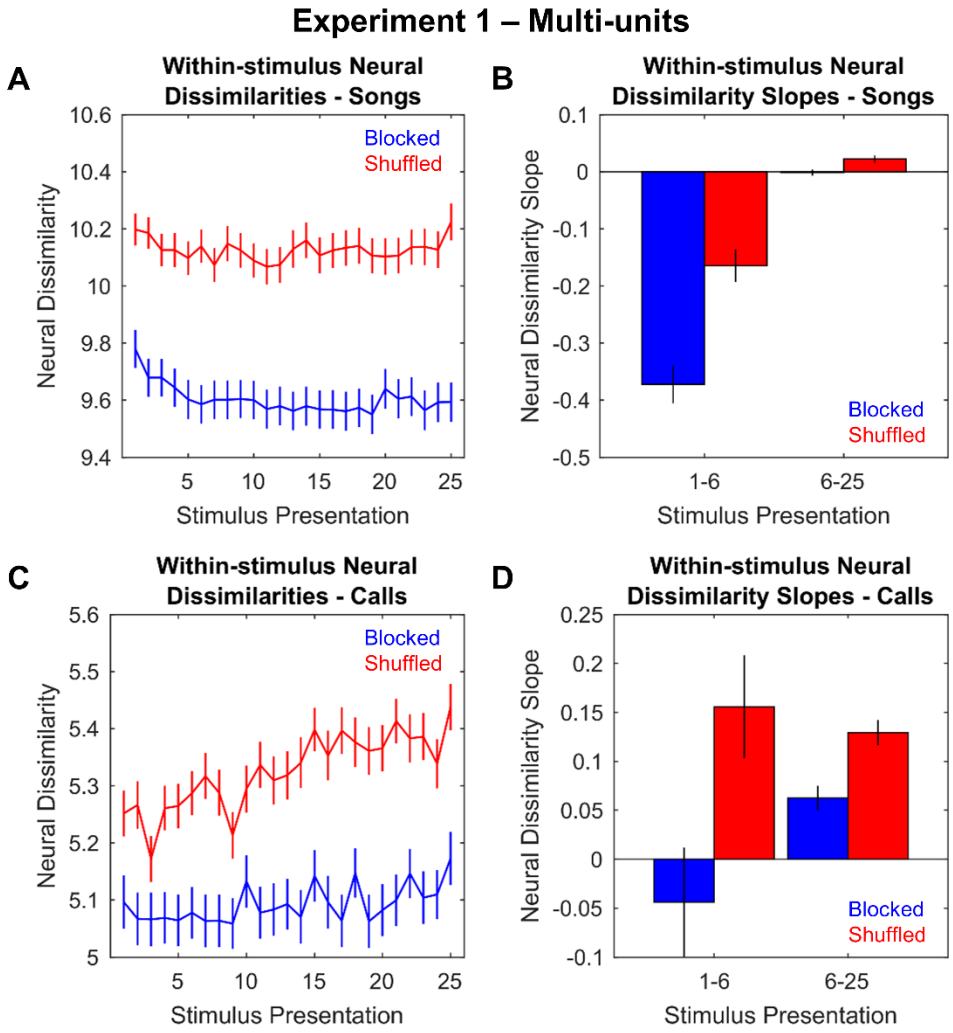

Figure 10. Multi-unit within-stimulus neural dissimilarities for songs and calls in blocked and shuffled sequences. **A** shows the within-stimulus neural dissimilarities as a function of stimulus presentation for songs in the two sequences. **B** shows the slopes of within-stimulus neural dissimilarities for presentations 1-6 and 6-25 for songs in the two sequences. **C** and **D** are similar to A and B, respectively, but they show the same metrics for calls.

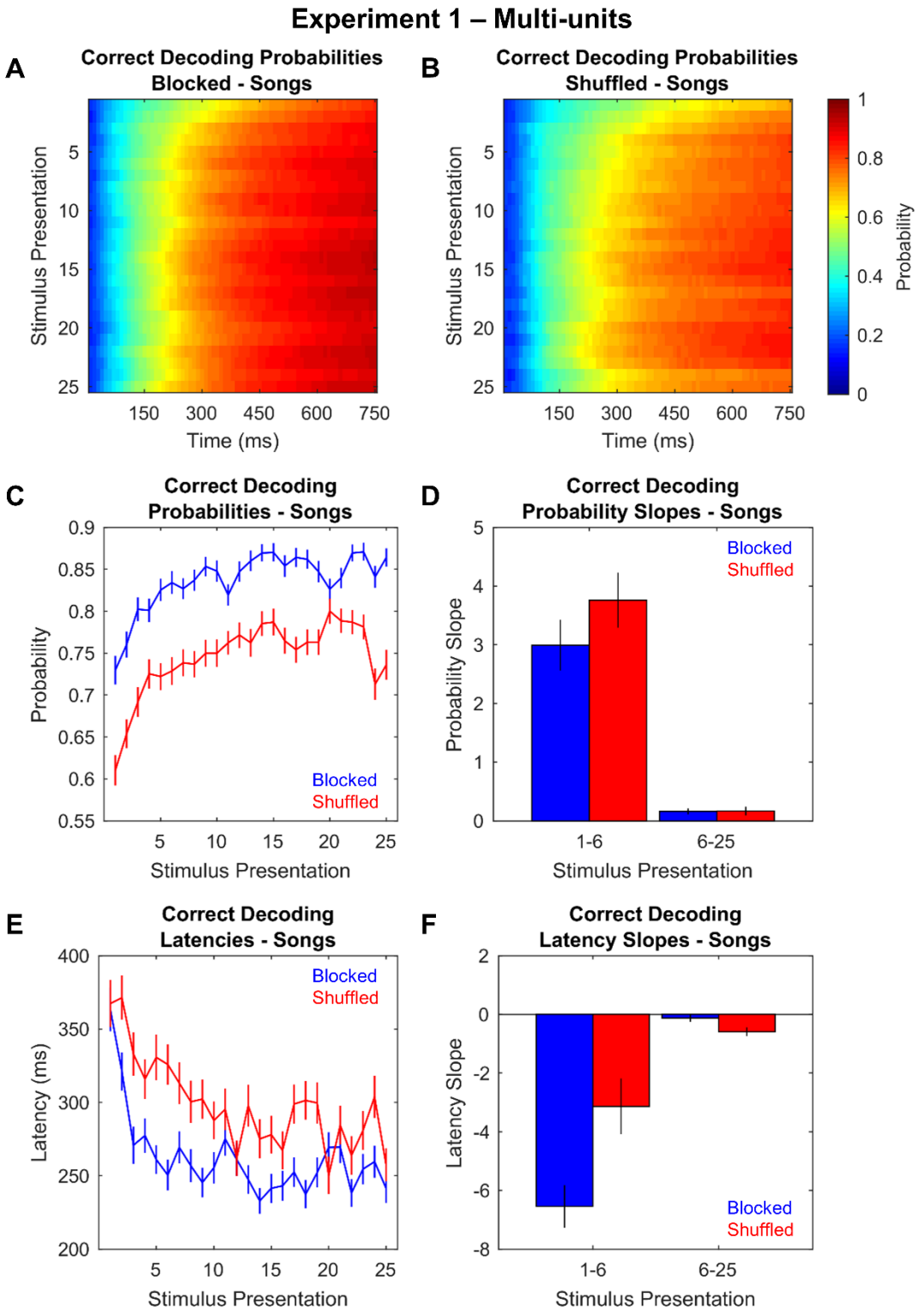

Figure 11. Multi-unit neural decoding accuracies for songs in blocked and shuffled sequences. **A** shows the correct decoding probabilities across time points along the stimulus duration and across stimulus presentations averaged across multi-units in the blocked sequence. **B** is similar to A, but it shows the same metrics in the shuffled

sequence. **C** shows the correct decoding probabilities (measured at 500 ms) as a function of stimulus presentation in the two sequences. **D** shows the slopes of correct decoding probabilities for presentations 1-6 and 6-25 in the two sequences. **E** and **F** are similar to C and D, respectively, but they show correct decoding latencies (to achieve a probability of .75) and their slopes.

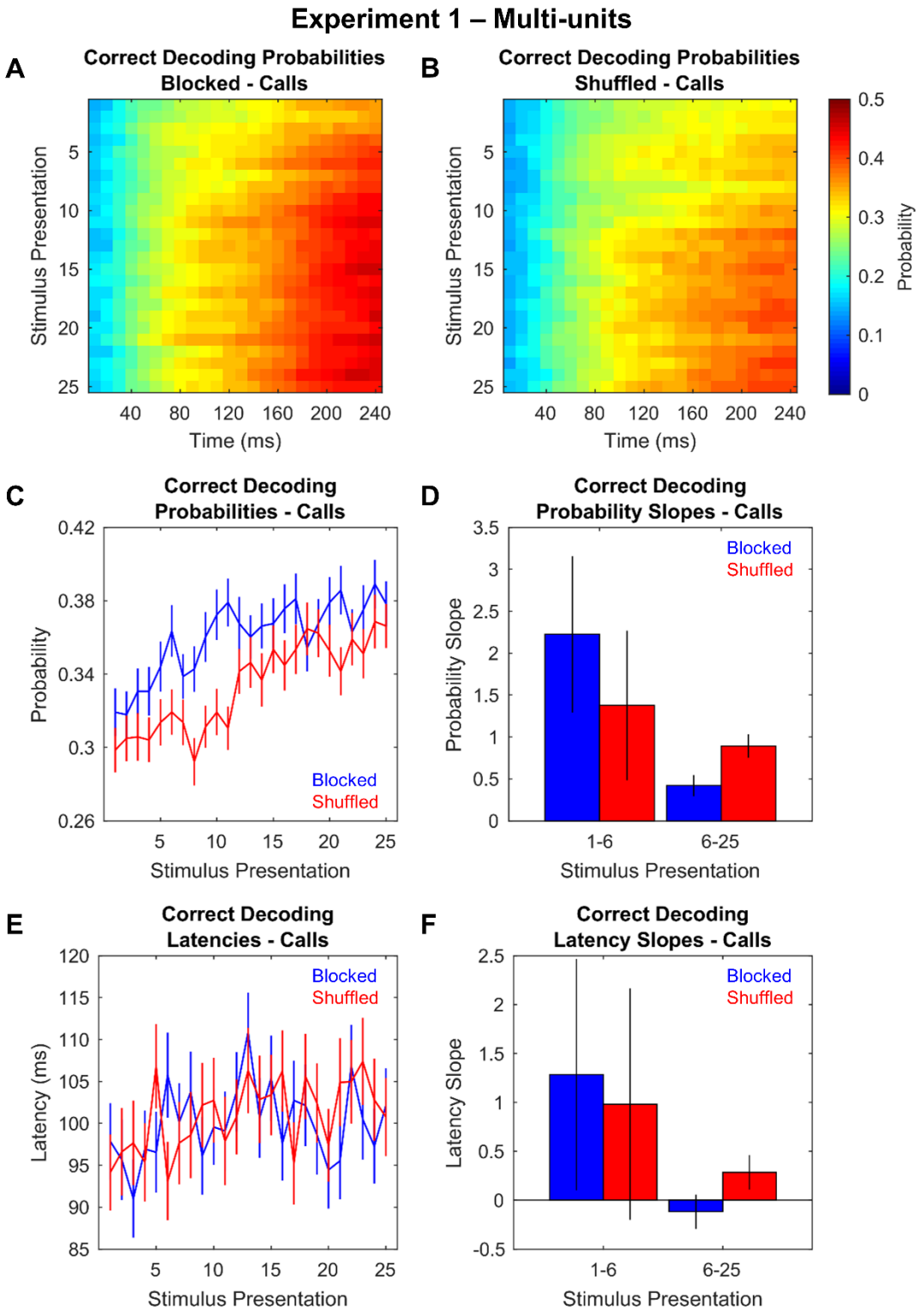

Figure 12. Multi-unit neural decoding accuracies for calls in blocked and shuffled sequences. **A** shows the correct decoding probabilities across time points along the stimulus duration and across stimulus presentations in the blocked sequence. **B** is similar to A, but it shows the same metrics in the shuffled sequence. **C** shows the correct

decoding probabilities (measured at 160 ms) as a function of stimulus presentation in the two sequences. **D** shows the slopes of correct decoding probabilities for presentations 1-6 and 6-25 in the two sequences. **E** and **F** are similar to C and D, respectively, but they show correct decoding latencies (to achieve a probability of .50) and their slopes.

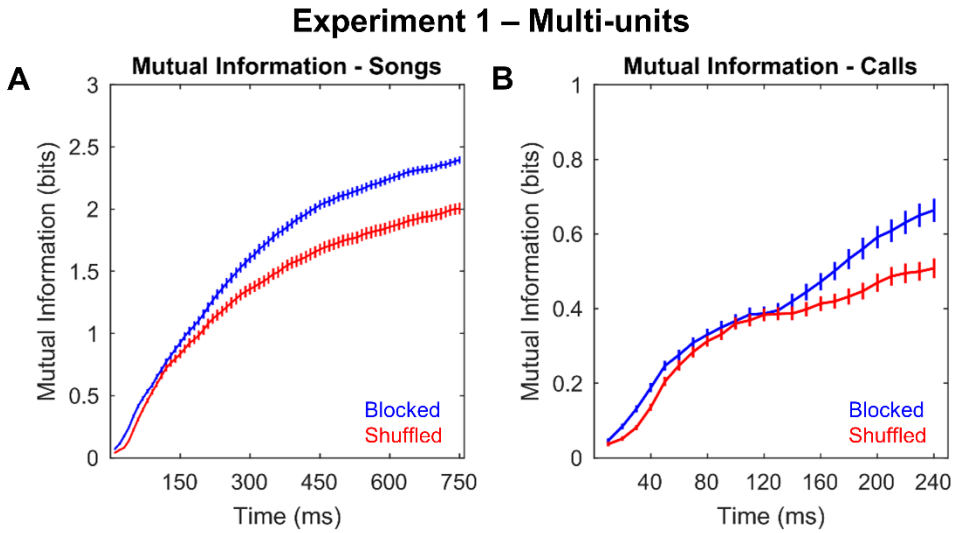

Figure 13. Multi-unit mutual information for songs and calls in blocked and shuffled sequences. **A** shows the bias-corrected mutual information estimations as a function of time along the stimulus duration for songs in the two sequences. **B** is similar to A, but it shows the same metric for calls. Plotted values are means ± SEMs across multi-units for each time bin of 10 ms.

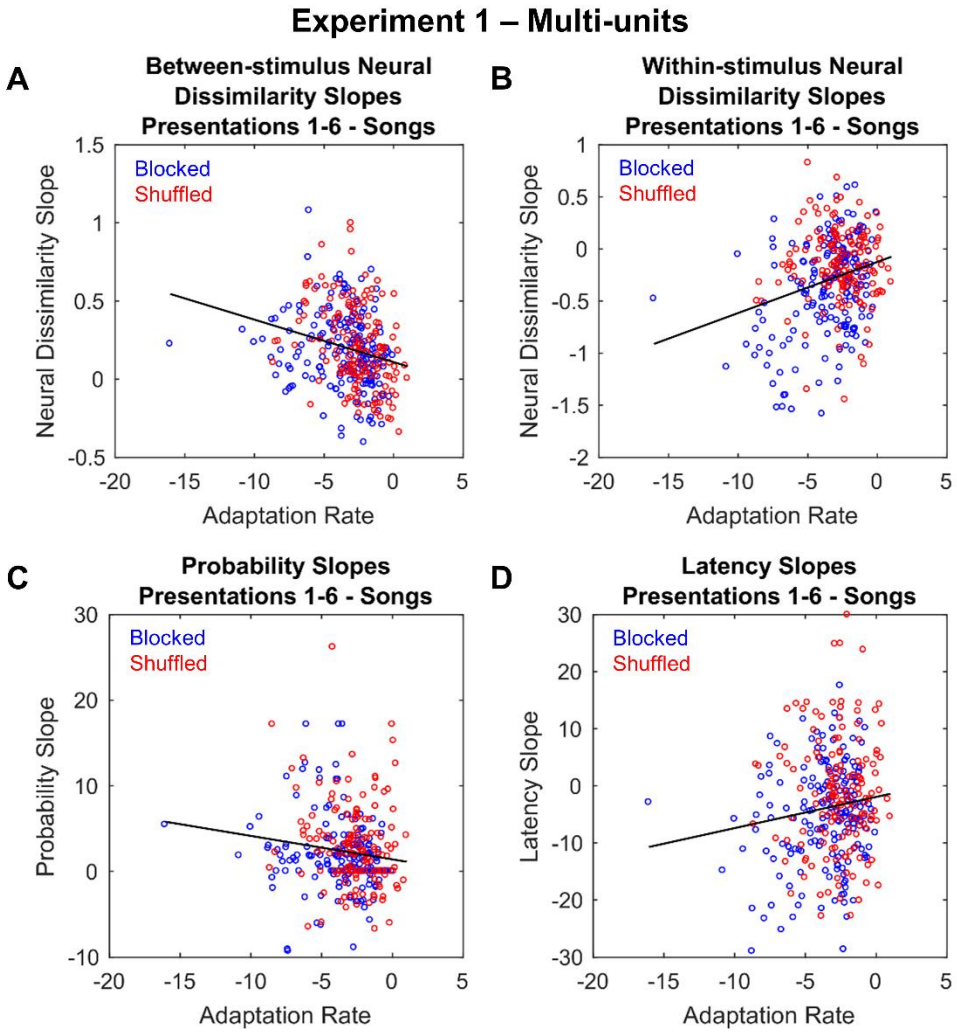

Figure 14. Multi-unit adaptation and neural discrimination relationships for songs for presentations 1-6. **A** shows the correlation between adaptation rates and the slopes of between-stimulus neural dissimilarities. **B**, **C**, and **D** are similar to A, but they show the correlations for the slopes of within-stimulus neural dissimilarities, correct decoding probabilities, and correct decoding latencies, respectively. Black lines show regression lines for blocked and shuffled sequences together.

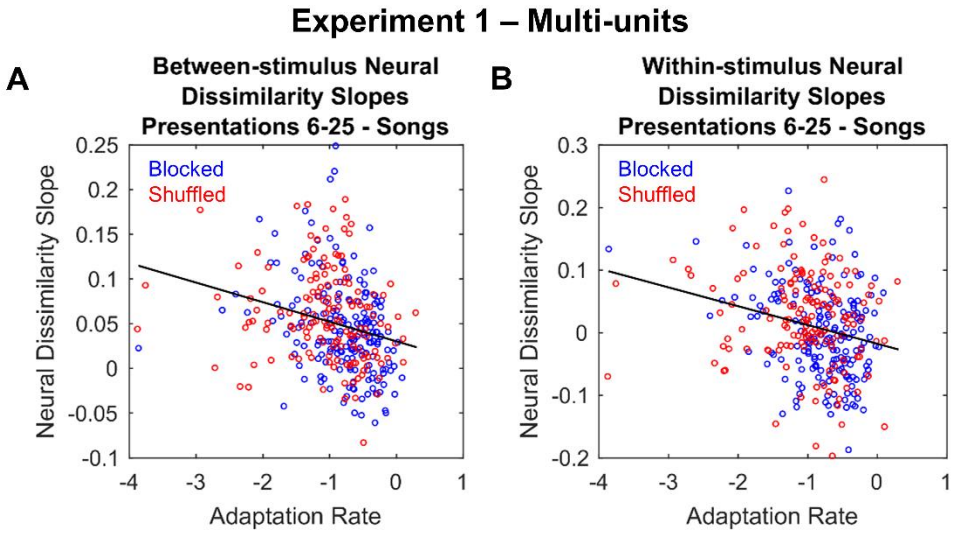

Figure 15. Multi-unit adaptation and neural discrimination relationships for songs for presentations 6-25. **A** shows the correlation between adaptation rates and the slopes of between-stimulus neural dissimilarities. **B** is similar to A, but it shows the correlation for the slopes of within-stimulus neural dissimilarities.

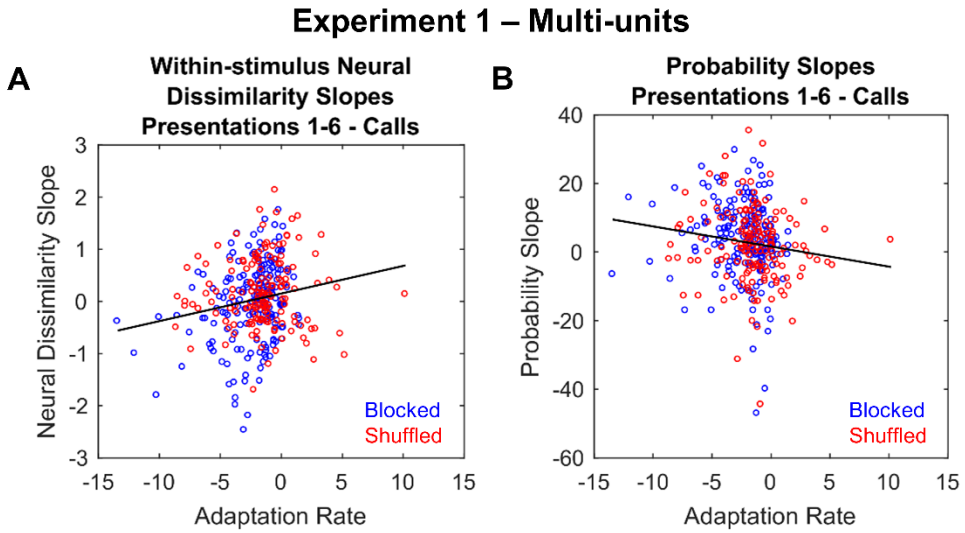

Figure 16. Multi-unit adaptation and neural discrimination relationships for calls for presentations 1-6. **A** shows the correlation between adaptation rates and the slopes of within-stimulus neural dissimilarities. **B** is similar to A, but it shows the correlation for the slopes of correct decoding probabilities.

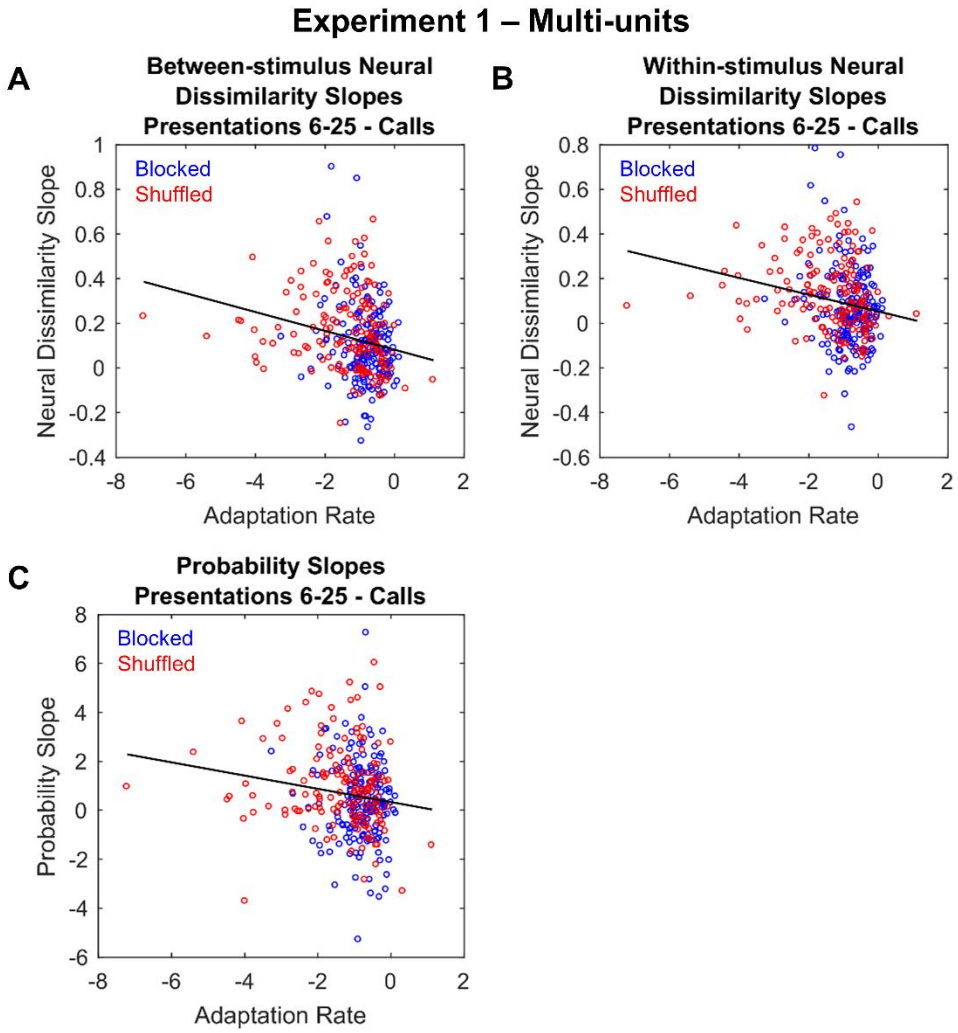

Figure 17. Multi-unit adaptation and neural discrimination relationships for calls for presentations 6-25. **A** shows the correlation between adaptation rates and the slopes of between-stimulus neural dissimilarities. **B** and **C** are similar to A, but they show the correlations for the slopes of within-stimulus neural dissimilarities and correct decoding probabilities, respectively.

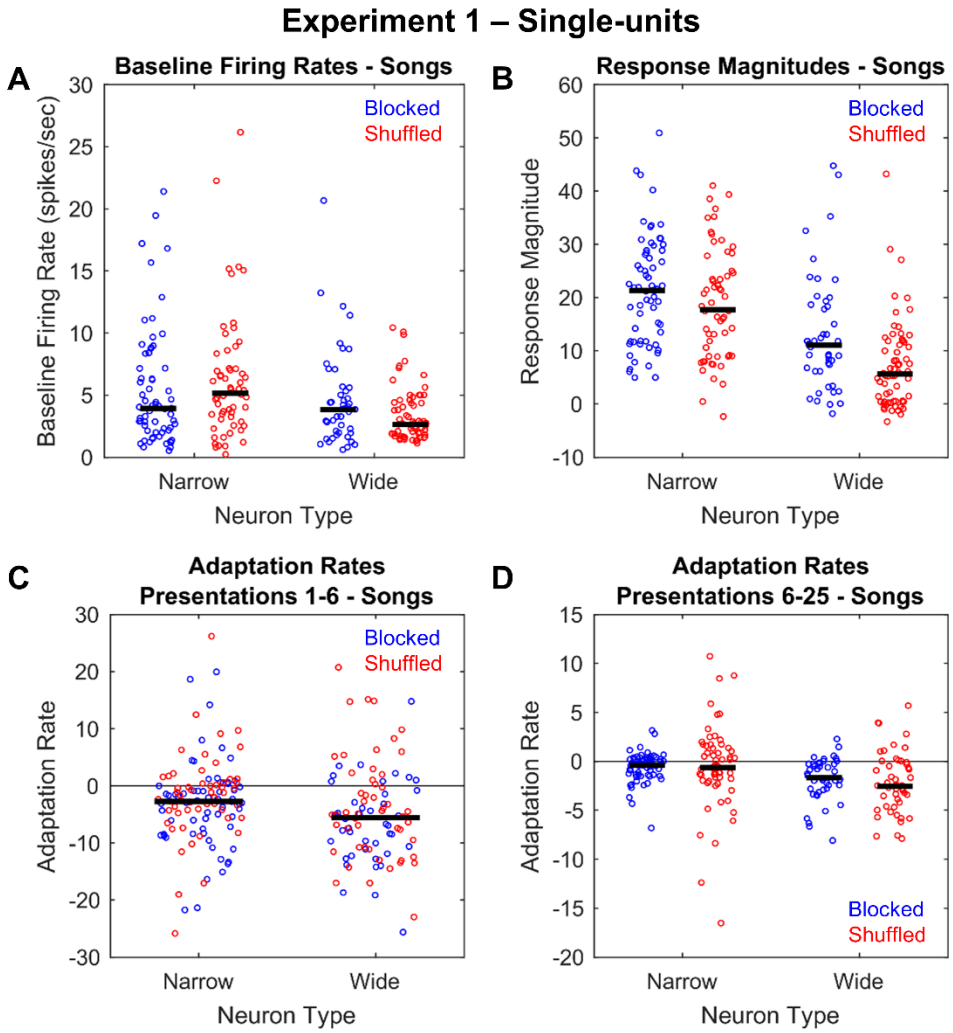

Figure 18. Single-unit response properties for songs for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the baseline firing rates for the two neuron types in the two sequences. **B**, **C**, and **D** are similar to A, but they show response magnitudes, adaptation rates for presentations 1-6, and adaptation rates for presentations 6-25, respectively. Individual data points are jittered along the x axis for illustrative purposes. Black lines show the medians.

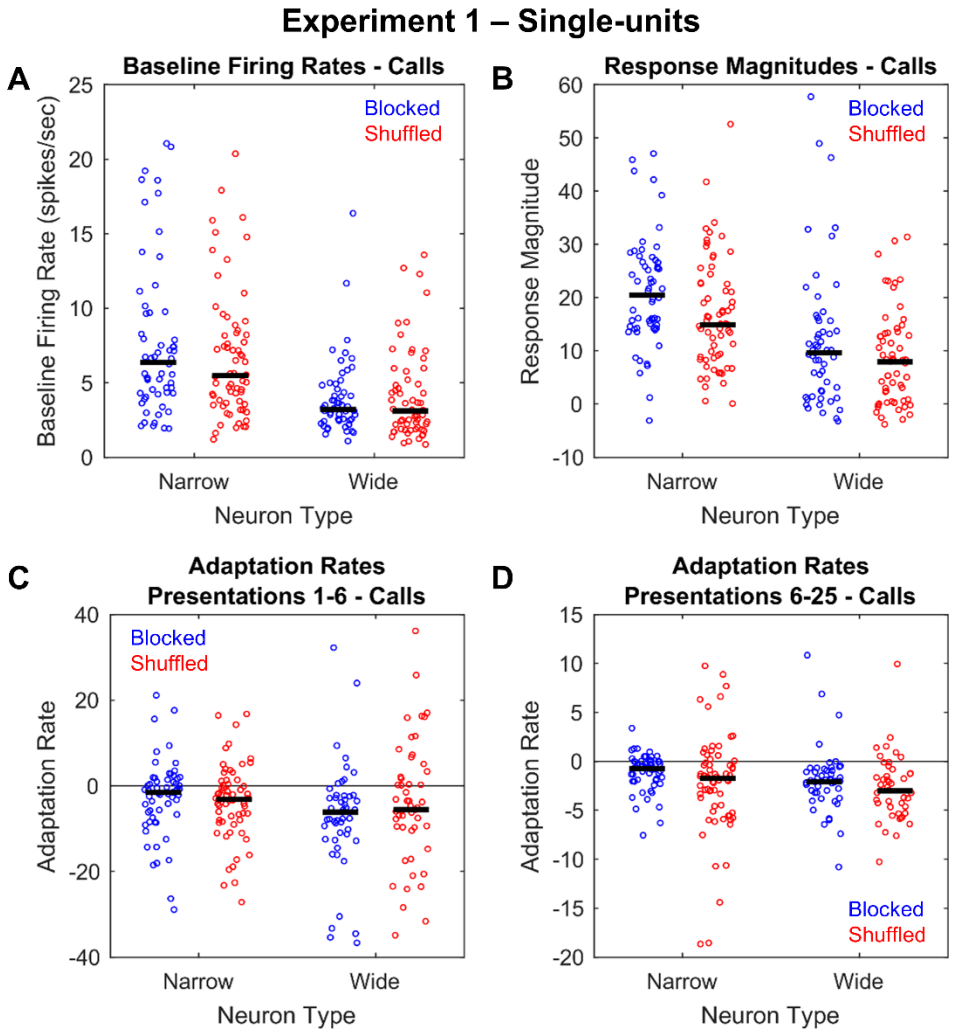

Figure 19. Single-unit response properties for calls for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the baseline firing rates for the two neuron types in the two sequences. **B**, **C**, and **D** are similar to A, but they show response magnitudes, adaptation rates for presentations 1-6, and adaptation rates for presentations 6-25, respectively.

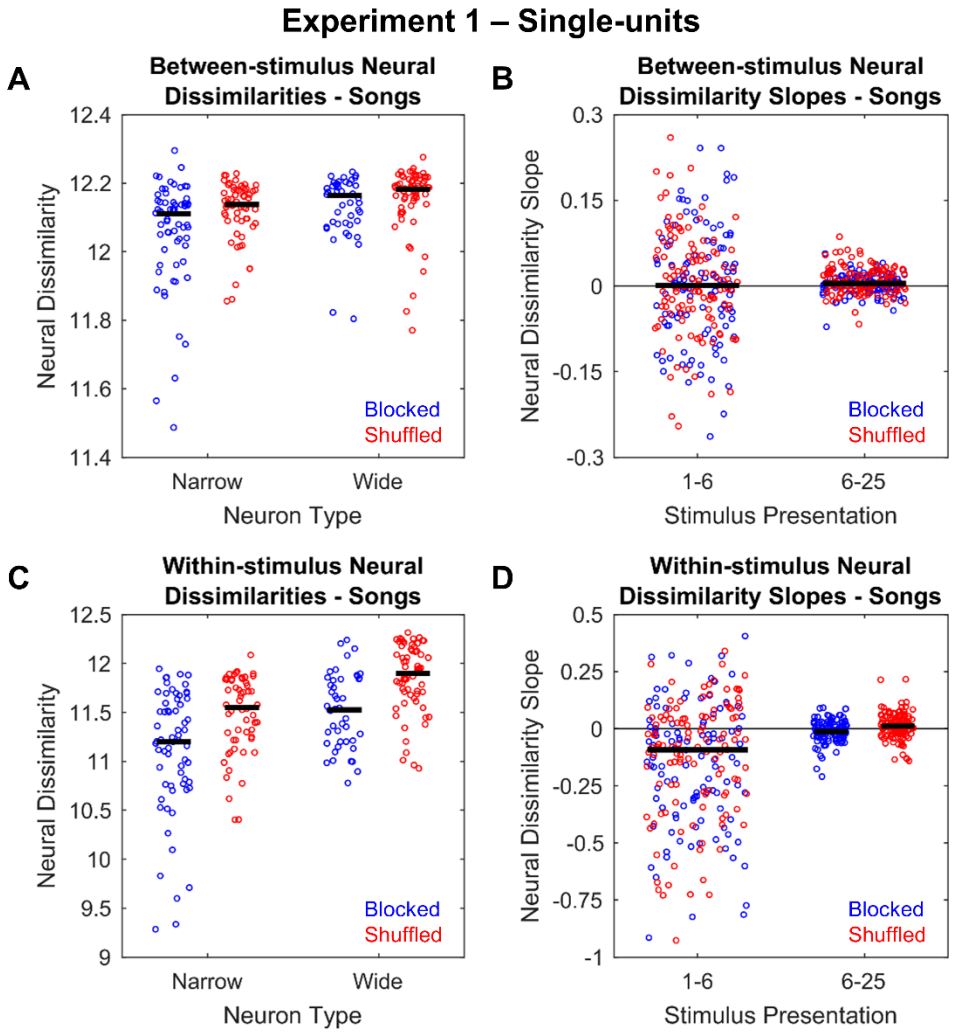

Figure 20. Single-unit neural dissimilarities for songs for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the between-stimulus neural dissimilarities for the two neuron types in the two sequences. **B** shows the slopes of between-stimulus neural dissimilarities for presentations 1-6 and 6-25. **C** and **D** are similar to A and B, respectively, but they show within-stimulus neural dissimilarities and their slopes.

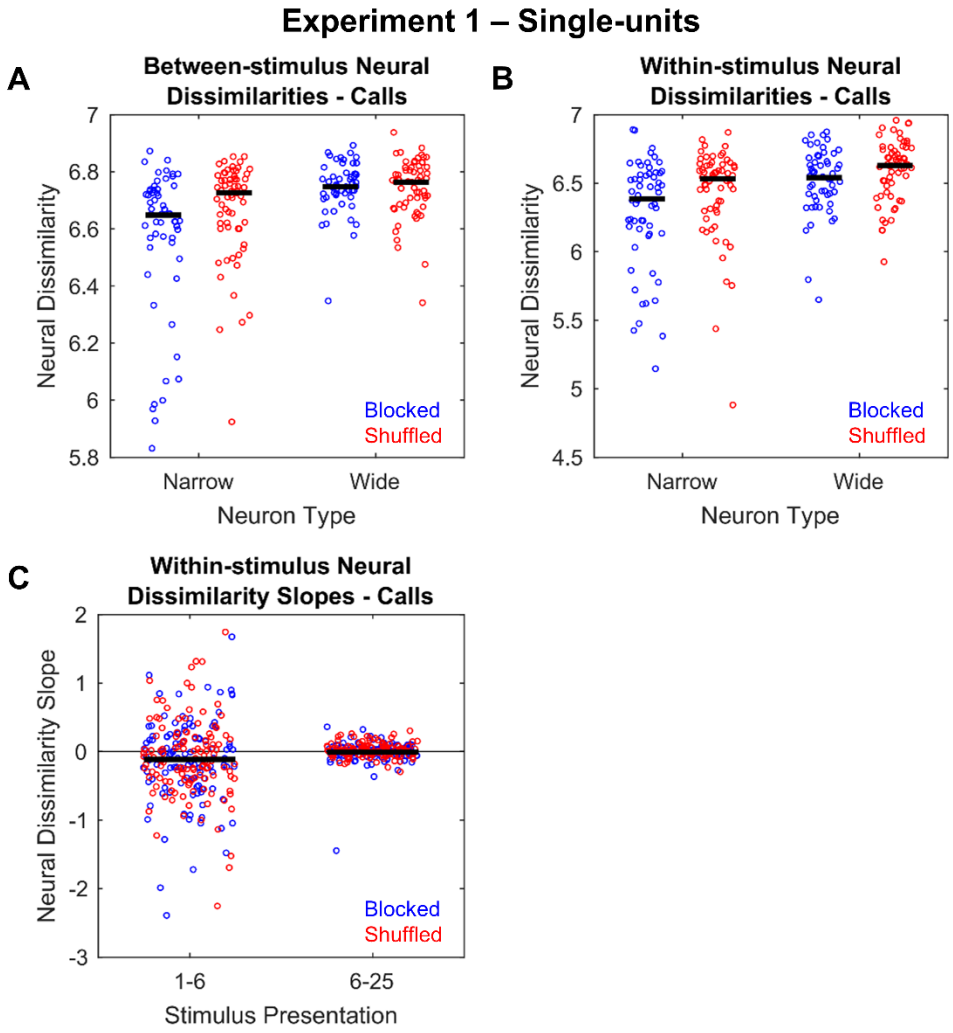

Figure 21. Single-unit neural dissimilarities for calls for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the between-stimulus neural dissimilarities for the two neuron types in the two sequences. **B** is similar to A, but it shows within-stimulus neural dissimilarities. **C** shows the slopes of within-stimulus neural dissimilarities for presentations 1-6 and 6-25.

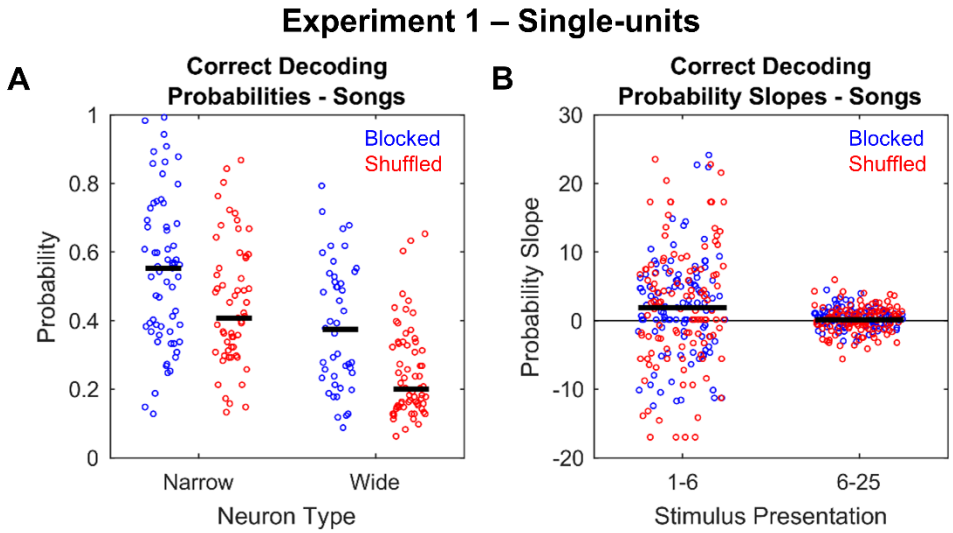

Figure 22. Single-unit neural decoding accuracies for songs for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the correct decoding probabilities for the two neuron types in the two sequences. **B** shows the slopes of correct decoding probabilities for presentations 1-6 and 6-25.

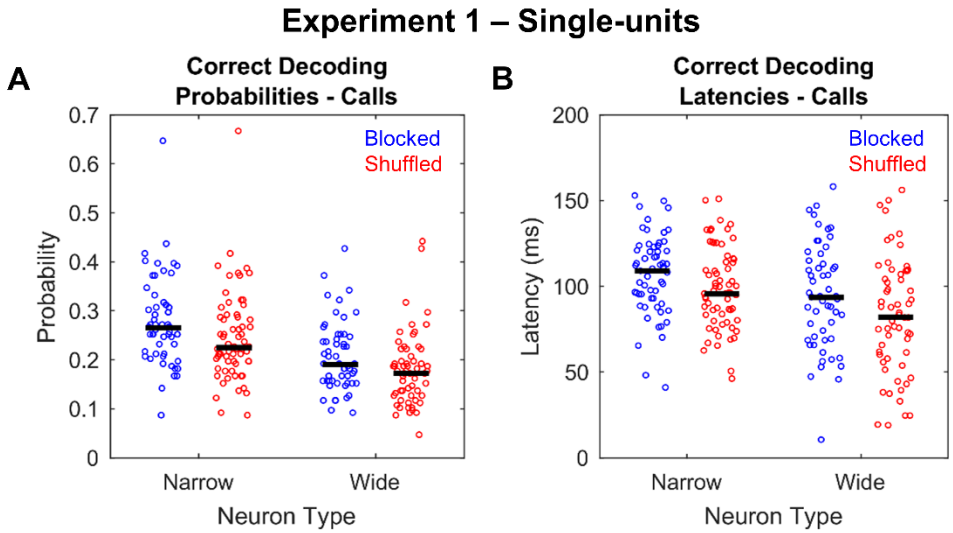

Figure 23. Single-unit neural decoding accuracies for calls for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the correct decoding probabilities for the two neuron types in the two sequences. **B** is similar to A, but it shows correct decoding latencies.

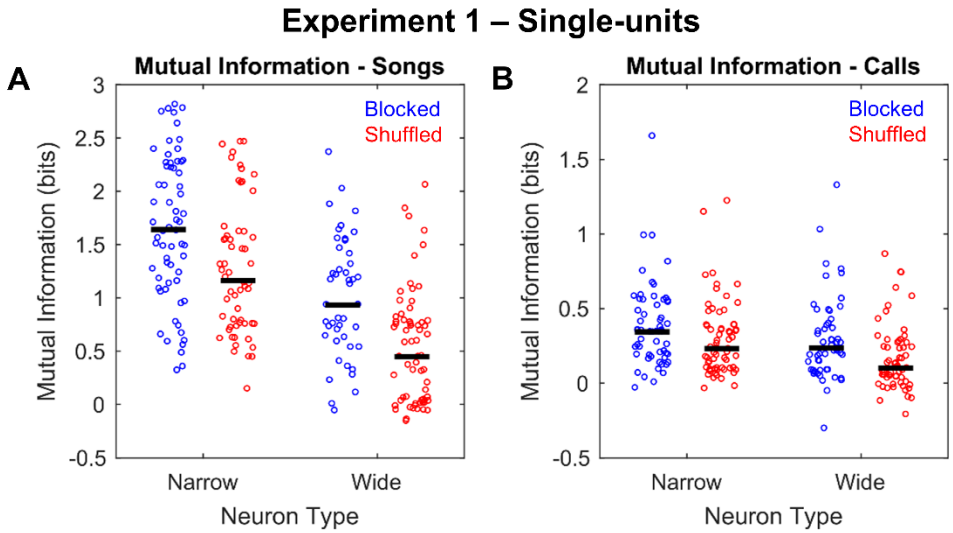

Figure 24. Single-unit mutual information for songs and calls for narrow and wide spike neurons in blocked and shuffled sequences. **A** shows the bias-corrected mutual information estimations for songs for the two neuron types in the two sequences. **B** is similar to A, but it shows the same metric for calls.

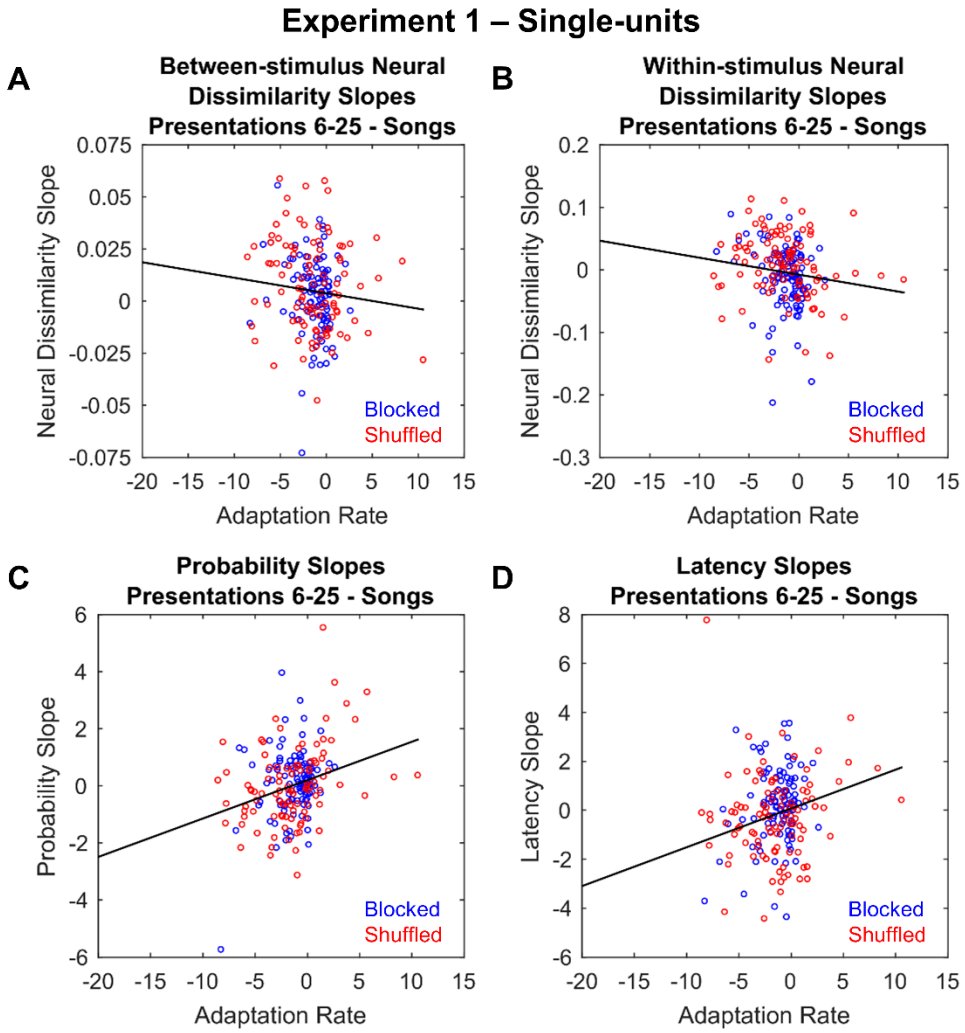

Figure 25. Single-unit adaptation and neural discrimination relationships for songs for presentations 6-25. **A** shows the correlation between adaptation rates and the slopes of between-stimulus neural dissimilarities. **B**, **C**, and **D** are similar to A, but they show the correlations for the slopes of within-stimulus neural dissimilarities, correct decoding probabilities, and correct decoding latencies, respectively.

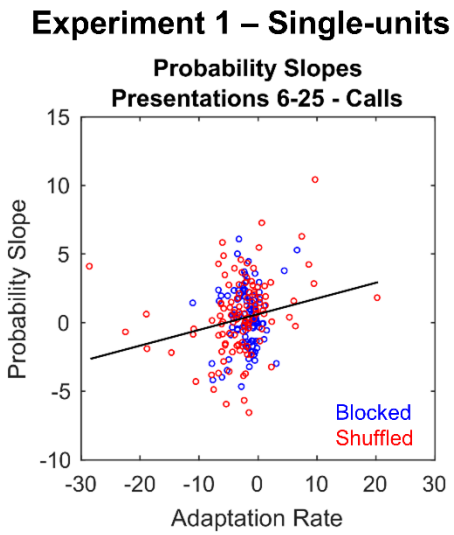

Figure 26. Single-unit correlation between adaptation rates and the slopes of correct decoding probabilities for calls for presentations 6-25.

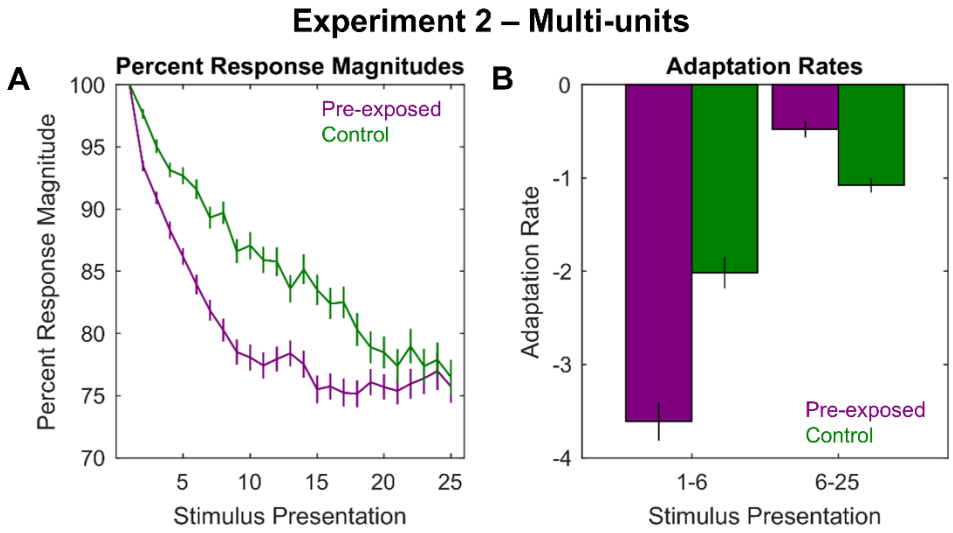

Figure 27. Multi-unit adaptation profiles for songs in pre-exposed and control exposure conditions. **A** shows the percent response magnitudes as a function of stimulus presentation in the two exposure conditions. **B** shows the adaptation rates for presentations 1-6 and 6-25 in the two exposure conditions.

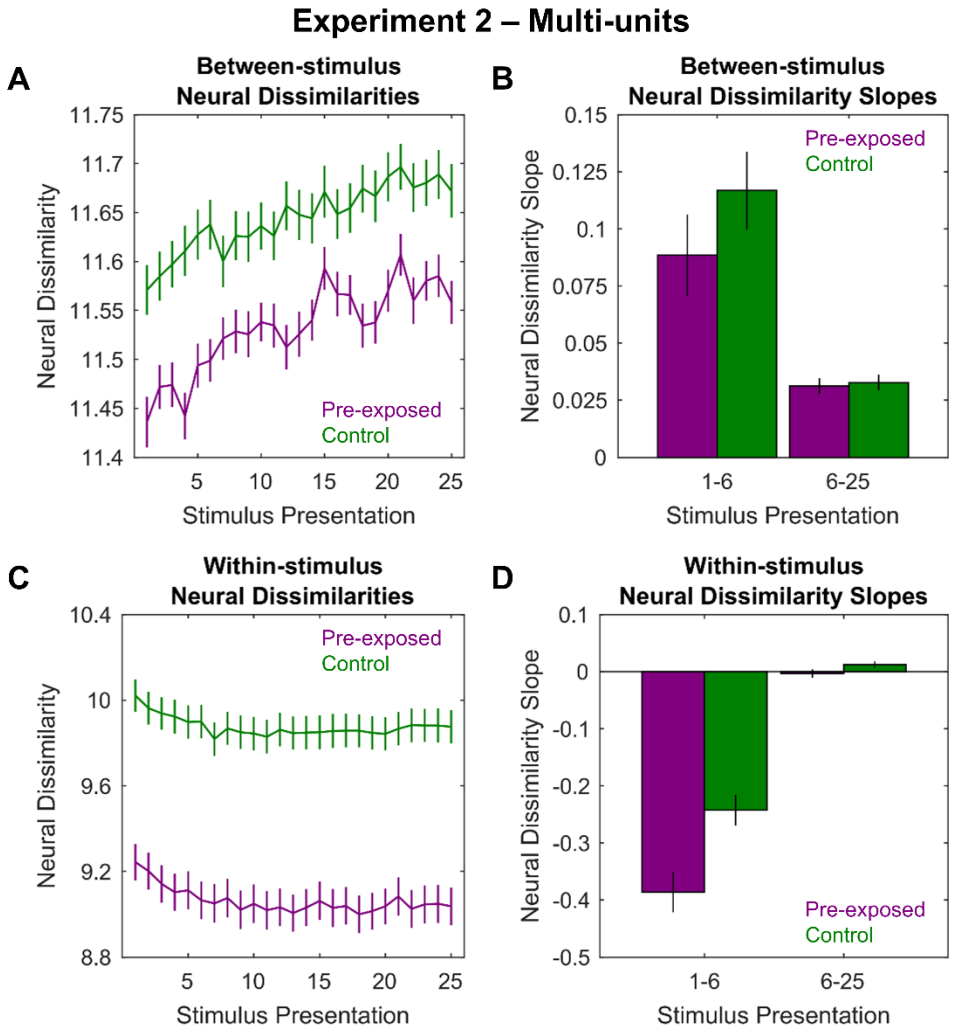

Figure 28. Multi-unit neural dissimilarities for songs in pre-exposed and control exposure conditions. **A** shows the between-stimulus neural dissimilarities as a function of stimulus presentation in the two exposure conditions. **B** shows the slopes of between-stimulus neural dissimilarities for presentations 1-6 and 6-25 in the two exposure conditions. **C** and **D** are similar to A and B, respectively, but they show within-stimulus neural dissimilarities and their slopes.

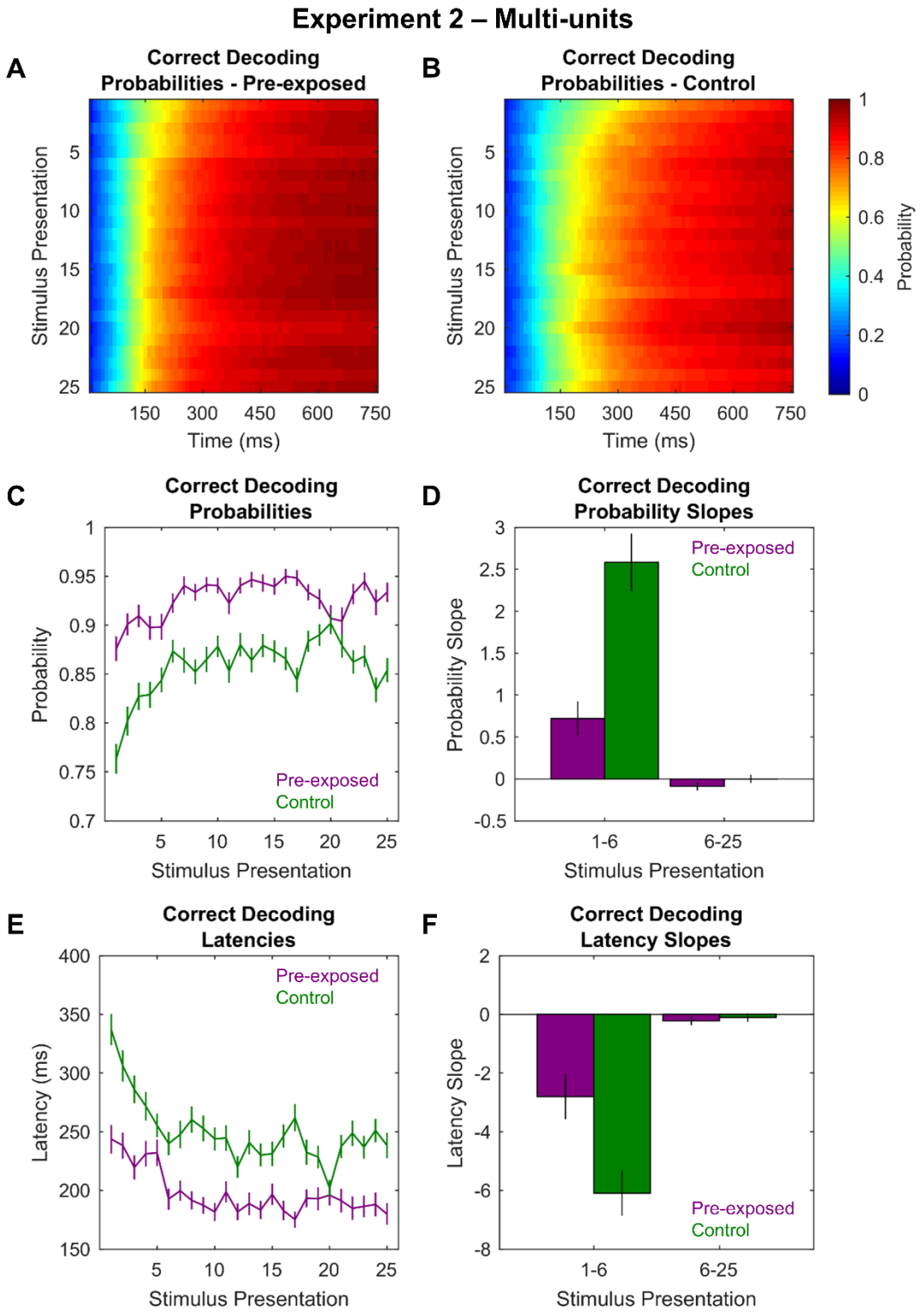

Figure 29. Multi-unit neural decoding accuracies for songs in pre-exposed and control exposure conditions. **A** shows the correct decoding probabilities across time points along the stimulus duration and across stimulus presentations in the pre-exposed condition. **B** is similar to A, but it shows the same metrics in the control exposure condition. **C** shows the

correct decoding probabilities (measured at 500 ms) as a function of stimulus presentation in the two exposure conditions. **D** shows the slopes of correct decoding probabilities for presentations 1-6 and 6-25 in the two exposure conditions. **E** and **F** are similar to C and D, respectively, but they show correct decoding latencies (to achieve a probability of .75) and their slopes.

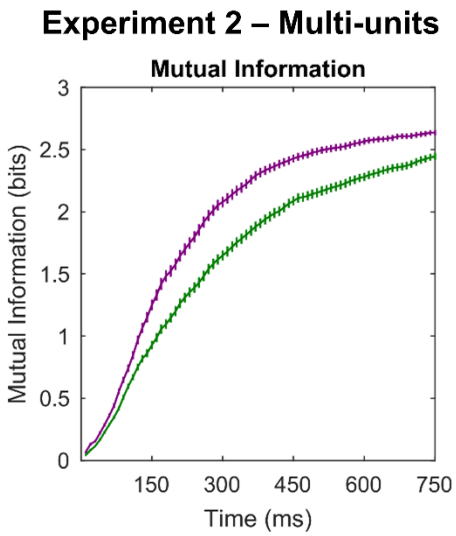

Figure 30. Multi-unit mutual information for songs in pre-exposed and control exposure conditions. The bias-corrected mutual information estimations are shown as a function of time along the stimulus duration in the two exposure conditions.

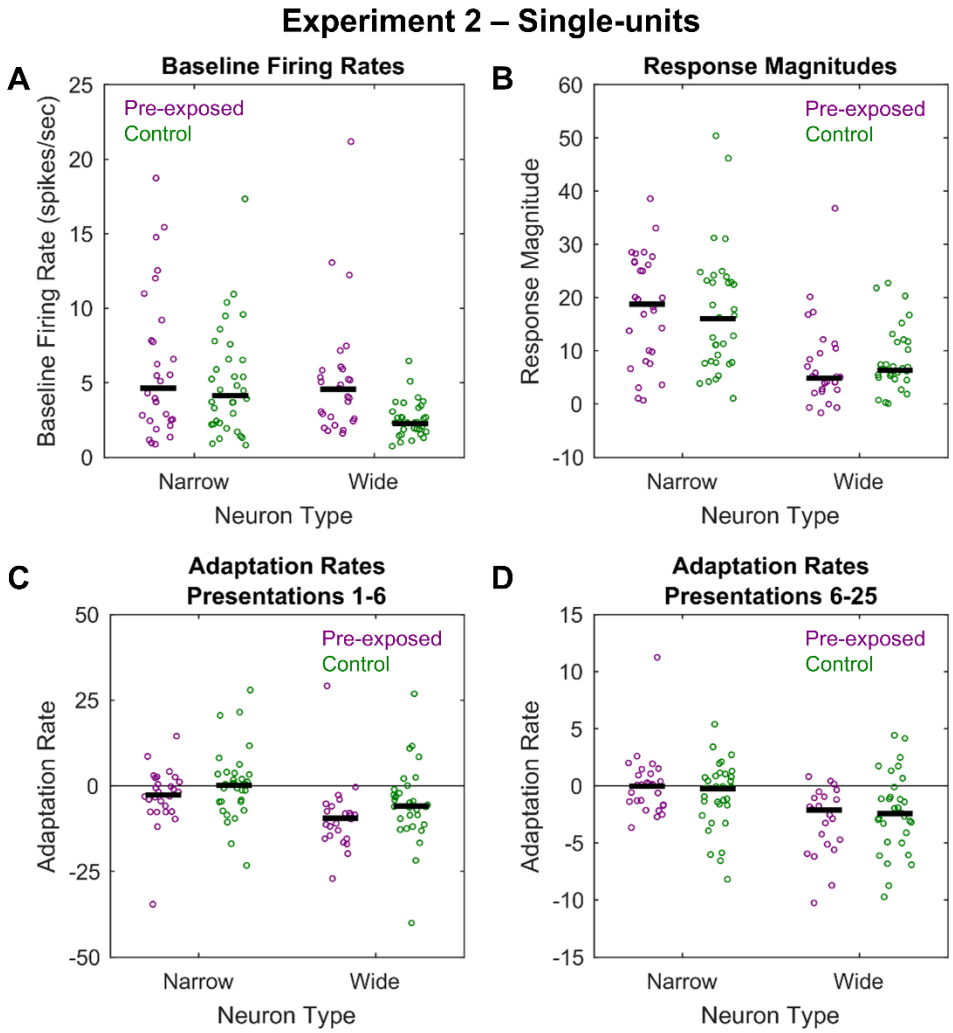

Figure 31. Single-unit response properties for songs for narrow and wide spike neurons in pre-exposed and control exposure conditions. **A** shows the baseline firing rates for the two neuron types in the two exposure conditions. **B**, **C**, and **D** are similar to A, but they show response magnitudes, adaptation rates for presentations 1-6, and adaptation rates for presentations 6-25, respectively.

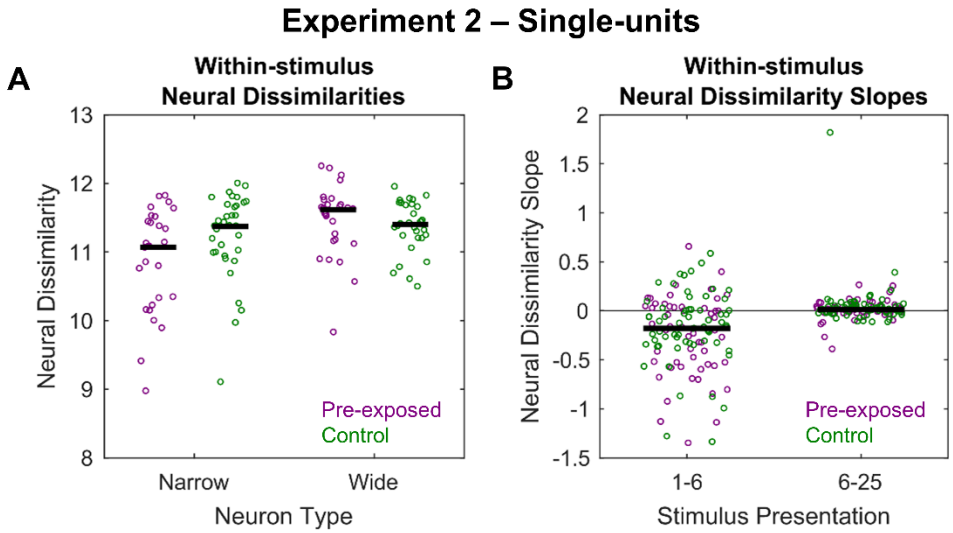

Figure 32. Single-unit neural dissimilarities for songs for narrow and wide spike neurons in pre-exposed and control exposure conditions. **A** shows the within-stimulus neural dissimilarities for the two neuron types in the two exposure conditions. **B** shows the slopes of within-stimulus neural dissimilarities for presentations 1-6 and 6-25.

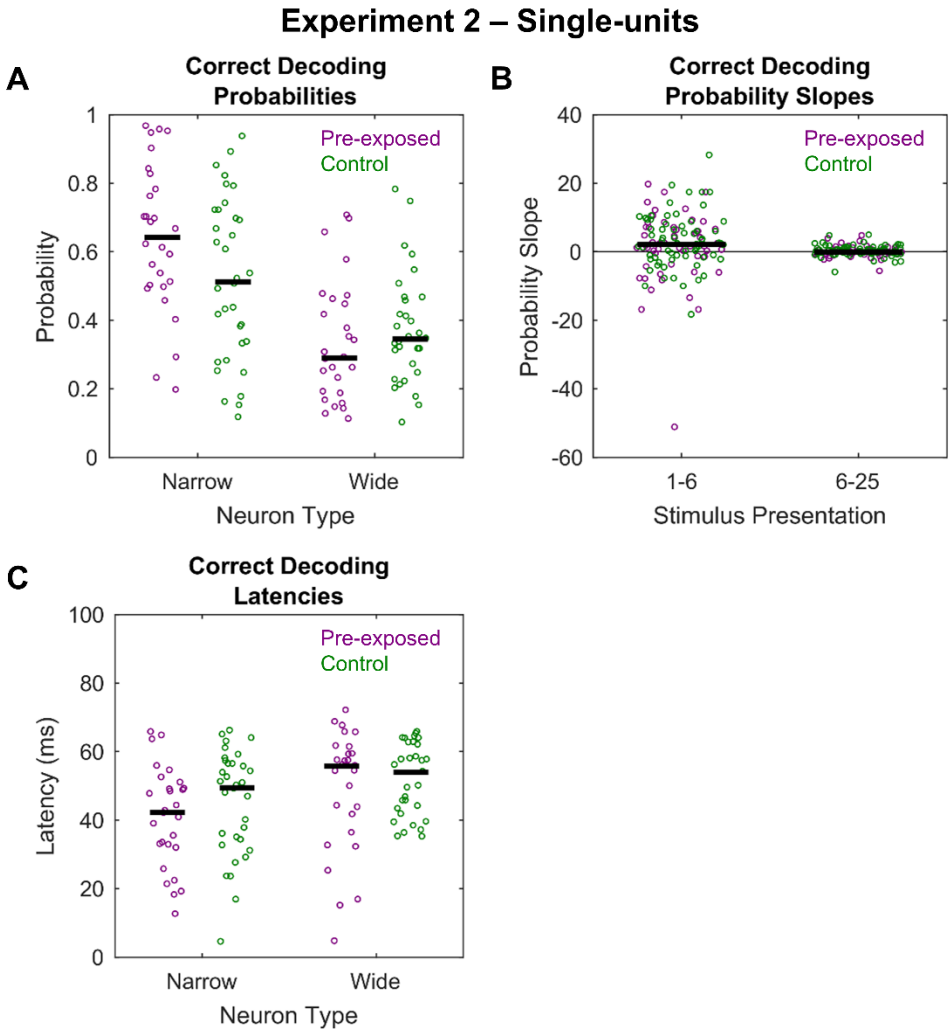

Figure 33. Single-unit neural decoding accuracies for songs for narrow and wide spike neurons in pre-exposed and control exposure conditions. **A** shows the correct decoding probabilities for the two neuron types in the two exposure conditions. **B** shows the slopes of correct decoding probabilities for presentations 1-6 and 6-25. **C** is similar to A, but it shows correct decoding latencies.

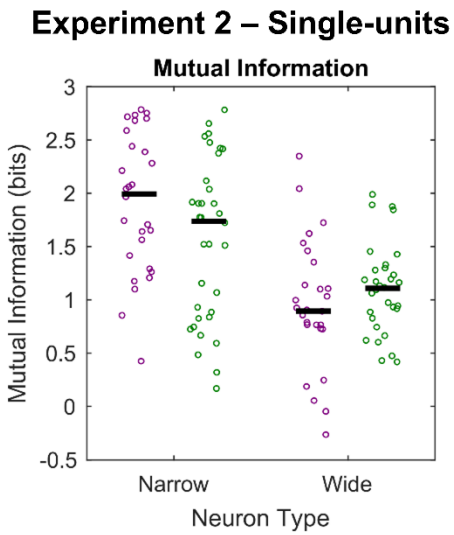

Figure 34. Single-unit mutual information for songs for narrow and wide spike neurons in pre-exposed and control exposure conditions.

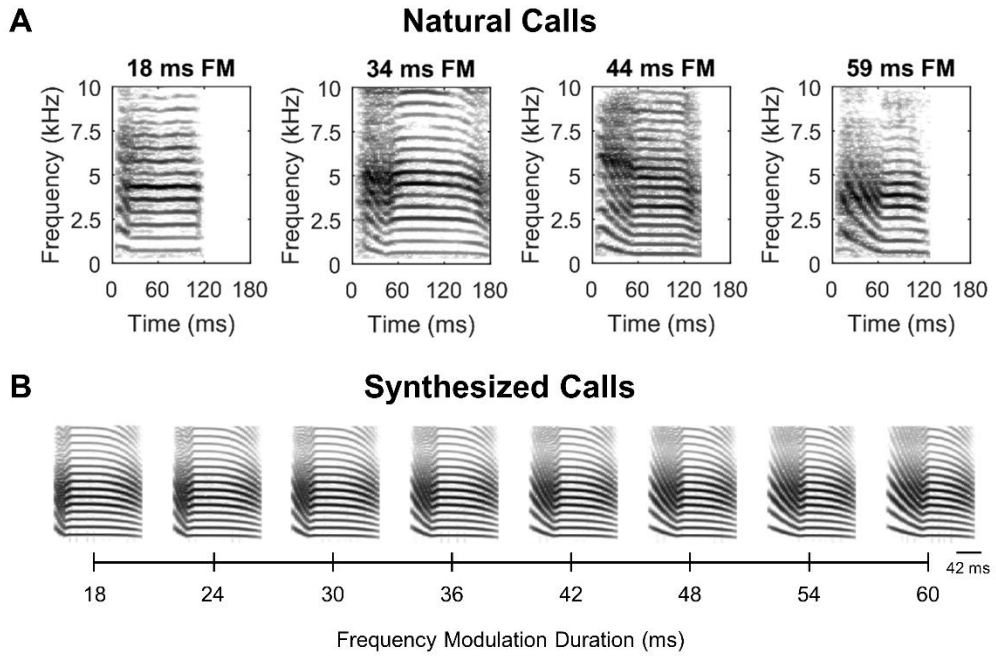

Figure 35. Natural and synthesized calls. **A** shows the spectrograms of natural calls varying in initial frequency modulation duration, in addition to other features. **B** shows the spectrograms of the 8 synthesized calls that were generated using mathematical functions to vary only along the initial frequency modulation duration dimension.

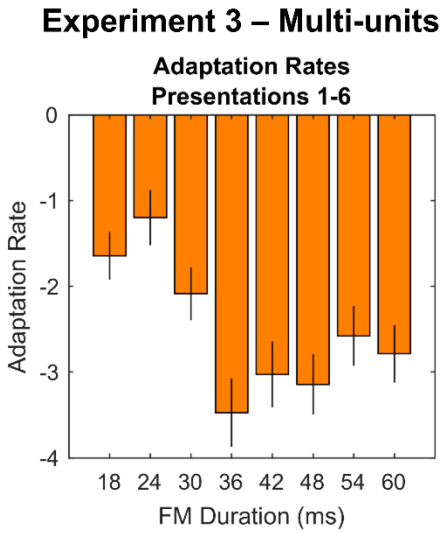

Figure 36. Multi-unit adaptation rates for synthesized calls for presentations 1-6.

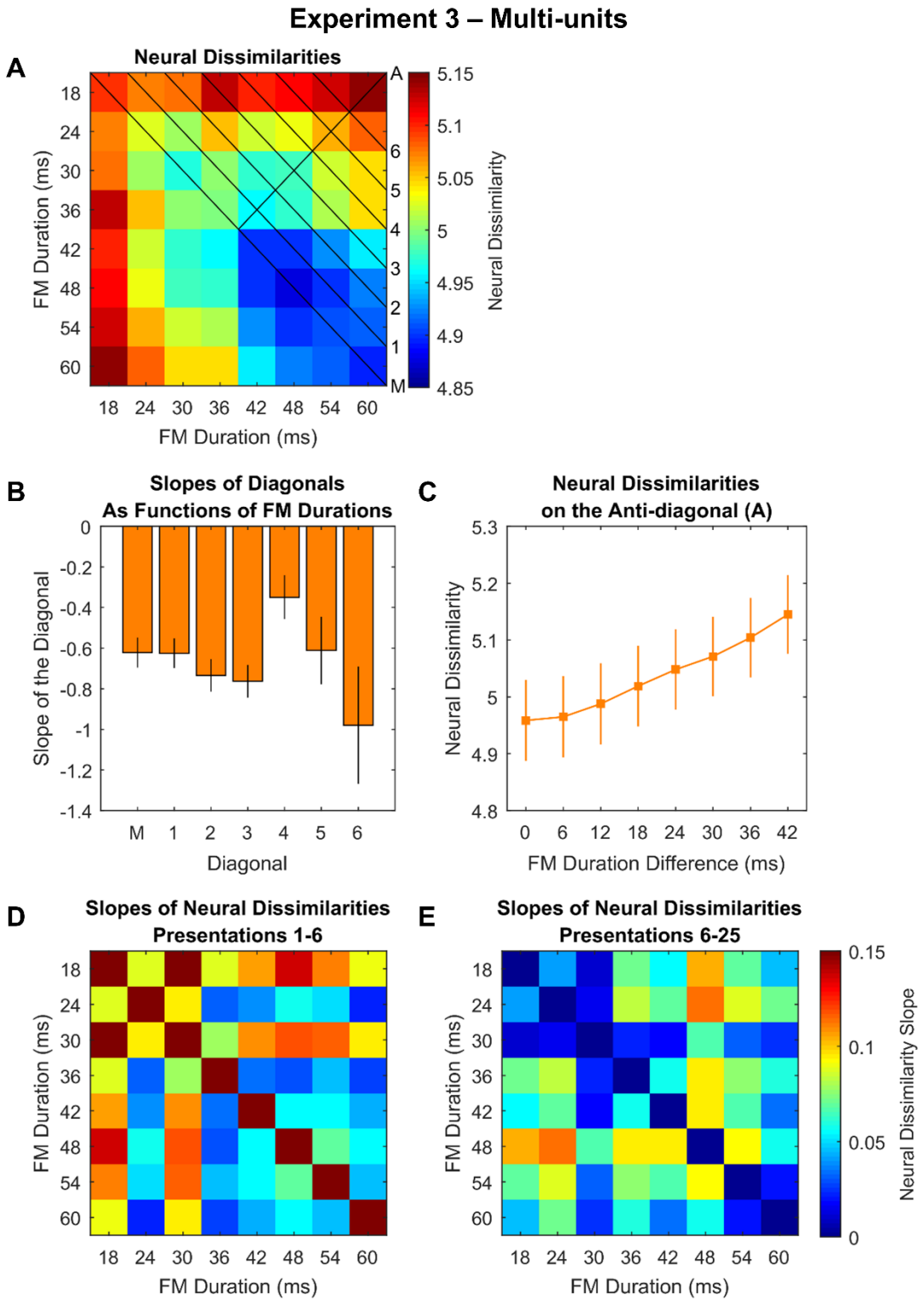

Figure 37. Multi-unit neural dissimilarities for synthesized calls. **A** shows the neural dissimilarities between the presentations of each possible synthesized call pair averaged across multi-units. The main (M) and the higher-order (1 through 6) diagonals of this matrix are indicated by lines. The anti-diagonal of this matrix is also shown. **B** shows the

slopes of these diagonals as functions of FM durations. **C** shows the neural dissimilarities on the anti-diagonal of the matrix in A. **D** and **E** are similar to A, but they show the slopes of neural dissimilarities for presentations 1-6 and 6-25, respectively.

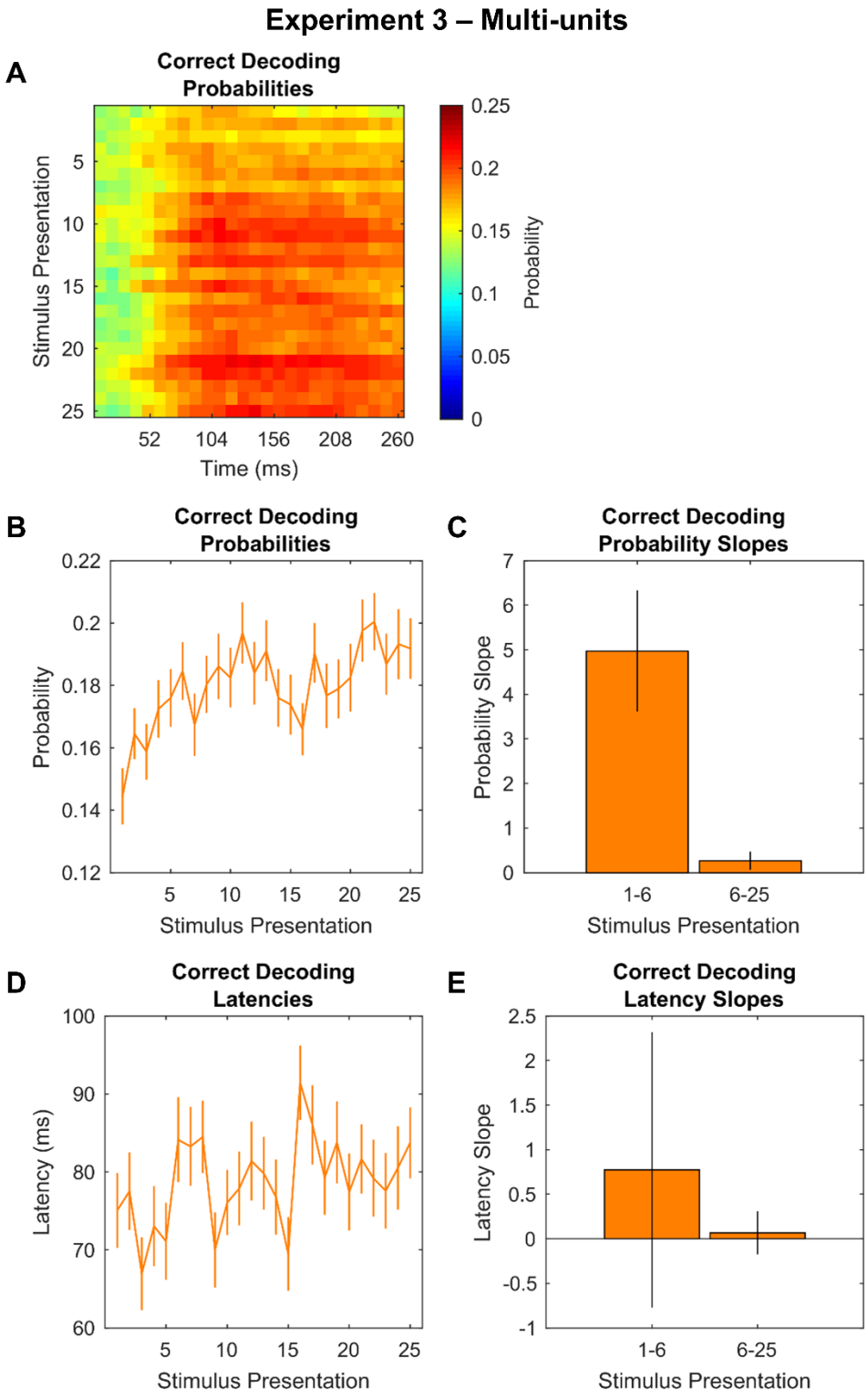

Figure 38. Multi-unit neural decoding accuracies for synthesized calls. **A** shows the correct decoding probabilities across time points along the stimulus duration and across stimulus presentations. **B** shows the correct decoding probabilities as a function of stimulus presentation. **C** shows the slopes of correct decoding probabilities for

presentations 1-6 and 6-25. **D** and **E** are similar to B and C, respectively, but they show correct decoding latencies and their slopes.

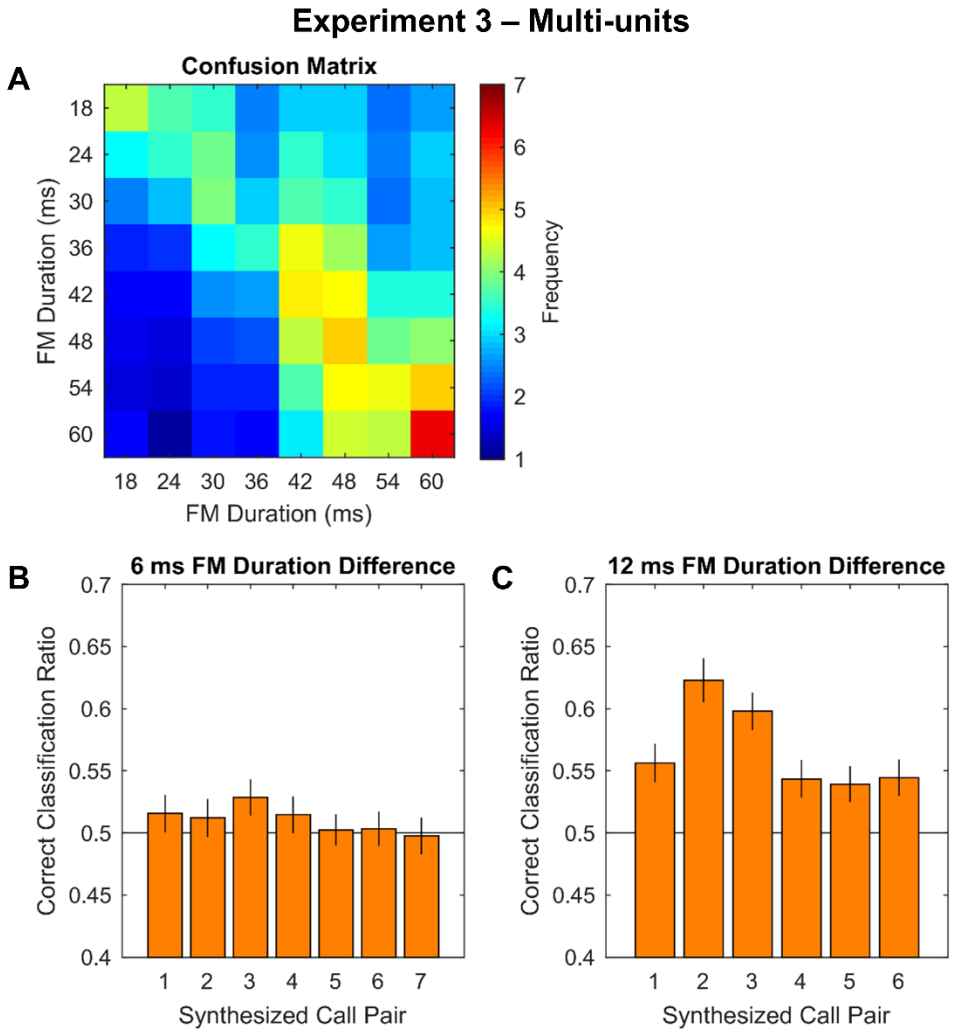

Figure 39. Multi-unit neural classifications for synthesized calls. **A** shows the confusion between true stimulus identities (rows) and neural classifications (columns) averaged across multi-units. **B** shows the correct classification ratios between pairs of synthesized calls that differ by 6 ms in their FM durations. **C** is similar to B, but it shows the same metric for synthesized call pairs that differ by 12 ms in FM durations.

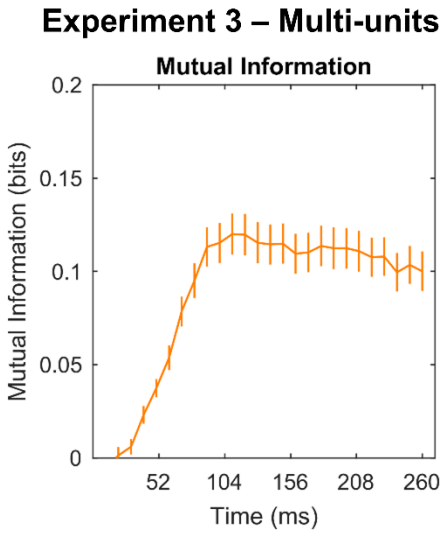

Figure 40. Multi-unit mutual information for synthesized calls. The bias-corrected mutual information estimations are shown as a function of time along the stimulus duration.

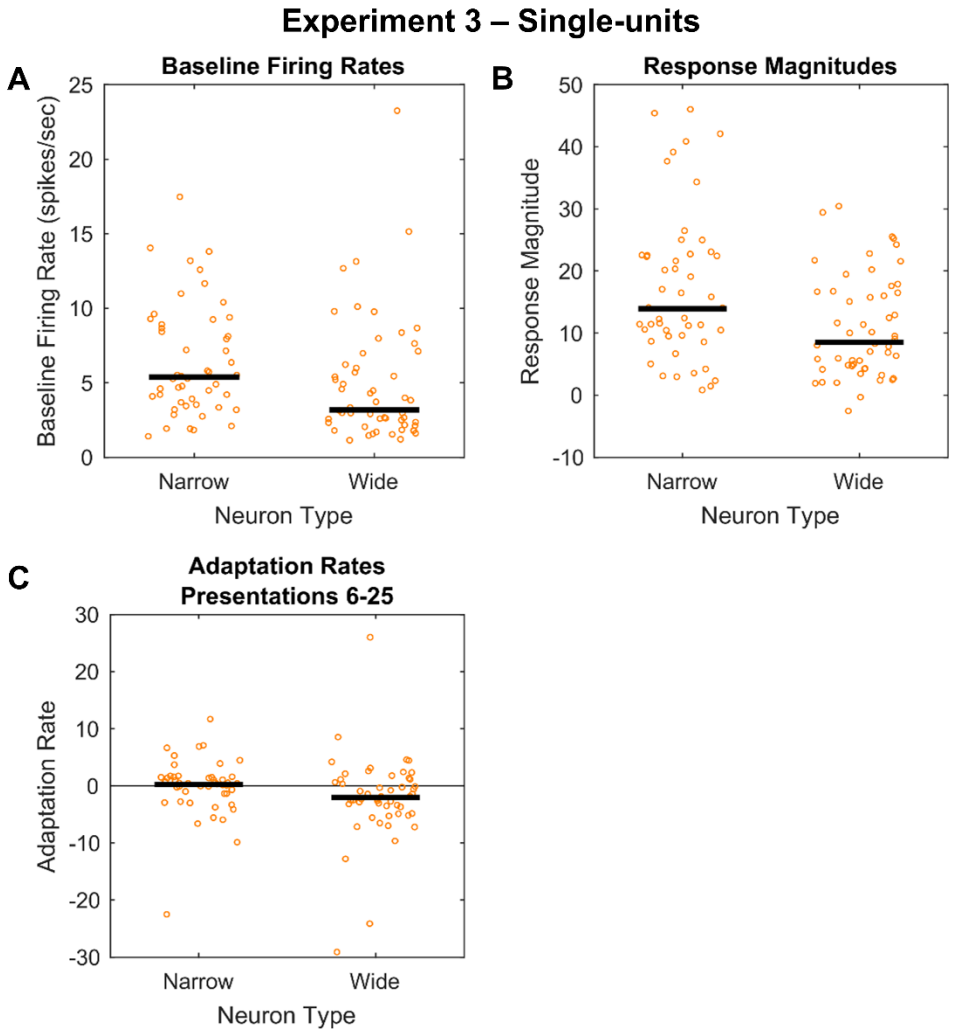

Figure 41. Single-unit response properties for synthesized calls for narrow and wide spike neurons. **A** shows the baseline firing rates for the two neuron types. **B** and **C** are similar to A, but they show response magnitudes and adaptation rates for presentations 6-25, respectively.

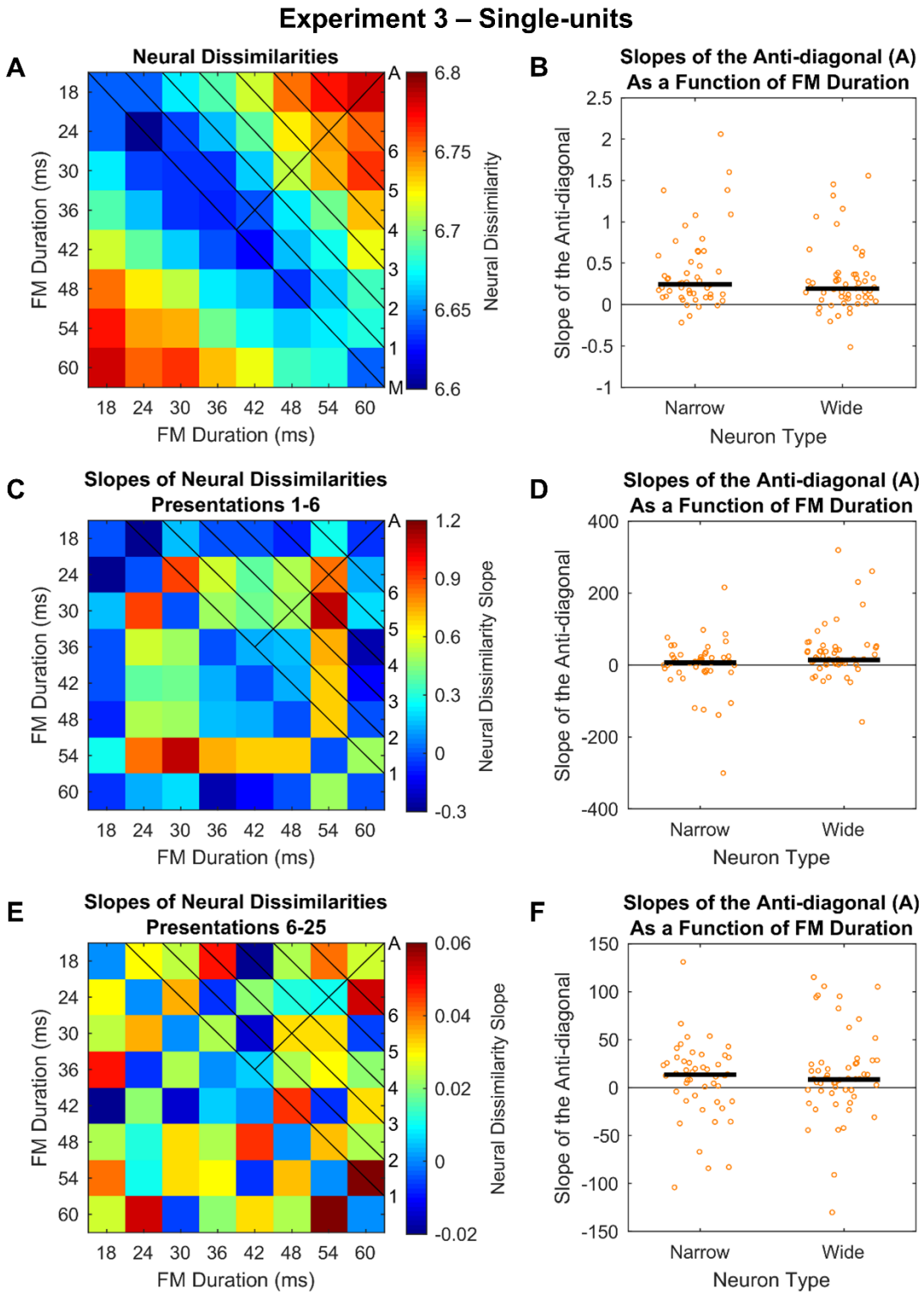

Figure 42. Single-unit neural dissimilarities for synthesized calls for narrow and wide spike neurons. **A** shows the neural dissimilarities between the presentations of all possible synthesized call pairs. The main (M) and the higher-order (1 through 6) diagonals of this matrix are indicated by lines. The anti-diagonal of this matrix is also

shown. **B** shows the slopes of this anti-diagonal as a function of FM durations for the two neuron types. **C** and **D** are similar to A and B, respectively, but they show the slopes of neural dissimilarities for presentations 1-6. **E** and **F** are also similar to A and B, respectively, but they show the slopes of neural dissimilarities for presentations 6-25.

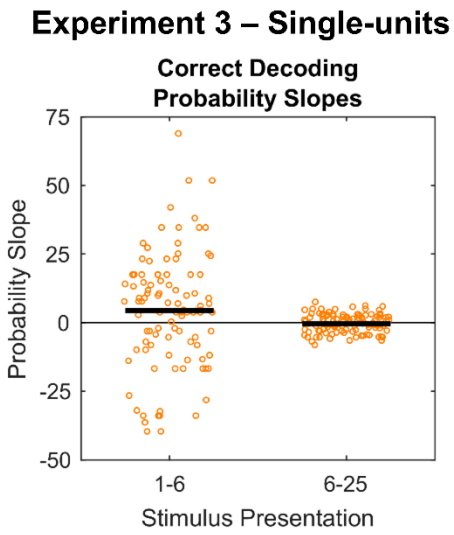

Figure 43. Single-unit correct decoding probability slopes for synthesized calls for presentations 1-6 and 6-25.

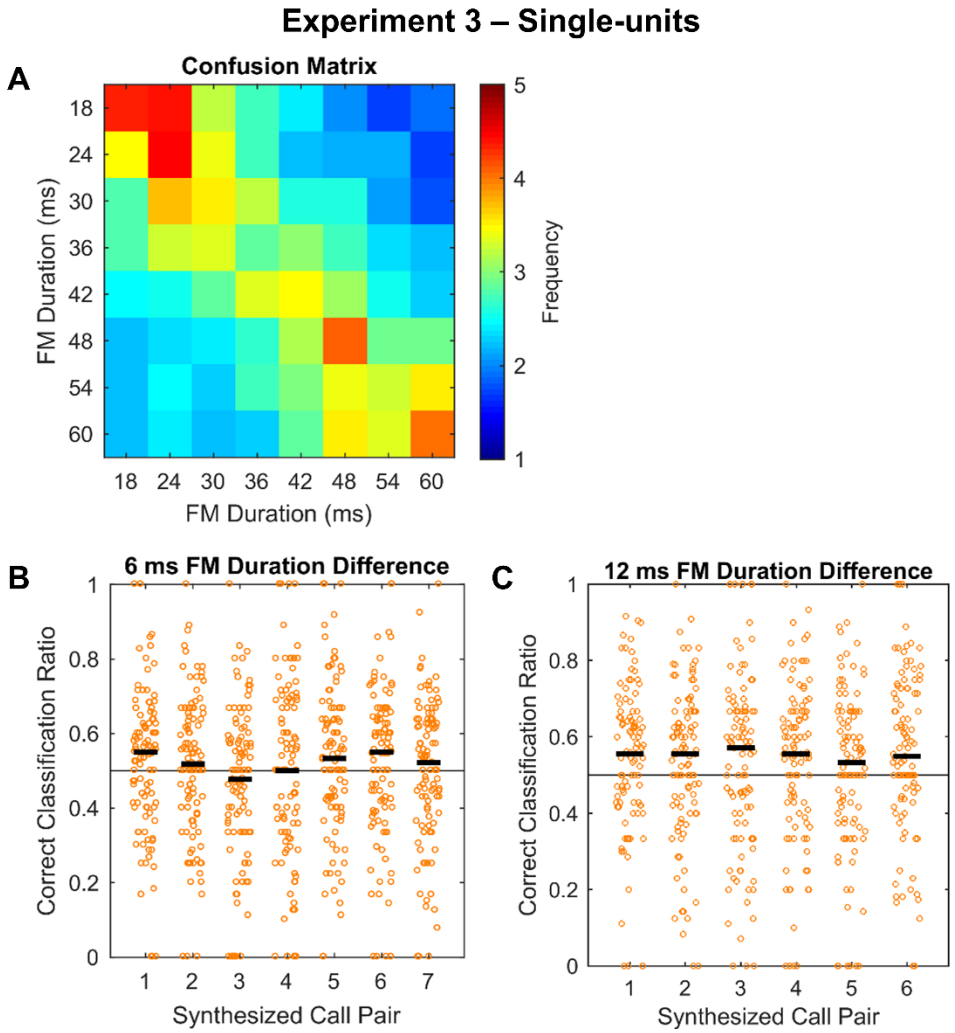

Figure 44. Single-unit neural classifications for synthesized calls. **A** shows the confusion between true stimulus identities (rows) and neural classifications (columns). **B** shows the correct classification ratios between pairs of synthesized calls that differed by 6 ms in their FM durations. **C** is similar to B, but it shows the same metric for synthesized call pairs that differ by 12 ms in FM durations.

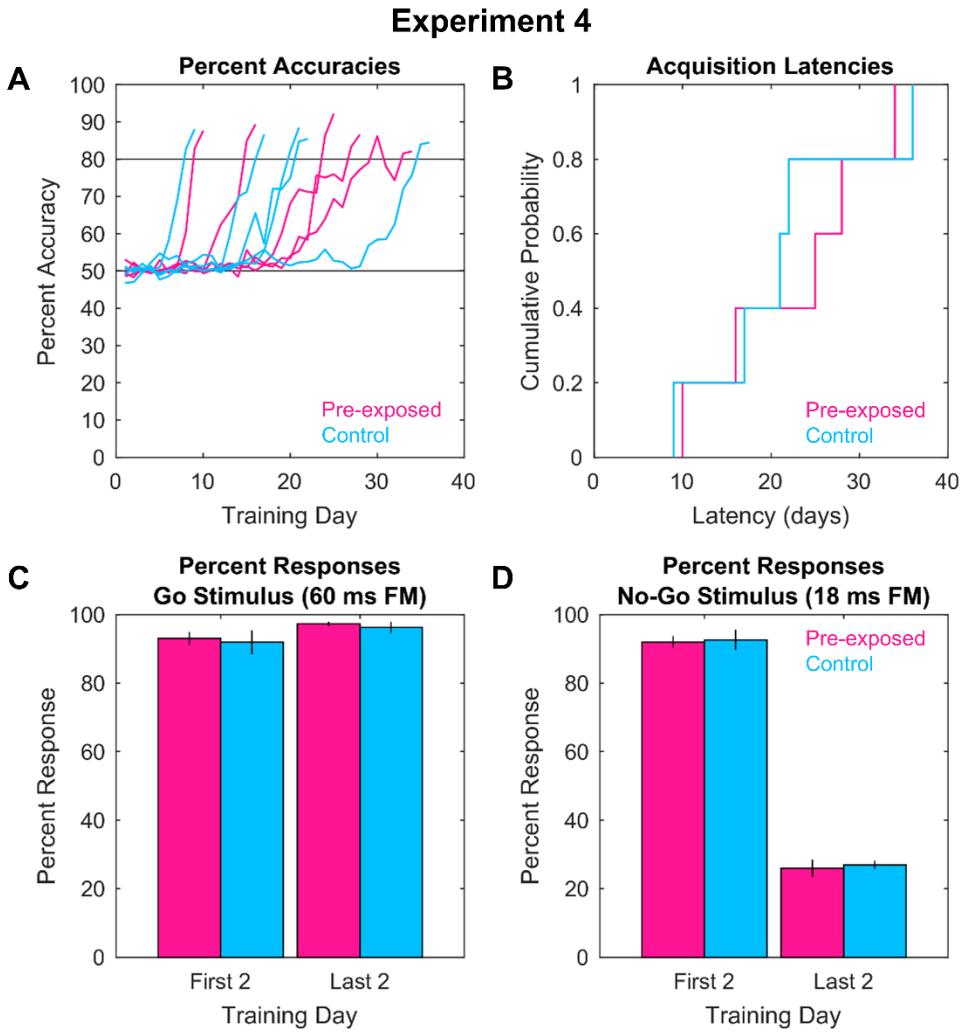

Figure 45. Behavioral responses in the Go/No-Go auditory discrimination training in pre-exposed and control exposure conditions. **A** shows the percent accuracies as a function of training days for all birds in the two exposure conditions. **B** shows the latencies in days of training to reach successful discrimination criterion in the two exposure conditions. **C** shows the percent responses to the Go stimulus in the first 2 and last 2 days of training in the two exposure conditions. **D** is similar to C, but it shows the responses to the No-Go stimulus. Plotted values are means ± SEMs across birds.

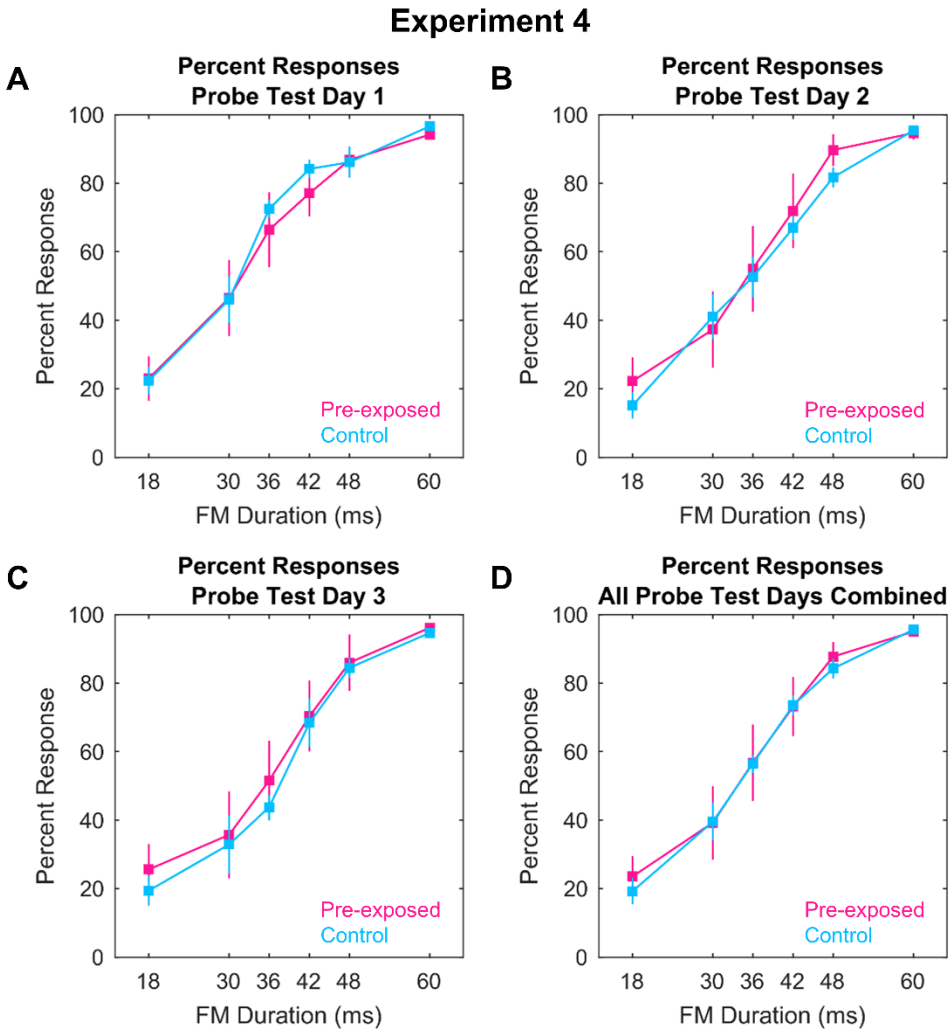

Figure 46. Behavioral responses in the probe test in pre-exposed and control exposure conditions. **A** shows the percent responses to the training (18 and 60 ms FM) and the probe stimuli (30, 36, 42, and 48 ms FM) in the first day of probe test in the two exposure conditions. **B**, **C**, and **D** are similar to A, but they show the responses in the second day, the third day, and across all days of probe test, respectively.

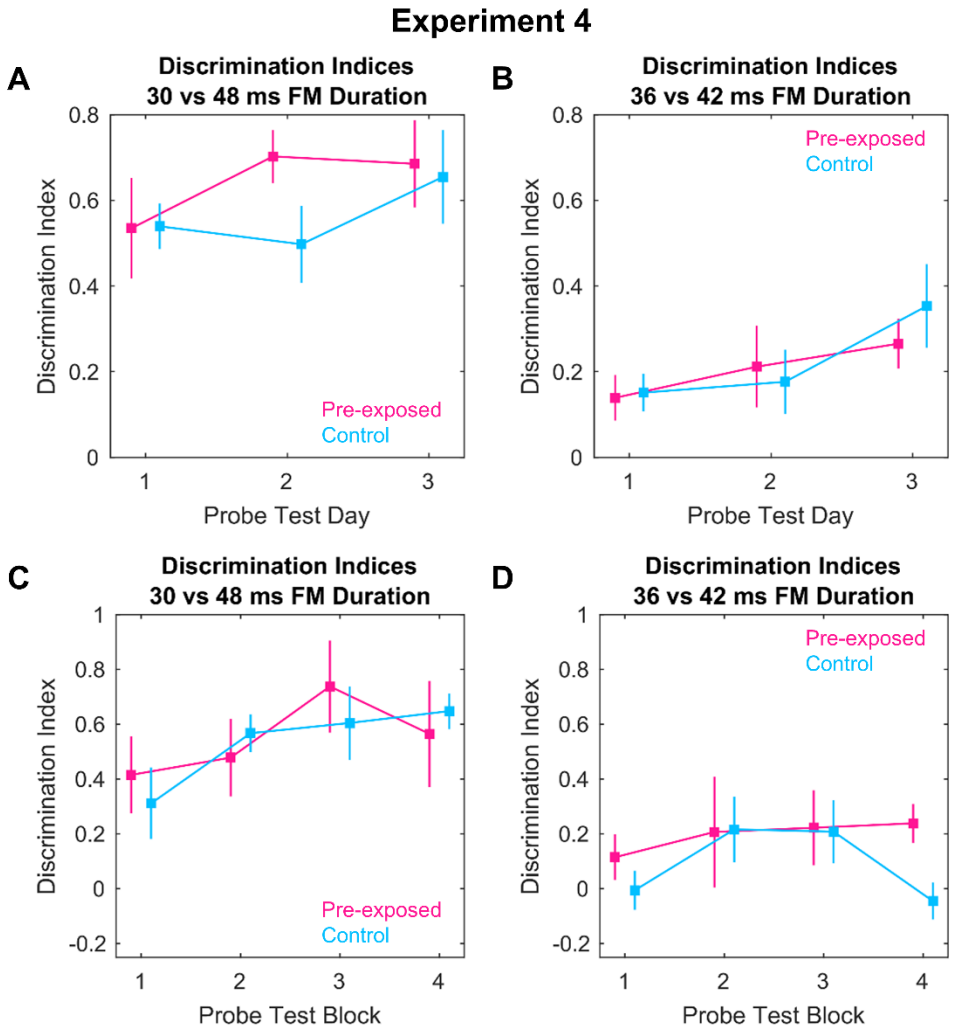

Figure 47. Discrimination of probe stimuli in pre-exposed and control exposure conditions. **A** shows the discrimination indices for the probe stimuli with 30 and 48 ms FM durations across probe test days in pre-exposed and control exposure conditions. These two stimuli were previously presented to the pre-exposed birds. **B** is similar to A, but it shows the same metrics for the probe stimuli with 36 and 42 ms FM durations. **C** and **D** are similar to A and B, respectively, but they show the same metrics across probe test blocks of 200 trials.

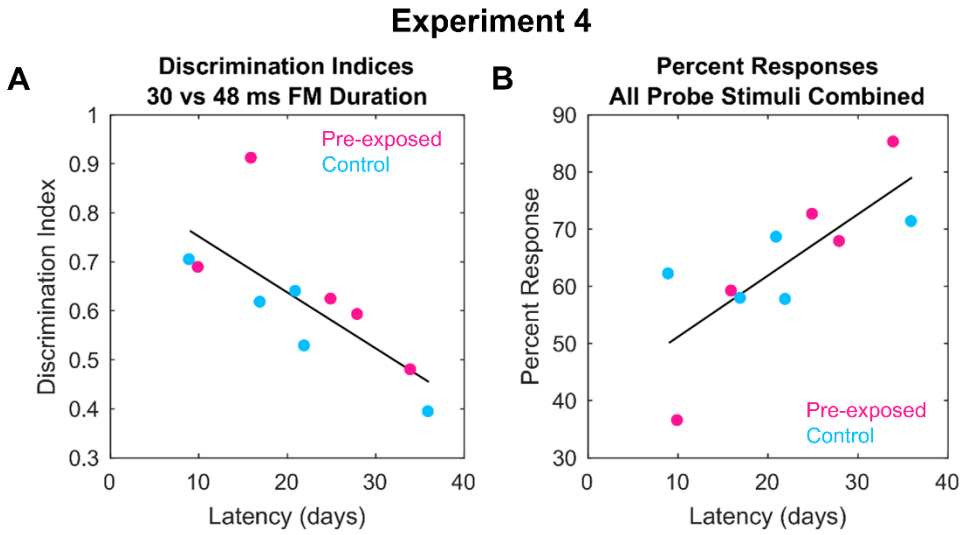

Figure 48. Initial auditory discrimination learning and probe test behavior relationships. **A** shows the correlation between latency in days of training to acquire the initial Go/No-Go discrimination and the discrimination indices for the probe stimuli with 30 and 48 ms FM durations. **B** shows the correlation between latency in days of training to acquire the initial Go/No-Go discrimination and the average levels of percent responses to all probe stimuli combined. Black lines show regression lines for pre-exposed and control exposure conditions together.

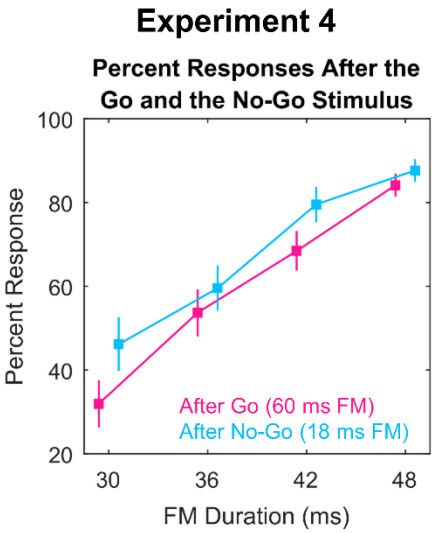

Figure 49. Percent responses to the probe stimuli after the Go (60 ms FM) and the No-Go stimulus (18 ms FM).