Encoding of complex sounds
in the auditory midbrain

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von

Dominika Lyzwa

aus

Braunschweig

Göttingen, 2014

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe.

Göttingen, Mai 2014

Dominika Lyzwa

Abstract

How complex natural sounds such as speech and vocalizations are encoded in the main converging center of the auditory midbrain is not yet fully understood.

For multi-units, which are composed of several single neurons, the representation of individual vocalizations is investigated across gradients of neural preferences in the inferior colliculus. The multi-units allow making inferences about collective neural behavior. The cross-correlation of averaged spiking responses is used to analyze if the vocalization representation depends systematically on the spatial distance between multi-units. The measure of response correlations and correlated variability indicates if interactions between the multi-units exist. Neural discrimination is employed to examine whether the whole set of spectrotemporally rich vocalizations can be reliably distinguished. The question is addressed whether along the frequency gradient of the inferior colliculus, individual vocalizations are encoded following their specific spectral content. This leads to the question whether vocalizations are accurately represented by individual multi-units, or, if groups of multi-units provide a more detailed representation, and if the multi-units interact to facilitate encoding. Some of these vocalizations display envelope correlations, and neurons might have neural preferences to these correlations, allowing them to encode the complex sounds in more detail. Reverse correlation analysis can capture nonlinear neural response properties. This analysis is used to probe preferences of single neurons to stimulus correlations, separately for different frequency carriers of the envelope, and to characterize the neural spectrotemporal tuning.

It is found that similarity of multi-unit spiking responses to vocalizations decreases linearly with spatial distance across the inferior colliculus, and extends to a few hundred micrometers. Neural correlations which are due to interactions between the multi-units are substantial but do not improve separability. Since interactions with other multi-units do not improve the neural separability, the multi-units act as independent encoders of vocalizations. However, neural discrimination improves substantially when the responses from more than one multi-unit are considered. In many cases, combining responses from a few multi-units results in a perfect discrimination of the whole set of vocalizations. The individual vocalizations are reliably encoded across wide frequency ranges, thus yielding a robust representation, which is relevant in a behavioral context. Optimal encoding does not strictly follow the spectral content of the vocalizations. The analysis of receptive fields reveals that the majority of the tested single neurons displayed preferences to envelope correlations. These are mostly found at the preferred frequency of the neuron and exhibit temporal tuning properties that enhance or modulate the neuron's linear response.

In conclusion, vocalizations are reliably encoded by independent multi-units across a wide frequency range of the inferior colliculus. Neural preferences to envelope correlations are present in the inferior colliculus and are likely to facilitate encoding of the vocalizations.

List of Figures

Table of Contents

1 Introduction

Audition is one of our major senses to interact with the environment in this world, allowing us to detect dangers (e.g. an approaching car), understand speech and enjoy music. A large amount of information is provided through auditory perception, e.g. if an object falls to the ground, based solely on the sound, we can estimate the consistency (e.g. wooden or metal), approximate size and location of the object.

1.1 General overview of the inferior colliculus

Propagating pressure waves, compressions and dilatations of air molecules, arrive at our ears and are perceived as sound. The waves that arrive at the outer ear travel through the ear canal, impinge on the tympanic membrane (ear drum) and cause it to vibrate. The three ossicles in the middle ear amplify and transduce the vibrations of the tympanic membrane into the fluid-filled cochlea. The ossicles serve as an impedance converter, their amplification is necessary in order for the sound to overcome the barrier from low impedance air to the higher impedance fluid. The cochlea decomposes the sound into its spectral components: via mechanical filtering, high, middle and low stimulus frequencies are sensed and encoded in a graded manner in different places on the basilar membrane. The place encoding of frequency, the *tonotopy*, is present throughout several stations of the auditory system. The hair cells on the basilar membrane transduce the mechanical deflection into a voltage which in turn synaptically activates the auditory nerve fibers. These project to the cochlear nucleus, the first nucleus located in the brainstem. The cochlear nucleus projects predominantly onto auditory nuclei of the contralateral hemisphere, however, projections to both hemispheres are present. The auditory information from the cochlear nucleus diverges into several processing pathways, which all project onto the central converging station, the inferior colliculus (IC). The combined information is projected to the thalamus and to the auditory cortex, which is believed to generate the perception of sound (see Fig. 1.1.1). However, almost all the information about the sound is already contained at a lower level, in the inferior colliculus (Fig. 1.1.2), which is a specific nucleus of the auditory midbrain that integrates projections from nearly all ascending brainstem nuclei [1, 2] (see Fig. 1.1.3). It has been shown that projections of specific brainstem nuclei terminate onto particular regions of the central nucleus of the inferior colliculus (ICC). They subdivide it into functional zones, e.g. the ventral division of the cochlear nucleus projects to the ventral part of the ICC, and respectively the dorsal division to the dorsal ICC.
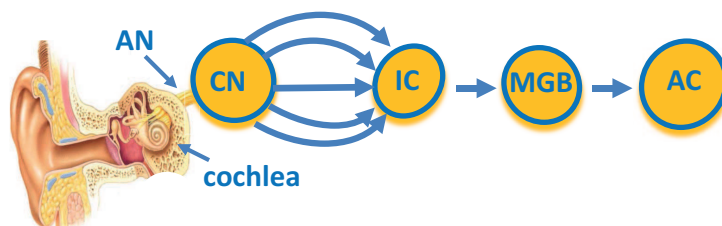
Figure 1.1.1: Schematic of ascending auditory pathway. Sound arrives at the outer ear; in the cochlea frequency decomposition is carried out, and the information is transduced via the auditory nerve (AN) to the cochlear nucleus (CN), the first nucleus in the auditory brainstem. From the CN, several pathways diverge that process the sound differently. All pathways converge in the IC (as schematized in Fig. 1.1.3). The IC projects to the medial geniculate body (MGB) in the thalamus; from the MGB the information is projected to the final station of the ascending auditory pathway, the auditory cortex (AC). [Ear drawing (Cochlear Ltd.)].

The ICC receives projections from several nuclei which have extracted different information from the sound, and thus integrates diverse inputs. This structure is named after its appearance and position: resembling a hill and being the lower hill out of two (see Fig. 1.1.4), it is called the *inferior colliculus*. The place coding of sound frequencies is also present in the ICC; the tonotopic gradient exists along the dorsolateral to ventromedial axis, with low frequencies being represented dorsal laterally and high frequencies represented ventral medially [3, 4], see Fig. 1.1.5. Along the tonotopic gradient, the stimulus frequency which elicits the highest spiking response, the *best frequency* (BF) of the neurons changes gradually. The ICC consists of isofrequency laminae, which are oriented approximately orthogonal to the tonotopic gradient. Within one isofrequency lamina, neurons have the same best frequency (within 1/3 octave [5]). Strong indications for a concentric gradient within one lamina of neural preferences to the modulation frequency of the sound envelope have been provided [6, 7, 8].
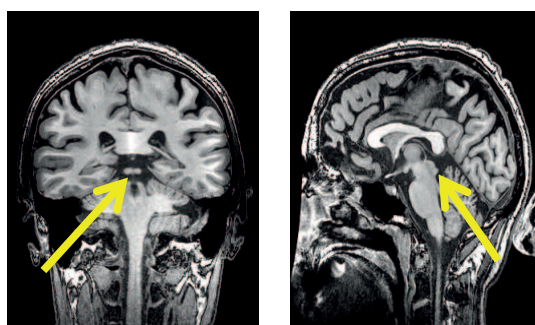
Figure 1.1.2: The inferior colliculus. Functional magnetic resonance image of a human brain with the inferior colliculus (indicated by arrow). Measured on a 3T GE Signa Excite HD scanner with a resolution of 0.7 mm^3 voxels (inversion prepared SPGR sequence); from the back (left) and from the side (right) [generously provided by David Ress, Texas University].
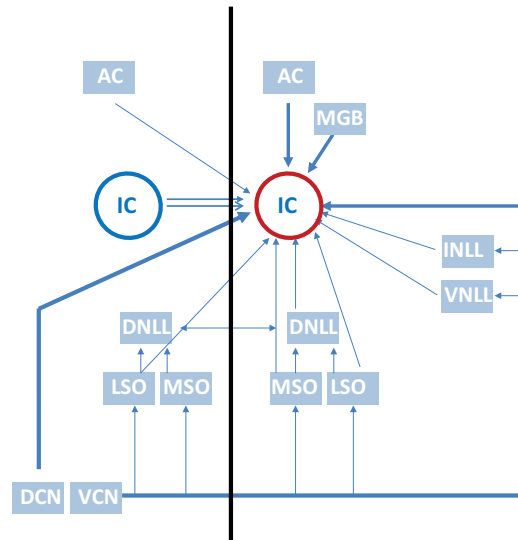
Figure 1.1.3: Projections to the inferior colliculus. The IC is the main converging station in the auditory midbrain. It receives several ascending and descending projections. The schematic is based on [9, 10], it shows projections to the right IC. The cochlear nucleus CN is the first brainstem nucleus and receives projections from the auditory nerve. (DCN-dorsal CN, VCN-ventral CN, MSO/LSO-medial and lateral superior olive, DNLL, VNLL- dorsal and ventral nucleus of the lateral lemniscus, MGB medial geniculate body, in the thalamus, AC auditory cortex, the final station of the auditory pathway.

The IC is the main processing station in the auditory midbrain [11], and apart from being a converging station, further sound feature extraction is presumably performed [12]; it is the metabolically most active structure in the brain [13]. The ICC is essential for extracting time-varying spectrotemporal information [14] and therefore might be important for processing complex sounds such as speech and vocalizations. Speech and vocalizations are spectrotemporally varying sounds which display a wide spectrum of acoustic properties.

Figure 1.1.4: Model brainstem. 3D-model of the brainstem imaged in Fig. 1.1.2. The inferior colliculi from both hemispheres are encircled in black and indicated by the arrow. The superior colliculi are situated above the inferior colliculi [provided by David Ress, Texas University].
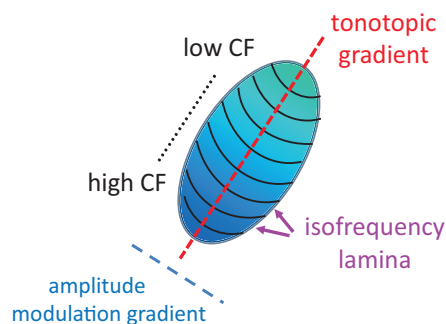
Figure 1.1.5: Schematic of the inferior colliculus. Along the tonotopic axis the characteristic frequencies (CF) change gradually from low to high frequencies (caudally to rostrally). The ICC consists of laminae which are approximately orthogonal to the frequency gradient. Within these laminae a concentric gradient of best amplitude modulation frequency exist.

The amplitude of their stimulus envelope and their spectral content change over time (amplitude and frequency modulation), and also the rates of change are dynamic. The vocalizations can display stimulus envelope correlations (e.g. Fig. 1.1.6a), and harmonics (e.g. Fig. 1.1.6b), which are frequency components that are integer multiples of the lowest periodic frequency which is present in the sound. These natural sounds are well suited for studying the auditory system since it was suggested that neurons are adapted to process them, and, that these sounds can trigger responses which are not elicited by artificial or simple acoustic stimuli [15].

The functional organization in order to efficiently represent and encode natural sound, that is, how neural preferences are organized within the ICC and whether regions, exist, that process specific acoustic properties, is one of the central questions in auditory neuroscience research. Further preferences for acoustic properties of natural sound that are systematically organized, other than the known gradients of neural preferences to stimulus frequency and amplitude modulation frequency, might exist.
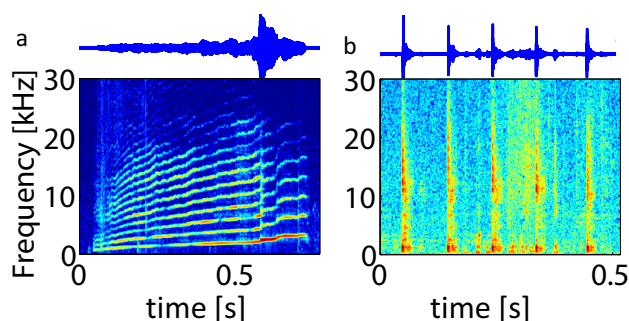
Figure 1.1.6: Example vocalizations. The waveform and the spectrogram of two guinea pig vocalizations are displayed. They contain harmonics and frequency modulations (a), and stimulus correlations of the envelope (b).

Understanding encoding mechanisms of natural sounds in this crucial midbrain nucleus is not only of fundamental scientific interest, but could also be beneficial to medical applications. Less than a decade ago, a hearing prosthesis was developed for hearing impaired people that cannot benefit from cochlear implants. The auditory midbrain implant [16] is inserted into the ICC; and has been implanted in five humans. In contrast to the cochlear implant, it did not yield improved speech comprehension [17]. This is partly due to the electrical stimulation algorithm, which is only slightly modified compared to the one applied in the cochlea, despite the advanced input processing and properties of the ICC.

1.2 Motivating questions

In this work, the representation of vocalizations across the ICC is investigated on the level of multi-unit clusters. These constitute the collective response from a group of neurons, i.e. a cluster of neighboring neurons that contains a number of single neurons, which can span several orders of magnitude. Using groups of neurons to study natural sound encoding has the advantage that this response is an integrated activity which might reflect local processing of the ICC. Furthermore, multi-units respond stronger to natural sound than single units [18].
The neural representation of acoustic stimuli across the ICC results from the interplay of the locally differing input and the spatially gradual changing of spectrotemporal neural preferences. Thus the question arises how similar the neural representations of vocalization are across the frequency gradient and across the amplitude modulation frequency gradient. Are the representations more similar across either one of these gradients, along which the sound is filtered either for spectral content or for amplitude modulations? The hypothesis that the preference gradients induce a gradually changing neural representation is tested. Strong deviations from such a gradually changing representation could point to confined functional sub-units. The similarity of neural representations of individual vocalizations is analyzed by cross-correlation of spiking activity. Interactions between the multi-units can lead to temporal correlations of their responses. The correlation analysis of simultaneous and non-simultaneous responses was used to derive whether similarity of neural representations is solely attributable to the stimulus, or if interactions exist between the multi-units. The neural representation of vocalizations across the ICC, and possible neuronal interactions between multi-units are investigated in Chapter 4, using cross-correlation.

The representation and similarity of responses to individual vocalization across the ICC yields insight into the distribution of responses and possible interactions between multi-units. Moreover, the encoding of vocalizations is of interest.

Motivating questions

Based on the separability of neural responses to different vocalizations, inferences about the encoding can be made when implying that a high accuracy of discrimination results from detailed encoding of the stimulus. For the discrimination analysis, the temporal information of spiking responses is used as it has been shown that spike-timing information is crucial for neural discrimination of vocalizations and intelligibility of speech [19, 20]. Previous work in the mouse ICC indicated heterogeneous neural responses to vocalizations [21] and spatial heterogeneity of receptive fields [22]. However, the specific encoding of individual vocalizations remains an open question because either heterogeneously distributed neural responses could lead to individual vocalizations being encoded equally well across the ICC, or the vocalizations might be encoded topographically, following the organization of spectrotemporal preferences in the ICC [3, 4, 6]. The latter was proposed to apply to neurons in the ICC of guinea pigs [23]. However, encoding of individual calls from a large set of vocalizations across the best frequency gradient at a fine spectral resolution has not been studied previously in ICC. Deviations from a purely spectrotemporal attribution of optimal encoding due to known and suggested nonlinear response mechanisms (e.g. cochlear distortions [24], spike generating nonlinearity [25] or suggested neural selectivity for single calls [26] are likely to exist. Thus, optimal encoding of individual vocalizations across the frequency gradient is compared to the spectral content of these stimuli (Chap. 5). This analysis is based on multi-units and it is tested whether the collective response from such a group of neurons allows reliable discrimination of the entire set of spectrotemporally rich vocalizations.

Single units have been shown to vary in their ability to discriminate from chance to even perfect discrimination [27, 28, 29]. In general, however, single neurons do not provide enough discriminative information to perfectly distinguish vocalizations. Groups of clusters and individual multi-units, were shown to outperfom discrimination ability of single neurons [30, 27]. However, a very large population of neurons responding simultaneously in order to encode a vocalization would not agree with the efficient encoding hypothesis [31].

It is tested whether the combination of successively more multi-units from different sites across the ICC improves discriminability of the set of vocalizations (Sec. 5.4.2). Neurons could interact in order to more efficiently encode sensory stimuli. However, neural representations have been proposed to decrease in redundancy from peripheral to cortical structures [32]. Whether signal correlations and correlated variability enhance or decrease encoding efficiency is a controversy [33, 34, 35, 36, 37], and might depend on the specific neural sensory system and its structure. The hypothesis that temporal correlations of simultaneously recorded multi-units facilitate neural discrimination was tested.

The encoding of vocalizations across the ICC, for individual multi-unit responses and for the combination of several responses that include correlations between the clusters with the use of spike-time based neural discrimination has been investigated as part of my research and the results of this analysis are presented and discussed in Chap. 5.

Vocalizations can display frequency and amplitude modulations, harmonics and envelope correlations. Neurons in the auditory midbrain might employ various processing mechanisms, linear and nonlinear ones, in order to encode these sounds. A neuron whose response to complex sound can be described as a linear superposition of the responses to its (independent) constituents has linear response properties, and a linear response is proportional to the input (e.g. increased spiking rate for enhanced stimulus intensity). A neuron which does not display these characteristics has nonlinear response properties.

Neurons were shown to display preferences (enhanced spiking) for certain frequency and amplitude modulations [38, 14], and it was suggested that such preferences for specific modulations of individual vocalizations facilitate encoding of these behaviorally relevant sounds [38]. Neurons might also display preferences to envelope correlations. Preferences to specific stimulus correlations would be due to nonlinear response mechanisms. In the ICC, $\sim 40\%$ of the neurons were shown to display nonlinear response properties [14], and different nonlinear response types exist [14, 24, 25, 39].

In this work, neural preferences to stimulus correlations are analyzed at the single cell level (Chap. 6). In order to assess in which frequency range the nonlinearities are dominant, correlation preferences were analyzed separately for each frequency carrier of the presented sound. Linear and nonlinear neural response properties of neurons can be visualized with receptive fields that can capture those preferences in response to a specific stimulus. Comparison of linear and nonlinear response properties might yield insight into whether the nonlinearity is due to intrinsic processing, in which case the receptive fields would be similar, or if it is already present in the input of the neuron, in which case the receptive fields would be disparate. Using first and second order reverse correlation, the linear and nonlinear receptive fields are computed and compared with respect to spectrotemporal preferences. For the reverse correlation analysis, an artificial sound that varies dynamically in frequency and amplitude modulation was used. This sound can be parametrically manipulated, thus, in contrast to natural sound, allowing for a systematic characterization of neuronal preferences over a broad spectrotemporal range. Single neuron preferences to envelope correlations are investigated by reverse correlation in Chapter 6.

1.3 Overview

The remaining thesis is structured as follows: The representation and similarity of multi-unit responses to vocalizations across the main central station of the auditory midbrain is investigated with cross-correlation in Chapter 4. Using neural discrimination, the encoding of these behaviorally relevant sounds by individual and interacting groups of multi-units is examined in Chapter 5. Preferences of single neurons to envelope correlations which are also present in natural sound are examined via reverse correlation in Chapter 6.

2 Related Work

The inferior colliculus has been studied with respect to its anatomy, neurophysiology, connectivities [40], and its excitatory and inhibitory synaptic inputs [41]. Using simple and artificial stimuli, e.g. pure tones [4, 42], white noise and amplitude-modulated noise [12], aspects of sound processing to extract specific features, for example amplitude and frequency modulation [12] and sound localization [43, 44] have been researched. However, the processing of natural, behaviorally relevant sound is the task daily undertaken by the auditory brain, and thus of interest but is not well understood. In this thesis, the representation and encoding of natural sound, vocalizations, and of acoustic sound properties in the ICC are investigated. This chapter provides a discussion of previous studies related to the research questions that are outlined in Sec. 1.2, and addressed in Chapters 4, 5, 6. This chapter is divided into three sections, in which previous work on the vocalization representation and correlation analysis (Sec. 2.1), neural discrimination analysis and encoding (Sec. 2.2), and nonlinear response properties and receptive fields (Sec. 2.3.3) are discussed. Each section is subdivided to discuss previous work on specific questions.

2.1 Correlations across the ICC

As a first analysis of this thesis the representation of vocalizations across the ICC, and correlation analysis to systematically investigate similarity of responses is carried out (Chap. 4). In this section, previous work on the vocalization representation across the best frequency gradient (Sec. 2.1.2), and analyses on response correlations and neural correlations in dependence on spatial and spectral distance are discussed (Sec. 2.1.2).

2.1.1 Neural representation along the tonotopic gradient

The representation of vocalization along the tonotopic gradient depends on the spectral content of the vocalizations. Suta and colleagues [23] recorded responses to four vocalizations (purr, chutter, chirp, and whistle) in 124 single neurons of the ICC in anesthetized guinea pigs. They analyzed these responses with trial-averaged temporal firing rate representations. They found that the poststimulus time histograms (PSTH) reflect the energy of the characteristic frequency of the neuron and that the spectrotemporal patterns of the vocalizations are encoded by the ICC neurons, but also that responses are heterogeneous. Thus, along the tonotopic gradient, the neural representation of a vocalization changes in dependence of the vocalization's spectral content.

In this work, the trial-averaged response representation by Suta et al. [23], likewise for vocalizations in the guinea pig ICC, is verified for the analyzed neural set, with the difference that not single neuron responses but multi-unit responses are used. These might respond differently, as they provide a collective response from a group a single neurons (Sec. 4.3.3). The analysis of Suta and colleagues is extended because a larger set of behaviorally relevant vocalizations with diverse spectrotemporal properties is investigated. In addition to the work of Suta and colleagues, the best frequency dependence of the response is demonstrated by comparison of the frequency filtered stimulus and the PSTH (Sec. 4.3.5). Neural responses to vocalizations change gradually with the neuron's best frequency.

2.1.2 Response correlations and neural correlations

The similarity of neural responses to a vocalization across the ICC is investigated with cross-correlation analysis, which is a more compact description than the PSTHs for the large set of neural responses (Sec. 4.4). The analysis also allows estimating the amount of correlations which are due to neural interactions. In the following, previous correlation studies in higher auditory structures that have investigated the similarity of responses in dependence of the neurons' spatial and spectral distance are discussed.

In the mammalian ICC, it has been shown previously that responses to complex sound are most correlated for close-by single neurons. Chen and colleagues [45] analyzed single neuron responses to artificial complex sound (dynamic moving ripple sound) in the cat ICC, using spectrotemporal receptive fields and spike-train cross-correlation for neural pairs with a maximal separation of ∼370 µm. They found that, on a grand scheme, the neurons that are neighbors display a higher probability of having similar receptive fields, thus similar spectrotemporal preferences. The best frequency (BF) was the most correlated parameter between neighboring neurons, receptive field similarity was largest for similar BFs, and correlated firing was stronger for neighboring neuron pairs and decreased with spatial distance and differing BFs. Chen and colleagues analyzed correlations between single neurons and based on their findings suggest local microcircuits of similar functionality in the ICC.

In this work, spike train cross-correlation on the level of groups of neurons is investigated (Sec. 4.4), since the collective response from a wider spatial volume can display different correlation dependencies than have been found for single neurons. Spectrotemporal receptive fields are not calculated for the multi-units (see Sec. 3.1.1), but correlations are analyzed with respect to spatial and spectral (difference in preferred frequency) distances of the neurons. Chen and colleagues used an artificial stimulus, however, in this thesis, correlations in response to natural sounds are tested. Natural sounds are realistic stimuli that are processed by the auditory system, and they might elicit responses that are not elicited by artificial stimuli [15]. In their analysis, Chen and colleagues probed spatial distances up to ∼370 µm, thus no statement could be made about correlations for neurons that are further away. In the present work, larger neural distances of up to 1600 µm are probed (Sec. 4.4.3).

Thus, taken together, in this thesis, in addition to the work of Chen and colleagues, the correlation dependencies of multi-units responses, and in contrast to their work natural sounds are investigated across large spatial distances.

Correlations of responses in dependence of spatial distance in the auditory midbrain have been described by a further study, however, not in the mammalian midbrain but for insects. Epping and Eggermont [46] analyzed neural activity of 150 multi-units in the auditory midbrain of the grass frog. Using cross-correlation, they found response correlations for 60 % of the units and 15 % of the units displayed correlations due to neural interactions and connections. These neural correlations were restricted to pairs with distances less than 300 μm, whereas the response correlations were independent of the spatial distance. The authors suggest that the high amount of response correlations relative to the small amount of neural correlations indicates that for generating the neural response, the stimulus inputs play a predominant role over neural mechanisms [46]. The authors attribute this to a spread projection of the stimulus input, rather than to restricted areas, which would be in line with a weak tonotopic organization in the grassfrog midbrain and to the finding that neighboring neurons often display different spectral preferences. Eggermont and Epping used multi-units and probed larger spatial distances than described by Chen et al. [45], which is also used in the analysis of this work. In constrast to the grasshopper midbrain, the mammalian ICC has a clear tonotopic structure with neighboring neurons displaying similar spectral preferences. Thus, the correlation structure as described by Epping and Eggermont might be different for the mammalian ICC, and will be investigated in this work (Sec. 4.4). Epping and Eggermont also found stimulus dependencies for half of the neural correlations, indicating that the functional organization is dynamic, thus the functional connections depend on the particular stimulus applied. Hence, in this thesis, correlations are analyzed separately for each vocalization, in order to account for possible stimulus dependencies.

In these previous studies no description of the spatial dependence was provided. In a higher auditory processing station, in the cortex, a correlation dependence on the spatial distance has also been shown. Eggermont [47] analyzed neural groups (that reflected patched activity and were termed 'clusters') in the cat primary auditory cortex with the use of cross-correlation matrices of spontaneous activity. The author found that the correlation followed an exponential decrease, $f(x) = a \cdot e^{-b \cdot x}$, with x the spatial distance in millimeter (a=0.05, b=0.24/mm). This dependence that was found for spontaneous activity of neural groups in the primary auditory cortex which is also tonotopically organized, and larger than the ICC, is compared to the correlation dependence of responses to vocalizations in the ICC.

Summary

To summarize, the analysis of the neural representation along the tonotopic gradient verifies and extends the finding of Suta et al. [23] for a larger vocalization set and for multi-unit responses (Sec. 4.3.3). Additionally, the best frequency dependence of the response is demonstrated with a filtering approach (Sec. 4.3.5). The relative amount of correlations due to the stimulus and due to neural interactions has been computed for the insect auditory midbrain [46], but not for the mammalian ICC and is investigated in this work (Sec. 4.4.5). Dependencies on spatial distance have not been found for the grasshopper midbrain [46], but have been shown in the mammalian ICC [45] and primary auditory cortex [47]. The dependencies on spatial and spectral distance have been analyzed for single neurons, showing that nearby neurons have a higher probability of displaying similar neural preferences and responses, and a microcircuitry has been proposed to exist in the ICC [45]. Contrary to these studies, in the here presented work, a larger spatial distance is covered, and responses of multi-units to natural sound stimuli are investigated (Sec. 4.4).

2.2 Neural discrimination of vocalizations

In Sec. 2.1, previous work on the representation and correlation of neural responses to the same stimulus were discussed. The auditory system can distinguish different complex sound stimuli. The discrimination of these different neural responses to stimuli can be used to study the encoding of the complex sounds. In this section, the findings of previous studies on encoding of individual vocalizations by a single neuron and groups of neurons and the effect of neural correlations on the discrimination are reviewed [29, 27, 26, 48].

It has been shown that spike trains can allow discriminating different stimuli with high accuracy. Discrimination between the same call from different producers in single neurons of insects was found to be good [29]. Machens and colleagues analyzed separability of 18 auditory periphery receptor single cell responses of a female grasshopper to the same song presented from eight different males and obtained neural discrimination values higher than 80 % and up to 100 %. They used spike trains convolved with an exponential function in order to mimic excitatory postsynaptic potentials (EPSP), and found temporal resolutions between 3-10 ms and spike train duration above 400 ms to yield the highest discrimination performances. Hence, the work of Machens and colleagues showed that the temporal resolution and spike duration affect discrimination. In general, findings for the grasshopper auditory periphery do not allow making direct inferences about processing of vocalizations in the mammalian auditory midbrain. Whereas processing might be similar across different mammal species [49], this is not necessarily true for other animals, such as insects or birds. The study by Machens and colleagues addressed discrimination between different stimulus trials. Whereas this is an important question it cannot be investigated in this thesis, as only one trial per vocalization is present.

Good discriminability between different trials of a call demonstrated by Machens et al. suggests that calls with different spectrotemporal properties should be discriminated even better. In contrast to the work of Machens and colleagues., in this work discrimination between different stimuli is performed, but also based on vocalizations and on spike-trains. Optimal spike-duration is investigated prior to the discrimination analysis (Sec. 5.3), as suggested by Machens et al. [29].

2.2.1 Preferred encoding of individual vocalizations

In this work, the optimal encoding of individual vocalizations with their specific spectrotemporal content is compared across the best frequency gradient. The total discrimination across a whole set of complex sounds does not depend on the frequency tuning, i.e. the best frequency of the neuron [27]. Using clustering of the squared Euclidean distance between pairs of spike trains, Schneider and Woolley have shown that single neuron responses of the mesencephalicus lateralis dorsalis in the zebra finch, the homologous to the inferior colliculus in mammals, could be used to discriminate among 11 bird songs with performances up to 100 % [27]. The authors did not find a correlation of total discrimination performance and spectral tuning for complex behaviorally relevant calls in the midbrain. However, preferred encoding of individual vocalizations in specific frequency regions might exist but is not detectable in the total discrimination. Thus, the encoding needs to be analyzed for individual vocalizations and is investigated in this work (Sec. 5.4.1).

Individual vocalizations in the guinea pig ICC have been shown to be encoded based on their spectrotemporal patterns [23]. The work by Suta et al. [23] which is also described in Sec. 2.1.2, showed that the poststimulus time histograms reflect the energy of the neuron's characteristic frequency and that the spectrotemporal patterns of the vocalizations are encoded. The authors did not find evidence for encoding of vocalizations based on call-selectivity of single neurons [50]. They show that responses are heterogeneous, which was later confirmed for mice [21]. Within a frequency segmentation of four intervals across the tonotopy, Suta and colleagues demonstrated for the four calls a dependence of the responsiveness on the neuron's characteristic frequency. Overall, their work indicated that the spike rate of ICC neurons depends on the relationship between the neuron's preferred frequency and the frequency spectrum of the vocalization.

Suta et al. [23] investigated responses to conspecific vocalizations in the ICC of guinea pigs, which is also investigated in the here presented work. However, substantial differences exist between the work of Suta et al. [23] and the here presented work (Sec. 5.4.1), and thus expand the previous study. In the here presented work, neural discrimination, which compares temporal spike train similarity across the whole set of 11 vocalizations is used, as opposed to the comparison of the trial-averaged response in form of PSTH in the work of Suta et al. [23].

This method can reveal discriminative response differences in more detail than the spike rate, and also incorporates trial-to-trial variabilities of the responses, which are realistic, but which are averaged out in the PSTHs. In the analysis of this work, neural discrimination is computed for relatively short segments (100 ms) of the vocalizations, as perceptually vocalizations are already discriminated at these time scales. Another difference to the study of Suta et al., is that in the here presented work responses from groups of neurons are investigated, thus neural activity at a wider angle, reflecting the output activity of a population of neurons. Furthermore, for the comparison of optimal encoding and the vocalization's spectral content along the best frequency gradient, a finer frequency resolution of 36 intervals is used than the one by Suta et al. which used only 4 frequency intervals to compare the neuron's responsiveness to individual vocalizations in dependence of their characteristic frequency. Additionally, in the here presented work, a spectrotemporally rich set of 11 behaviorally relevant vocalizations is used, and might reflect more realistically the entire set of guinea pig communication calls, and hence allow investigating their auditory processing, than the four calls used by Suta and colleagues. A minor difference to the work of Suta et al. [23] is that here rather than the preferred frequency at the lowest spike-eliciting intensity (the characteristic frequency), the preferred frequency at the *actual* presented stimulus intensity (the best frequency) is used and might yield a more accurate comparison of preferred and presented frequencies. However, trends of dependencies should be the same for both, characteristic and best frequency. In conclusion, the here presented work differs from the work by Suta and colleagues because multi-unit responses are compared based on short individual spike train trials and a much finer frequency resolution is used for the comparison of encoding and preferred frequency across a larger set of vocalization stimuli.

Deviations from a purely spectrotemporal mapping of vocalizations may be attributed to nonlinear response properties of the neurons. Nonlinear responses could be displayed by suggested highly selective neurons that respond only to a specific vocalization, but not to others even though those contain spectral properties within the neuron's excitatory tuning range ('call-selectivity' [26]), or other nonlinear responses in the ICC [24, 25, 51].

Not only a topographic representation as demonstrated by Suta et al. [23], but also heterogeneous responses have been shown to exist in the mammalian ICC, and contribute to efficient encoding. Holmstrom et al. [21] recorded from single neurons in the mouse ICC, while presenting four natural vocalizations and modified versions. Using a correlation-based similarity measure, the authors computed neural discrimination for a natural vocalization against its modified versions, based on measures of spike train similarity and information theoretic methods. They showed that neurons display heterogeneous responses to each perturbation of acoustic features in these stimuli, and different neurons responded differently to the same vocalization. Their work indicates that heterogeneous neural responses in the mouse inferior colliculus efficiently encode vocalizations [21]. This group also showed spatial heterogeneity of receptive fields in the mouse ICC [22].

Thus, the encoding of individual vocalizations remains an open question because either heterogeneously distributed neural responses [21] could lead to individual vocalizations being encoded rather equally well across the ICC, or the vocalizations might be encoded more topographically [23], following the organization of spectrotemporal preferences in the ICC [3, 4, 6].

2.2.2 Population coding

Single neurons have been shown to vary in their ability to discriminate natural stimuli such as vocalizations from chance to perfect discrimination [27, 28]. Even though in some cases they yield neural discrimination values of up to 100 % [27, 29], in general, single neurons do not provide enough discriminative information to perfectly distinguish vocalizations. It seems probable that several neurons are necessary to discriminate these complex natural sounds. However, a very large population of neurons responding simultaneously in order to encode a vocalization would not agree with the efficient encoding hypothesis [31]. Multi-units, groups of neurons, have been shown to outperform the discrimination ability of single neurons [30]. In this work it is investigated whether discrimination accuracy changes when combining responses from a successively increasing number of multi-units, and if the accuracy is altered by whether multi-units with similar or dissimilar frequency tuning are combined (Sec. 5.4.2).

Previous studies have investigated the encoding of combined responses with neural discrimination analysis [27, 30]. Schneider and Woolley [27] analyzed separability of combined responses from simulated read-out neurons which received the input from several individual neurons, and they analyzed separability of concatenated responses from several neurons. They combined responses from 2-5 individual neurons with either similar or dissimilar tuning. Improvement was observed for combining up to 3-5 neurons, with the highest gain for neurons with similar frequency tuning. The authors suggest that pooling reduces trial-to-trial variability of single neuron responses and therefore increases separability [27]. In contrast to the work of Schneider and Woolley, in the present work, multi-units are used, which are likely to display different trial-variabilities than single neurons. Therefore it is not obvious how the findings of Schneider and Woolley are applicable to combining responses of multi-units. Possibly different encoding mechanisms are present in the mammalian ICC and in the bird ICC-analog. In the present work responses to vocalizations were combined by concatenation of spike trains, as preformed by Schneider et al. [27], but also by addition of spike trains and concatenation of spike rates. These different pooling methods allow to separately investigate encoding for combining either only spectral or temporal information, or the joint information from different multi-units.

It has been shown that the discrimination ability of groups of neurons, which are composed of several single neurons or individual multi-units outperforms the discrimination ability of a single neuron [30, 27]. Engineer and colleagues [30] have shown this for neurons in the primary auditory cortex of awake and anesthetized rats which were presented human speech sounds, shifted into the rats' hearing range. They demonstrated this by correlating neural discrimination with behavioral discrimination. Neural discrimination of combined responses was performed on 1 ms binned spike trains using the summed Euclidean distance of the bins, and provided to the classifier as a matrix. The authors found that multi-unit responses outperform single neuron responses in discriminating temporally precise onsets of complex stimuli, which has also been shown in previous work [52]. When discriminating responses from 16 combined multi-units, neural discrimination was very high, and correlation with behavioral discrimination was better for an individual multi-unit than for the combined response. The authors attribute this to a ceiling effect of the high discrimination performance for combined multi-units. Engineer and colleagues investigated encoding strategies in the auditory cortex, which need not be the same ones employed by the inferior colliculus, also, their presented complex sounds are not behaviorally relevant for the rats. This might alter encoding of these sounds. In the here presented work, behaviorally relevant vocalization stimuli are used. Furthermore, by combining subsequently the response from one multi-unit, a more gradual change in population size is provided, and allows studying the joint encoding by these populations in the inferior colliculus.

2.2.3 Effect of neural correlations

The neural activity of the units involved in the representation of the vocalization could be coupled in such a way as to lead to a more efficient encoding. However, neural representations have been proposed to decrease in redundancy from peripheral to cortical structures [32]. The hypothesis that neural correlations facilitate discrimination of vocalizations is tested in this work (Sec. 5.4.3). Both, response and neural correlations have been shown theoretically [34, 53] and experimentally [33] to be destructive or invariant [33], and favorable [34, 53, 35, 36] to encoding of sensory stimuli [37]. The effect of the correlations might depend on the specific neural sensory system and its neuronal structure. Nirenberg et al. [33], for example, showed for single cells of the mouse retina using information theoretic techniques that over 90 % of the information about the visual stimulus can be obtained when removing correlations, and therefore suggest that the ganglion cells act largely independently to encode information.

In the auditory midbrain neural correlations of relatively long spike trains have been shown to improve discrimination. Wang et al. [28] performed discrimination on responses to 20 conspecific songs from groups of single neurons and individual neurons in the L-field in awake and anesthetized zebra finches, the analog of the primary auditory cortex in mammals.

To test neural codes, they used a correlation-based similarity measure of spike trains [54] and the Victor-Purpura spike timing metric. They found that 60 % of the units showed an increase in discrimination performance of about 10 %, whereas the rest showed a smaller decrease. The authors state that neural correlations can improve discrimination accuracy. However, these results were only valid for relatively long spike train durations. For spike train durations below 250 ms, no difference for simultaneous and non-simultaneous recordings was visible, and even for durations below 600 ms differences were not significant. Wang and colleagues used a correlation-based similarity measure of spike trains to analyze responses from groups of neurons to vocalizations. This was also performed in the present work for shorter spike trains (Sec. 5.4.3). However, Wang et al. investigated the primary auditory cortex of birds which might display substantial differences in the neural correlational structure compared to the mammalian ICC which is investigated in this study (Sec. 5.4.3).

Summary

The ICC has been suggested to encode vocalizations based on their spectral and temporal acoustic properties [50, 23], but also heterogeneous encoding has been shown [21]. In this work (Sec. 5.4.1), the analysis of Suta et al. [23] is extended in that not the responsiveness of the neurons but their spike train trials are analyzed and tested for separability. Combining responses has been shown to improve discrimination accuracy in higher auditory structures [30, 27]. In contrast to previous work on neural discrimination, in this analysis the combined responses to behaviorally relevant stimuli from multi-units in the mammalian inferior colliculus are investigated by subsequently increasing the population size and separately comparing the temporal and spectral contributions (Sec. 5.4.2). It has been shown that neural correlations do not alter encoding in the zebra finch auditory cortex [28], but not in the mammalian ICC and this is investigated in the present work (Sec. 5.4.3).

2.3 Nonlinear response properties

In Section 2.2, previous studies on the ability of neurons to encode vocalizations were discussed. Neurons might be tuned to specific properties of the vocalizations [38]. This tuning could be formed e.g. by preferences to several features [55], and the joint processing of linear and nonlinear response mechanisms [56]. Neural response properties can be probed with receptive fields. The receptive field which describes the neural preference to a feature is also termed a *filter*, because the neuron filters the sensory stimulus for this feature. In this work, single neurons are probed for preferences to stimulus envelope correlations and these nonlinear properties are compared to linear spectrotemporal preferences (Chapter 6).

2.3.1 Classical and non-classical receptive fields

Receptive fields (RF) describe preferences of the neuron to features of the sensory stimuli. They summarize the encoding characteristics of the particular neuron by displaying the spiking activity in dependence of the stimulus properties. At first in the visual system, receptive fields have been classified into classical and non-classical receptive fields, capturing either linear or nonlinear neural response properties [57]. A neuron whose response can be described as a linear superposition of the responses to the stimulus' independent constituents has linear response properties and these can be characterized by a classical receptive field. Response properties which do not follow this linear integration but exhibit nonlinear summation properties can be described by non-classical (or extra-classical) receptive fields [57, 58, 59, 60, 61, 62]. They were found for the visual [57, 63, 60] and the auditory system [64]. Schulze and Langner found regions which could be excited outside the neuron's classical receptive field [64]. The functional construction of the non-classical receptive field (nCRF) and its interaction with the classical receptive field (CRF) remains an open question. The CRF and nCRF might act together as a single processing unit, optimized for natural stimuli. It has been suggested that stimuli in the nCRF can modulate the responses evoked by CRF stimulation [65, 66]. Natural sensory stimulation of the nCRF increases selectivity and decorrelates the responses of neurons, thus increasing sparseness and information transmission [56]. Thus, in which way nonlinearities in the ICC relate to these definitions of non-classical receptive fields needs to be further investigated.

Non-classical receptive fields have been suggested to exist in the ICC [67]. Schneider and Woolley [67] found that major nonlinear mechanisms in auditory midbrain neurons can be attributed to non-classical receptive fields. These are composed of side-band excitation and/or inhibition and modulate spiking responses to stimuli that fall within the classical receptive fields. The authors recorded responses from 84 single neurons in zebra finches in response to noise, pure tones and vocalizations, and analyzed the responses using spectrotemporal receptive fields (STRFs). The authors suggest that extra-classical receptive fields with a static spike threshold nonlinearity contribute to the formation of stimulus-dependent STRFs.

2.3.2 Spike-triggered covariance analysis to probe response properties

The identification of nonlinear neural response properties can be realized by different analysis methods. One method is the reverse correlation analysis which was pioneered by De Boer and Kuyper [68] and generalized to characterize spectrotemporal filtering properties of neurons throughout the auditory system [69, 70, 71, 72, 55]. From the reverse correlation analysis the spike-triggered average (STA, STRF) and the spike-triggered covariance (STC) can be obtained.

Another method to characterize relevant sensory stimulus features are maximally informative dimensions (MID) [73]. The maximally informative dimensions are obtained by maximizing the mutual information between the neural responses and projections of the stimulus onto low-dimensional subspaces. This procedure is applied iteratively by increasing the dimensionality of this subspace [73]. Methods which combine both approaches also exist. Pillow and Simoncelli described an information-theoretic approach with a linear-nonlinear-Poisson cascade model that combines the STRF and STC [74]. Each of the two approaches, the reverse correlation and the information-based one have advantages and disadvantages. Whereas both methods, STA/STC and MID, can be applied to Gaussian stimuli (e.g. ripple noise, dynamic moving ripple), only MIDs can be directly applied to natural sound and correct for correlations of any order [73, 75]. However, with MIDs only a limited number of relevant features can be extracted. In this work, the spike-triggered covariance method will be used for detecting nonlinear response properties, because it allows capturing all relevant stimulus features and because the dynamic moving ripple sound is used as a stimulus, thus no corrections for global stimulus correlations need to be applied. The second order reverse correlation, the spike-triggered covariance [76, 77] has been applied in the visual [78, 74] and in the auditory system [79, 80, 81, 82, 83, 84, 71]. Relationships between the receptive fields obtained from the first and second order reverse correlation have been found for auditory nerve fibers [79, 80]. Yamada and Lewis [79] applied the STC to predict temporal responses of bullfrog auditory neurons to complex acoustic waveforms, using broadband white noise. They found that the discrete Fourier transform of the highest-ranking singular vector was comparable to the frequency response map of the analyzed neuron, and hence can give information about the neuron's spectral tuning.
Lewis and Dijk [80] have shown for primary auditory afferent fibers using broad band white noise, that the STRF can be derived from the second order reverse correlation. These studies analyzed responses from auditory nerve fibers, thus the results cannot be generalized to the ICC which has been shown to display intrinsic nonlinear processing mechanisms [25]. In these earlier studies, the spike-triggered average and covariance have been computed by averaging segments of the stimulus waveform, whereas in the present study, instead of the whole stimulus waveform, the stimulus amplitude modulations for each frequency carrier of the sound stimulus are used for the analysis (Chap. 6).

2.3.3 Nonlinearities in the ICC

Nonlinearities of the neural response which originate in earlier auditory structures and are present in the input to the ICC, such as the cochlear rectification or cochlear distortions [24] have been found. Nonlinear response mechanisms that are intrinsic processing properties might exist, originating in the ICC, such as spike generating nonlinearities [25], feedback kernels [51] or the suggested neural selectivity for single calls [26]. Whether neurons in the ICC display nonlinear response properties is a controversy.

Using spectrotemporal receptive fields, and measuring preferences for frequency and amplitude modulations, Escabí et al. [14] found that ~ 40 % of ICC neurons in cats (out of 81 single neurons) display nonlinear response properties. Neurons which displayed no preferences to ripple noise but to dynamic moving ripple (DMR) sound that contained specific frequency and amplitude modulations were classified as nonlinear.

Atencio et al. [85] argue that the receptive field dimensionality, i.e. the number of features a neurons is tuned to, increases from the auditory midbrain to the cortex and that at the level of the ICC only one linear stimulus filter is enough to describe neural responses. The authors recorded from single neurons of cats in response to DMR sound and pure tones. They analyzed the responses with spectrotemporal receptive fields (STRF) and the first and second maximally informative dimension (MID, see Sec. 2.3.2). The authors found that the STRF and first MID are very similar in the ICC. Using an information theoretic approach they showed that combining the STRF or first MID with a second MID does not further increase the transmitted information. In the auditory cortex, however, this was not the case, and the second MID yielded substantial increase in the transmitted information. This suggests that more than one filter is necessary to capture processing of auditory cortex neurons, but that the processing of ICC neurons is well described by a linear receptive field. In their study the number of analyzed neurons is not stated and might be small. Concluding from the displayed graphs it appears that around 40 neurons were analyzed which is a rather small set. Thus it cannot be ruled out that the finding applies only to a subset of ICC neurons and that neurons in the ICC display nonlinear response properties.

On the other hand, Andoni and Pollak indicated in their work that ICC neurons encode multiple spectrotemporal features [39]. Using pure tones, logarithmic frequency modulated (FM) sweeps centered around the neuron's best frequency and conspecific calls, they showed that single neurons ($n =136$) in the awake Mexican free-tailed bat ICC are selective for direction and velocity (termed 'spectral motion') of sweeping signals. For their analysis they employed an information-theoretic approach in which the spectrotemporal receptive field and spike-triggered covariance were used to find the most informative subspace that maximized the information between the raw stimuli and the stimuli that evoked a neural response [74]. The authors found that the most informative features were selective for the spectral motion of FM sweeps. Based on their findings, Andoni and Pollak suggest that two mechanisms exist which induce this selectivity, and which apply respectively to half of the neurons. These mechanisms were proposed earlier for neurons in the visual system [86]. The first mechanism for the processing of neurons consists of two linear filters with a quadratic phase shift whose output is squared and summed. In the second proposed mechanism neurons are tuned to opposing directions, which either increase or suppress the spiking activity. This mechanism consists of two linear filters with opposite orientations and a spiking response which corresponds to the difference between their squared output [86]. They authors state that the neurons are tuned to multiple features and that selectivity for natural communication calls can already be observed at the level of the inferior colliculus.

Summary

Non-classical receptive fields which lie outside the excitatory frequency range in the form of side-band excitation and/or inhibition, and which modulate the neuron's response have been suggested for the ICC [67]. In their work, Schneider and Woolley used spectrotemporal receptive fields which capture the linear response, whereas in the present work, nonlinear response properties are derived from the neural response using the spike-triggered covariance. It has been found for auditory nerve fibers, that the spectrotemporal receptive field can be derived from the spike-triggered covariance [79, 80], but this cannot be generalized to the auditory midbrain as different processing mechanisms might exist. It has been suggested that the processing of ICC neurons is well described by a linear receptive field [85], however, another study suggested that ICC neurons are tuned to multiple (also nonlinear) features [39]. These two studies used information-theoretic approaches, whereas in the present study the spike-triggered covariance is employed. In contrast to previous studies which used the entire waveform or stimulus envelope, in the present study, the stimulus amplitude modulations for each frequency carrier of the sound stimulus are used to compute the spike-triggered covariance. This allows identifying variations of nonlinear responses across frequencies and to compare the temporal tuning of the STC to that of the spectrotemporal receptive field. In this analysis, neural preferences to envelope correlations in the mammalian ICC are probed (Chap. 6).

2.4 Summary

In summary, this chapter gives an overview of previous studies that have analyzed the neural representations of vocalizations, the correlations of neural responses, and encoding of these natural sounds by individual neurons and groups of neurons. Furthermore, nonlinear responses and analysis methods to probe these responses are reviewed.

The present analysis of the neural representation along the tonotopic gradient verifies and extends the finding of Suta et al. [23] for a larger vocalization set and for multi-unit responses (Sec. 4.3.3). Response correlations and neural correlations have been analyzed for single neurons in the mammalian ICC only for relatively short spatial distances [45], and it has been shown that nearby neurons have a higher probability of displaying similar neural preferences and responses, suggesting a microcircuitry. In contrast to previous studies, in this work, a large spatial distance is covered, and responses of multi-units to natural sound stimuli are investigated (Sec. 4.4).

The ICC has been suggested to encode vocalizations based on their spectral and temporal acoustic properties [50, 23], but also heterogeneous encoding has been shown [21]. The work of Suta et al. [23] is extended in this study (Sec. 4.4) because it is not based on the responsiveness of the neurons but spike train trials are analyzed and tested for separability. Previous studies have shown that combining responses from several neurons improves neural discrimination [30, 27] and that neural correlations improve discrimination [28]. Contrary to these previous studies, in the here presented work, the combination of neural responses is investigated for the mammalian ICC. This is achieved by subsequently increasing the neural population size and separately comparing the temporal and spectral contributions (Sec. 5.4.2). In contrast to previous work [28], the effect of neural correlations is analyzed for short spike trains (Sec. 5.4.3).

The encoding of vocalizations might result from neural preferences to specific properties of these natural sounds [38]. Nonlinear response properties have been analyzed in the ICC [14, 67, 85, 39, 39], but the analyses did not yield consistent results. Nonlinear side-band excitation and/or inhibition to modulate the neuron's response has been shown [67], and that ICC neurons are tuned to multiple features [39]. However, a different study suggested that ICC responses can be fully described by a linear filter. In contrast to these studies, in this work, the spike-triggered covariance is derived from the amplitude modulation spectrum of the sound stimulus, separately for each frequency carrier, in order to investigate nonlinear response properties in the ICC (Chap. 6).

3 Basics and neural data

In this chapter, the experimental set-ups and the general statistical analysis methods used in the presented work are described. The specific application of the methods to the data and modifications, where appropriate, are included in each Chapter 4, 5, 6. In the first section, the two experimental set-ups to record the neural data, which are used in this work, are described. In the second section, the methods to process the neural guinea pig recordings in order to extract characteristic properties such as the frequency tuning and spike-timing are described. The remaining three sections introduce the statistical methods and models employed in Chapters 4, 5 and 6.

3.1 Experimental set-ups

Neural recordings from two different set-ups are analyzed in this work. These set-ups differ in animal subject, sound stimulus and presentation mode, anesthesia and electrode type, and display different advantages. Guinea pig multi-unit recordings in response to conspecific behaviorally relevant vocalizations (from Set-up 1) are explored in Chapters 4 and 5. Single-unit recordings from the cat inferior colliculus in response to a synthetic spectrotemporally varying sound (from Set-up 2) are employed in Chapter 6. In the following, these two set-ups and the approach to separate the single neuron spiking responses from a multi-neuron recording are described.

3.1.1 Multi-unit recordings in response to guinea pig vocalizations (Set-up 1)

Recording and calibration of guinea pig vocalizations, and recording of neural data were performed by T. Rode, T. Hartmann, and H. Lim and are also described in [87]. The advantages of this set-up are that a representative, spectrotemporally rich set of natural, behaviorally relevant stimuli was presented. Responses were recorded simultaneously from 32 channels and allow studying whether neuronal interactions exist between these sites. However, the recorded data is multi-unit activity with non-separable units, which allows analyzing collective behavior of groups of units, but does not allow drawing inferences about single neuron processing. Furthermore, recording time is relatively short (1-1.6 s) and trial size (n_{trial}=20) relatively small, and therefore does not allow the application of information theoretical and reverse correlation analyses.

Vocalization stimuli

Vocalizations are spectrotemporal dynamically varying sounds, which display a wide spectrum of acoustic properties such as harmonics, temporal correlations, amplitude and frequency modulations. These natural sounds are well suited to study the auditory system as neurons are adapted to process them. They can trigger responses which are not elicited by artificial or simple acoustic stimuli [15].

Guinea pigs are very communicative and display a rich repertoire of behaviorally relevant vocalizations. The eleven vocalizations studied here constitute a representative set of guinea pig communication calls and give information about the animal's behavioral state [88]. Figure 3.1.1 shows the waveform, spectrogram and power spectrum of the vocalizations used in this study. These complex sounds display a variety of frequency ranges and envelope types. The spectrograms of the vocalizations in Fig. 3.1.1e-k display harmonics. The waveform has a simple periodic shape for some vocalizations ('tooth chatter', 'purr', drr', Fig. 3.1.1a-c) and is quite complex for others (e.g. 'squeal', 'low whistle', Fig. 3.1.1j,k). Some vocalizations have a frequency content of up to 30 kHz ('tooth chatter', 'scream long', 'scream short'). Vocalizations were recorded with a sampling rate of 97656.25 Hz from male and female Dunkin Hartley guinea pigs. The vocalizations were played 20 ms after start of recording and vary in duration between 300 ms and 1300 ms.

Pure tone stimulation was used to create frequency response maps (see Sec. 3.2.2). A total of 40 stimulus frequencies, ranging between 0.5-45 kHz, with a ramp rise and fall time of 5 ms each and a duration of 50 ms, were presented 20 ms after recording onset.

Stimulus calibration

During the experiments, the stimuli were presented through loudspeakers via a tube (ca. 2 mm×1 mm) directly to the ear drum, monaurally to the left ear of the guinea pig. To account for this passage and modification of the sound wave, the original vocalizations were calibrated. The impulse response function of the system (loudspeaker-tube-mold of ear canal) was recorded with an ear mold of the ear canal from one guinea pig. The original vocalizations described in Sec. 3.1.1 were filtered with the impulse-response function with an adaptive normalized least mean squares filter prior to presentation ('calibrated'). These calibrated vocalizations were played at a sampling rate of 195310 Hz.

Out of 11 vocalizations, four showed substantial deviations between the original and the calibrated version ('long chutter', 'low chutter', 'drr', 'purr'). All vocalizations showed deviations in onset by several milliseconds. In Figure 3.1.2 waveforms of the original and calibrated vocalizations are given for two examples. For the 'tooth chatter', the original vocalization shows good agreement with the calibrated version (Fig. 3.1.2a). However, in the case of the 'purr', the waveform of the calibrated version is very different to the original one, as no periodic structure is present anymore (Fig. 3.1.2b).
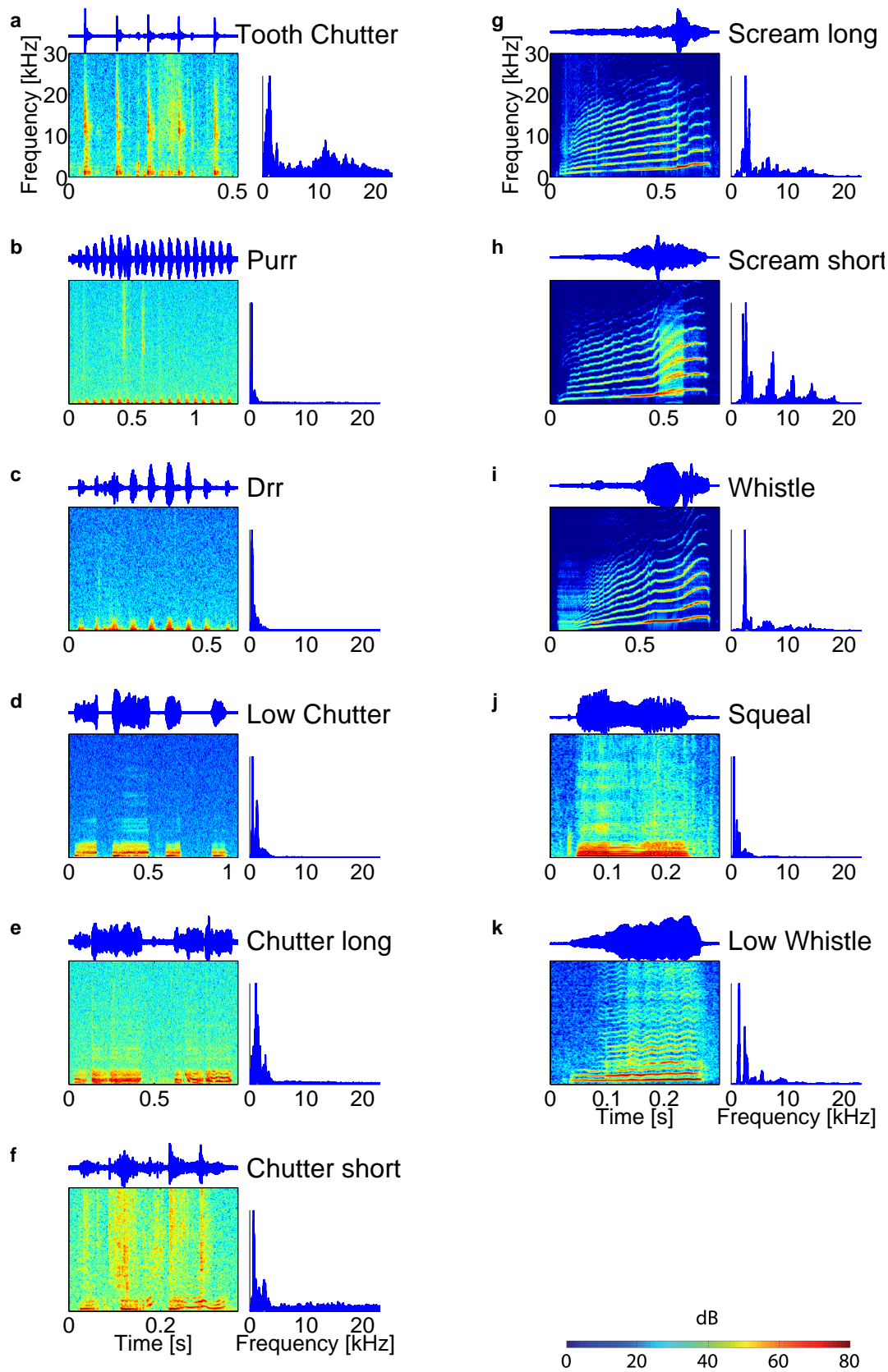
Figure 3.1.1: Guinea pig vocalizations. Waveforms, spectrograms and power spectrums for each of the 11 studied vocalizations. Vocalizations vary in duration from 0.3-1.3 s (j, b), the envelopes of vocalizations shown in a-c have periodicities of respectively 10.3 Hz, 15.4 Hz and 14 Hz. Both 'screams' and the 'whistle' (g-i) show distinct harmonics and broad frequency distributions.
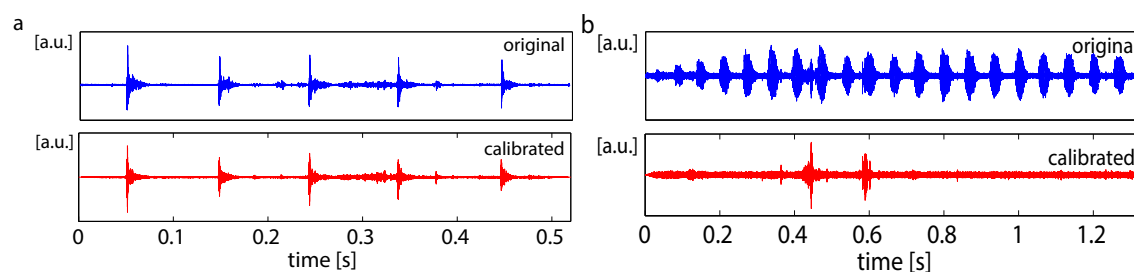
Figure 3.1.2: Calibrated vocalizations. Original (-) and calibrated (-) vocalizations of the 'tooth chatter' (a) and 'purr' (b).

Data acquisition

Neural recordings from the contralateral ICC of 11 adult male and female Dunkin Hartley guinea pigs were taken while acoustically presenting conspecific vocalizations with calibrated loudspeakers via stereotactic ear tubes to the right ear. The experimental set-up is schematized in Fig. 3.1.3. The guinea pigs were anaesthetized with Ketamine and stereotactically fixed with ear tubes through which the sound was presented directly to the eardrum. A total of 11 different vocalizations (see Fig. 3.1.1) were played with intensities of 30-70 dB SPL in steps of 10 dB SPL. For each vocalization, 20 trials were recorded at a given intensity. Recordings of 1 s and 1.6 s duration were taken with a TDT Tucker Davis System with a sampling rate of 24414 Hz. In each animal, the multi-site electrode array was inserted into 3-4 different positions for recording. For the recording, either linear double-shank arrays (Fig. 3.1.4a) or 4-double-tetrode arrays (Fig. 3.1.4b) with impedances between 0.5-1 MΩ at 1 kHz were used (NeuroNexus Catalog, Ann Arbor, MI 48108, 2014). The first one captures multi-unit activity across a wide best frequency range, across several isofrequency laminae, whereas the second captures the activity of several multi-units from few isofrequency laminae. With these arrays the neural activity was recorded simultaneously from 32 different channels (recording sites).
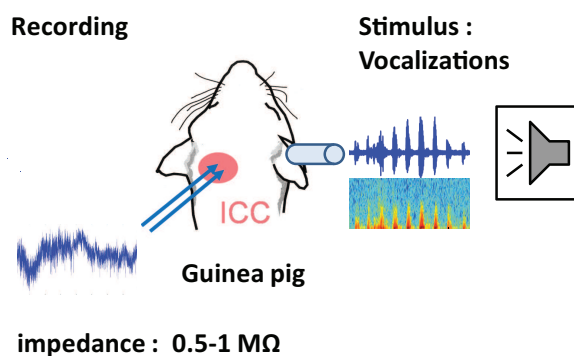
Figure 3.1.3: Set-up 1: Conspecific vocalizations are acoustically presented via a tube to one ear, while recording with a multi-site electrode from the contralateral ICC of the guinea pig (Head drawing adapted from [49].)

Figure 3.1.4: Dimensions of recording electrodes. Multi-site electrodes used in Set-up 1, that simultaneously record neural activity from 32 channels. a) double-shank electrode, consisting of two shanks, with 16 linearly spaced recording channels on each shank; b) 4-double-tetrode electrode, consisting of four shanks with two tetrodes on each shank. The blue inset shows the distances between the channels on one tetrode. Impedances are between 0.5-1 MΩ (Dimensions taken from NeuroNexus Catalog, Ann Arbor, MI 48108, 2014).

The electrode array was introduced under an angle of 45° dorsolateral along the gradient of best frequencies, the tonotopic gradient. Frequency response maps were calculated during the experiment in order to verify the tonotopic gradient [4]. Recordings were taken from 11 guinea pigs in 3 or 4 electrode insertion positions, taken all together 36 recording positions. From each position the electrode array (either double-shank or double-tetrode) that consisted of two shanks recorded simultaneously the activity from 32 multi-units. This yielded 72 sets which consist each of simultaneously recorded activity from 16 multi-units from one shank.

3.1.2 Single-unit recording in response to DMR sound (Set-up 2)

Recordings in the ICC of cats in response to dynamic moving ripple (DMR) sound, and isolation of single neurons were performed by C. Chen, M. Escabí and H. Read and are also described in [14, 89, 45, 90]. The DMR sound is not natural, but contains amplitude and frequency modulations, which are present in natural sound. Recording times are relatively long (2×10 min) and the sound has the advantage that it can be parametrized, which is not easily possible with natural sounds. This allows capturing response preferences with the reverse correlation method (Chapter 6). However, the dynamic moving ripple sound might lack other properties of natural sounds.

Dynamic moving ripple sound stimulus (DMR)

The dynamic moving ripple sound stimulus is a synthetic, parametrizable sound, which contains frequency and amplitude modulations that are important components of natural sound. Its time-varying envelope is used to probe spectrotemporal preferences of neurons.
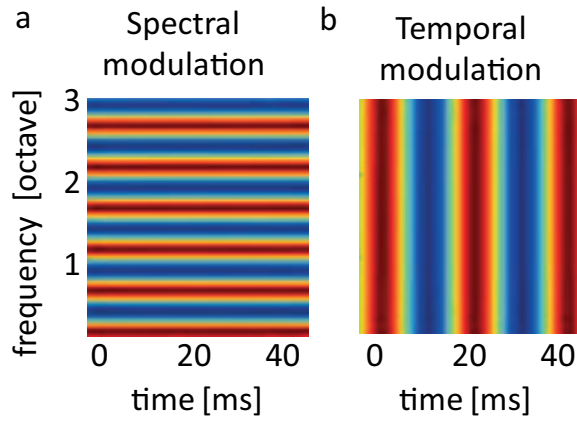
Figure 3.1.5: Frequency and temporal modulation. Sound energy modulation in 1 dimension. Red and blue indicate respectively high and low intensity. a) sound energy changes across frequencies but not with time, only spectral modulation (2 cycles/octave) is present; b) sound energy changes with time, but not across frequencies, only temporal modulation (50 Hz) is present (adapted from [14]).

The DMR sound which was used for the recordings described in this work has first been introduced and applied by Escabí et al. [14], and has since then been widely used [91, 92]. Modulations of the sound envelope in only one dimension are displayed in Fig. 3.1.5. In Fig. 3.1.5a the sound envelope is constant over time but displays variations in energy for different frequencies, the *spectral modulation* is 2 cycles/octave. The sound envelope energy in Fig. 3.1.5b is the same for all frequencies, but the amplitude changes with time (1 cycle in 20 ms $\widehat{=} \frac{1}{0.02 \text{ s}}$), the *temporal modulation* (amplitude modulation) is 50 Hz.

In order to create the DMR sound, energy modulation in time *and* in frequency is applied to a bank of sinusoidal ($n = 659$) carriers of frequency f_k. The frequency carriers span a range between 0-5.8 octaves, with a spectral resolution of 0.0085 octaves. The ripple density $\Omega(t)$ defines the number of spectral peaks per octave at a time instant, and $F_m(t)$ defines the instantaneous temporal modulation. These two values are independent and vary slowly in time: $\Omega(t)$ varies within 0-4 cycles/octave with a maximum rate of change of 3 Hz, and $F_m(t)$ varies between 0-500 Hz at a maximum rate of 1.5 Hz. Within these parameter ranges, amplitudes are uniformly distributed in order to cover the stimulus space in a statistically unbiased manner. The rates of change had been taken from observed rates for similar features in speech and vocalizations [93, 14]. The parameters vary at a slow rate compared with the integration time of ICC neurons, which is about ∼10 ms [90]. The peak-to-peak amplitude of the DMR was set to 30 dB since this produces robust responses for the vast majority of IC neurons [94]. The spectrum of a 400 ms long segment of the DMR sound envelope is displayed in Fig. 3.1.6.
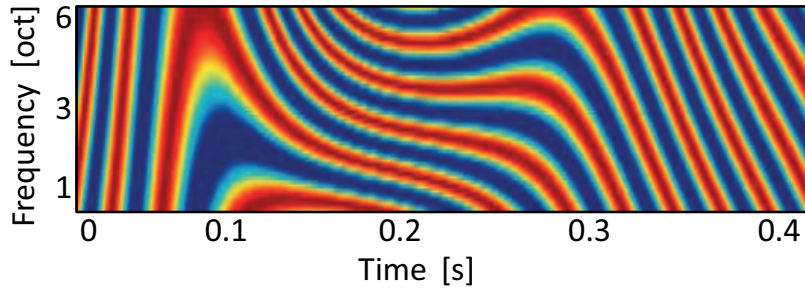
Figure 3.1.6: Dynamic Moving Ripple Sound. The spectrogram of the dynamic moving ripple sound (DMR) envelope (adapted from [14]).

The formula of the envelope is given by

$$S_{\mathsf{DMR}}(t, X_k) = M/2 \cdot \sin\left(2\pi\Omega(t)X_{\mathrm{k}} + \Phi_{\mathrm{k}}\right) \tag{3.1.1}$$

where M is the modulation depth ($M{=}30$), $\Omega(t)$ the ripple density, Φ_{k} is the integrated modulation rate of $F_{\mathrm{m}}(t)$ and $\Omega(t)$, and X_{k} is the frequency axis in octaves relative to the lowest stimulus frequency ($f = 1000$ Hz). The acoustic waveform which was played to the animals is obtained by adding all modulated sinusoidal frequency carriers f_{k}:

$$s(t) = \sum_{k=1}^{L} S_{\mathrm{Lin}}(t, X_{\mathrm{k}}) \cdot \sin(2\pi f_{\mathrm{k}} + \phi_{\mathrm{k}}) \tag{3.1.2}$$

with ϕ_{k} a randomly chosen phase between 0-2π, which creates a noise-like character that is necessary for the reverse correlation analysis, and the modulating spectrotemporal envelope S_{Lin}. This is a transformed representation of the spectrotemporal envelope (Eq. 3.1.1, in decibels) in linear amplitude units bound between $10^{-M/20}$ and 1:

$$S_{\mathrm{Lin}}(t, X_{\mathrm{k}}) = 10^{\frac{S_{\mathsf{DMR}}(t, X_{\mathrm{k}}) - M/2}{20}}. \tag{3.1.3}$$

Compared with natural sound, the DMR offers the advantage that it can be parametrically manipulated, thus allowing for a systematic characterization of neuronal preferences to specific amplitude and frequency modulations. The reverse correlation method requires the input to be white noise [95]. However, it has been shown that neurons in higher auditory areas in primates are not well driven by white noise or simple stimuli [96]. The DMR fulfills several requirements for use with the reverse correlation methods, while containing the frequency and amplitude modulations and coherent modulations in the range characteristic for vocalizations [14, 97]. The local envelope statistics of the DMR are dynamic, they continuously vary with time. Short-term correlations exist and have been evaluated for several envelope segment durations [14]. The global (long-term) autocorrelation consists of an impulse-like central peak of width 3 ms and $\frac{1}{4}$ octave [14]. A narrow impulse-like character of the autocorrelation is a prerequisite to use reverse correlation analysis for the derivation of receptive fields [72, 97]. A 10 min sequence of the DMR was played twice (20 min in total) at a fixed intensity of 80 dB SPL, dichotically in reverse order. The acoustic presentation is described in detail in [14].

Data acquisition

From five adult female cats recordings from the right ICC were taken. Following surgery (described in [14, 89, 45, 90]), the animal was kept anesthetized in a nonreflexive state by continuous infusion of Ketamine and Diazepam. Throughout the experiment heart rate, temperature, breathing rate, and reflexes were monitored and the infusion rate was adjusted accordingly. Neural recordings were performed over a period of 24-72 h. The probes were first positioned on the surface of the IC with a stereotaxic frame (Kopf Instruments) at an angle of 30° relative to the sagittal plane (orthogonal to the frequency-band lamina) [5], and successively inserted deeper in the ICC. Acute 4-tetrode (16 channel) recording probes (two shanks with two tetrode sites on each, 150 μm spacing, impedance 1.5-3.5 MΩ at 1 kHz, NeuroNexus Technologies) were used to record neuronal activity from the ICC. Best frequencies spanned a range of 1.3-16.7 kHz.

Data processing

Neural responses had been recorded with a RX5 Pentusa Base station (Tucker-Davis Technologies) followed by offline analysis in MATLAB (MathWorks Inc.). The continuous neural traces were digitally bandpass filtered (300-5000 Hz) and were used to detect spikes exceeding 5 standard deviations of the spontaneous activity. The candidate action potentials and spike waveforms were aligned and sorted using peak values and first principle components with an automated clustering software (KlustaKwik software) [98].
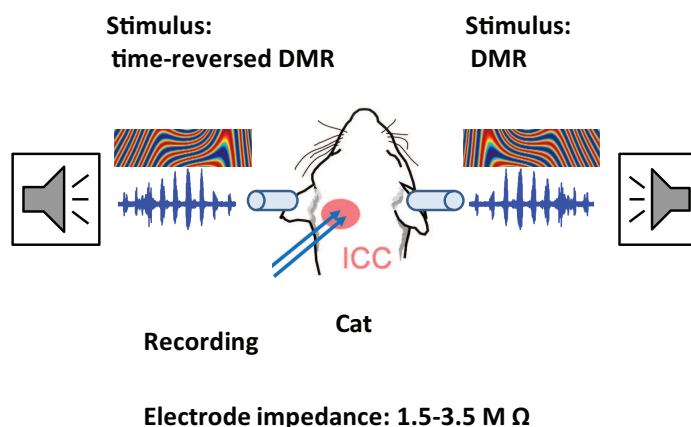
Figure 3.1.7: Set-up 2: DMR sounds (reversed for left ear) are acoustically presented via tube-loudspeaker systems to both ears, while recording with a multi-site electrode (impedance 1.5-3.5 MΩ from the right ICC of a cat. (Head drawing adapted from [49].)

3.1.3 Processing of neural activity

In this section the approach to isolate single neurons from the recordings of Set-up 1 is described.

Multi-unit recordings

The contact of the electrodes (Sec. 3.1.1, 3.1.2) in general captures the neural activity not of a single neuron ('single unit') but of several neighboring neurons ('multi-units'). The lower the impedance, the more neurons are recorded from, thus spanning a wider radius around the electrode contact. The recording radius of the used electrodes (impedances: 0.5-1 MΩ at 1 kHz) was estimated by the manufacturer to be 50-125 μm (NeuroNexus, Ann Arbor, MI 48108, personal comm.). A representative recording of Set-up 1 is given in Fig. 3.1.8b, and a representative one from Set-up 2 in Fig. 3.1.8a. Whereas spikes are clearly visible in the recording from Set-up 2, in the recording from Set-up 1 they are not clearly distinguishable.
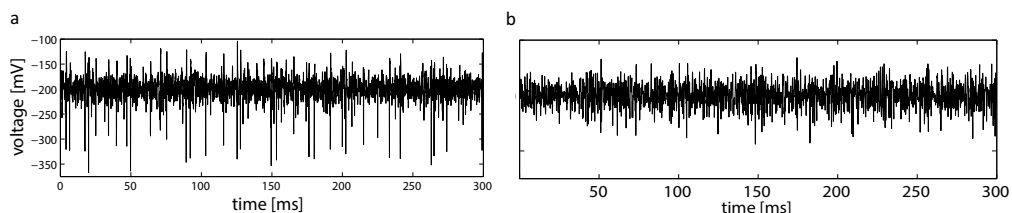
Figure 3.1.8: Recorded neural data. Bandpass-filtered recordings from a) Set-up 2 and b) from Set-up 1, in which spikes are not clearly visible.

Spike-sorting

The spikes from multi-unit recordings may be assigned to different neurons based on their shape and amplitude. In this way several single neuron responses can be obtained from a multi-unit recording (*Spike-sorting*). The histogram of the inter-spike-intervals (ISI) is a good indicator of the spike-sorting quality, and spikes from one neuron with ISI <1 ms are not realistic. The offline spike-sorting program *WaveClus* [99] was used to spike-sort recorded activity from Set-up 1. In this spike-sorting program, a wavelet-analysis is carried out on the recordings, and based on the wavelet coefficients, action potentials are clustered, following the Ising-model. Clustering is based on spin-spin correlations in a super-paramagnetic regime in which aligned domains appear [100].

For the spike-sorting procedure all trials (n=20) were concatenated, filtered between 300-3000 Hz and assigned as spikes for voltages exceeding a certain threshold (Θ). The used parameters for the minimum and maximum value for the temperature for the super paramagnetic clustering were $T_{min} = 0$ and $T_{max} = 0.2$, with 100 iterations for each temperature. The used minimum and maximum value for the number of clusters were $C_{min} = 3$ and $C_{max} = 33$ and for the threshold of spike detection $\Theta_{min} = 3\sigma$ and $\Theta_{max} = 20\sigma$. Further parameters include the sampling rate of 195310 Hz, the number of pre- and post-event data points ($D_{pre} = 15$, $D_{post} = 20$), the detector dead time 3 ms, and the high and low-pass filter frequencies for the spike detection ($F_{low} = 300$ Hz, $F_{high} = 3000$ Hz).
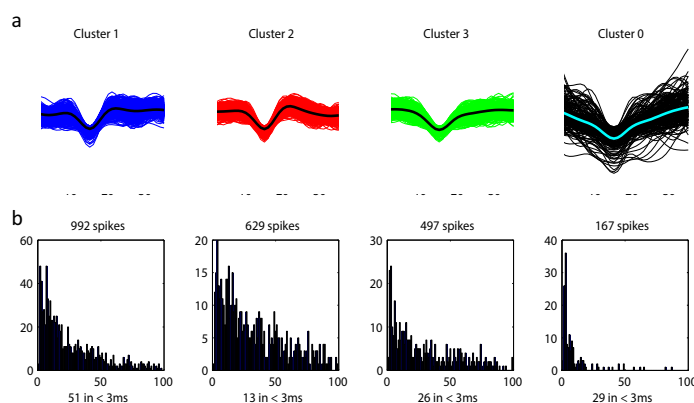
Figure 3.1.9: Spike-sorted waveforms. a) Waveforms assigned to three clusters (Cluster 1-3), and waveforms that could not be clustered (Cluster 0). b) Inter-spike-Intervals for the spikes of the three clusters

Figure 3.1.9a shows an example of the spike-waveforms found with *WaveClus* and ISI histograms for one recording. Three clusters were found to which the waveforms could be assigned, but a high percentage of waveforms could not be assigned to these groups or further clustered. The distribution of inter-spike-intervals for each cluster can give information about the quality of the clustering. Clusters with assigned spikes with intervals of less than 1 ms indicate sub-optimal clustering, because ISI <1 ms are not realistic, and thus spikes originating from different neurons were assigned to the same cluster. All clusters have ISI of less than 1 ms. The neural recordings cannot be well separated. The spike-sorting was performed on a subset of the large set of analyzed multi-units, and was exemplified here for a representative recording.

3.2 Methods

In this section the processing methods, which are applied to the raw voltage recordings of Set-up 1, in order to obtain spiking activity and local field potentials and to obtain basic neural characteristics, are described. The neural data from Set-up 2 were provided in the form of single neurons (Sec. 3.1.2) with neural properties, such as the best frequency, already determined.

3.2.1 Pre-processing

From the voltage recording, which had mean absolute values around ∼80 µV, two different response types, spiking and local field potential activity (LFP), were obtained (Fig. 3.2.1). The mean was subtracted before filtering the responses. In order to obtain spiking responses, a high frequency passband for filtering is applied, and a low passband is applied for LFPs (see Fig. 3.2.1). Separation of single units was not possible (Sec. 3.1.3), therefore the spiking activity of multi-units was employed.
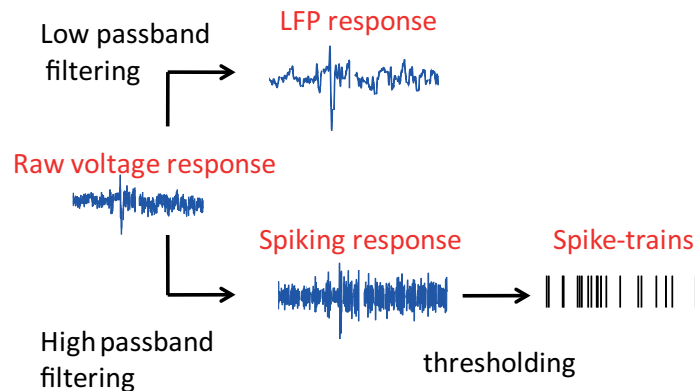
Figure 3.2.1: Filtering for spiking and LFP activity. Low and high frequency passband filtering is applied to the voltage recordings in order to obtain respectively spiking and LFP responses. The spiking response is thresholded to obtain binary spike values (see Fig. 3.2.2).

The multi-unit thresholded spiking activity is referred to as *spiking activity* in the analysis of the recordings from Set-up 1 (Chap. 4, 5). From the spontaneous activity, the mean and standard deviation (σ) is computed. Activity exceeding the threshold, which is a linear function of the standard deviation ($\Theta = z\sigma$), is counted as a spike. No refraction time between spikes was assumed, as these likely originate from different single units. The average spiking response is obtained by binning (temporal window Δt) of individual spike-trains and summation across trials. This yields poststimulus time histograms (PSTH).
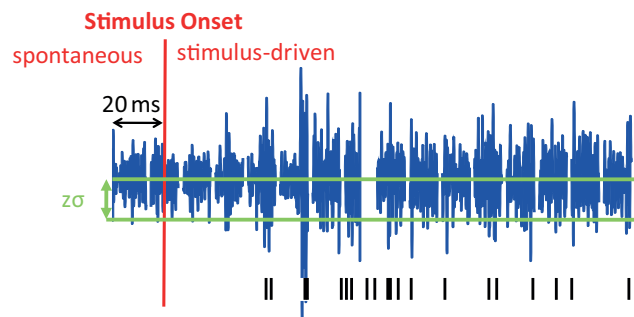
Figure 3.2.2: Multi-unit thresholded activity. Schematic to obtain multi-unit thresholded activity from the filtered recording. During the first 20 ms no stimulus was presented. From these 20 ms, the spontaneous activity, the mean and standard deviation (σ) is computed. Activity exceeding the threshold ($\Theta = z\sigma$) is counted as a spike.

3.2.2 Frequency response maps (FRM)

Frequency response maps (FRM) for a neuron or multi-unit cluster show the average spiking response to pure tones, in dependence of stimulus frequency and intensity (examples are given in Fig. 4.3.1). The stimulus frequency eliciting the highest spike-rate at a given intensity, the tone-evoked best frequency (BF), ranged from 0.5 to 45 kHz in the recordings from Set-up 1.

The best frequency at the lowest intensity that still elicited a neural response is the characteristic frequency (CF). The distributions of characteristic and best frequencies at a stimulus intensity of 70 dB for all analyzed multi-units from Set-up 1 are displayed in Fig. 3.2.4, separately for double-shank and tetrode experiments. The tetrode experiments show a multi-modal distribution of BFs and additional peaks of distinct frequencies for the CF distribution. Tetrode arrays record activity from multi-units of few isofrequency laminae, hence several multi-units have the same preferred frequency. The tetrode arrays were not inserted as deep into the ICC as the double-shank arrays. Therefore mostly responses from multi-units of low and middle frequency laminae more dorso-laterally were recorded from. Double-shank arrays record responses across several isofrequency laminae, and thus cover a wider range of preferred frequencies.
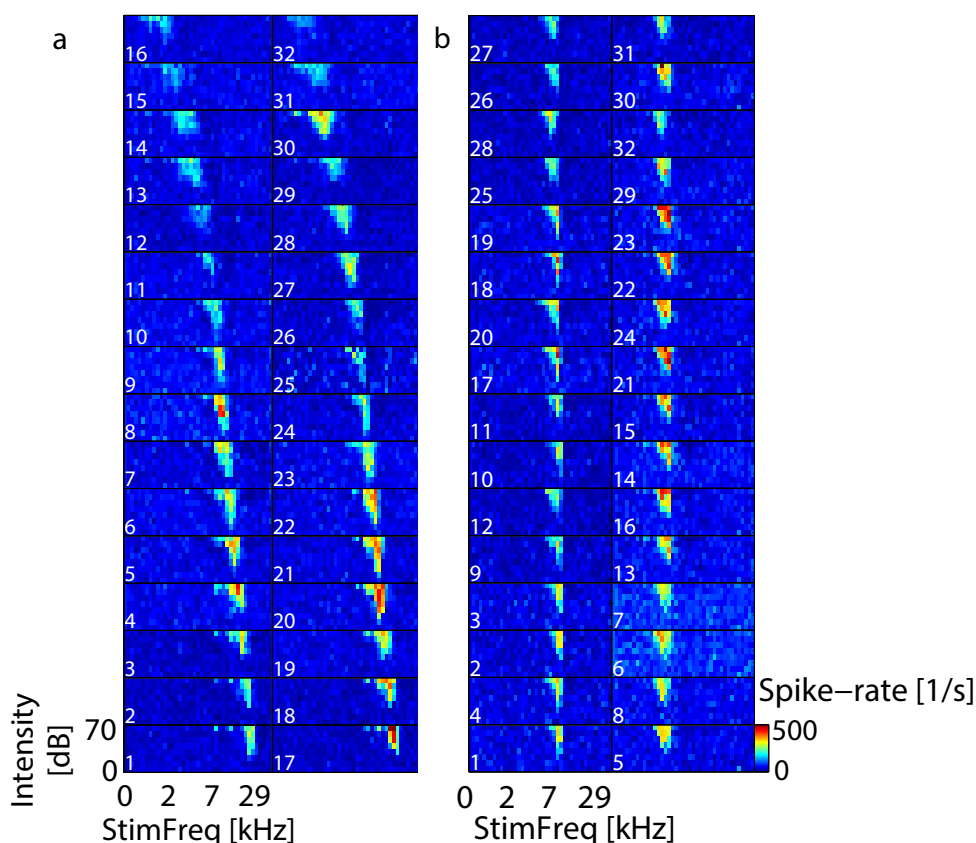
Figure 3.2.3: Frequency tuning along the tonotopic gradient and within isofrequency lamina. a) FRMs of 32 multi-units recorded with a double-shank electrode. The characteristic frequency increases gradually from sites higher up (dorso-lateral) to lower ones (ventro-medial). The FRM shape also changes from a broader symmetric shape to an elongated and skewed one for higher frequencies. The topmost sites do not show strong responses and might be lying outside the ICC (CF range from 0.5-29 kHz). b) FRMs of multi-units recorded with a double-tetrode electrode from two isofrequency laminae with CFs of ∼26 kHz (left column) and ∼15 kHz (right). The CF does not change visibly within one isofrequency lamina, but the frequency tuning varies in sensitivity, e.g. spike-rates from site 23 and site 27 vary by over 250 Hz.

This yields a more continuous distribution of CFs and BFs than for the tetrode experiments. Preferred frequencies from the middle frequency range are more frequent than low or high ones. This could be due to electrodes not all being inserted equally deep into the ICC and covering different frequency ranges at the extremes, but all passing through the middle frequency range.

Low preferred frequencies are not very frequent, which could be due to the anatomy of the low frequency laminae being skewed, and thus not offering a wide surface orthogonal to the tonotopic gradient to record from.
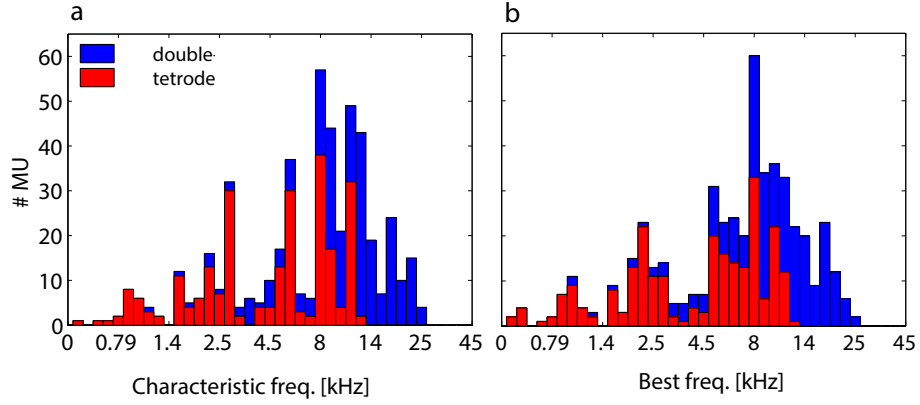
Figure 3.2.4: Distribution of characteristic and best frequencies. The distribution of a) characteristic (CF) and b) best (BF, at 70 dB stimulus intensity) frequencies of all analyzed multi-units is displayed for multi-units from double (-) and tetrode (-) experiments.

3.3 Correlation analysis

The recordings from Set-up 1, presented in the previous sections are analyzed with respect to their similarity across the ICC (Chap. 4). The similarity is quantified by cross-correlation of spike trains. The correlation analysis carried out in this work includes response and neural correlations for responses from multi-unit pairs.

3.3.1 Cross-correlation of neural responses

Correlations between responses are computed in the following manner: pairwise, responses from two multi-units $x(t), y(t)$, of length n, are cross-correlated, and the highest correlation value within a maximum allowed delay of $\tau_{\max}$ between the responses is selected.

$$Corr(\tau) = \max \left(\frac{\sum_{t=1}^{n-\tau}(x(t+\tau) - \langle x \rangle) \cdot (y(t) - \langle y \rangle)}{\sqrt{\sum_{t=1}^{n}(x(t) - \langle x \rangle)^2 \sum_{t=1}^{n}(y(t) - \langle y \rangle)^2}} \right) \qquad (3.3.1)$$

The correlogram is averaged across all trials, before selecting the highest correlation within a delay $\tau_{\max}$. The maximum possible delay $\tau_{\max}$ differed for the specific analysis carried out and is noted in the respective sections.

3.3.2 Cancelling temporal correlations

Neuronal interactions can lead to temporal correlations of their responses. These add to the correlations, which are due to similar neural processing properties and inputs to the neurons (*stimulus correlations*). The stimulus correlations and the correlations due to neural interactions constitute the response correlations which are measured for two neurons. The neural (or temporal response) correlations are only present if recordings were taken simultaneously. Hence, in order to cancel them, simultaneously recorded trials of different neurons are randomly shuffled before the pairwise correlation [101]. The trial-averaged correlogram of shuffled trials is called the *Shift predictor* and represents the stimulus correlations. Subtracting the shift predictor from the measured correlations yields the contribution which is due to neural interactions. This analysis can indicate whether neural interactions exist and whether they are prominent or minor.

The correlations due to interactions between neurons or multi-units, in general, are accompanied by a covariation of the neurons' trial-variabilities, and are therefore also termed *noise correlations* [37]. In order to visualize the effect of noise correlations on the encoded stimulus information, scatter plots are often used. They display the distribution of averaged spike-rates of two neurons or neural groups to different stimuli. The more the response distributions to the different stimuli overlap, the less information is carried by them. Comparing the scatter plot of simultaneous trials to the one with shuffled trials yields information whether the neural correlation affect encoding. If, for example, separability of responses to different stimuli increases when removing neural correlations, these are detrimental for encoding.

3.3.3 Biophysically detailed auditory model (Meddis Model)

The averaged neural spiking responses are compared to predictions of two models (Sec. 4.3.4). The first model consists of filtering and rectification of the stimulus waveform (Sec. 4.2.3), the second one is a biophysically detailed auditory model (Sec. 4.2.3). Several models exist that simulate auditory neural responses [102]. For simulating responses from the ICC, the models by Carney [103] and Meddis [104] are often used.

Here, the Meddis Model was chosen, because it simulates in detail biophysical processing of the auditory periphery. This detailed replication will be compared to a rather simple filtering approach (Sec. 4.2.3). In the following, this model is briefly described along the auditory pathway which was introduced in Sec. 1.1. Sound that arrives at the outer ear is filtered by the ear canal. This is simulated by a bandpass-filter (1-4 kHz with a gain of 10 dB). The amplification of the ossicles is simulated with a low-pass filter (<50 Hz) for high frequencies (>2 kHz) and with a high-pass filter (>1 kHz) for low frequencies. Even the attenuation of sound levels that are very high (>90 dB SPL) by feedback to the middle ear by efferent motor neurons from the medial superior olive is simulated by the model.

The basilar membrane responds linearly at low frequencies and almost linearly at high frequencies. However, at middle frequencies it responds nonlinearly and produces a compressed output. The processing of the basilar membrane is simulated by a dual-resonance-nonlinear filter, which adds the outputs of two parallel pathways, a linear and a nonlinear one. The linear pathway consists of a constant gain and three subsequent identical gamma-tone filters and a second simulated feedback from the brainstem. This second simulated intensity feed-back loop reduces the displacement of the basilar membrane and is activated for loud sounds independently for each best frequency. The nonlinear pathway consists of three identical first order gamma-tone filters, a psychophysically observed compression function, followed by three identical first order gamma-tone filters.

The displacement of the inner hair cells of the basilar membrane, the receptor potentials and conductance changes are modeled in detail. The inner hair cells release vesicles to the auditory nerve fibers. The transmitter quantity released by these vesicles is estimated by using a quantized, a full spiking model. The model can generate independent projections from low and high spontaneous rate auditory nerve fibers. The auditory nerve fibers with their specific best frequency are simulated to converge on cochlear nucleus cells which have the same best frequency. In the model, the cochlear nucleus contains 10 times less cells than there are auditory fibers. The Meddis model simulates the processing of the cochlear nucleus and of one further auditory brainstem nucleus. The simulated processing of this nucleus is not specialized to any particular auditory brainstem nucleus and for the purpose of this work represents the ICC. The cochlear nucleus cells are simulated to project to ICC neurons with the same best frequency, and 10 projections converge onto one ICC neuron. Hence, the model simulates a convergence such that 100 auditory nerve fibers converge onto one ICC neuron.

3.4 Neural discrimination

Neural discrimination has been widely used [30], with a broad range of techniques [105]. Statistical and machine-learning techniques are applied in order to discriminate recorded neural responses which were elicited by different stimuli. It is a powerful tool in analyzing encoding of vocalizations by groups of neurons, as opposed to receptive field analysis which requires further processing if natural sound is used [106].

The neural responses are classified based on a chosen distance metric. In the following, when referring to 'classification' it is meant to imply 'neural discrimination'. A confusion matrix, $Conf$, gives the correct and false classifications, rows representing assigned stimulus classes, columns representing the actual presented stimulus classes, Fig. 3.4.1.
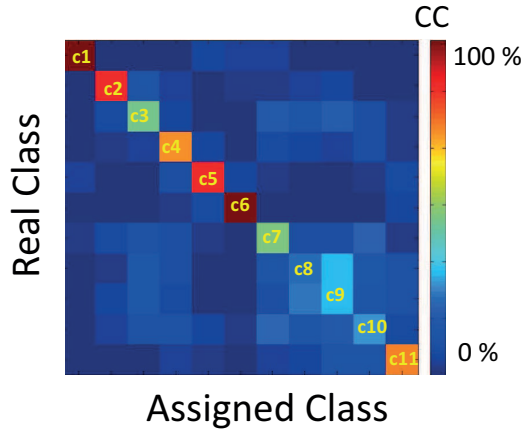
Figure 3.4.1: Example confusion matrix for N_{voc} =11 classes. Entries on the diagonal show the correct classification; all other entries represent false classifications. The average across the diagonal yields the total correct classification (CC).

Perfect classification would yield entries only on the diagonal. The entries on the diagonal of the confusion matrix represent the correct assignments for each individual vocalization $k = 1 \ldots N_{\text{voc}}$, averaged across all cross-validation iterations N_{xval}.

$$CC_i(k) = \frac{1}{N_{\text{xval}}} \sum_{x=1}^{N_{\text{xval}}} Conf_x(k,k) \tag{3.4.1}$$

The average across the diagonal yields the total correct classification. The neural discrimination method applied to the specific data set is described in detail in Sec. 5.2.

3.5 Reverse correlation: Spike-triggered average and covariance

This section introduces the reverse correlation method (spike-triggered average, spike-triggered covariance), testing for significant stimulus variances, as well as two further methods to capture neuronal response preferences (ripple transfer function, conditioned response histogram). These methods are applied in order to investigate nonlinear neural response properties (Chap. 6).

3.5.1 Introduction

The receptive fields are derived from the kernels of the Winer expansion [95], an expansion of functionals which approximates nonlinear systems [76, 77, 69].

It was shown that the Winer kernels up to the second order (Eq. 3.5.1), respectively correspond to the following expressions (Eq. 3.5.2), adapted from [84]:

$$r(t) = k_0 + k_1 + k_2 \tag{3.5.1}$$

$$= \langle y(t) \rangle + \frac{1}{A} \langle y(t) \, x(t - \tau) \rangle + \frac{1}{2!A^2} \langle [y(t) - k_0] \, x(t - \tau_1) \, x(t - \tau_2) \rangle \tag{3.5.2}$$

$$\cong \frac{1}{T} \int_0^T dt \sum_{i=1}^N \delta(t - t_i) + \frac{1}{AT} \int_0^T dt \, x(t - \tau_1) \sum_{i=1}^N \delta(t - t_i) + \frac{N_0}{AN} \sum_{i=1}^N x(t_i - \tau)$$

$$+ \frac{1}{2A^2 T} \int_0^T \left(\sum_{i=1}^N \delta(t - t_i) - N_0 \right) \times x(t - \tau_1) \, x(t - \tau_2) dt \tag{3.5.3}$$

$$= N_0 + \frac{N_0}{AN} \sum_{i=1}^N x(t_i - \tau)$$

$$+ \frac{N_0}{2!A^2} \left[\frac{1}{N} \sum_{i=1}^N x(t - \tau_1) \, x(t - \tau_2) - \frac{1}{T} \int_0^T x(t) \, x(t - \tau) dt \right] \tag{3.5.4}$$

where N is the total number of spikes, T is the entire recording time, A is the power spectral density of the stimulus, $x(t)$ is the temporal input to the system (the stimulus), $t_{i=1...N}$ are spike times, $y(t)$ is the temporal output of the system (the measured spike train), $N_0 = N/T$ is the average spiking rate and τ, τ_1 and τ_2 are time delays from the interval $[0, T]$. The first term is the average system's response. The second term is obtained by cross-correlating the stimulus $x(t)$ with the response $y(t)$; the third term is obtained via a second-order cross-correlation of the stimulus and spike train (Eq. 3.5.2). From this expression, the terms in Eq. 3.5.3 are only approximated due to finite recording times. Spike occurrences at times t_i were represented by δ-function pulses ($\delta(t - t_i)$), and the whole spike train is the summation of all δ-pulses. In the final expression (Eq. 3.5.4), the first term represents the average spike rate, the second term the spike-triggered average, and the third term represents the spike-triggered covariance.

3.5.2 Spectro-temporal receptive field (STRF)

The linear spectrotemporal receptive field (STRF) is obtained by averaging all N_{spikes} stimulus segments preceding a spike [77, 14]; it is also termed the *spike-triggered average* (STA). This linear receptive field corresponds to the second term in Eq. 3.5.4.
The STA in response to the envelope of the DMR sound stimulus (Sec. 3.1.2) is given by:

$$\mathsf{STA}(f_k) = \frac{1}{\sigma_{\text{DMR}}^2 T} \sum_{i=1}^N S_{\text{DMR}}(t_i - \tau, f_k) \tag{3.5.5}$$

with the normalizing variance of the envelope S_{DMR} (Eq. 3.1.1), $\sigma_{\text{DMR}}^2 = \text{M}^2/8$, recording time ($T = 600$ s), and the delay τ relative to the spike timing t_i. This delay determines the temporal width of the window.

The STAs for all frequency carriers f_k correspond to the 2-dimensional linear spectrotemporal receptive field. The STRF is the best linear model that transforms any time-varying stimulus into a prediction of the firing rate of the neuron [69, 71].

3.5.3 Spike-triggered covariance (STC)

The third term of the Winer expansion (Eq. 3.5.1) takes into account nonlinear interactions, and corresponds to the spike-triggered covariance. The formula for the spike-triggered covariance (STC), derived in Eq. 3.5.4, is given by:

$$\text{STC} = \frac{N_0}{2!A^2}\left[\frac{1}{N}\sum_{i=1}^{N}x(t-\tau_1)x(t-\tau_2) - \frac{1}{T}\int_0^T x(t)x(t-\tau)dt\right] \tag{3.5.6}$$

The first term is the covariance of the stimulus. The second covariance is computed from the stimulus segments preceding a spike. The unit of the spike-triggered covariance is spikes per second per squared decibel.

3.5.4 Ripple transfer function and Conditioned response histogram

The STA describes spectrotemporal neuronal preferences, and the STC describes preferences to stimulus correlations. Further descriptions of neuronal preferences are employed in this work. Neuronal preferences to specific combinations of amplitude and frequency modulations are captured by the ripple transfer function (RTF) and the conditioned response histogram (CRH).
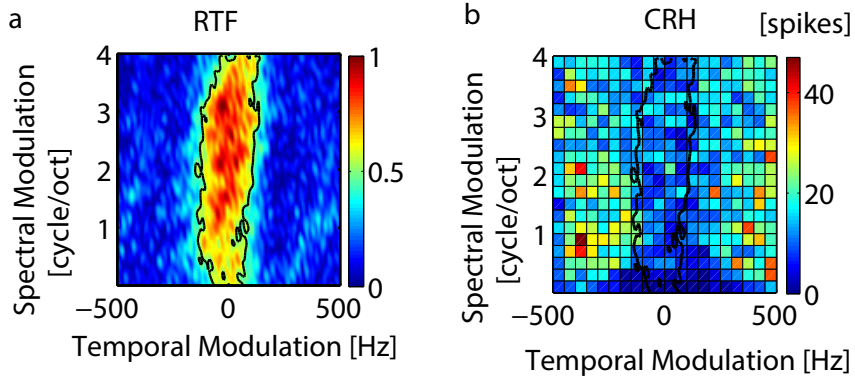
Figure 3.5.1: Ripple transfer function (a); and Conditioned response histogram (b) for one neuron. Positive temporal modulations represent upward sweeps and negative temporal modulations downward sweeps.

The RTF is obtained from the statistically significant STRF ($p > 0.002$) by performing a two-dimensional Fourier transform [107, 14]. From the RTF, the best modulation parameters (best temporal modulation and best spectral modulation) can be derived. The RTF can also reveal the neuron's bandpass or lowpass properties of spectral and temporal information. The neuron's filtering properties are also indicated by the STRF. Neurons that display interleaved patterns of excitatory (*on*) and inhibitory (*off*) subfields in their STRFs, have in general bandpass response characteristics. If such interleaved patterns are absent, then the neurons generally have lowpass response characteristics. Neurons whose activity is not phase-locked to the envelope, and, for which thus no statistically significant STRF can be derived because responses to different spectrotemporal features average out, the RTF-derivation is not possible. Therefore, the RTF only depicts time-locked energy in the neural response.

Another method to map neuronal preferences as a function of the temporal (modulation rate) and spectral (ripple density) stimulus parameters, is the conditioned response histogram (CRH) [14]. In order to obtain the CRH, the range of applied temporal and spectral modulation rates is binned. Any bin of the 2-dimensional conditioned response histogram is incremented by 1, each time the temporal and spectral modulation parameter combination of this bin elicits a spike. This way, one approximates the probability distribution function of observing a specific set of instantaneous parameters given a spike. The CRH is obtained by counting stimulus parameters in dependence of spike occurrence and is insensitive to spike-timing jitter, as opposed to the RTF, which is obtained from the STRF, thus from averaging the stimulus waveforms. Averaging can also eliminate preferences to specific features. The CRH can describe nonlinear neuronal responses which cannot be captured by the STRF or the RTF [14]. The CRH also displays nonspecific activity, which in general does not contribute to the construction of the STRF and the RTF. In general, the STRF and CRH are very similar [14]. In some cases, however, opposite preferences are shown by the RTF and CRH; such an example for one neuron is given in Fig. 3.5.1.

3.5.5 Significance testing of distributions

Significance was assessed using Student's *t*-test for normal distributions, and the Wilcoxon-Mann-Whitney test for comparison of non-normal distributions (significance level $\alpha = 0.05$). The *p*-value is given.

In summary, in this chapter, the experimental set-ups which were used to obtain the neural data, and general processing methods for neural activity were described. The basic statistical methods and models used in Chapters 4-6 were introduced. Specific methods which are applied to the analyzed data sets and questions are detailed in each chapter.

4 Spatial response correlations

4.1 Introduction

The collective response from a group of neurons, i.e. clusters of neighboring neurons that contain a number of single neurons spanning one order of magnitude, is termed *multi-unit* activity. It measures neural activity at a level between patch-clamping, in which single neuron responses are recorded, and larger-scale neurophysiological signals, such as local field potentials (LFP), and e.g. functional magnetic resonance imaging [108], which are used to observe how the system might work as a whole. Investigating the encoding of natural sounds on the level of multi-units has the advantage that this response is an integrated activity which might reflect local processing of the ICC. Furthermore, multi-units respond stronger to natural sound than single units [18] and natural sound stimuli can be more accurately discriminated based on these responses than based on single unit responses [30]. It also has been suggested that an enhanced selectivity to stimulus characteristics over single-unit coding by reduction of background noise might be achieved by association of multi-unit events that are in close temporal correlation to each other [46]. Thus, multi-units might capture collective neural behavior which is complementary to single-unit recording analysis. In this work, multi-unit responses to vocalizations in the ICC are investigated. The ICC receives different pre-processed inputs from ascending nuclei that terminate on different locations within the ICC [109]. Neural preferences to stimulus frequency and modulation are mainly organized gradually within the ICC [4, 6, 7]. The neural response depends on the received input and on the neuron's processing, on its stimulus preferences. The representation of acoustic stimuli which results from the interplay of the heterogeneous input and the gradual organization of neural preferences is investigated for vocalizations which display a rich set of spectrotemporal properties. The hypothesis that despite the various inputs, the best frequency and modulation axis induce a gradually changing neuronal representation of vocalizations across the ICC is tested. To this end, neural responses are compared by cross-correlation with respect to spatial and spectral (difference in preferred frequency) distance of the multi-units. The neural responses are compared for directions along and across isofrequency laminae which correspond respectively to the best frequency and modulation gradient.

Within one isofrequency lamina, neurons have the same preferred stimulus frequency, within a 1/3 octave [5], but a gradient in best amplitude modulation frequency exists [6, 7], whereas across isofrequency laminae, the preferred stimulus frequency changes with spatial distance. The hypothesis is tested that neural responses are more similar within an isofrequency lamina, which was suggested to be a functional processing unit [5], than across the tonotopy. The correlation analysis is performed on the spiking activity, i.e. the processed output of the ICC neurons. The similarity of the spiking response will be compared to the similarity of local field potential activity (LFP). The LFP is long-range activity and captures the presynaptic input [110].

Furthermore, simultaneous and non-simultaneous recordings are compared as this yields indications whether similarity of spiking responses is mainly induced by the stimulus or by neuronal interactions. The latter might suggest cooperative processing of the multi-units. Whether correlated variability of spiking responses enhances or decreases encoded information is a controversy [37]. The effect of noise correlations between two multi-units on the separability of vocalizations in the ICC is displayed.

In this chapter, at first, neural responses to vocalizations are displayed and predicted based on the multi-unit's best frequency, using two different approaches. After description of the responses, similarity in dependence of spatial and spectral distance is obtained by pairwise multi-unit cross-correlation analysis. Finally, response and noise correlations are compared.

4.2 Methods

The provided neural multi-unit data which was analyzed in this chapter and in Chapter 5, are recordings from the ICC of several guinea pigs, to whom a set of diverse conspecific vocalizations were presented acoustically to one ear. The experimental set-up (Set-up 1) and basic methods were described in Sec. 3.1.1, 3.2. In this section, the specific application to the presented multi-unit data is described.

4.2.1 Thresholding, binning and convolution of multi-unit spikes

In order to investigate responses to vocalizations in the ICC, the spiking activity is employed for the analysis, because this is the processed output of the neurons [110]. In some cases, the analysis is extended to the local field potentials (LFP), which is long-range activity and considered to contain i.a. the synaptic input to the neurons [110].

Spiking multi-unit activity was obtained from the recorded voltage traces by subtracting the mean, applying a Butterworth filter with a passband of 300-3000 Hz and thresholding at 3 standard deviations ($\Theta_s = 3\sigma$) above the ongoing activity. Thresholds between $\Theta_s = 1.5 - 3.5\sigma$ were evaluated. A threshold of 3σ yielded the best match of spiking activity and the presence of a stimulus, and is also a common value for thresholding.

This value was kept constant across all analyzed recordings from Set-up 1 in order to detect spikes relative to the spontaneous activity of the specific multi-unit. In this way, different spontaneous rates of the multi-units and adaptation effects over time are accounted for. The spontaneous or ongoing activity was acquired from the first 20 ms of each recording, in which no stimulus was applied.

Bin sizes between Δt=0.5-5 ms were evaluated by comparing the obtained poststimulus time histograms and the eliciting stimulus waveform. A bin size of Δt=1 ms yielded the least noisy, yet detailed representation of the stimulus waveform. Averaging the binned spike trains across all $n_{\text{trial}} = 20$ trials, yields the poststimulus time histogram (PSTH).

In order to reduce trial-variability of these binary spike trains and mimic the time course of excitatory postsynaptic potentials (EPSP), the binned multi-unit spike trains were convolved with a filter function, $f_{\text{EPSP}}(t) = t \cdot e^{-\alpha t}$, with time t [111, 29, 112]. The full width at half maximum α of the filter function was chosen to be 3 ms; as this is the smallest time scale, thus the highest temporal resolution of the time window found to yield maximum discrimination performances between spike trains in response to vocalizations in the (grasshopper) auditory system [29]. Responses to pure tones were binned but not convolved, as these spike trains are only used to estimate average firing rates but are not employed for comparison of the temporal courses. The local field potentials were obtained by Butterworth filtering the voltage traces in a range between 0.5-500 Hz [110].

4.2.2 Stimulus intensity

The poststimulus time histograms were compared for stimulus intensities of 40, 50, 60 and 70 dB SPL. Responses to intensities of 40 and 50 dB were weaker than for the higher intensities. In most cases, responses to 60 and 70 dB were similar, however, in some cases an intensity of 70 dB elicited more pronounced responses that reflected the stimulus waveform. Thus, responses to a stimulus intensity of 70 dB SPL were employed for the analysis.

4.2.3 Predicting the neural response

In order to predict neural responses to a stimulus for multi-units with different best frequencies, two approaches are employed: 1) filtering the stimulus around the multi-unit's best frequency, and 2) simulation of the response with the Meddis Model.

Filtering of stimulus around the best frequency

To test predictability of the multi-unit response, based solely on the stimulus and the best frequency of the unit, the stimulus waveform was filtered. The vocalizations (displayed in Fig. 3.1.1 on page 25) were Butterworth filtered around the best frequency of each multi-unit, and then compared to the PSTH of this multi-unit.

The applied passband was 1/3 octave and positioned symmetrically around the best frequency, as this is the frequency range contained within one isofrequency lamina. To allow for comparison, the filtered stimulus was rectified and normalized to the maximum value of the PSTH. To quantify similarity of the filtered stimulus and the stimulus-elicited segment of the PSTH, cross-correlation was used (see Sec. 3.3). A delay of maximum $\tau=10$ ms was allowed, in order to account for possible temporal shifts due to stimulus calibration (see Sec. 3.1.1).

Meddis Model parameter

The biophysically detailed auditory periphery model, the Meddis model was introduced in Sec. 3.3.3. In order to compare the model prediction for vocalization stimuli to the measured responses, some changes to the Meddis Model had to be made, and specific model parameter values were used. The Meddis Model was modified to yield simulated PSTHs, rather than one only single unit output. Furthermore, the spiking threshold was adapted to be the one of a guinea pig. An added current of 28 nA per spike at the level of the cochlear nucleus yielded an induced spontaneous activity of neurons in the ICC, which was observed for the physiological data (see Fig. 4.3.3). Only high spontaneous rate auditory nerve fibers, which are more abundant than the low spontaneous rate fibers, were considered as a first approximation of the output, as the model currently allows only one type of fiber for simulation of ICC membrane potential. The model simulates one single spike train for a neuron with a given preferred frequency. To approximate the multi-unit activity, the responses of five single neurons with closely spaced characteristic frequencies (with a maximum difference of 0.1 octaves) around the CF of the recorded multi-unit were simulated. These five spike trains were binned and averaged as the recorded data (Sec. 4.2.1). The simulation and recorded data were cross-correlated as described in the previous paragraph.

4.2.4 Neuronal pairs from the same or different isofrequency laminae

Responses from multi-units from the same lamina are compared to those from different laminae in this chapter. In general, tetrode arrays record from several sites within the same isofrequency lamina, whereas double-shank arrays mainly record responses from different isofrequency laminae, but deviations exist for both cases. Multi-unit pairs were assigned to the same or to two different isofrequency laminae, based on their spectral and spatial distance. Pairs differing by more than 1/3 octave in their characteristic frequency were assigned to different isofrequency laminae. The correlation analysis was performed on pairs from the same recording yielding ideally $(32 \times 31)/2 = 496$ pairs, which do not include autocorrelations or correlations counted twice.

However, only pairs with a minimum distance of 200 µm were considered for the analysis, in order to assure that no multi-units are taken from adjacent sites, which possibly share the same neurons and therefore yield excessively high correlation values. First, the distances between the multi-units of all pairs were mapped according to the channels on the electrode. Distances between all 32 channels were obtained either directly from the NeuroNexus manual (NeuroNexus Catalog, Ann Arbor, MI 48108, 2014) or calculated using the Pythagorean theorem, yielding a 32×32 matrix, respectively for the double-shank and the double-tetrode array. Minimum and maximum distances for the double-shank and double-tetrode array were respectively $D_{min} = 100$ µm, $D_{max} = 1581$ µm and $T_{min} = 25$ µm, $T_{max} = 1372$ µm. The obtained spatial distances were also used to investigate if correlation values depend on the spatial distance between two multi-units (Sec. 4.4.3, 4.4.4).

4.2.5 Averaging across multi-units

Correlation is computed between multi-unit responses, either to spike trains convolved with the EPSP-filter function or for LFP activity, to the same vocalization, from one recording set. One recording set consists of responses from 32 multi-units. The findings are verified to be generally true for all sets. The average across all recording sets is computed as well. In the following, results for an individual recording set and the average across all recording sets are displayed. When averaging correlation values across multi-unit pairs, electrode positions and animals, the displayed error is the standard deviation and depicts the correlation variability across multi-unit pairs. The error was also computed via error propagation, thus assuming independence of all results, which in general is true, except for correlations of close-by multi-units from one recording set. These errors are close to zero in many cases. However, the displayed error was chosen to be the standard deviation to indicate the variability across multi-unit pairs. The distribution of spatial distances was not uniform and also varied for double-shank and tetrode recordings. Therefore, the number of multi-units, for which averages were taken for one spatial distance, varied across recordings. Differences in response similarity based on different locations and thus different dominant inputs, as well as amplitude modulation and frequency preferences are averaged out when taking the mean across all multi-units and yield a result, generally valid across the ICC.

4.3 Neural responses

The spectral tuning of the multi-units determines their responses to vocalizations [23]. In the following, examples of averaged responses are displayed, and two approaches to predict the neural response based on the spectral tuning are presented.

4.3.1 Multi-unit frequency response maps

The preferred stimulus frequencies of the multi-unit, its characteristic (CF) and best frequency (BF), were determined from the frequency response map (FRM), as described in Sec. 3.1.1. The frequency response maps show the compound response of several single neurons to pure tones. Examples of differently shaped multi-unit FRMs are displayed in Fig. 4.3.1. The multi-unit FRMs are similar to single neuron FRMs and previously reported multi-unit FRMs [113, 114]. The characteristic frequency was used to define spectral distances between two multi-units (Sec. 4.2.4). The preferred frequency at the specific stimulus intensity, the best frequency, and the overall spectral tuning of the multi-unit determines the response of the neuron to vocalizations.
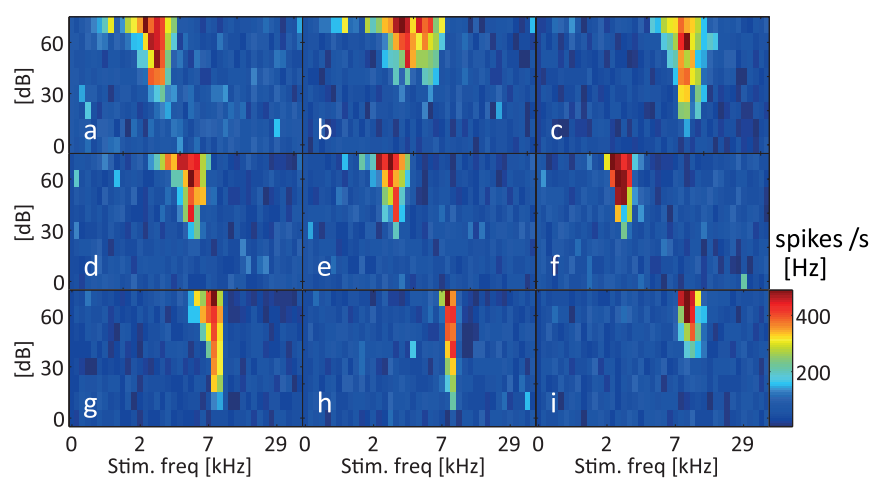
Figure 4.3.1: Multi-unit frequency response maps. The frequency response maps display different shapes: (e, f) symmetric v-shaped, (a, g) skewed v-type shape and (d) intermediate shapes between these two; (b, c) show broad frequency tuning, (h, i) very narrow frequency tuning. Spike-rates are high throughout several dB-levels, the highest spike-rate being 450 Hz.

4.3.2 Distribution of neural response energy across best frequencies

The vocalizations evoke enhanced spiking activity in several isofrequency laminae (see Fig. 5.2.4). The white noise, which contains energy in a wide frequency range, also drives neurons to spike. In contrast, pure tones usually only elicit responses in a few isofrequency laminae that match their stimulus frequency. Based on the firing rate, the similarity of responses across the ICC could be investigated. However, temporal information is crucial for recognition and discrimination of complex dynamically varying sounds such as speech and vocalizations [20, 30]. In this work, the temporal response courses of the multi-units across the ICC will be compared.
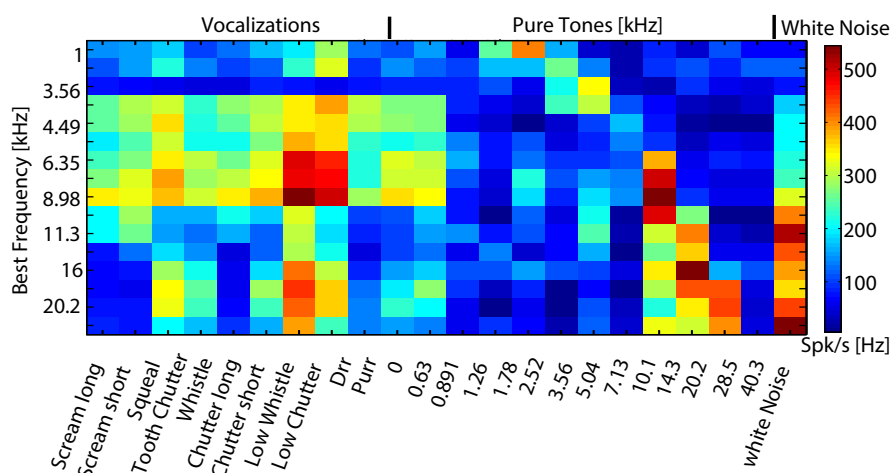
Figure 4.3.2: Response energy across the frequency gradient. The average spike rate across the duration of the stimulus for all 11 vocalizations, 10 pure tones and white noise is displayed for multi-units along the tonotopic gradient from one recording, spanning a best frequency range 1-25.4 kHz.

4.3.3 Poststimulus time histograms

The post-stimulus time histograms (PSTH) display the trial-averaged (n_{trial} =20) temporal response. Responses to vocalizations vary for differently frequency-tuned multi-units and follow the spectrally matching components in the stimulus, see Fig. 4.3.3. Although less strongly, responses from multi-units with the same best frequency also vary. This has been previously found for a subset of the vocalizations used in this work [23]. Figure 4.3.3 displays the PSTHs of multi-units along the best frequency gradient in response to three vocalizations. Responses to the 'tooth chatter' phase-lock to the stimulus envelope for multi-units throughout the whole best frequency range (Fig. 4.3.3a), because the stimulus has spectral energy in this range. However, the responses in general become broader for high BFs. Responses to the 'purr', on the other hand, phase-lock accurately only for low BFs, but then become broad and unspecific (Fig. 4.3.3b). For the 'purr', spectral energy is present for frequencies up to 3 kHz (Fig. 3.1.1b), except for a high frequency contribution at 600 ms, which is probably due to the filtering of the stimulus prior to presentation (Sec. 3.1.1). To this feature, almost all multi-units phase-lock. At the start and at the end of the stimulus presentation, onset and rebound responses are pronounced for middle and high BFs. The response's dependence on the match of the best frequency and the spectral content of the stimulus [23] is clearly illustrated by responses to the 'scream' (Fig. 4.3.3c). In the beginning of the stimulus, only low frequencies are present, and only low-BF multi-units respond. Subsequently, the stimulus contains frequencies up to 25 kHz and middle-BF multi-units respond. High-BF multi-units respond to a high frequency peak at 600 ms. Similarity of responses to the same stimulus by different multi-units can be directly obtained by comparing the PSTHs.
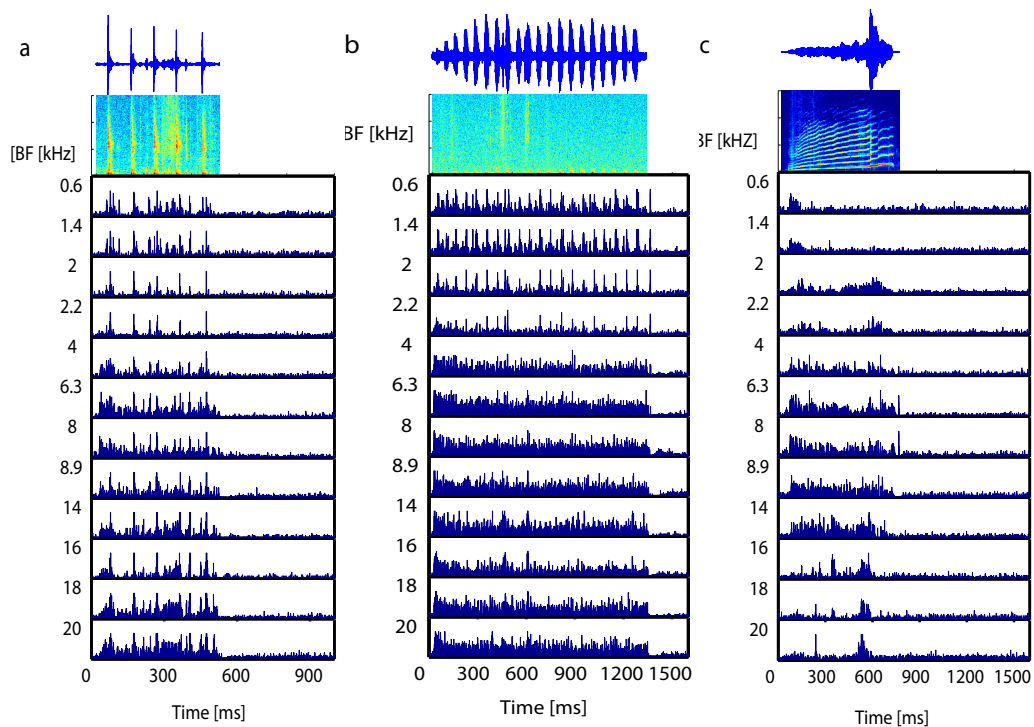
Figure 4.3.3: Poststimulus time histograms in response to the vocalizations a) 'tooth chatter', b) 'purr', and c) 'long scream' for multi-units of a double-shank recording, spanning a best frequency range of 0.6-20 kHz. On top, the waveforms and spectrograms are displayed (detail in Fig. 3.1.1).

However, in order to quantify similarities across the large data set of 1120 multi-units used in this work, a more compact measure is employed. To this end, responses are cross-correlated and the correlation value indicates the degree of spike train similarity (Sec. 4.4).

4.3.4 Prediction of neural response

The match between the multi-unit's best frequency and the stimulus' spectral content substantially influences the neural response. In the following, the predictive power of the neuronal response based only on the stimulus waveform and the multi-unit's best frequency is displayed with two models. At first, a simple filtering approach of the stimulus waveform around the best frequency is used, and in a second approach an auditory periphery model, the Meddis Model is applied.

4.3.5 Stimulus filtering around best frequency

The dependence of the neural response on the spectral content of the vocalization and the neuron's preferred frequency was exemplified in the previous section. This is further demonstrated when predicting the neural response for multi-units with different BFs.

The eliciting stimulus waveform is filtered around the best frequency of the multi-unit (Sec. 4.2.3). Figure 4.3.4 displays the neural responses and the filtered vocalizations, for the same three examples as given in Fig. 4.3.3. The overlap is good in some cases, but less good for others (indicated in Fig. 4.3.4). Further examples are given in Fig. 4.3.5, which show multi-units from a tetrode recording. The filtered response is the same for units with identical best frequencies. The correlation values are also provided, but these do not always well capture the overall match, which is best evaluated visually. Phase-locking is replicated by this filtering approach, but the strong onset responses of the neurons are not captured in some cases. The finding that this filtering approach yields good overlap for some of the multi-units is surprising, given that the ICC is a major converging station and receives input from various auditory processing stations. This has been found independently from a similar study within a collaboration [87], which attributes the responses with a good overlap to multi-units that receive direct projections from the cochlear nucleus, which is the first brainstem nuclei and processes responses from the auditory nerve fibers. In some cases, multi-unit responses to vocalizations in the ICC are well predicted by best-frequency filtering, rectification and normalization of the stimulus waveform.
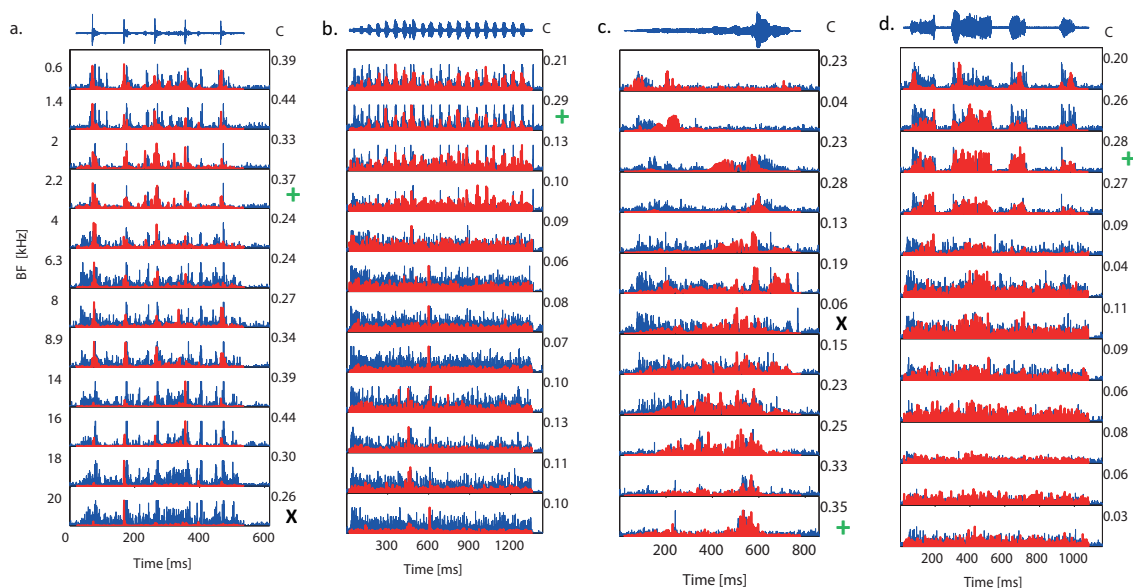
Figure 4.3.4: Stimuli filtered around the best frequency. Comparison of filtered vocalizations (-) with the neural response in form of PSTHs (-), for the a) 'tooth chatter'; b) 'purr'; c) 'scream'; d) 'low chutter' for multi-units with best frequencies: 0.6-20 kHz. On top, the original waveforms are displayed. The correlation values (C) of the filtered stimulus and neural response are given, and examples of good (+) and less good overlaps (×) are marked.
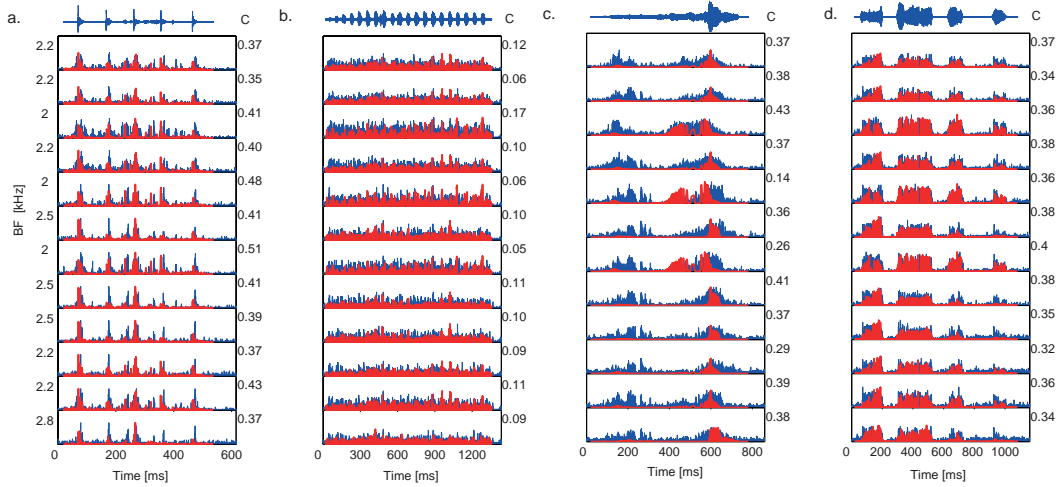
Figure 4.3.5: Stimuli filtered around best frequency for a tetrode recording. Comparison of filtered vocalizations (-) with the neural response in form of PSTHs (-), for the a) 'tooth chatter'; b) 'purr'; c) 'scream'; and d) 'low chutter' for multi-units with similar best frequencies BF:2.2-2.8 kHz. On top, the original waveforms are displayed. The correlation values (C) are given.

4.3.6 Comparison of neural responses with prediction of Meddis Model

In the previous section, an approach which consisted of stimulus filtering, rectification and normalization yielded, in some cases, good overlap of the neural response and the prediction. In the following, predictive power of a second approach is tested. This is a biophysically detailed model, which takes into account individual processing of auditory periphery stations, to simulate neural responses (Meddis Model, Sec. 4.2.3). Figure 4.3.6 shows the recorded and simulated responses. The simulated response strength varies with the spectral content, which was also observed for the neural recordings (Fig. 4.3.3). However, phase-locking is not well replicated and responses are broad and unspecific (Fig. 4.3.6a). On average, correlation values are lower for this approach than for the filtering approach in Sec. 4.3.5; e.g. correlation values are for the 'tooth chatter' are $M_{Meddis} = 0.19 \pm 0.05$, $M_{Filter} = 0.33 \pm 0.07$, and for the 'scream' $M_{Meddis} = 0.15 \pm 0.09$, $M_{Filter} = 0.206 \pm 0.097$. The simulated mean firing rates are higher than those of the neural data. Phase-locking, especially for low best frequency multi-units is not replicated in the simulation. However, it is present for lower stimulus intensities of 40 dB. Thus, the lack of phase-locking might be due to an increased unspecific response at low BFs which hides the phase-locking. The threshold of the model might require further adjustment of simulation and neural data, especially to match both, stimulus-elicited and spontaneous rates. The latter ones are not displayed as they were not used for the correlation analysis.
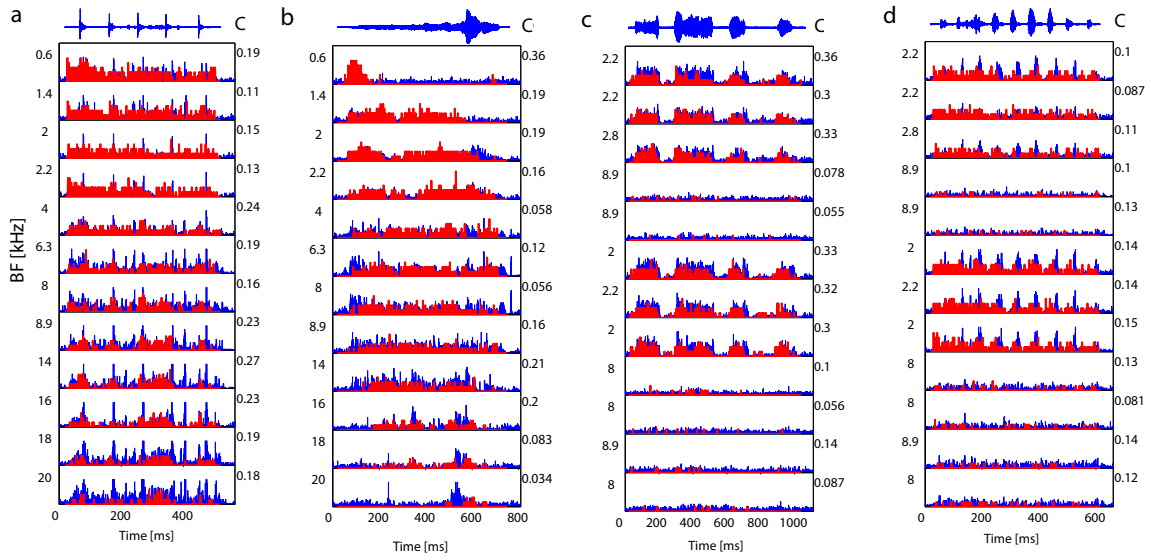
Figure 4.3.6: Simulated neural responses with the Meddis Model. Comparison of simulated response (-) with the neural response (-), for the a) 'tooth chatter'; b) 'scream'; c) 'low chutter'; and d) 'drr' for multi-units with best frequencies between 0.6-20 kHz. (a, b) show responses from a double-shank recording, (c, d) those from a tetrode recording, with several similar best frequencies. On top, the waveform is displayed. The correlation values (C) of the simulated and recorded response are given.

Furthermore, the modified Meddis model, even though very biophysically detailed, in the present form has some limitations, e.g. only one pathway from the auditory nerve to the ICC is modeled, although several projections to the ICC exist (Sec.1); inhibitory interactions between neurons, populations of neurons, or from different processing stations (e.g. the contralateral ICC) are omitted; the cell type used in the model are chopper cells, while e.g. primary-like and pauser cells also exist in the ICC [109]. However, deviations might also be caused by the stimulus calibration (Sec. 3.1.1). Thus, the specific application of the biophysically detailed Meddis model to predict multi-unit responses to vocalizations in the guinea pig ICC without further modifications does not exceed the prediction provided by the filtering approach (Sec. 4.3.5, 4.2.3). Filtering the stimulus waveform around the best frequency yields a good prediction of the neural response for some multi-units.

4.4 Correlation analysis

In Sec. 4.3, recorded neural responses to vocalizations and predicted responses, based on the neurons' best frequencies, were displayed. In the following section, the similarity of responses across and within isofrequency laminae is investigated by cross-correlation (see Sec. 3.3.1), which yields a compact description for the large set of multi-units analyzed in this work. Responses, whose temporal course are compared, are either EPSP-convolved spike trains, hereafter referred to as 'spiking activity', or LFP-activity (see Sec. 4.2.1).
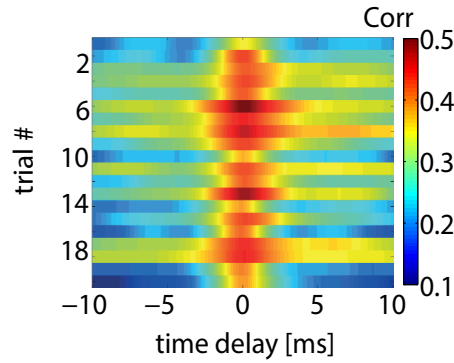
Figure 4.4.1: Trial-variability of cross-correlation values. Cross-correlation of spiking activity from two multi-units ($BF_1 = 14$ kHz; $BF_2 = 20$ kHz) within a delay window of 10 ms, for all $n_{trial} = 20$ trials. The spiking activity is in response to the 'long scream'.

When pairwise cross-correlating the spike-trains of two multi-units, the maximum correlation value varies in magnitude and relative delay for the different trials, see example for correlating spiking activity Fig. 4.4.1. Similarity of responses was compared using cross-correlation (see Eq. 3.3.1) with a maximum delay $\tau_{max} = [-10 \text{ ms}, 10 \text{ ms}]$ for neuronal pairs across and within isofrequency laminae. This delay is within the range of maximum response latencies found in the ICC [115]. The average peak value across all trials was computed to quantify response similarity.

4.4.1 Intra versus inter-lamina response similarity

In order to test if responses to vocalizations are more similar across the best frequency gradient or within an isofrequency lamina, which were suggested to be functional processing units [5], response correlations across and within laminae are compared. The Figure 4.4.2a displays correlation values for pairs of multi-units from one recording, for each vocalization, for spiking activity and local field potentials. Correlation values show large variability within one recording set, as depicted by the error bars which correspond to one standard deviation. However, correlation values are significantly higher within isofrequency laminae than across laminae (two-sided Wilcoxon-Mann-Whitney test, $p < 0.05$) for almost all vocalizations. When averaging correlation values across all recordings (n = 4,223 multi-unit pairs), differences are significant for all vocalizations for LFP activity, but are much smaller for spiking activity and not significant for two vocalizations, Fig. 4.4.2b. Correlation values are significantly bigger for LFP than for spiking activity, for each vocalization, within and across laminae ($p < 0.05$). The LFP responses are long-range activity, and spread throughout wider spatial and frequency regions, than spiking responses which are confined to one multi-unit. Response similarity varies across vocalizations, partly significantly. Vocalization-pairs (2-4,10-9) do not vary significantly, respectively for spiking (and LFP) responses across and within laminae.
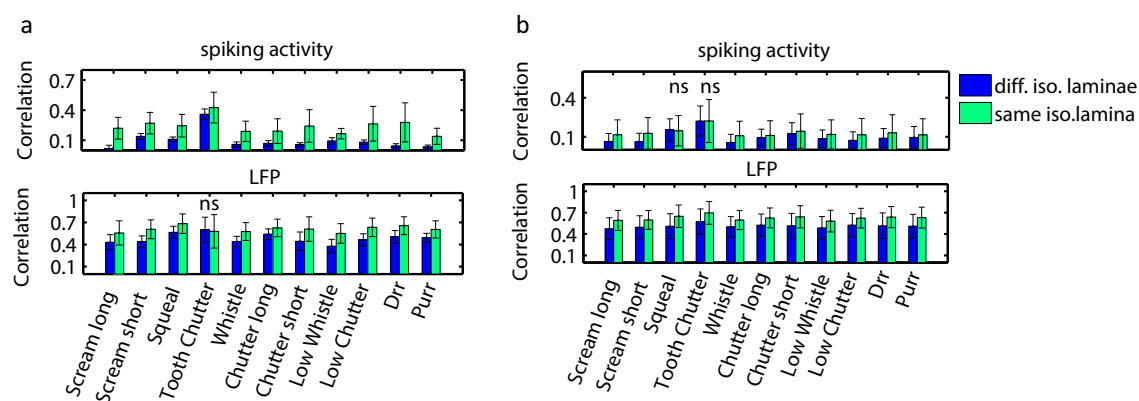
Figure 4.4.2: Correlations within and across isofrequency laminae. Averaged correlation values for multi-units a) from one recording (n = 155), and b) averaged across all recordings (the number of pairs was kept constant across and within laminae for comparison, n = 4223), are displayed for spiking activity (top) and local field potentials (bottom). Responses within isofrequency laminae are mostly significantly more correlated than across laminae (two-sided Wilcoxon-Mann-Whitney test $p = 0.05$).

The responses to the 'tooth chatter' are significantly larger than for all other vocalizations, which is due to the stimulus' spectral content across a wide frequency range and the responses' phase-locking throughout this range (see Fig. 4.3.3a).

Higher correlation within isofrequency laminae might point to vocalizations being processed by isofrequency laminae as functional units, however, even though the differences are significant, differences are minor and variability across multi-units is high.

4.4.2 Dependence on spectral distance

The analyzed vocalizations show diverse spectral contents (see Fig.3.1.1). In general, multi-units respond strongly to stimulus segments whose spectral content matches their best frequency (Sec. 4.3). Thus the question arises how similar responses are from neurons with different preferred frequencies, and which mainly respond to different segments of the stimulus? Matrices displaying the correlations between all 32 channels of one recording were computed for each of the 11 vocalizations, for spiking activity (and LFP), yielding 660 matrices (four examples are shown in Fig. 4.4.3). Neighboring channels, with similar preferred frequency show relatively high correlation values. However, no systematic pattern of high correlations in dependence of spectral distances was found for individual vocalizations. The patterns differ across recordings and across vocalizations. In the example of Fig. 4.4.3a, the 'tooth chatter' elicits very high correlation values for neighboring multi-units which display similar frequency tuning, and response similarity decreases with spectral distance and distance of the recording sites.
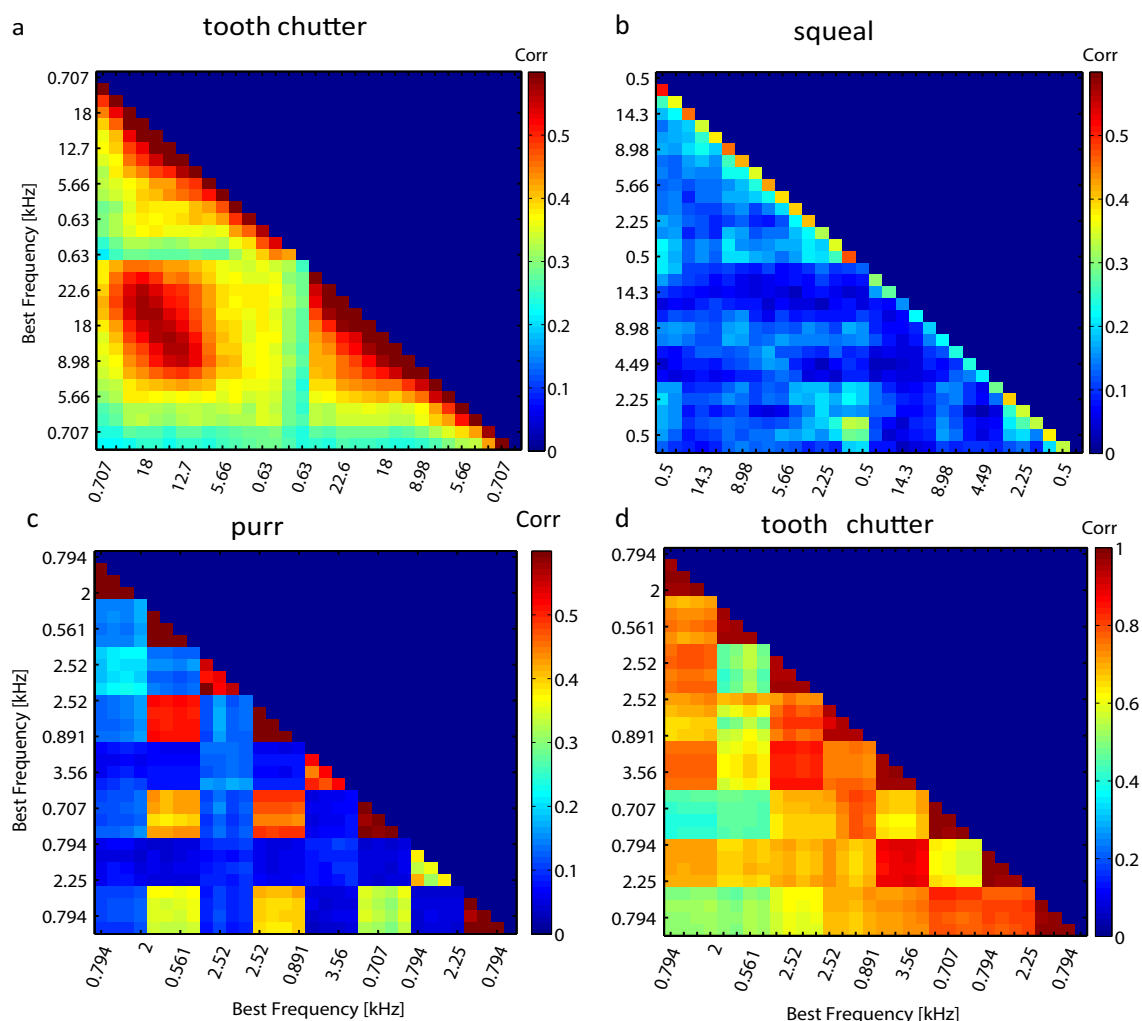
Figure 4.4.3: Correlation dependence on spectral distance. Correlation values between the 32 multi-units of one recording and the best frequencies are displayed for the a) 'tooth chatter', b) 'squeal', (a,b) from double-shank recordings; c) 'purr', and d) 'tooth chatter', (c,d) from tetrode recordings. (a,b,c) represent correlations of spiking activity, (d) of LFP activity.

A further peak of high similarity is present at middle-high frequencies, which could be due to strong phase-locking to the stimulus throughout a broad frequency range (Fig. 4.4.3a). In several cases, higher correlation values are only observed for adjacent multi-units, which display small differences in spectral tuning (Fig. 4.4.3b). The correlation matrices computed from recordings with double-tetrodes show high correlation values for patches of 4×4 multi-units (Fig. 4.4.3c, d), that correspond to the four sites on the tetrode. Response similarity of LFPs is generally higher than for the spiking activity (Fig. 4.4.3d). A systematic dependence of response similarity only on the spectral difference was not found. The lack of systematic patterns is probably due to each recording being taken from a different location within the ICC, and best frequencies, hence spectral distances differed.

Furthermore, other components such as modulation tuning and locally differing input influence the response at the specific position. The matrices show high correlation values mainly for multi-units with small differences in frequency tuning, only for nearby multi-units (Fig. 4.4.3). Thus, the correlations could be dependent on the spatial distance between the multi-units. It has been shown previously that neighboring (single) neurons display high correlation values of spike trains [45].

4.4.3 Dependence on spatial distance

In order to display the relation between correlation values and the spatial distance, the values were averaged for multi-units for each spatial distance of one recording set. The relations are shown for spiking activity and LFPs, exemplary, for one recording set for correlations across (Fig. 4.4.4) and within isofrequency laminae (Fig. 4.4.5). Correlations decrease with spatial distance and are almost zero for distances of 400 µm for the spiking responses; LFP correlations display a flatter decrease with distance, and correlation values of around 0.5 are still present for the maximum measured distance of 1600 µm (Fig. 4.4.4).
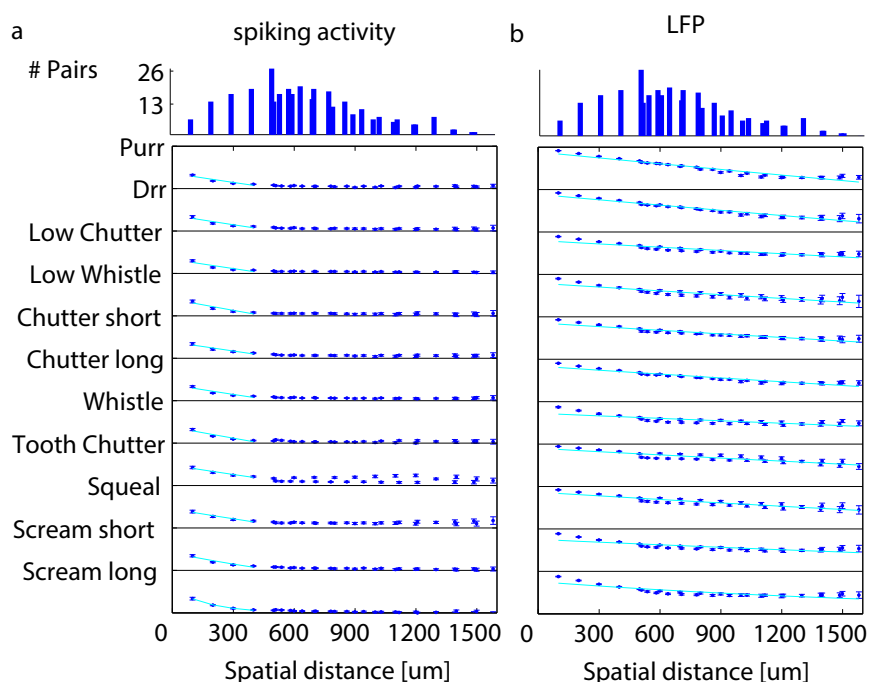
Figure 4.4.4: Correlation dependence on spatial distance across isofrequency laminae. Averaged correlation values for each distance from one recording set are displayed for all vocalizations, for a) spiking activity, and b) LFPs. The y-axis of the subplot for each vocalization shows a correlation range of 0-1. In this plot correlation values are also shown for spatial distances of 100 µm. Exponential fits ($f(x) = a \cdot e^{-bx}$, x is the spatial distance) for each distribution are shown (−); with overlaps of 61-100 %. On top, the number of multi-unit pairs for each spatial distance over which the average correlation was computed is displayed.
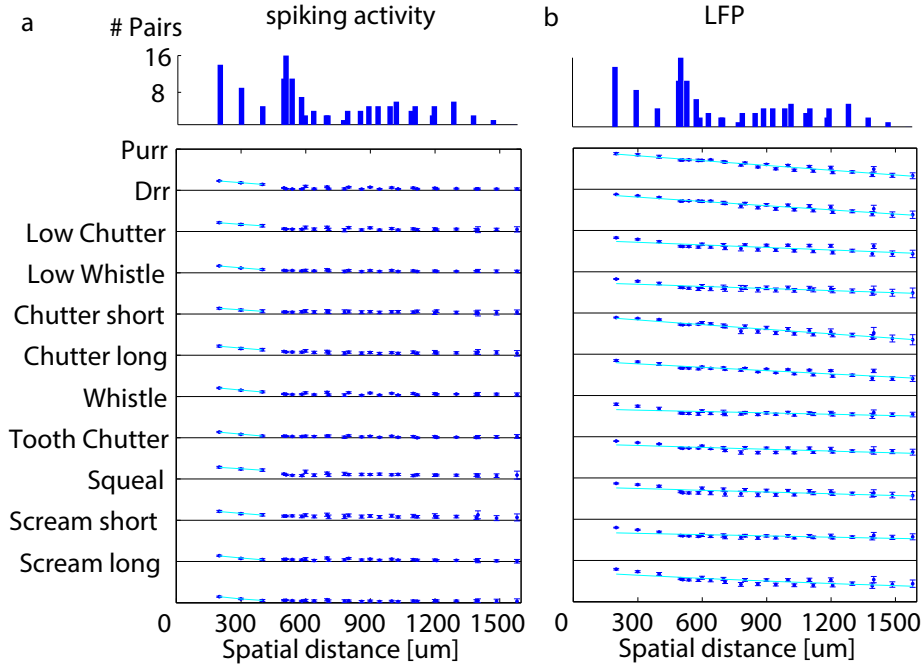
Figure 4.4.5: Correlation dependence on spatial distance within isofrequency laminae, similar to Fig. 4.4.4. Averaged correlations values for each distance from one recording set are displayed for all vocalizations, for a) spiking activity, and b) LFPs. The y-axis of the subplot for each vocalization shows a correlation range of 0-1. Exponential fits ($f(x) = a \cdot e^{-bx}$, x is the spatial distance) for each distribution are shown (-); with overlaps of 41-100 %. On top, the number of multi-unit pairs for each spatial distance over which the average correlation was computed is displayed.

For multi-units within one lamina from one recording set, the same is observed. The distributions can be approximated with an exponential function $f(x) = a \cdot e^{-bx}$, x is the spatial distance). For the eleven examples displayed in Fig. 4.4.4, 4.4.5), the range of values for a, b and the match of data and fit (χ) was found to be:

$a = 0.01 - 0.59$, $b = 0.57 - 0.0069$, $\chi = 84 - 100$ %, $\langle\chi\rangle = 91$ % (Fig. 4.4.4a),
$a = 0.76 - 1$, $b = 0.012 - 0.003$, $\chi = 61 - 94$ %, $\langle\chi\rangle = 80$ % (Fig. 4.4.4b),
$a = 0.01 - 0.39$, $b = 0.48 - 0.0041$, $\chi = 95 - 100$ %, $\langle\chi\rangle = 98$ % (Fig. 4.4.5a),
$a = 0.73 - 1$, $b = 0.042 - 0.002$, $\chi = 42 - 93$ %, $\langle\chi\rangle = 67$ % (Fig. 4.4.5b).

Dependencies on the spatial distance differ across vocalizations. The correlations displayed in Fig. 4.4.4, 4.4.5 are shown for distances up to 680 μm for all vocalizations (Fig. 4.4.6). A clear correlation decrease is observed from 200-500 μm in all four displayed cases. The dependence of the correlation on the spatial distance varies across recording sets, and thus across insertion locations within the ICC. Figure 4.4.7 shows an example for one vocalization ('scream long') of correlations from all recording sets.
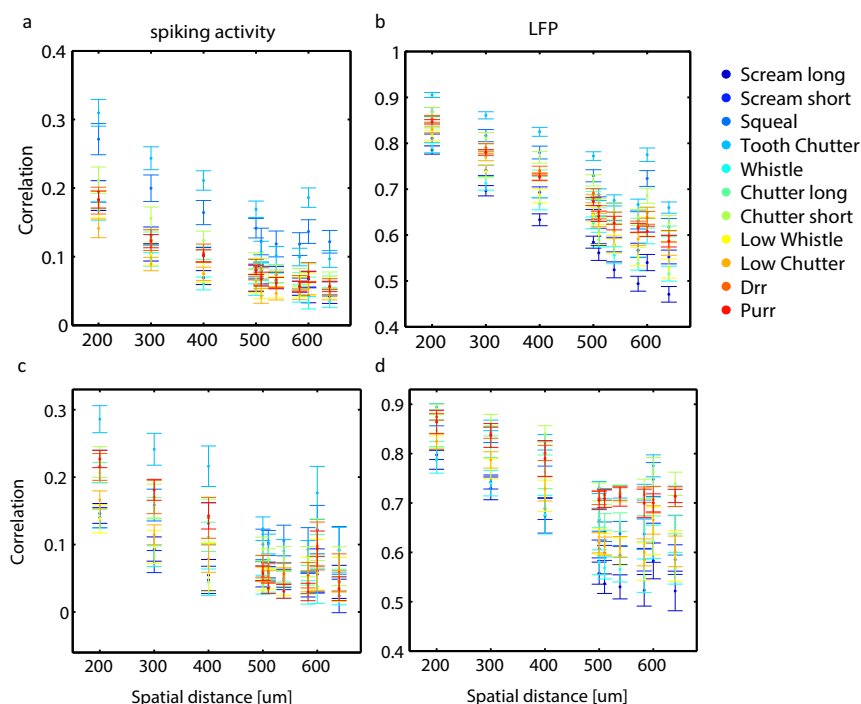
Figure 4.4.6: Comparison of spatial dependence of all vocalizations for small spatial distances. Correlations are shown for multi-units (from the recording sets displayed in Fig. 4.4.4, 4.4.5), across (a,b) and within laminae (c,d), for spiking activity (left) and LFPs (right).

Spiking response similarity decreases systematically with spatial distance, and is almost zero for multi-units with a spatial distance of more than 300 µm. Correlations of LFP activity are still present at large distances of 1600 µm, and show an almost linear dependence on the spatial distance.
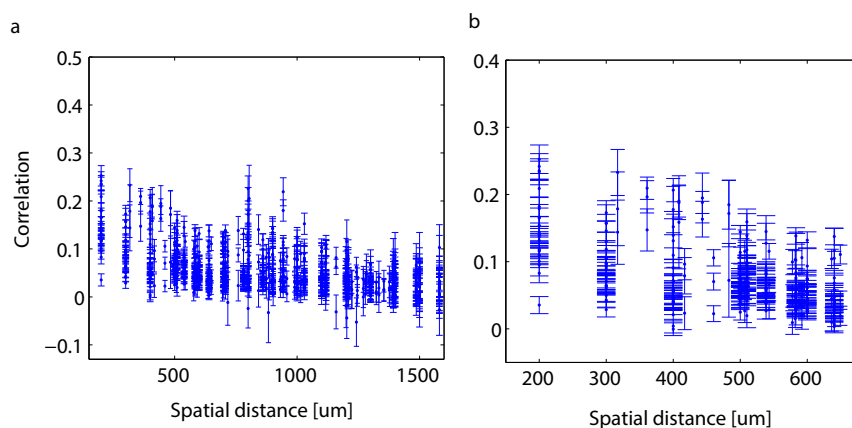
Figure 4.4.7: Correlation dependence on spatial distance for all multi-units. The correlations from all recording sets in response to one vocalization ('scream long') are displayed across laminae for spiking activity across the entire probed spatial range (a), and for smaller distances (b).

As it has been shown above, that spatial distances substantially influence response correlations between multi-units, the analysis in Sec. 4.4.1 is extended by evaluating the distribution of spatial distances of the recording sets that were used (Fig. 4.4.8). The distributions are the same for spiking and LFP activity. The distributions for the single recording set differ substantially, and on average the multi-units from the same laminae had smaller spatial distances than across isofrequency laminae (Fig. 4.4.8a). Thus, higher correlation values within the same lamina for the single recording set (Fig. 4.4.2a) are mainly attributed to smaller distances. Therefore the higher mean values for correlations within the same lamina can be explained by this biased distribution. The spatial distributions for the multi-units of all recordings, on the other hand, are similar across and within laminae; distances of 200 μm are more abundant for within lamina correlations, which might explain that mean correlation values are significantly higher within laminae.
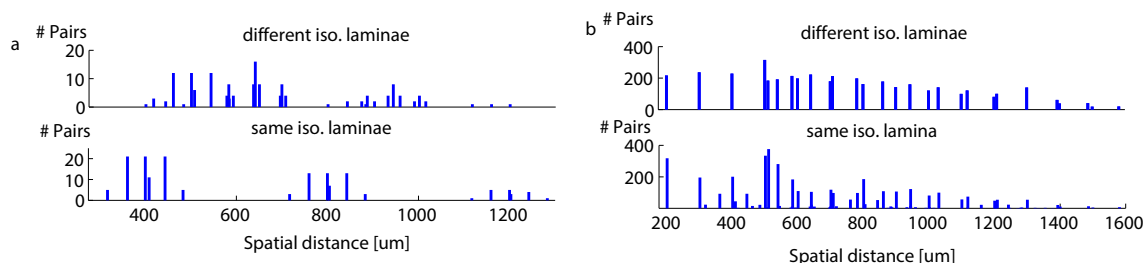
Figure 4.4.8: Distribution of spatial distances for the recording sets used in Fig. 4.4.2. The distribution of spatial distances of multi-units from different and same isofrequency lamina is given for a) the single recording set used in Fig. 4.4.2a, and b) for the multi-units from all recording sets used in Fig. 4.4.2b.

4.4.4 Joint dependence on spectral and spatial distance

In the previous sections it had been shown that a clear dependence of the response correlations on the spatial distance exists, but a dependence on spectral distance has not been found. To evaluate if co-dependencies on spatial and spectral distance exist, correlation values across all recording sets are displayed with respect to spatial and spectral distance of the multi-units. This was performed for each vocalization, and correlation values were normalized by the distribution of spatial and spectral distances of all analyzed multi-units (see Fig. 4.4.10). The joint correlation dependencies on spatial and spectral distance are shown for three vocalizations in Fig. 4.4.9. In general, highest correlation values are present for small spatial and spectral differences; exceptions of high correlation values for larger distances exist (e.g. 2 octaves and 475 μm in Fig. 4.4.9a). They reflect correlations of one or few multi-units, which were averaged out in the spatial dependence analysis of Sec. 4.4.3. No systematic joint dependence on the spatial and spectral distance is found for the multi-unit response correlations.
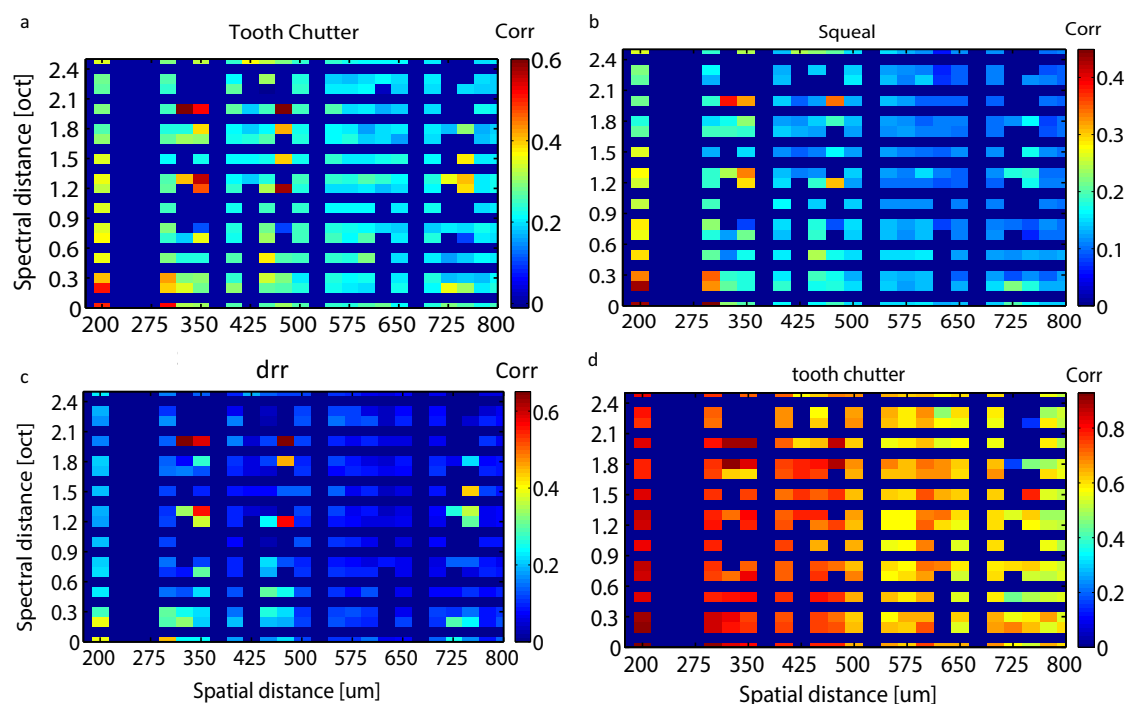
Figure 4.4.9: Response correlation in dependence of spatial and spectral distance. Correlation values, averaged across all recording sets, are displayed with respect to spatial distance and difference in best frequency of the multi-units, for the a), d) 'tooth chatter, b) 'squeal', and the c) 'drr'. (a,b,c) are response correlations for spiking activity and (d) of LFPs.

4.4.5 Response correlations and neural correlations

Comparison of correlations from simultaneous to those of non-simultaneous responses gives an indication of the amount of correlations due to neural interactions between the multi-units (Sec. 3.3). It has been shown previously that the sum of the stimulus-driven and neural correlations does not necessarily yield the measured response correlation [116]. Simultaneous spiking response correlations are larger than non-simultaneous correlations. However, within one recording set, the difference is not significant for all vocalizations, neither within nor across isofrequency laminae (Fig. 4.4.11). Stimulus-driven correlations are significantly larger than neural correlations. The difference between response, stimulus-driven and neural correlations is significant for all vocalizations for LFP activity, and all differences are significant when averaging across recordings (Fig. 4.4.12). Correlations display variations as visualized by the standard deviation. Differences between response, stimulus-driven and neural correlations are larger for LFP than for spiking activity. In conclusion, spiking activity interactions between multi-units in the ICC exist.
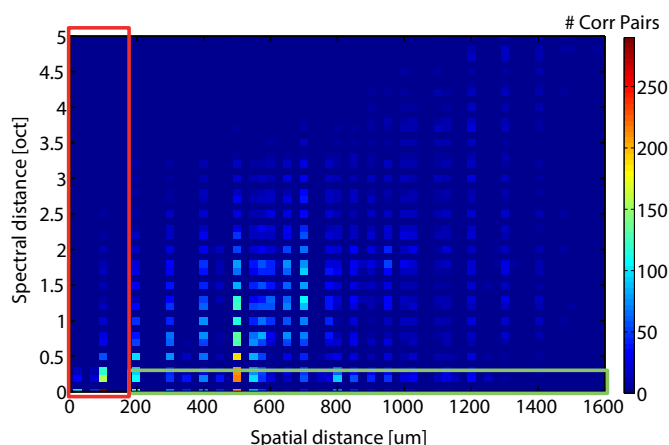
Figure 4.4.10: Distribution of spectral-spatial distances for all multi-unit pairs used in Fig. 4.4.7, 4.4.9, (n=15,692). The distribution is broad across the whole frequency and spatial distance range, with peaks for almost no spectral distance and spatial distance of 100 μm, and at spectral distances up to 0.5 oct and spatial distances of 500-100 μm. These are mainly due to responses from respectively the same tetrode of the double-tetrode recording and to responses from the same isofrequency lamina of double-shank recordings (Sec. 3.1.1). Marked regions were either excluded for all other correlation analysis (▫), assigned to the same isofrequency laminae (▫), and the rest to different isofrequency laminae.
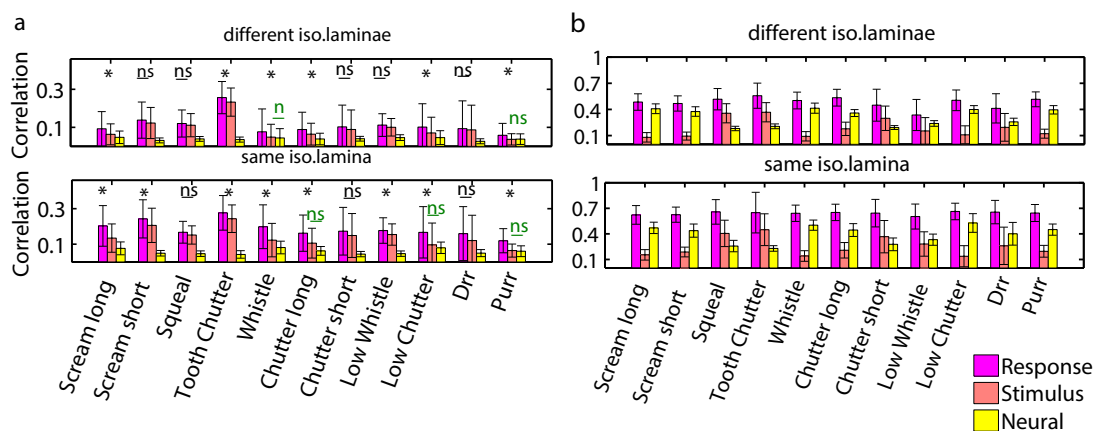
Figure 4.4.11: Response, stimulus-driven and neural correlations from one recording set. a) correlations for spiking activity, respectively across different and within the same isofrequency lamina (n=110 pairs); and b) correlations for LFP activity (n=122 pairs), all differences are significant. Significant differences between the response and stimulus-driven correlations are denoted by *, non-significance by **ns** ($p = 0.05$), and are displayed in green for the comparison between stimulus-driven and neural correlations.
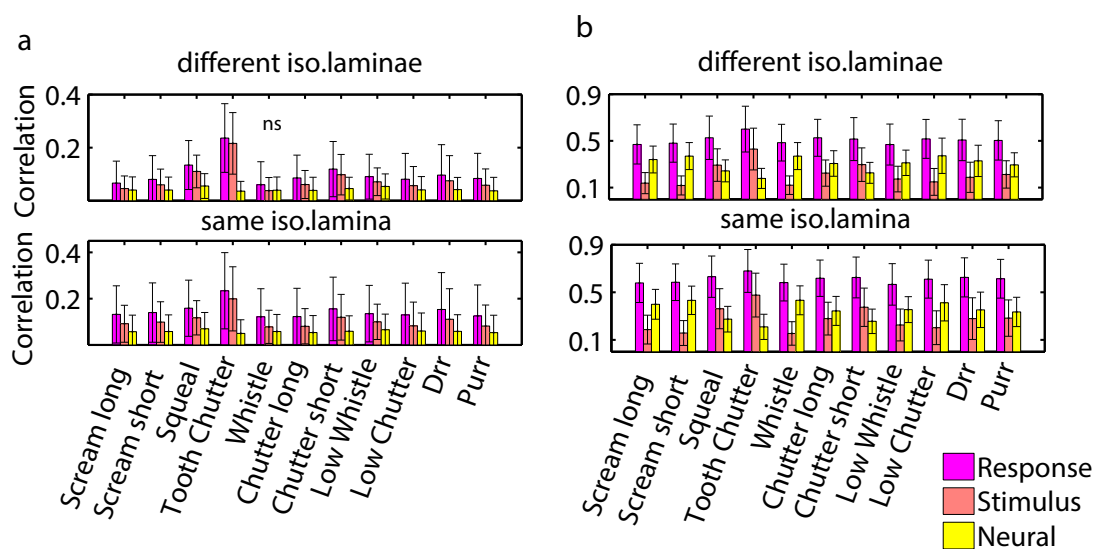
Figure 4.4.12: Response, stimulus-driven and neural correlations, averaged across all recordings. a) correlations for spiking activity, respectively across different and within the same isofrequency lamina (n=32,457 pairs); and b) correlations for LFP activity (n=12,699 pairs). Non-significant differences are denoted by **ns**, all other differences are significant ($p = 0.05$).

4.4.6 Effect of correlated variability on separation of spike-rates

In the previous section, it was shown that temporal response correlations of spiking activity, although differing for each recording set, are overall significantly higher for simultaneous than for non-simultaneous responses.

Interactions between neurons can lead to a covariation of their trial-to-trial variability of the temporal spiking responses, which might also manifest in a trial-variability of their average spike-rates. The averaged stimulus-elicited spike rate for each trial (n=20), for pairs of two multi-units are compared for simultaneous and non-simultaneous responses to investigate whether a decorrelation (shuffling) induces better separability of the spike-rates to different vocalizations. Examples for multi-unit pairs from the same or different isofrequency laminae, and for distant and relatively close-by multi-units are displayed in Fig. 4.4.13. Although shuffling changes the distribution of responses, separability of the 11 vocalizations does not change substantially in any of the cases. Correlated trial-variability due to interactions between neurons in the ICC does not alter separability of the multi-units' spike-rates to vocalizations, although at a single neuron level this might be different.
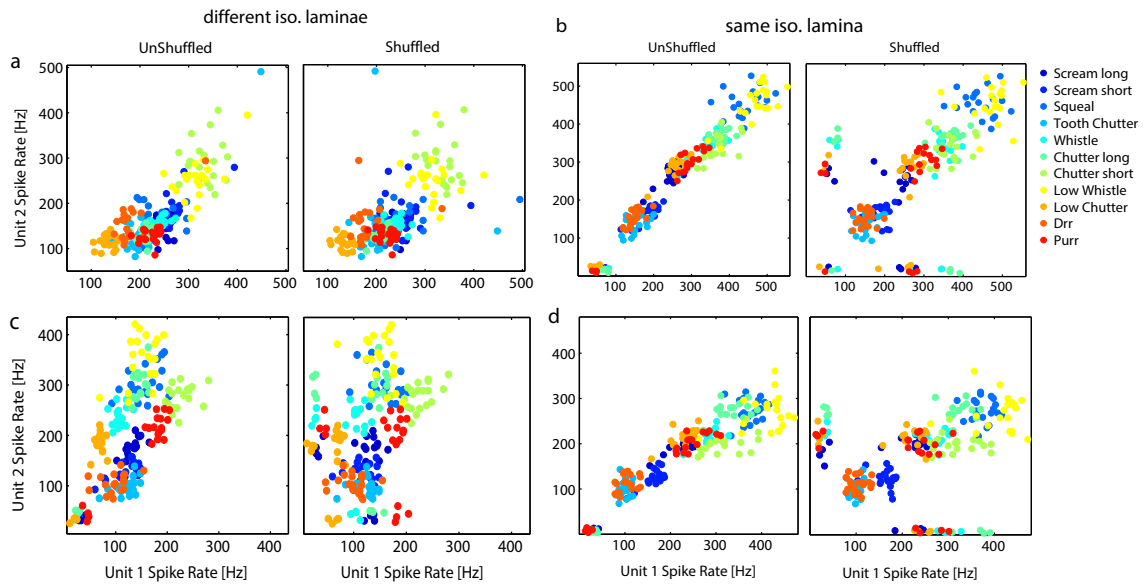
Figure 4.4.13: Separation of spike-rates with correlated trial-variability. The scatter plots of the average spike-rates in response to the 11 vocalizations for simultaneous (**Non-Shuffled**) and non-simultaneous (**Shuffled**) trials are compared for 4 multi-unit pairs, for units from different (a, c) or the same isofrequency laminae (b, d); and which are spatially relatively close (a, b) or far away (c, d). a) $BF_{1,2}$ = 5 kHz, 8 kHz; 360 μm; b) $BF_{1,2}$ = 2.25 kHz, 2.83 kHz; 400 μm; c) $BF_{1,2}$ = 8 kHz, 2.25 kHz; 1270 μm; d) $BF_{1,2}$ = 2 kHz, 2.25 kHz; 800 μm.

4.5 Discussion

In this chapter, neural multi-unit responses to a set of 11 spectrotemporally rich guinea pig vocalizations within and across isofrequency laminae of the ICC were analyzed. In some cases, these multi-unit responses can be approximated by the bandpass filtered waveform of the vocalization, filtered around the best frequency of the multi-unit.

Response correlations for multi-units within isofrequency laminae were found to be significantly higher than those across laminae. However, differences of the mean correlation are minor and the correlation variability is high. Therefore, it cannot be concluded that response similarity is substantially higher along either the best frequency or amplitude modulation frequency gradient, across the entire ICC. Also, no systematic patterns of correlation dependence on spectral distance were found across vocalizations or recording sets. Response similarity depends on the spatial distance between the two multi-units, and decreases exponentially with increasing distance. In general, for distances above 400 μm, no correlation of spiking responses is present, whereas LFP-responses show a more flat, also linear, decrease and display correlations for distances as high as 1600 μm.

Simultaneous response correlations are significantly higher, thus correlations due to interactions between the multi-units exist, but are minor. The separability of the 11 vocalization stimuli is not altered by the absence of neural correlations.

Discussion

In this analysis, only correlations for two multi-units that are separated by at least 200 μm were analyzed, in order to ensure comparison of different signals. However, neural responses from a radius of more than 200 μm might be picked up by the recording electrode. In order to exclude this possibility, and to investigate correlation dependence for smaller distances (<200 μm), single neuron recordings could be used. Furthermore, this would allow making inferences of connectivities. However, limitations of the shift predictor exist for single neurons such as the obscuring of neural interaction due to deterministic responses or temporal overlap of stimulus and neural response [46]. Probabilistic models based on pairwise interaction, describing the weights of interactions in a network could be established [117]. Furthermore, knowing the positions of the neurons would allow mapping out and testing the dependence of the response similarity on the specific location within the ICC. Anesthesia has been shown to affect neural responses [118], and by fluctuating or slowly changing anesthesia levels also changes in the brain state might affect the cross-correlation strength. Thus, response similarity might differ in awake animals, however, recordings in awake animals also bear difficulties and biases.

Here it was shown, that despite diverse inputs to the ICC from different ascending projections, terminating on different parts in the ICC [109], and existing response heterogeneities [21], representations of vocalizations gradually decrease in similarity across the best frequency and modulation gradients. Thus, despite the vocalizations displaying very diverse and inhomogeneous spectral contents, and the spectrotemporal gradients present in the ICC, the vocalizations exhibit gradual linear response similarity.

Multi-unit clusters with similar responses are spatially spread for distances <400 μm, however, differences might exist for single neurons and cannot be captured at this resolution level. These findings might give indications for applications in auditory midbrain prostheses, e.g. distances for spatial separation of stimulating electrodes. It was found that multi-unit clusters display largely independent encoding, correlations due to neural interactions are minor and they do not contribute to a better separability of the spike-rate responses to vocalizations.

In conclusion, the interplay of the total input to the neurons and their spectrotemporal response preferences, which both depend on the neurons' location within the ICC, result in a gradual change of multi-unit responses with distance. Thus, when comparing responses of neuronal groups to complex sound, such as vocalizations, across the ICC, it is important to take their spatial separation into account and not only their frequency tuning.

5 Neural discrimination of vocalizations

5.1 Introduction

In the previous Chapter 4, the representation and similarity of neural responses to individual vocalizations, across and within isofrequency laminae of the ICC, has been described. While this analysis yielded insight into the spread of individual vocalizations, and into the presence of signal correlations, it could not answer how well responses to different vocalizations can be distinguished. However, the discrimination of an individual vocalization against a whole set of vocalizations, thus the encoding of vocalizations is relevant in a behavioral context and its underlying mechanisms are open questions in auditory neuroscience.

In this chapter, the encoding of 11 species-specific vocalizations is investigated with the use of neural discrimination. In this method, the separability of recorded neural responses to different stimuli is tested, and assuming that good discrimination results from detailed encoding of the stimulus, inferences about neural encoding can be drawn. It is a powerful tool in analyzing encoding of vocalizations by groups of neurons, as opposed to receptive field analysis which requires further processing if natural sound is used [106]. Neural discrimination is applied to determine how well the spectrotemporally rich set of vocalizations is encoded by multi-unit clusters in the ICC. The optimal encoding of vocalizations across the best frequency gradient of the ICC is compared to the spectral content of the individual vocalizations. In a purely linear system, spectral content and optimal encoding could match perfectly, and a large overlap of stimulus frequency and strongest neural response at this best frequency were found [4]. However, the ICC is highly nonlinear [14] and might employ (on the level of single neurons as well as multi-units) nonlinear mechanisms to encode behaviorally relevant natural sound. Thus, the encoding of individual vocalizations along the frequency gradient is investigated. Furthermore, the necessary order of multi-units to perfectly discriminate the set of vocalizations is addressed. Responses from several multi-units were combined to investigate whether groups of multi-units represent the whole set of vocalizations more faithfully than only one unit. In the previous chapter, it was shown that correlations between different multi-units are mainly stimulus correlations and not due to neuronal interactions (Sec. 4.4.5). Even though correlations that are due to interactions are minor, they could still have a modulatory effect (facilitation or decrease) on the encoding of vocalizations. It was tested if temporal response correlations between the multi-unit clusters facilitate the unique neural representation of the vocalizations.

In this chapter, at first, the neural discrimination analysis which was applied to the multi-unit recordings is described in detail (Sec. 5.2). Then, the optimal discrimination procedure and response duration are obtained. Subsequently, the encoding of vocalizations by multi-units in the ICC with respect to the above mentioned questions is investigated.

5.2 Methods

For the analysis of this chapter, the neural guinea pig recordings which are described in Sec. 3.1.1 are used. These recordings are pre-processed as described in Sec. 4.2.1. In this section, the neural discrimination procedure which was applied to individual and combined multi-unit responses is described in detail.

5.2.1 Spectral content of vocalizations

The waveform, spectrogram and power spectrum of the vocalizations used in this study are shown in Fig. 3.1.1 on page 25. The waveform has a simple periodic shape for some vocalizations and is quite complex for others. Some vocalizations have a frequency content of up to 30 kHz ('tooth chatter', 'scream long', 'scream short'). The main spectral content of the vocalization is obtained by computing the power spectrum and integrating the power for frequency bands f_N, which are centered around the recorded multi-units' best frequencies, $f_N = \{0 - 0.25\}, \{0.25 - 0.55\}, \ldots, \{38 - 45.5\}$ kHz. The power for each vocalization is normalized and its logarithm is displayed (Fig. 5.4.4).

5.2.2 Discrimination of single responses

The procedure to test separability between responses to different stimuli from one multi-unit is exemplified in Fig. 6.2.3a for two vocalization stimuli, and consists of data pre-processing, feature extraction and classification. This classification procedure is performed independently for each multi-unit with its specific frequency tuning. Recordings from one multi-unit to the different stimuli are preprocessed as described in Sec. 4.2.1. To separate responses to different stimuli, features obtained from the time courses were employed. The tested features were: mean response rates (for LFP) and firing rate (for spiking activity) across 300 ms; a 6-dimensional feature vector, containing respective mean rates over six response segments of 50 ms which make up the 300 ms, yielding a finer resolution of the response rate over the previous mean rate, and finally, the five most prominent frequencies of LFP-recordings obtained from the power spectrum. The first two metrics have been applied previously for successful discrimination between responses to vocalizations [29, 26]. Features were obtained for each trial ($n_{\mathrm{trial}} = 20$). All 20 trials were divided into test and training sets and entered to a classifier. A 10-fold cross-validation with 10 % test data was performed for each classification, and yielded an estimate of the error.
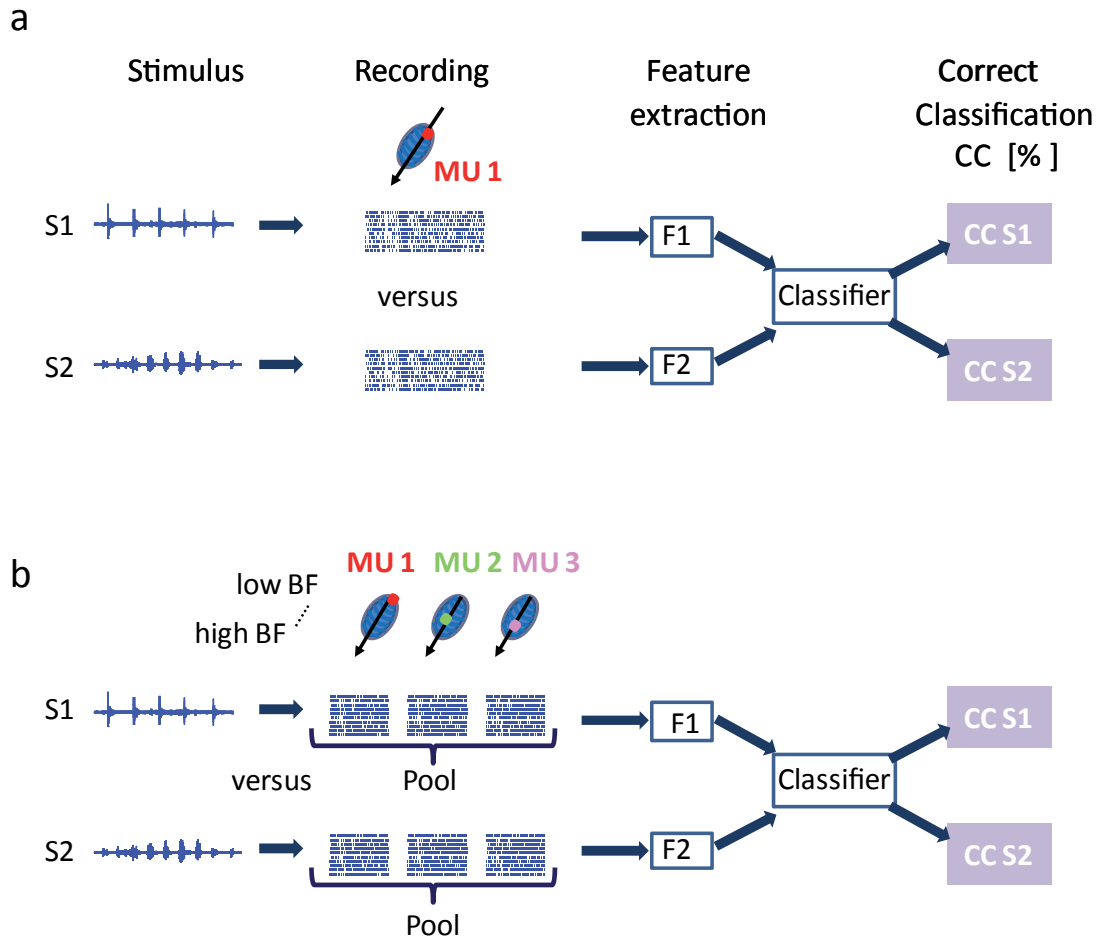
Figure 5.2.1: Schematic of neural discrimination. The classification procedure between responses elicited by two different vocalization stimuli, S1 and S2 is exemplified. a) Recordings from one multi-unit (MU 1) to stimuli S1 and S2 are preprocessed, features (F1 and F2) are extracted and entered to a classifier which yields the percentage of correctly assigned responses for stimulus 1 and 2 (correct classification CC). b) Responses from three multi-units (MU1, MU2, MU3) along the tonotopic gradient are combined. From these combined responses, features (F1 and F2) are extracted and given to the classifier to obtain the percentage of correct assignments of the responses.

Tested classifiers included linear discriminant analysis (LDA, which fits a normal density to each class and estimates the combined covariance matrix), a naive Bayesian classifier (which estimates a diagonal covariance matrix) and a nearest neighbor classifier, using Euclidean distance [105]. As a further approach, correlation of spike trains to be classified (test trials) with the labeled spike trains (training trials), is employed. Thus correlation values serve as features. Correlation values are computed as the cross-correlation (see Eq. 3.3.1 on page 35) at zero lag between the responses ($\tau_{\max} = 0$). It was computed at zero lag, because even though response latencies vary across multi-units with different spectrotemporal preferences [115], here, responses from the same multi-unit, thus with the same response latency, were compared. The correlation approach allows comparing temporal responses at a resolution of 3 ms, which is the width α of the filter function (see Sec. 4.2.1).

The test trial is correlated with all training trials of each class. The class which, on average, yields the highest correlation value is assigned to the test trial. The classification of all test trials for each class yields a confusion matrix (see Sec. 3.4). The mean across the diagonal of the confusion matrix, averaged all cross-validation iterations, yields the total correct classification between all vocalizations ($N_{\text{voc}} = 11$).

$$CC = \frac{1}{N_{\text{voc}}} \sum_{k=1}^{N_{\text{voc}}} CC(k, k) \tag{5.2.1}$$

The classification yields the percentage of correctly assigned responses to each stimulus (Correct Classification, CC [%]). The classification error is the standard deviation obtained from all cross-validation steps.

The confusion matrix and total correct classification is computed for each multi-unit with its specific frequency tuning. This allowed investigating whether the ability to discriminate between individual vocalizations (or between groups of vocalizations), differs for multi-units with their specific frequency tuning and best frequency (Sec. 5.3.3, 5.4.1).

5.2.3 Combined responses

In order to assess whether combining responses from several multi-units improves discrimination between different stimuli, responses that were recorded simultaneously from different multi-units to the same stimulus were pooled. Figure 5.2.1b illustrates the procedure for combining responses in an example for three multi-units with different best frequencies. Pooling was performed for multi-units from either along the best frequency gradient (across several isofrequency laminae) or from the same isofrequency lamina (see Fig. 5.2.2). Separability was tested for these combined responses to different vocalizations. Pooling methods have been applied earlier and are discussed in more detail in [27]. Schneider and Woolley added single neuron spike trains in zebra finches in response to songs. In this work, spike trains obtained from simultaneous recordings of different multi-units were either concatenated (Fig. 5.2.3a) or added (Fig. 5.2.3b).

Across isofrequency laminae   **Within one isofrequency lamina**

frequency/tonotopic gradient   modulation gradient

Unit 1   Unit 2   Unit 3   Unit 1   Unit 2   Unit 3

different best frequencies   same best frequencies
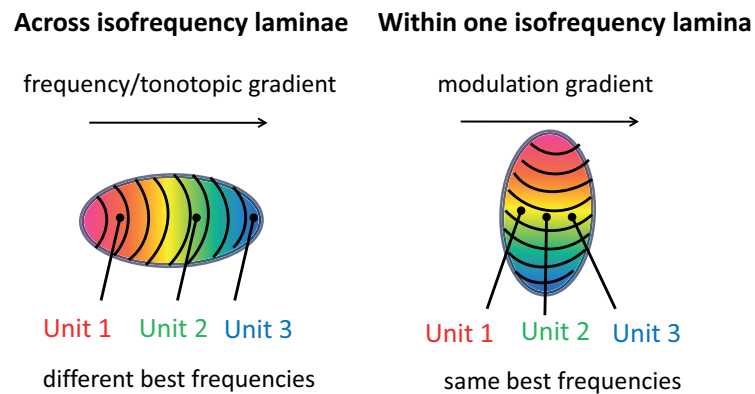
Figure 5.2.2: Combination of neural responses from across and within isofrequency lamina is illustrated for a schematic ICC. Neurons across isofrequency laminae (along the tonotopic gradient) have different best frequencies. Neurons within an isofrequency lamina have similar best frequencies.

Concatenation preserves the temporal spiking response separately for each multi-unit (Fig. 5.2.3). Concatenation also preserves the frequency information because the order in which responses from multi-units are concatenated is the same for all combined responses that are compared. Thus, during classification, spike trains from the same multi-units with their specific frequency tuning, but in response to different vocalizations, are compared. For recordings with a tetrode array, several multi-units lie within the same isofrequency lamina. They have similar best frequencies, but might have preferences for different amplitude modulations depending on their spatial distance within the lamina [6]. Therefore, in this case, concatenation preserves information of amplitude modulation (AM) preferences. Addition of spike trains (Fig. 5.2.3) maintains temporal information. However, as responses from differently frequency-tuned multi-units are added and not distinguishable anymore, the spectral information is lost.

Concatenation   Addition

Unit 1   Unit 2   Unit 3

S1   +   **S1**
100 ms   300 ms   +   100 ms
100 ms

Unit 1   Unit 2   Unit 3

S2   +   **S2**
100 ms   300 ms   +   100 ms
100 ms

Figure 5.2.3: Combining multi-unit responses. Example for one trial and two vocalization stimuli (S1 and S2). Concatenation: Temporal responses of length 100 ms from $n = 3$ different units are concatenated, yielding a combined response of length n×100 ms. Addition: Temporal responses of length 100 ms from $n = 3$ units are added up, yielding a combined response of length 100 ms.

Multi-units were combined in a third manner to preserve only spectral information. For this purpose, mean firing rates for each multi-unit and trial are computed from spike trains of 100 ms duration. The spike-rates are concatenated (Fig. 5.2.4) in the same way as the spike trains had been concatenated (Fig. 5.2.3). The information from different units with different spectral tuning, thus the spectral information, is preserved, but not the temporal information. The concatenation of spike-rates yields a feature vector whose length corresponds to the number of combined multi-units.
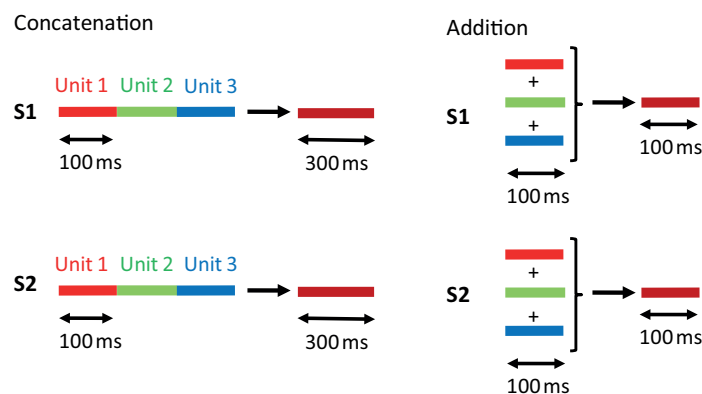
The test trial is assigned to the class for which its feature vector yields the minimum squared difference in spike-rate of test and training trial. Discrimination is computed for a successively increasing number of combined multi-units (Sec. 5.4.2). The additional multi-unit to be combined was either the spatially nearest neighbor, or, alternatively was chosen at random. Whereas in the first case information from similarly frequency-tuned units is successively combined gradually along the BF-gradient (or similar preferences for amplitude modulations gradually along an isofrequency lamina), in the second case, information from multi-units with very different frequency tuning (or AM preferences) is combined. Performance values are averaged over two iterations for gradual pooling and over three iterations for random pooling.
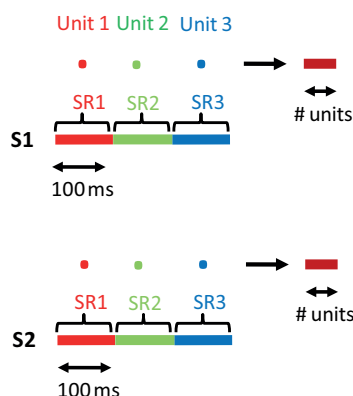
Figure 5.2.4: Concatenation of spike-rates. From the temporal responses (length 100 ms) the spike-rates are computed for each of the $n = 3$ different multi-units, in response to stimulus S1 and S2. The spike-rates are concatenated to yield a 1×n vector.

5.2.4 Canceling neural correlations

Multi-units may interact in order to more efficiently encode complex sounds. This interaction can lead to temporal correlations of their responses. To test if these neural correlations have an effect on encoding vocalizations by groups of multi-units, the discrimination was compared when they are present and when they are absent.

Hence, in order to cancel the neural correlations, simultaneously recorded trials of different multi-units are randomly shuffled (see Sec. 3.3.2) before combining them. The effect of temporal response correlations on the neural discrimination of combined multi-unit responses was investigated. To this end, the neural discrimination of simultaneous and non-simultaneously combined responses is compared. In a further comparison, the temporal correlation values are explicitly included into the feature vector. Thus, correlations between the multi-unit trials which are pooled are computed (*CorrRes*). The spiking responses from two multi-units are cross-correlated (see Eq. 3.3.1) and the highest value within a maximum possible delay of 10 ms between the responses is selected. This delay is within the range of maximum response latencies in the ICC [115]. For combining the responses of (n_{comb}) multi-units, a correlation matrix which displays the correlation between each multi-unit trial is computed. The average correlation value for the response of one multi-unit with all other responses of that combined trial is computed, and yields a $1 \times n_{\mathrm{comb}}$ vector which is concatenated with the feature vector that contains the classification values based on the time courses (*Corr*). Test trials were assigned to the class for which the training and test trials had, on average, maximum correlation of the time courses, *Corr*, and average minimum squared differences of response correlations *CorrRes*. Neural discrimination is compared for simultaneous responses and non-simultaneous responses when explicitly including these correlations (Sec. 5.4.3).

Averaging across multi-units

One main goal of this work is to study neural discrimination of individual vocalizations with their specific spectrotemporal properties across multi-units with different frequency-tuning. However, the average neural discrimination between all behaviorally relevant vocalizations –of which a representative set was used for this work – is of interest, since this is the task the auditory system of the guinea pigs solves perfectly, allowing them to react to these communication calls [88]. Hence, it is possible that preferences for specific vocalizations might be averaged out.

In this work, discrimination is compared for individual or combined multi-units of one electrode shank and recording (n=72), Fig. 5.4.1, 5.4.5, 5.4.8. Findings are in general true for all such sets. Averages across these sets are also given (Fig. 5.4.2, 5.4.3 5.4.6). When averaging classification across multi-units, electrode positions or animals, the classification error was computed via error propagation. The range of best frequencies of the multi-units recorded from varied for each animal. Therefore the number of multi-units for which averages were taken for one best frequency (Fig. 5.3.3c, 5.4.2, 5.4.2) varied across animals. Differences in neural discrimination based on different amplitude modulation preferences would be averaged out when taking the mean across multi-units with the same best frequency. Validity of averaging across different animals and positions was tested (Fig. 5.3.3).

5.3 Data pre-analysis

First of all, a classification procedure and spike train length were selected which yielded the best classification performance for these multi-unit responses. Additionally, classification performances were compared across animals, electrode insertion positions and for different intensities at which the vocalizations were presented. In the first part of the analysis a sufficient classification procedure, optimal response length and temporal location within the recording are selected, Fig. 5.3.1, 5.3.2. For this purpose, a recording set with an overall high classification performance which covers a wide range of best frequencies was used. The double-shank array was chosen as it covers a much wider range of best frequencies than the tetrode array which usually just covers two to three isofrequency laminae.

5.3.1 Comparing performance for different classification procedures

In order to determine which data type (EPSP-spiking response or local field potential), feature of the response and classifier yield the maximum correct classification for neural discrimination between all 11 vocalizations, the performance of (data type, feature, classifier)-tuples on responses to 300 ms segments of the vocalizations was compared. The focus of this study was the separability between multi-unit responses from the central inferior colliculus, the output, thus the spiking activity and not the synaptic input, the local field potential [110]. Hence, spiking activity was used for the analyses. Nevertheless, classification performance for local field potential responses are also compared. Figure 5.3.1 displays the classification performance for the 3-tuples. The performances were averaged across 10 multi-units of one recording (Fig. 5.3.1 top), and are also shown individually for these multi-units with differing best frequencies. These span a range between 0.5-18 kHz (Fig. 5.3.1 bottom). Using correlation to discriminate 300 ms long EPSP-spike trains yielded the highest correct classification (85 %). But differences to some of the other 3-tuples were minor. The naive Bayesian classifier and linear discriminant analysis on a 6-dimensional feature vector with firing rates across 50 ms also yielded high correct classification values (75 %). However, a specific focus of this study is to compare the finer temporal structure of the multi-unit responses (<10 ms) which has been shown to be crucial for neural discrimination [19]. This is achieved by the correlation approach which takes a 3 ms-resolved temporal structure into account. Classification performances for discrimination between 11 vocalizations varied for multi-units with different best frequencies. However, maximum correct classification values did not systematically correlate with the multi-unit's best frequency. Using correlation for discrimination of spike trains holds the advantage of comparing the temporal structure of responses and achieved the highest correct classification. Therefore, this method was used subsequently.
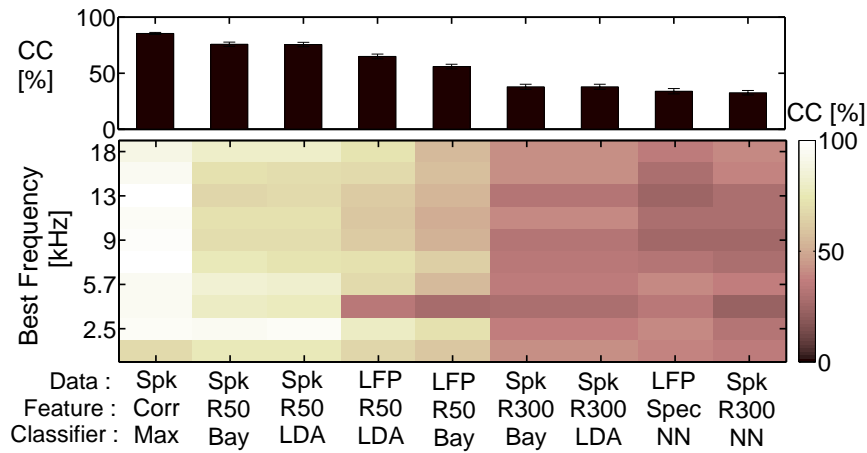
Figure 5.3.1: Comparison of classification performances for 9 (data type, feature, classifier)-combinations, for classification between all 11 vocalizations on 300 ms of recording. Data types: EPSP-spike trains (**Spk**), and local field potentials (**LFP**); features are: correlation values (**Corr**), average (firing) rate of 300 ms (**R300**), 6 average (firing) rates of each 50 ms (**R50**); 5 most important response frequencies (**Spec**). Classification was performed by choosing the maximum value (**Max**), naive Bayes classifier (**Bay**), linear discriminant analysis (**LDA**) or nearest neighbor classification using Euclidean distance (**NN**). **Top**: Correct classification averaged across 10 multi-units are displayed for each 3-tuple. **Bottom**: Classification performance displayed individually for each multi-unit, which span a best frequency range of 0.5-18 kHz.

5.3.2 Neural discrimination of different vocalization segments

In order to test from which starting point within the 1 second recording and which length of the spike trains should be employed for the analyses, correct classification was compared between all 11 vocalizations for 1) consecutive spike trains of 10 ms before and after stimulus onset; 2) spike trains of (5, 10-50, 100-900) ms from the same starting point; and 3) for 100 ms long consecutive spike trains covering the whole recording time. The analysis was performed on multi-units from one recording set covering a BF-range of 0.63-20.1 kHz, and the average values of all 16 multi-units are used for comparison.

Figure 5.3.2 displays the percentage of correct classification of 10 ms-long spike trains. Before stimulus onset, classification performance is at chance level ($1/11 \approx 9.1$ %) and increases gradually until about 50 ms after stimulus onset up to 25±5 % without further increase. Thus, only after stimulus onset discriminative information exists in the responses. The increase after stimulus onset is gradual and not sharp because the actual starting times of the individual vocalizations differ (Fig. 3.1.1 a and b). After about 70 ms recording time, classification does not increase further, hence no additional discriminative information is present in the responses; all vocalizations have begun at this time (Fig. 3.1.1). Performance values are relatively low due to the short spike train length of only 10 ms.

Correct classification gradually increased for increasing spike train length, starting at 70 ms of the recording.

Performances were respectively 19 ± 2 %, 55 ± 4 %, 83 ± 4 % and 94 ± 3 %, for lengths of 5, 100, 300 and 900 ms. This increase of correct classification with spike train length was shown previously for recordings in zebra finches and grasshoppers [27, 48, 29]. Machens and colleagues found spike train lengths above 400 ms to yield near to perfect discrimination [29].

Classification performance for consecutive 100 ms long spike trains across the recording of 1000 ms did not vary significantly, ranging from 53 ± 3 % to 58 ± 3 %. Hence, the exact starting point of the spike trains within the 1 second recording is immaterial. Thus, vocalizations do not contain a single specific feature within a temporal window, which distinguishes them from the other vocalizations, but rather provide information throughout their duration, allowing for instantaneous discrimination of the vocalizations.

For the present study, the same starting point for all analyses was employed consistently. To analyze classification performance across animals and positions, 300 ms and not 900 ms long spike trains were used, since only the first one included the responses to all ongoing 11 vocalizations (Fig. 5.3.3). In order to increase complexity of discrimination and clearly see differences in performance, when analyzing individual vocalizations in dependence of the multi-unit's best frequency, and when combining responses of several multi-units, 100 ms spike trains were used (Fig. 5.4.1, 5.4.2, 5.4.3, 5.4.5, 5.4.7, 5.4.8).
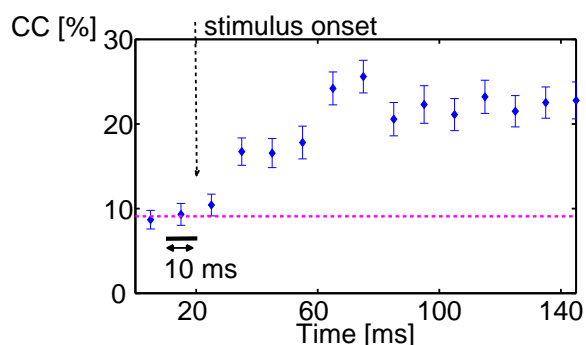
Figure 5.3.2: Onset of discrimination. Percentage correct classification between all 11 vocalizations of consecutive spike trains of 10 ms duration. Low correct classification values are due to the relatively short spike train length. Before stimulus onset, classification was at chance level (9.1 %, pink dashed line). After stimulus onset classification performance gradually increased to maximum 25 % correct classification. Discriminative information is present only after stimulus onset.

5.3.3 Variability across animals and electrode insertion locations

In order to test the validity of averaging classification values across all recordings, correct classification was compared between 11 vocalizations of 300 ms long spike trains, across average classification values from each animal.

Classification values were also compared for different insertion locations in one animal (32 multi-units), and it was investigated how they vary with the intensity at which the vocalizations were presented, 30-70 dB SPL, in steps of 10 dB SPL. Correct classification varies significantly across animals (Fig. 5.3.3a), the best discrimination being 85 %, for animal 1 (Exp. 1). Performances for some animals are very similar (e.g. Exp. 1 and Exp. 2). Classification values vary across electrode insertion locations (Fig. 5.3.3b), thus across the ICC with different preferences for frequency and amplitude modulation, but are less pronounced than differences due to stimulus intensity. The overall classification performance increases significantly with stimulus intensity, as spike rates increase for higher intensities. Figure 5.3.3 displays the average correct classification for each animal for each best frequency. Responses from some animals are overall better discriminated than others (e.g. # 2). Correct classification does not depend on the multi-unit's best frequency and varies across animals.
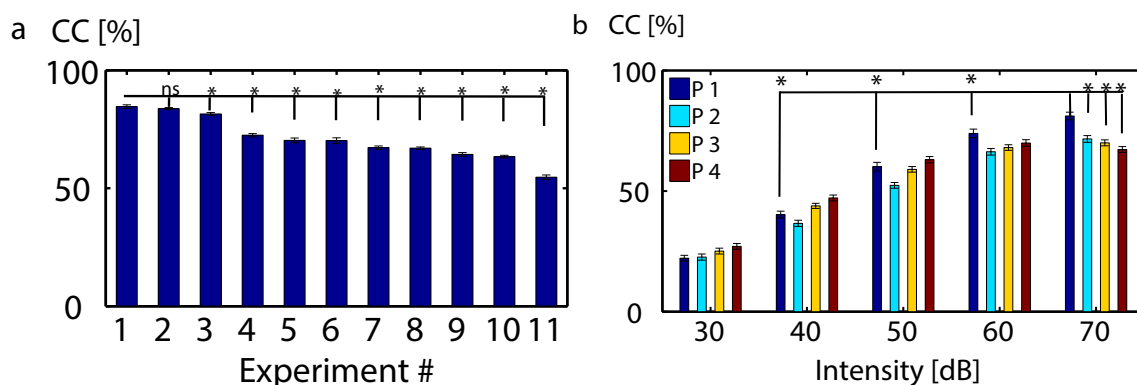
Figure 5.3.3: Discrimination values vary across animals, electrode insertion locations and increase with stimulus intensity. a) average correct classification across all recordings from one experiment (animal). b) average classification of responses of all 32 multi-units for each of the four electrode insertion positions (P1-P4), for different intensities at which the vocalizations were presented. (***** denotes significant, **ns** denotes non-significant differences; mean, Wilcoxon-Mann-Whitney test $p = 0.05$).

The total classification does not depend on the best frequency (see Fig. 5.3.4). This has been shown previously for songs in single neurons of zebra finches [27]. Thus, in order to avoid smearing out differences in discrimination, in the subsequent analyses the classification values were not averaged across animals and electrode positions but taken from one recording of one animal. Recordings for stimulus intensities of 70 dB SPL were employed for the analyses, as they yield the highest classification performance.
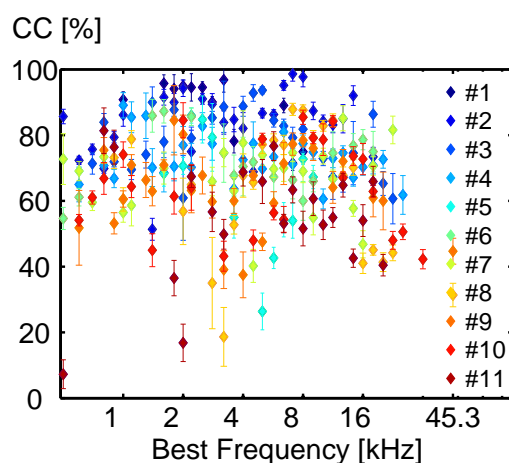
Figure 5.3.4: Total neural discrimination across best frequencies. Average correct classification value of each experiment ($N = 11$) for each best frequency. Higher classification values do not correlate with best frequencies but vary for different animals.

5.4 Results

Neural discrimination between a spectrotemporally rich set of 11 species-specific vocalizations was analyzed for 1152 multi-units across and within isofrequency laminae of the central inferior colliculus of 11 guinea pigs. It was tested whether encoding of individual vocalizations varies across the best frequency gradient of the ICC. Combined spike train responses from several multi-units were employed to investigate whether groups of multi-units result in a better discrimination than individual multi-units, and whether neural correlations between the multi-units contribute to a yet better discrimination.

### 5.4.1 Dependence on best frequency

Encoding averaged across several vocalizations does not vary systematically with best frequency [27]. The total neural discrimination between responses to all 11 vocalizations has been shown not to vary systematically with the multi-unit's best frequency (Fig. 5.3.4). However, depending on their spectral content, neural discrimination of individual vocalizations might differ with the neuron's preferred frequency.

To address this question, discrimination of 100 ms-long spike train responses to four vocalizations was compared for multi-units with different best frequencies. The four vocalizations were divided into two groups; vocalizations of the same group have similar spectral content. A relatively short spike train length of 100 ms was chosen in order to raise the separability threshold and enable the detection of subtle differences in discrimination. Figure 5.4.1 shows correct classification between responses to the 'drr', 'low chutter', 'whistle' and 'long scream'. These have a main spectral content either below 1.5 kHz ('drr', 'low chutter', see Fig. 3.1.1c,d), thus a low frequency content, or above 1.5-2 kHz ('whistle', 'long scream', see Fig. 3.1.1g,i), thus a broad spectrum of frequencies.

Responses were recorded with a tetrode-array (Fig. 5.4.1a) and a double-shank array (Fig. 5.4.1b), from multi-units with different best frequencies. The 'drr' and 'low chutter' are nearly perfectly encoded for BFs up to 0.7 kHz, whereas the 'whistle' and 'long scream' are poorly discriminated in this range. For BFs above 4.5 kHz, the 'whistle' and 'long scream' are perfectly discriminated but the 'drr' and 'low chutter' are frequently confused. Responses of middle-BF multi-units to the two vocalizations which contain a broad spectrum of frequencies, including middle and high frequencies, are perfectly discriminated, whereas the vocalizations with low frequency content are poorly discriminated (Fig. 5.4.1a.) Vocalizations with low frequency content are nearly perfectly discriminated for low-BF multi-unit responses. A clear transition towards a perfect discrimination of vocalizations containing middle and high frequencies is visible as the multi-units' BFs increase, Fig. 5.4.1b. The transition from nearly perfect discrimination of low to middle-high frequency vocalizations is present between 0.7-4.5 kHz. This clear preference for discriminating certain groups of vocalizations over others was also observed for different combinations, e.g. 'drr', 'low chutter', 'squeal', and 'low whistle'.

The performances of these four individual vocalizations for classification between all 11 vocalizations, averaged across all 1152 multi-units are displayed in Fig. 5.4.2 for BFs between 0.5 kHz and 25.4 kHz.

Classification values of responses from within one lamina or across laminae showed a similar dependence on the best-frequency and were averaged to give only one value per best frequency. A clear trend of higher correct classification of vocalizations containing low frequencies from responses of low-BF multi-units, and of vocalizations with broad spectral distribution from responses of middle-and high BF multi-units is apparent.

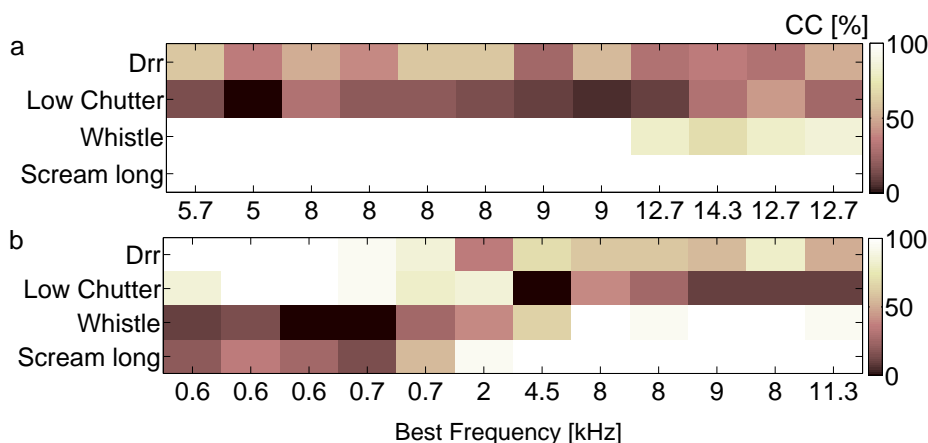**Figure 5.4.1:** Tradeoff of optimal neural discrimination with best frequency. Neural discrimination between vocalizations with a low ('drr', 'low chutter') and middle-high ('whistle', 'scream long') frequency content for multi-units from a recording with a) several multi-units from the same lamina BF = 5.7-9 kHz; and b) from a recording along the BF-gradient with a broader range of best frequencies BF = 0.6-11.3 kHz.

Displayed are also relative frequency contents for these four vocalizations which are displayed separately for all 11 vocalizations in Fig. 5.4.4. These show the same trend as the classification performances, and are for some frequencies very similar to the classification performances (e.g. 'long scream'). Deviations exist, which could be due to preferred encoding of fast amplitude modulations by low BF neurons [92], and hence a better discrimination of vocalizations containing fast amplitude modulations than would be predicted solely by their frequency content. The match between the averaged correct classification and relative frequency content for all 10 vocalizations varied between 74-80 %, using the Euclidean distance. The 'tooth chatter' displayed only a match of 53 %. Figure 5.4.2 shows correct classification between 11 vocalizations averaged across all 1152 multi-units for further vocalizations ('chutter long', 'low whistle', 'squeal', 'scream short'). Trends of preferred encoding for the individual vocalizations along the BF gradient exist. The 'tooth chatter' and 'short chutter' do not show pronounced preferred encoding by certain best frequency multi-units, the 'purr' displays bimodally distributed preferred encoding. Thus, with the exception of two vocalizations, best discrimination of individual vocalizations is spatially distributed across the ICC. Accurate encoding of a vocalization in the spiking response allows its good discrimination against other vocalizations. Perfect discrimination between responses to similar vocalizations indicates that these vocalizations are encoded in detail and hence can be separated based on minor spectrotemporal differences. If, on the other hand, vocalizations are coarsely encoded, responses to similar vocalizations are frequently confused.

The 'tooth chatter' and 'long chutter' display important frequency content across the whole best frequency range (see Fig. 3.1.1), thus they are not preferentially encoded by a certain range of best frequency multi-units. Vocalizations with low frequency content are preferentially encoded by low-BF multi-units and vocalizations with middle-high frequency content are preferentially encoded by middle-high BF multi-units.
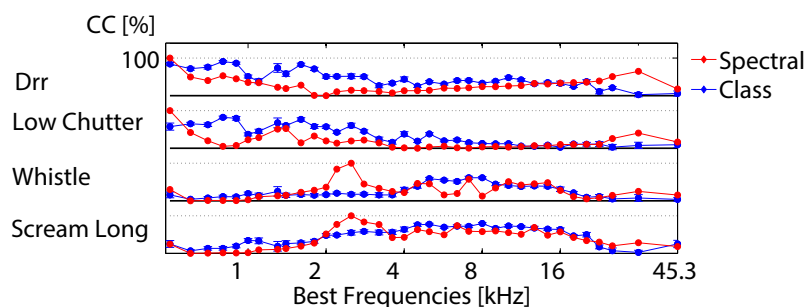
Figure 5.4.2: Correct classification and spectral content for the four vocalizations shown in Fig. 5.4.1, across the best frequency gradient. The correct classification (•) of individual vocalizations, for discriminating between all 11 vocalizations (averaged across 1152 multi-units) is displayed for the 'whistle', 'scream long', 'drr' and 'low chutter'. The relative frequency content is displayed for each vocalization (•). The black line denotes 100 % correct classification for each vocalization. The match between relative frequency content and correct classification is perfect for some frequencies.
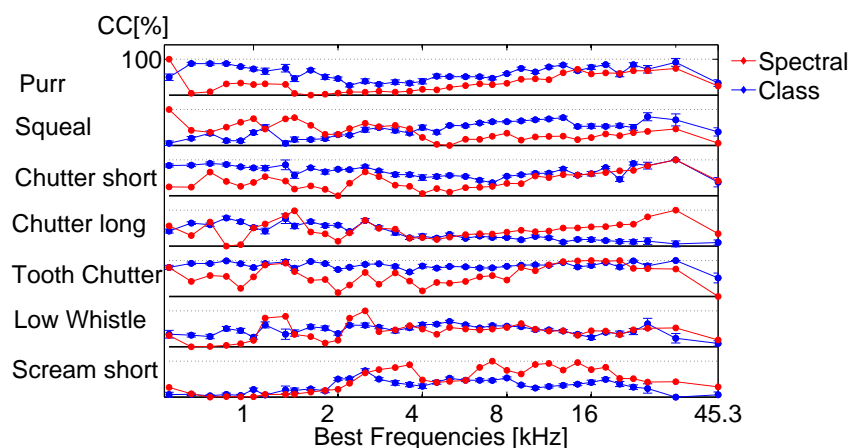
Figure 5.4.3: Correct classification and spectral content for individual vocalizations. The correct classification (●) of individual vocalizations, for discriminating between all 11 vocalizations (averaged across 1152 multi-units), and the vocalizations' relative spectral content ●) is displayed for the remaining 7 vocalizations. Substantial differences between spectral content and optimal encoding exist for some vocalizations.

This preference for encoding groups of vocalizations is spatially distributed. Main spectral contents of the vocalizations also display a spread arrangement across the BF-range. These results are consistent with earlier work showing that neurons in the ICC encode spectrotemporal acoustic patterns of vocalizations [23]. Figure 5.4.4 illustrates the relative frequency content for the 100 ms segment of each vocalization which elicited the responses that were used for the classification. Figure 3.1.1 displays spectral contents for the entire vocalizations. Whereas the 'purr', 'drr' and 'low chutter' have main spectral contents at low frequencies, the 'whistle', 'long scream' and 'short scream' have almost no energy at these frequency ranges, but have important relative power at frequencies above 2 kHz. Main spectral contents of the 100 ms vocalization segments are distributed across the whole BF-range.
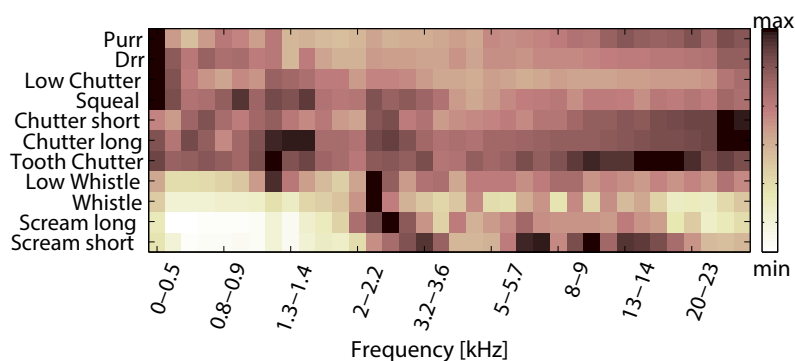
Figure 5.4.4: Spectral content of vocalizations. Relative frequency content for the 100 ms segment of each vocalization which elicited the responses used for the classification. The main frequency contents of the vocalizations are spread across the entire analyzed best frequency range.

5.4.2 Combined responses

To test whether the combined response of several multi-units yields better neural discrimination of all 11 vocalizations, 100 ms-long spike trains from a successively increasing number of multi-units were combined by either concatenating or adding the spike trains. Concatenation preserved the spectral and temporal information whereas addition only preserved the temporal information.

Additionally, discrimination of average firing rates of the combined units was performed. Concatenating the units' spike-rates preserved only spectral information. To test if combining spectral and temporal information affects neural discrimination in a similar way, the following three cases were compared: combining both, spectral and temporal information, and combining only temporal or spectral information. Differently frequency-tuned multi-units from along the BF-gradient were combined and also multi-units from one isofrequency lamina with similar best frequencies combined. To answer the question whether differently tuned multi-units (in frequency or amplitude modulation) provide more information than similarly tuned ones, responses were combined in two different sequences. The additional multi-unit to be combined was either the nearest spatial neighbor or was chosen randomly. Whereas in the first case one successively combines information from similarly tuned multi-units gradually along the BF-gradient (one isofrequency lamina), the second case allows combining information from multi-units with different frequency tuning (different amplitude modulation preferences).

The displayed data are taken respectively from a double-shank (Fig. 5.4.5a-c) and a double-tetrode (Fig. 5.4.5d-f) recording of one animal, and are representative for the whole study of 11 animals. Perfect classification is achieved when using temporal and spectral information across laminae (Fig. 5.4.5a) and nearly perfect classification is achieved when using temporal and spectral information and also when using only temporal information within few laminae (Fig. 5.4.5d, e). Figure 5.4.5a demonstrates that correct classification continuously increases significantly for combining temporal and spectral information of up to 5 multi-units.

Combining responses from several multi-units significantly increases discrimination (Fig. 5.4.5). Figure 5.4.5 depicts correct classification for combining information from successively 1-16 multi-units, along the BF-gradient (a-c) and within few adjacent isofrequency laminae (d-e). Combining 5 multi-units led to perfect discrimination, and was not degraded by adding further multi-units. Across all experiments, discrimination performance increased for up to 3-6 multi-units. Combination yielded near-to-perfect (on average 90 %) correct classification for all recordings, in several cases perfect classification. Thus, combining short responses of 3-6 multi-units is sufficient to uniquely represent 11 different vocalizations.

Increase in correct classification for combining an additional multi-unit was higher when using temporal and spectral information (10 % per unit) than when using only temporal information (5 % per unit), Fig. 5.4.5a,b.

For the latter case, correct classification did not increase significantly when combining more than 5 multi-units but stayed constant around a value of 80 %. Correct classification of spike-rates was low for one multi-unit (20 %), increased for combining up to 5 multi-units (8 % per unit), then continued to increase at a lower rate (2 % per unit) up to a performance level of 80 % and stayed constant for combination of more than 10 multi-units (Fig. 5.4.5c). Using only temporal spiking information, although saturating at a low number of combined multi-units, did not recover the entire information about the encoded signal. Using only spectral information also did not recover the entire encoded information, and a much larger number of units might need to be combined to achieve comparable correct classification. Thus, spectral and temporal spiking information are necessary for near-to-perfect discrimination. Gradually and randomly combined responses yield similar values of correct classification. This suggests that combined responses from differently frequency-tuned as well as from similarly tuned multi-units yield a better discrimination performance than the response from only one multi-unit. Not only large differences in tuning and in responses but also small differences provide new information to uniquely represent vocalizations in the ICC.
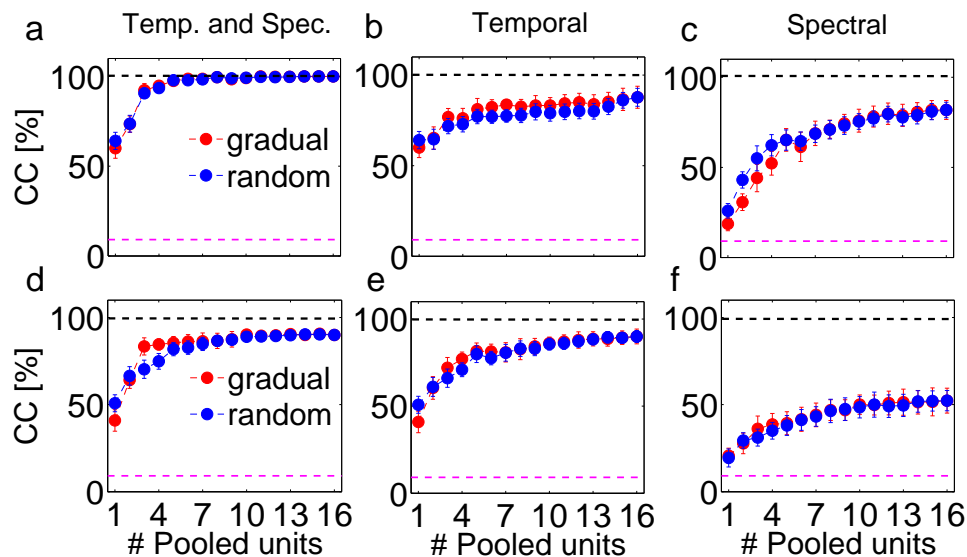
Figure 5.4.5: Combined multi-unit responses. Correct classification of spiking responses to 11 different vocalizations for an increasing number of combined multi-units. Spiking responses are concatenated or added to preserve (a, d) spectral and temporal; (b, e) temporal; or (c, f) spectral response information. Combining responses significantly increases correct classification. Multi-unit responses were combined either across several laminae (a-c, with BF=0.5-23 kHz) or within few adjacent isofrequency laminae (d-f, BF = 6.3-8 kHz), and were taken respectively from one simultaneous recording. Multi-units that are successively combined were selected either gradually in tuning preferences (•) or randomly chosen (•). Averages over 2-3 iterations are shown. The dashed pink line denotes chance level, the black top dashed line is drawn at 100 % correct classification.

The above stated results are largely valid when combining responses from multi-units within few adjacent isofrequency laminae (Fig. 5.4.5d-f.) However, the following exceptions exist: the courses of classification for using both, temporal and spectral, and for using only temporal spiking response information are similar and reach the same maximum performance of 90 %, although more units are needed to reach this value for the latter case. Combining spectral information only yields a maximum performance of 50 %. Responses from few adjacent isofrequency laminae do not contain enough information to allow for perfect discrimination between all vocalizations. Using only temporal information is sufficient to nearly perfectly represent vocalizations. Thus, mostly temporal information is encoded within an isofrequency lamina. The average of all 77 pooling procedures confirms the above stated result (Fig. 5.4.6). Temporal information is encoded in more detail within few adjacent isofrequency laminae than across isofrequency laminae and thus leads to higher correct classification. Correct classification is almost identical for gradually and randomly combined responses.
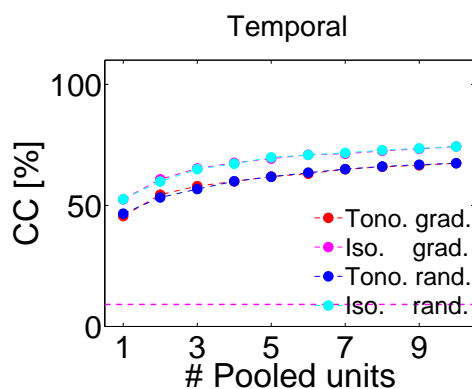
Figure 5.4.6: Average correct classification across all recordings. Correct classification for combining temporal information of spiking responses. Correct classification is higher for combination of multi-units from within few adjacent isofrequency laminae than across several laminae. Gradual (●,●) and random (●,●) combination yield similar correct classification values, along and across isofrequency laminae. The dashed pink line denotes chance level.

5.4.3 Contribution of neural correlations

Simultaneously responding multi-units can interact in order to more efficiently encode vocalizations. This interaction can lead to neural (temporal) correlations of their responses. The correlations between the multi-unit responses could improve separability in several ways. One possibility might be that the interaction strength between the multi-units and consequently their neural correlations varies for responses to different vocalizations and thus aids discrimination between the vocalizations. Another possibility would be that the multi-units' simultaneous responses are more similar to each other than non-simultaneous ones.
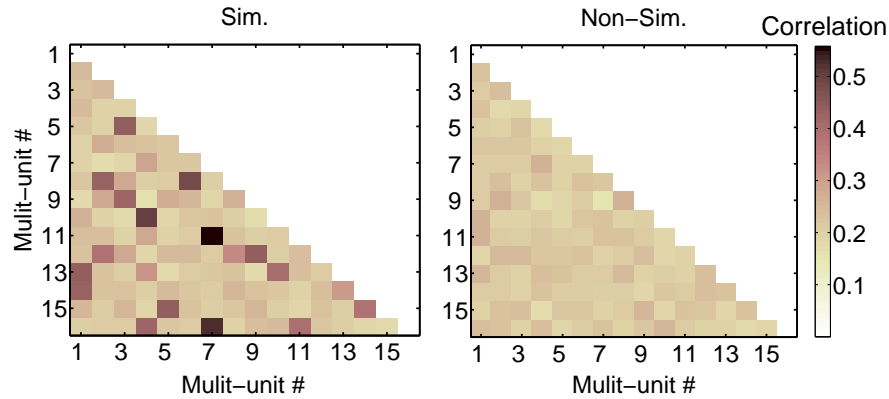
Figure 5.4.7: Correlations of combined multi-unit responses. (a) Half-matrices of correlation coefficients (within a maximum possible delay of $\tau =10$ ms) between responses recorded simultaneously (**Sim**, left) and non-simultaneously (**Non-Sim**, right), in response to the 'squeal' presented at 70 dB SPL, from 16 multi-units spanning a best frequency range of 0.7-22.6 kHz. For simultaneous responses, correlation values differ across multi-unit pairs. Some multi-unit pairs have higher correlation for simultaneous responses than for non-simultaneous ones. Other multi-unit pairs show similar values for simultaneous and non-simultaneous responses. The correlation matrices are symmetric and the diagonal is unity, hence only their lower halves are displayed.

Hence their neural correlations are stronger and by means of redundancy, vocalizations are more faithfully represented. These possibilities were tested by comparing the averaged correlation values between all multi-unit responses of one pooling, to each vocalization, for simultaneously and non-simultaneously recorded responses. Responses of multi-units were combined by concatenation since this approach preserves temporal and spectral information and yielded higher correct classification values than addition of responses, as stated above. Combination of non-simultaneously recorded responses was performed by randomly shuffling the 20 response trials of each multi-unit before concatenation of trials from different multi-units. Figure 5.4.7 shows the pairwise correlation of spiking responses for all 16 multi-units of one recording, for one vocalization. Correlations of simultaneously recorded responses are higher or equally high as correlations of non-simultaneous responses. They vary for different pairs of multi-units. In Fig. 5.4.8a the averaged correlation values for simultaneously and non-simultaneously recorded responses between multi-units are displayed for each vocalization.

For each vocalization correlation values of simultaneously recorded responses are significantly higher than those from non-simultaneously recorded ones (mean, $p =0.05$, Student's t-test). Correlations do not vary significantly across vocalizations, neither for simultaneous nor for non-simultaneous responses. The only exceptions are responses to the 'tooth chatter' which yield significantly higher correlations than to all other vocalizations. Responses to the 'tooth chatter' are phase-locked to the envelope for multi-units across almost all studied best frequencies. This leads to higher correlation values of those responses.

However, these significantly higher correlation values do not yield significantly higher classification values for the 'tooth chatter', but are comparable to other vocalizations such as the 'long scream'. Thus, should neural correlations facilitate a more unique representation of vocalization, this might be achieved via redundancy.

If neural correlations contribute to a unique representation of the vocalizations and hence a better neural discrimination, then their cancellation will lead to a decrease in neural discrimination. The hypothesis that neural correlations improve separability of responses was tested by comparing neural discrimination when combining simultaneously and non-simultaneously recorded 100 ms long spike trains. Correlation values between the multi-unit responses of a combined set were also explicitly included in the feature vector. However, classification values of combined responses for which correlation values were not explicitly included was also compared. In Fig. 5.4.8b correct classification between all 11 vocalizations is compared for combined simultaneously and non-simultaneously recorded responses.
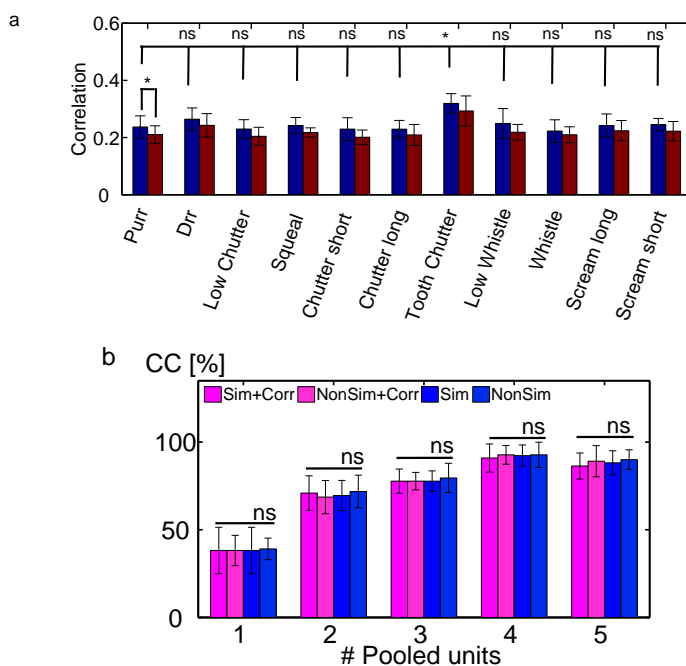
Figure 5.4.8: Correlations of combined of multi-unit responses and their effect on neural discrimination. (a) Average correlations of all (n_{comb} =5) combined multi-units for each vocalization are compared for simultaneously and non-simultaneously recorded responses, for a stimulus intensity of 70 dB SPL. Simultaneous and non-simultaneous response correlations differ significantly for each vocalization. Correlation values across vocalizations do not vary significantly, except for responses to the 'tooth chatter' which are significantly higher than response correlations elicited by all other vocalizations. (b) Comparison of correct classification between all 11 vocalizations for combining 1-5 responses, recorded simultaneously and non-simultaneously, for explicitly including response correlations (**Sim+Corr**,**NonSim+Corr**) and when not including them (**Sim**, **NonSim**), at 70 dB SPL. (* denotes significant results, **ns** non-significant; mean, p =0.05, Student's t-test).

Simultaneity of recordings does not produce a detectable difference in correct classification. Correct classification does not vary significantly for combining simultaneous or non-simultaneous responses, for the case of explicitly including the multi-units' correlations and for the case of not including them. Combining responses increases neural discrimination as shown previously (Sec. 5.4.2), and did not increase further when combining more than 5 multi-units. To increase the complexity of discrimination, thus to raise the threshold, correct classification for combined simultaneous and non-simultaneous responses was compared for stimulus intensity levels of 30-70 dB SPL, in steps of 10 dB SPL. Overall, correct classification values were lower, however, classification did not differ significantly between simultaneous and non-simultaneous combined responses. Correlations were computed allowing for a delay between responses from different multi-units. If no delay is assumed, the computed non-simultaneous response correlations are much lower than simultaneous ones, and classification performance is significantly higher for non-simultaneous recordings than for simultaneous ones. Thus in this case, neural correlations are found to be detrimental for discrimination (mean, $p = 0.05$, Student's t-test). This suggests that the interplay of multi-units from different sites does not aid unique representation of the vocalizations.

5.5 Discussion

Summary

In this chapter, encoding of vocalizations in multi-units of the ICC was analyzed using neural discrimination of spike trains.

It was found that within the multi-unit's neural response to the vocalization, it did not substantially influence the performance which specific temporal segment was chosen for the discrimination analysis. It was demonstrated that, in general, individual vocalizations are optimally discriminated across a rather broad range of the best frequency gradient. The spatial distribution of the capability to accurately encode the stimulus differs for the individual vocalizations. Preferred encoding along the BF-gradient follows in most cases a similar trend as the spectral content distribution of the vocalization. A vocalization which contains important spectral energy across the whole studied best frequency range ('tooth chatter') is not preferentially encoded by some multi-units, but is encoded uniformly well across best frequencies. For some vocalizations and frequencies, the match of preferred encoding and spectral content is perfect ('long scream', Fig. 5.4.2), for others, deviations exist. However, trends of preferred encoding for the individual vocalizations and spectral content are often similar.

Furthermore, the work shows that groups of a few multi-units yields improved total discrimination over only one multi-unit. Combining responses from up to 3-6 multi-units significantly increases discrimination between the 11 vocalizations. It did not significantly alter the discrimination performance whether multi-units with similar frequency tuning (alternatively amplitude modulation tuning) or with different tuning were combined. The combined spectral and temporal information of responses from 3-6 multi-units across frequency laminae allowed, on average, for near-to-perfect discrimination for all studied 77 recording sets, and was even perfect for several of them. Thus, a few multi-units are able to reliably encode a representative set of behaviorally relevant vocalizations. It was demonstrated that neural correlations of combined multi-units do not contribute to a higher accuracy of neural discrimination between vocalizations, but have either no impact or are detrimental. These results are independent of the multi-unit's best frequency and, hence, of the response strength to individual vocalizations. That neural correlations do not facilitate encoding is further supported by the finding that discrimination improves regardless of whether neighboring or distant neurons were combined, with neighboring neurons having a higher probability of interacting and hence displaying neural correlations.

Critical Review and Future Work

The discussion of the neural data in Sec. 4.5 also applies to the analysis discussed here. Neural discrimination, especially for multi-units that did not yield perfect classification when being combined, is likely to improve in awake animals, as anesthesia has non-negligible effects on the neural activity [118]. Still, the neural discrimination values obtained in this study were very high, in several cases even perfect.

The findings presented in this chapter are valid for multi-units. It would be interesting to investigate whether a multi-unit responds differently to different vocalizations because it either recruits different groups of single neurons for each vocalization, or because the spike times of the multi-unit and thus of the single neurons are different for each vocalization; the same holds for averaged spike-rates. A single neuron resolution within the multi-unit cluster would allow answering this question.

In the presented work, the overlap of the stimulus' spectral content and optimal encoding is not perfect, deviations exist and might point to nonlinear processing. Discrimination performances were averaged across all multi-units within a best frequency interval. Thus, preferences of individual multi-units might be averaged out. These deviations could be due to amplitude and frequency modulation preferences within isofrequency laminae [6], or due to higher temporal but poorer spectral resolution for low BF neurons (vice versa for high BF neurons [92]). Deviations could also be explained by nonlinear processing properties, such as a possible selectivity for specific vocalizations as suggested by [26], or by spatial heterogeneity of receptive fields in the ICC [22]. A possible high call-selectivity of single neurons would be averaged out, and only if regions of high call-selective neurons for individual vocalizations existed, would these have been perceivable as an enhanced discrimination. This, however, was not the case, and even the existence of single neurons in the ICC which are call-selective is controversial. The deviations need to be analyzed in more detail. Single neuron recordings, for which not only the best frequency, but also best amplitude modulation frequency and the location within the ICC is known, could be employed. Single-unit recordings would allow to test for call-selectivity to individual vocalizations. They would also allow for a direct comparison of the vocalization's spectral content and its best discrimination in dependence on the neurons' best frequency (Sec. 5.4.1), since the analyzed multi-unit's BF is a compound BF of several single neurons. Such an analysis could be used to quantify deviations from a linear mapping of the spectrotemporal content of the vocalizations across the ICC and could point to increased regions of nonlinear processing. Furthermore, single-unit recordings might allow quantifying the amount of correlations due to neuronal interactions, and possibly infer structural connectivities between neurons, which is not possible for multi-units [116]. Temporal correlations which are due to neural interactions might exist within the multi-units and could support a microcircuit functional arrangement.

A distortion in the neural representation, which might hamper and corrupt inferences that are obtained from direct comparison of the stimulus and the neural response, results from the presentation of calibrated vocalizations (Sec. 3.1.1). The neural responses, in general, follow the temporal and spectral form of the original vocalizations. However, in order to avoid any ambiguity, it would be necessary to record the sound exiting the tube and compare it to the neural response.

It would be interesting to examine encoding not only for different vocalizations, but for natural different exemplars of the same vocalization. The hypothesis that discrimination of these variants is optimal in frequency regions tuned to parts of the spectrum that are acoustically more discriminative could be tested, and also how the variability of one vocalization manifests in the neural representation, thus how small differences are represented in the neural system.

Conclusion

The findings on the representation of vocalization are valid on the level of multi-units without assuming that the same group of single neurons responds to different vocalizations. Neural discrimination of individual vocalizations remains very good over broad frequency regions and this should be beneficial in a behavioral context.

Responses from a few multi-units are sufficient to faithfully represent the whole set of studied vocalizations with very diverse spectrotemporal properties. Improvement of neural discrimination when combining responses due to reduction of trial-to-trial variability might be less predominant for multi-units than for single-units, and suggests complementary information from the different multi-units. Multi-units in the ICC act as independent encoders in the sense that correlations with other neurons do not facilitate encoding and redundancy is minimized. No direct interaction beneficial to encoding between the multi-units is present, and might suggest parallel processing. Possibly, the independent spiking responses are projected to and combined in higher processing stations. This work supports an efficient encoding strategy, suggesting that higher level neural representations match the statistical and behavioral qualities of the stimuli.

6 Nonlinear single neuron responses

6.1 Introduction

In the previous Chapter 5 the encoding of vocalizations by clusters of neurons was investigated. It was demonstrated that within each neuronal cluster, there is responsiveness to almost all vocalizations. Vocalizations and speech can display temporal and spectral correlations including harmonics, frequency and amplitude modulations. Hence, how does this responsiveness to vocalizations manifest at the single neuron level? Do preferences to specific acoustic features of the vocalizations exist? The neurons in the auditory midbrain might employ various processing mechanisms –linear and nonlinear ones– to encode these behaviorally relevant sounds. In the inferior colliculus an important part of the neurons ($\sim$ 40 %) display nonlinear response properties [14].

In this chapter, nonlinear response properties of single cells are analyzed. For the visual system it has been suggested that simple receptive fields which display linear summation properties form a unit with a surrounding complex receptive field that displays nonlinear summation properties [119, 65, 66], and that this might increase selectivity of neurons [56]. In this work, linear and nonlinear receptive fields are investigated. Do nonlinear auditory responses lie within the linear spectrotemporal receptive field? Or do such nonlinearities lie outside, thus forming a non-classical receptive field with nonlinear summation properties analogous to cells in the visual V1 cortex [56]? Since natural sounds are composed of several temporal and spectral correlations, they do not allow directly linking the neural response to a particular stimulus property, but only to a composition of a spectral and temporal component. A systematic approach to identify neuronal preferences, on the other hand, is offered by the use of the dynamic moving ripple (DMR) sound which contains combinations of frequency and amplitude modulations and dynamically covers a certain range of the parameters [14].

In order to link the stimulus properties and the neuronal response, the single cell spiking activity is recorded while presenting the artificial DMR sound. The reverse correlation method is well suited for the derivation of nonlinear stimulus preferences since it does not put any limit on the number of extracted stimulus features, as opposed to the maximally informative dimension method [73, 91, 55]. Furthermore, the DMR stimulus meets the requirements for use with the reverse correlation method.

Thus, the averaged spiking activity in response to the stimulus frequency and delay time (spectrotemporal receptive field) and the averaged spiking activity to stimulus correlations (the spike-triggered covariance) can be computed. The spike-triggered average (STA) and covariance (STC) are derived from the Wiener kernel analysis [95], an expansion of functionals which approximates nonlinear systems [76, 77, 69]: $r(t) = k_0 + k_1 + k_2 + ...$. The zeroth order Wiener kernel term corresponds to the average spike rate, the first term corresponds to the STA and the second order Wiener kernel, which corresponds to the spike-triggered covariance, captures correlations within the stimulus which trigger the neuron to spike [120, 71]. Neural ICC responses were analyzed with respect to the second order Wiener kernel in order to test whether they display preferences to these nonlinear stimulus interactions.

In previous studies, the entire auditory sound waveform was considered for the reverse correlation analysis [84, 79] or for derived information theoretical approaches [39]. In the present work, however, the reverse correlation method is applied to the spectrotemporal envelope separately for each frequency carrier. Using the amplitude modulation spectrum of the DMR sound for the reverse correlation analysis allows studying the representation of amplitude modulations in the ICC. Amplitude modulations are essential for recognition of complex sounds such as speech [20] and vocalizations. At the level of the ICC, frequencies greater than about 1 kHz are not encoded due to the inability of the neurons to follow the fast changes [42]. The frequency carriers, the fine structure, which were used for the DMR sound, span a frequency range of 1-48 kHz.

Nonlinear response properties to the envelope have been observed in several processing stations of the ascending auditory pathways and become increasingly complex and specialized as one progresses up the auditory hierarchy [121, 91, 97]. However, it has not been investigated previously in which frequency range these nonlinearities are contained. By computing the reverse correlation for each frequency carrier of the DMR sound, it is possible to assess in which frequency range the nonlinearities are dominant and how these frequencies relate to the best frequency of the neuron. Thus, instead of the whole stimulus waveform, only the individual stimulus amplitude modulations for each frequency carrier of the sound stimulus are used for the analysis.

In the following Section 6.2, the derivation of the spike-triggered covariance to DMR sound for each frequency channel is described in detail. Examples of derived STCs and the distribution of significant nonlinear response properties across the frequency channels and across the population of studied neurons are given in Sec. 6.3. The significant spectrotemporal receptive field and the significant spike-triggered covariance are compared. The findings are discussed in the last part of this chapter.

6.2 Methods

The basic methods employed for the characterization of nonlinear response properties were described in Section 3.5.3. Those were extended for the use of the dynamic moving ripple sound envelope to obtain spike-triggered covariances for each frequency channel and ascertain significance of these nonlinear receptive fields.

6.2.1 Stimulus and neural data

In order to identify neuronal response preferences, dynamic moving ripple sound was delivered independently to both ears of the animal. The DMR sound was played to the right ear and the time-reversed DMR sound was played to the left ear. Recordings were made from the left ICC of anesthetized cats, and processed to yield single unit responses which were used for the reverse correlation analysis. The experimental set-up and the DMR stimulus are described in detail in Sec. 3.1.2 on page 27.

6.2.2 Spike-triggered average (STA)

The spike-triggered average (STA), also denoted as the linear spectrotemporal receptive field (STRF), was obtained by averaging the temporal windows of the DMR sound envelope that elicited a spike [14]. This receptive field is 2-dimensional, with a maximum time delay of 50 ms, a temporal resolution of 0.5 ms and a frequency range of 0-5.8 octaves, with 659 frequency channels, thus providing a fine spectral resolution of 0.0085 octaves. The maximum temporal delay of 50 ms was chosen because integration and delay times of ICC neurons are enclosed within this temporal window [14, 115]. The statistically significant portion of the STA is obtained by keeping all values of the STA that exceed 3σ (1.6σ) of the control noise STRF (which was obtained by adding random sound waveform segments) and setting all other values to zero [14]. The STRFs have excitatory and/or inhibitory regions, which indicate respectively enhanced and suppressed spiking activity for these specific temporal and spectral parameters, see Fig. 6.2.1. For the STRF and for the STC-derivation the full 20 min of recorded spiking activity and DMR sound are used. The STRF is well suited to map linear response properties [69, 70, 97, 107]. However, to describe nonlinear response properties, the second order reverse correlation method, the spike-triggered covariance is necessary.

6.2.3 Spike-triggered covariance (STC)

Nonlinear neural response preferences can be derived with the spike-triggered covariance (STC). It is obtained by comparing the variance of the entire stimulus distribution and the spike-eliciting variance.
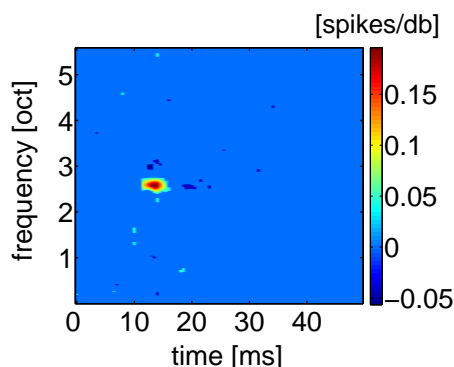
Figure 6.2.1: The spectrotemporal receptive field displays spectral and temporal preferences of a neuron. Parameter regions with high spiking activity indicate neuronal preferences, whereas spiking activity below the average (number of spikes is zero) points to regions of suppressive parameter regions.

The STC describes the average (second order) correlations within the stimulus that lead to an elicitation of a spike. While the average spike rate and spectrotemporal receptive field are rather intuitive, the spike-triggered covariance is not. In Figure 6.2.2, an example of a schematic STC is shown. The matrix is symmetric around the diagonal, but for better visualization the full matrix is displayed. The schematic shows an enhanced spike rate for stimulus correlations at time delays $(\tau_1, \tau_2) = (15,5)$ ms and $\tau_1, \tau_2 = (5,15)$ ms, thus at a relative delay of 10 ms. The spike rate is suppressed for stimulus correlations between time delays $(\tau_1, \tau_2) = (10,30)$ ms and $(\tau_1, \tau_2) = (30,10)$ ms, thus at a maximum relative delay of 20 ms. In order to illustrate the relation of the preferred stimulus correlations and the shape of the spike-triggered covariance, a schematic with three examples is given in Fig. 6.2.3.
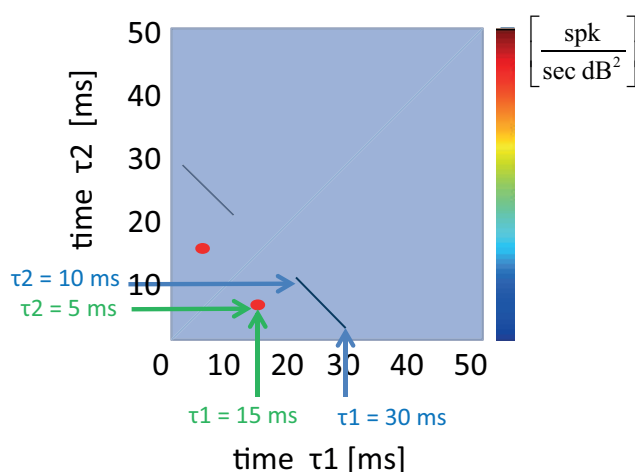

Figure 6.2.2: Schematic example of spike-triggered covariance with excitatory (red) and inhibitory (blue) stimulus correlations.
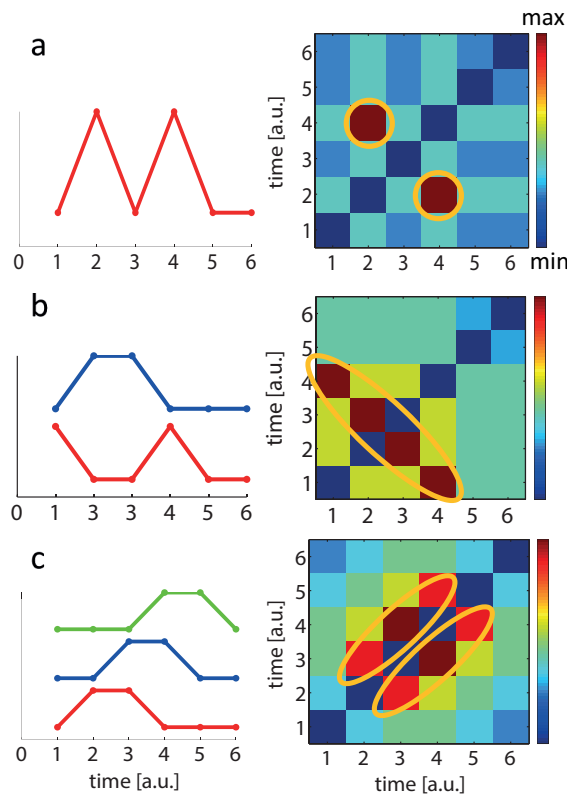
Figure 6.2.3: Relation of stimulus correlations and STC shape. Three examples of preferences to different stimulus correlations (left) and the derived outer products (right) are given a), b) and c). Diagonals which correspond to zero delay have been subtracted in the outer products.

Correlations within the stimulus (Fig.6.2.3a left) exist at delay times $(\tau_1, \tau_2) = (2,4)$ and $(4,2)$, and higher values are present at these delay times in the outer product (Fig.6.2.3a right), which is derived from the stimulus. In the second example, stimulus correlations are present at $(\tau_1, \tau_2) = (1,4)$ and $(4,1)$, and at $(\tau_1, \tau_2) = (2,3)$ and $(3,2)$, with relative delays of 3 and 1 bins. These preferences give rise to high values along the orthodiagonal. In the third example three different stimulus correlations which occur at different times τ but all have the same relative delay of 1 bin are present. These stimulus correlations give rise to high values along the paradiagonals, which are 1 bin away from the diagonal. For a larger relative delay, e.g. 5 bins, the paradiagonals will be 5 bins away from the diagonal. Shapes like the one in the third example (at a distance of 0.5-1.5 ms from the diagonal) were found previously in the auditory nerve when computing the spike-triggered covariance in response to white noise [81]. The pattern of paradiagonals was interpreted by the authors to represent a neuronal response component that is phase-locked to the square of the envelope of a filtered version of the stimulus waveform, or resulting from a quadrature pair of eigenvectors. Unlike prior work in the auditory nerve, the STC (see Eq. 3.5.6) derived in this work is obtained from the DMR sound envelope and calculated separately

for each frequency channel *f*. This approach allows identifying dominant nonlinearities across frequency carriers.

Derivation

The STC is obtained from the second order Wiener kernel (Eq. 3.5.6 on page 40). For each frequency channel *t*, the covariance matrix that is obtained from the whole stimulus is subtracted from the covariance matrix that was computed only from stimulus segments preceding spikes (spike triggered segments):

$$\text{STC}(f) = \text{STC}_{\text{spike}}(f) - \text{C}_{\text{int}}(f) \tag{6.2.1}$$

with the spike-triggered covariance $\text{STC}_{\text{spike}}$:

$$\text{STC}_{\text{spike}}(f) = \frac{N}{T} \frac{1}{(2(\frac{m_{\text{db}}^2}{8})^2)} \sum_{i=1}^{N} m_{\text{db}} \frac{1}{N} \left(S(f, t-\tau) - \frac{m_{\text{db}}}{2} \right) \otimes \left(S(f, t-\tau)' - \frac{m_{\text{db}}}{2} \right) \tag{6.2.2}$$

The outer product is computed for each temporal DMR envelope segment of length τ which elicited a spike at time *t*, and averaged across all *N* obtained matrices for each frequency channel (schematically shown in Fig. 6.2.4).
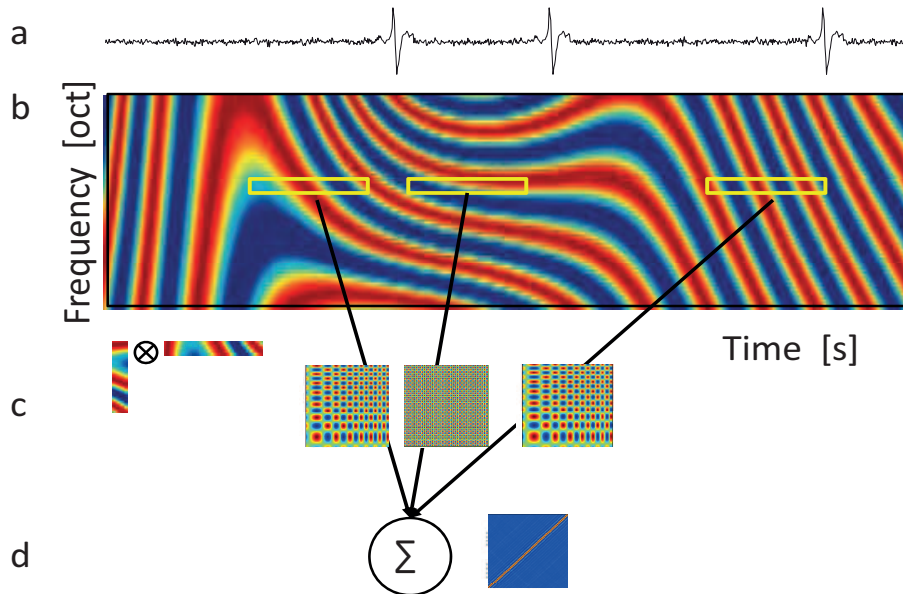

Figure 6.2.4: Schematic derivation of spike-eliciting covariance. a) recorded spike train; b) DMR stimulus amplitude modulation spectrum; c) outer products of *N* τ-long segments preceding a spike; d) averaged $\text{STC}_{\text{spike}}(f)$ for frequency channel *f*. The top shows the first 500 ms of a single spike train obtained from recording while the DMR sound with the amplitude modulation spectrum shown in b) was presented. When a spike occurred, an immediately preceding time window of τ=50 ms was taken from the spectrum for a single frequency carrier *f*. c) From this vector the outer product was computed. The spike-triggered covariance is the average of all obtained outer products.

The difference between the covariance and the outer product is that for the covariance, for each (τ_1, τ_2)-combination the mean of each segment is subtracted before multiplication, whereas for the outer product this is not the case. The outer product can be used for the spike-triggered covariance analysis [79]. The number of spikes, N, varied from neuron to neuron, but was always higher or of the order of 10^3, and the maximum time delay for each segment ($\tau = 50$ ms) was the same as for the STA. The frequency resolution is 0.017 octaves, with 220 frequency channels. It is adjusted to densely sample ($20\times$) within the average STRF bandwidth of the ICC which is 1/3 octave. Each stimulus segment is multiplied by the mean intensity $m_{db} = 30$ dB, and centered. The outer products are averaged across all N matrices. The averaged STC is normalized by the mean firing rate $\frac{N}{T} = N_0$, with a total recording time of $T = 600$ s, and with the squared standard deviation of the DMR sound $\sigma^2 = \frac{m_{db}^2}{8}$ [14]. The STCs from both trials A and B are averaged. In Figure 6.2.5a, an example of an obtained spike-eliciting STC (STC_{spike}) is displayed. In order to obtain the correlation matrix which is only due to stimulus correlations that elicit a spike, the intrinsic stimulus correlations of the DMR envelope have to be removed. This is achieved by subtracting the average outer product of the whole DMR stimulus from STC_{spike} (Eq. 6.2.1). The DMR sound displays no global correlations, however, it does display short-term correlations (see Sec. 3.1.2 on page 27). This has an effect on the computed spike-triggered outer product, see Fig. 6.2.5a. Several paradiagonals of fading intensity with increasing distance from the diagonal can be seen. These are due to the short-term correlations of the DMR stimulus and are sometimes not removed by averaging the outer products for all recorded spikes. The effect of these short-term correlations becomes dominant for neurons which are highly feature selective. For these neurons that almost exclusively fire in response to a specific combination of a temporal and spectral modulation, the local stimulus correlations captured in STC_{spike} are pronounced. They are not removed by subtracting by the global mean, which is obtained by randomly selecting segments from the stimulus and averaging their outer products.
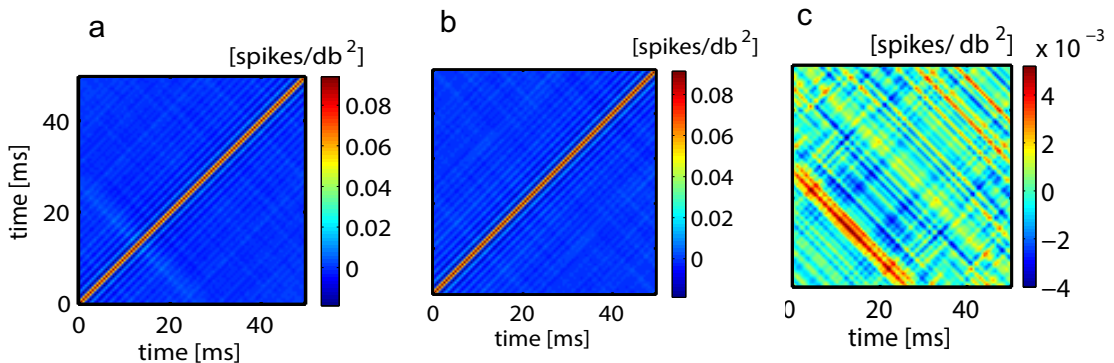

Figure 6.2.5: The a) spike-triggered outer product, the b) outer product obtained with jittered spike times and the resultant c) spike-eliciting correlation matrix, computed for one frequency channel of one neuron.

To account for these local correlations, the spike times were added random temporal jitter from a range of -12,5 ms to 12,5 ms, because within this range short term correlations are still present, however, envelope phase-locking is abolished. The matrix which captures the correlations of the stimulus C_{int} is described by a similar formula as for STC_{spike}, with the difference that the spike times have been added jitter from a confined temporal range:

$$C_{int}(f) = \frac{N}{T} \frac{1}{(2(\frac{m_{db}^2}{8})^2)} \sum_{i=1}^{N} m_{db} \left(S(f, ts - \tau) - \frac{m_{db}}{2} \right) \otimes \left(S(f, ts - \tau) - \frac{m_{db}}{2} \right) \quad (6.2.3)$$

The average outer product was computed for these jittered spike times ts from the DMR envelope and the obtained correlation matrix C_{int} was subtracted from the spike-triggered correlation matrix STC_{spike}. Figure 6.2.5b shows the outer product for which temporal jitter has been added to the spike-times for one neuron. The paradiagonals are also visible in this averaged outer product, and the intensity range is the same as for the spike-triggered correlation matrix. The only perceivable differences are the light lines orthogonal to the diagonal at delay times around 20 ms in the spike-triggered correlation matrix. Removing the intrinsic stimulus correlations (Fig. 6.2.5b) from the spike-triggered outer matrix (Fig. 6.2.5a) yields only the correlations which trigger the spiking of the neuron. The resultant correlation matrix is displayed in Fig. 6.2.5c. The paradiagonals due to short-term correlations are not perceivable anymore, but the previously light orthogonal line is very pronounced. The intensity of this difference matrix is more than one order of magnitude smaller than either one of the two spike-triggered matrices. For the remainder of this chapter this matrix will be referred to as STC.

Several whitening procedures have been described for non-white Gaussian stimuli [122, 123]. A general approach is to normalize the obtained significant dimensions by the pseudo-inverse covariance matrix of the stimulus [106]. However, the DMR being Gaussian white noise [14], these whitening procedures do not have to be taken into account.

6.2.4 Significance testing

In order to determine if the obtained STC matrix is indicative of neuronal preferences for stimulus correlations, significance testing was performed. Singular value decomposition of the STC matrices for each frequency channel was computed, and distributions of singular values were obtained. To estimate the noise level from the recorded data, an STC is computed using Eq. 6.2.1 from randomly chosen 1 min segments of the 10 min spike-train recording ('bootstrapping'). From this matrix STC_{boot}, the singular values and vectors are derived, and averaged for trials A and B. This procedure is repeated 100 times for different segments, and the mean (μ) and standard deviation (σ) of these 100 singular value distributions were computed. This yields an estimate of the noise level, the estimated error with a standard deviation. The singular values of STC and STC_{boot} are ordered accordingly to decreasing value.
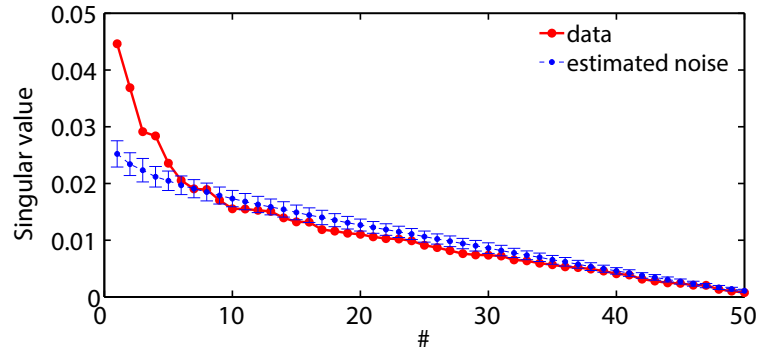
Figure 6.2.6: Significance testing. The distribution of singular values which are derived from the STC (displayed in Fig. 6.2.5a) for the frequency carrier f = 2.6 oct are shown. Values exceeding the estimated noise level by at least $\Theta = 1.6\sigma$ (p=0.05) are considered significant.

The distribution of singular values obtained from the STC and the estimated noise distribution for one frequency channel (f = 2.6 oct) of a neuron are given in Fig. 6.2.6. The spike-triggered covariance was considered significant in this work, if its singular values exceed the noise by at least $\Theta = 1.6\sigma$ standard deviations (p=0.05). These singular values and corresponding vectors were significant. STCs were estimated significant if they displayed at least one significant singular vector. The analysis has also been performed for a higher threshold, using $\Theta = 3\sigma$ (p=0.001). The significance level will be given in each analysis. Unless stated otherwise, only significant STCs are further analyzed in this work. The significant vectors and their corresponding values can be used to reconstruct a STC for which the noise is removed, STC_{Rec}.

6.2.5 Regularized STC

The regularized (STC_{Reg}) for each frequency channel is reconstructed by summing the outer products of only the *significant* singular vectors which are weighted by their singular values. The STC STC_{Reg} for an STC(f) with n significant singular values $\lambda_{i=1..n}$ of sign $p_{i=1..n}$ and singular vectors $\overrightarrow{S_i}$ is given by [82]:

$$\text{STC}_{\text{Rec}} = \sum_{i=1}^{n} p_{\text{i}} \cdot \lambda_{\text{i}} \cdot \overrightarrow{S}_{\text{i}} \otimes \overrightarrow{S}'_{\text{i}} \tag{6.2.4}$$

The original and reconstructed STC for one neuron and frequency channel are displayed in Fig. 6.2.7.
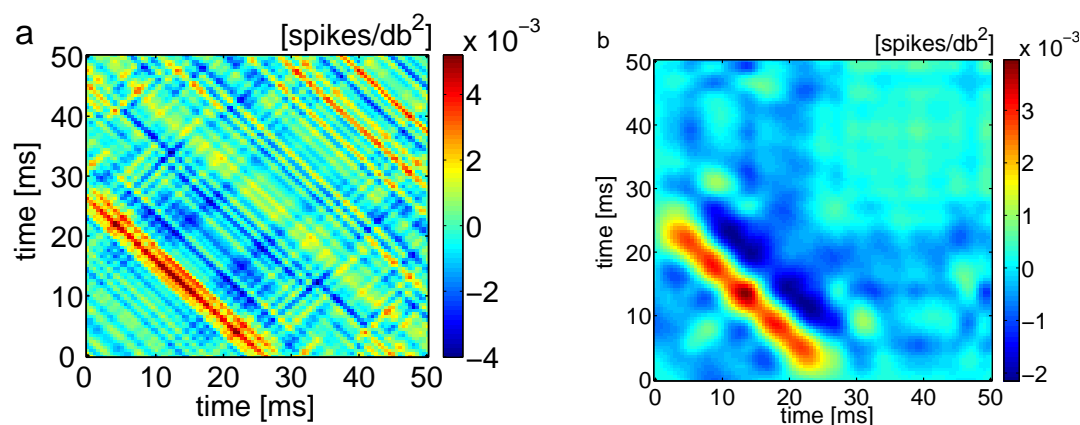
Figure 6.2.7: Reconstructed STC for a neuron and frequency channel (for which the singular value distribution is displayed in Fig. 6.2.6). a) original STC; b) reconstructed STC, using all 5 significant singular vectors.

6.2.6 Spectrotemporal representation of STC

In order to be able to compare a significant STC to the linear STRF, the STA, a 'STRF-like' representation is proposed: for each frequency channel, the highest significant singular vector is plotted in dependence of time. If no significant values were found for this frequency channel, the vector is simply a zero-vector. This yields a matrix of singular vector value in dependance of the frequency channel and the time delay. This representation is of the same dimensions as the linear STRF.

6.3 Results

The analysis described in the previous Section 6.2 was applied to 178 single neurons. Significant singular values and thus significant STCs for at least one frequency channel were found in 132 neurons (75%). When testing with an elevated threshold of $\Theta = 3\sigma$ still 102 (57%) displayed significant STC.

6.3.1 Significant values across frequencies

The values and the total number of singular values (SV) derived from the STC varied across frequency channels. In Fig. 6.3.1, the distribution of significant singular values across all 220 frequency channels is shown for three neurons. In most cases, the highest significant singular values and also the largest quantity of singular values were obtained from STCs at and around the best frequency. STCs of other frequency channels still yielded significant values, but of smaller magnitude and quantity (Fig. 6.3.1a). For some neurons, only the STC of one frequency channel at or close to the best frequency yielded significant values (Fig. 6.3.1b). In some cases, only a few significant values were found, at the best frequency but also at frequency channels more than one octave away (Fig. 6.3.1c).
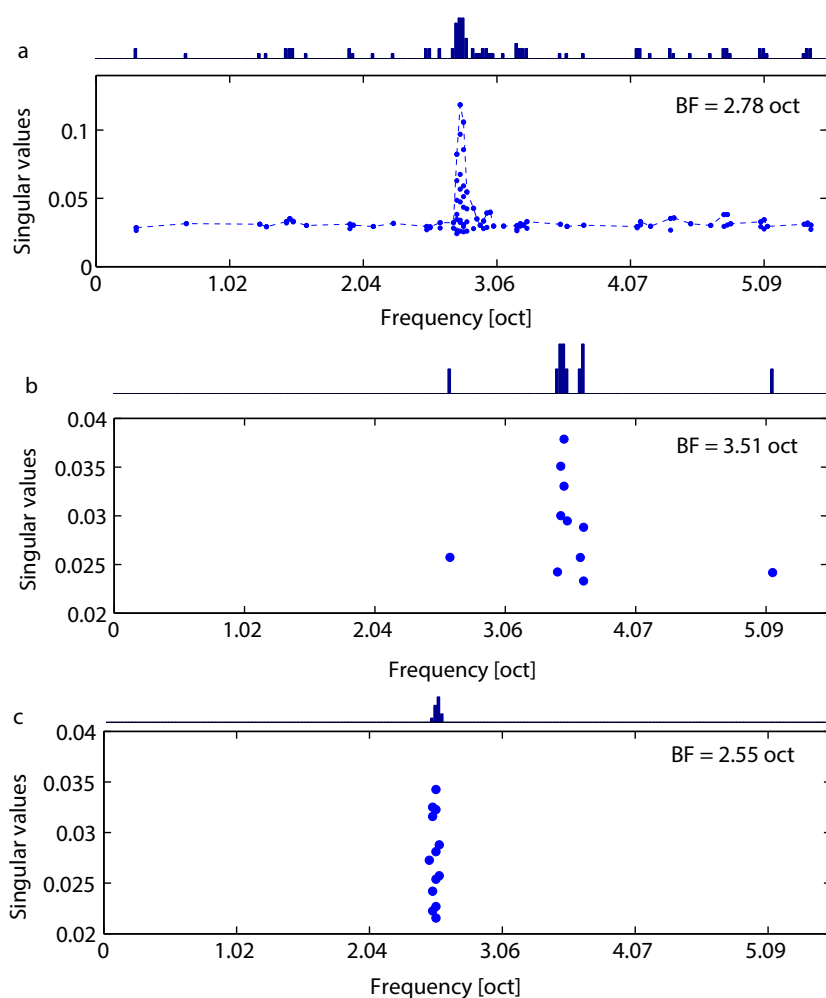
Figure 6.3.1: Distribution of significant singular values across frequency channels. The distributions for three different neurons a), b) and c) are given ($\Theta = 1.6\sigma$, p=0.05). The blue dashed line is a guide to the eye. On top of each graph the relative number of significant values for each frequency channel is displayed.

In most cases, the highest singular values (and often the largest quantity) was found for the STC at the best frequency (BF) channel, see Fig. 6.3.2. The dominant STCs are at the best frequency, and the magnitude of the singular values generally falls off symmetrically to both frequency sides of the BF. The width of the coherent frequency range around the best frequency which displays significant singular values was computed for each neuron. For 84 % of the neurons, this coherent bandwidth around the BF equals or is less than 2/3 of an octave (see Fig. 6.3.3). The singular values differ across frequency channels. The STCs from which they were derived also differ. They differ not merely in magnitude but also in their 'shape'.
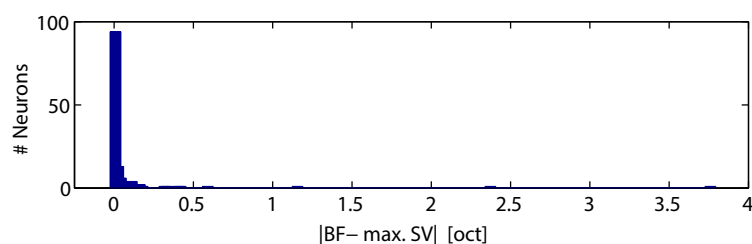
Figure 6.3.2: The STCs at the best frequency channel have the highest singular values. For each neuron, the absolute spectral difference (in octaves) between the frequency channel of the STC displaying the highest singular value and the best frequency are shown. For 92 neurons the best frequency yields the highest singular value, and for 97 % of the neurons, the frequency channel with the highest singular value and the best frequency differ by less than half an octave ($\Theta = 1.6\sigma$, p=0.05).

6.3.2 Significant STCs

Spike-triggered covariances vary across frequency channels and across neurons (Fig. 6.3.4, 6.3.4). Figure 6.3.4a, e show representative STCs at the best frequency with orthodiagonals. Stimulus correlations with maximum delays of 30 ms affect spiking activity. The stimulus correlations follow an excitatory-inhibitory (Fig. 6.3.4a) or an inhibitory-excitatory-inhibitory (Fig. 6.3.4e) pattern. However, STCs which display neuronal preferences for stimulus correlation with a constant time delay at around 10 ms (for detailed explanation see Sec. 6.2.3) are frequent, although these are present mostly for STCs off the best frequency (Fig. 6.3.4b). The displayed STC show a temporal series of inhibitory and excitatory stimulus correlations with a constant delay. When averaging the STCs across all frequency channels, these stimulus correlations are dominant (Fig. 6.3.4c,g). In some cases they are also pronounced when only considering significant STCs (Fig. 6.3.4d) but averaged out in other cases (Fig. 6.3.4h). The reconstructed STC from of Neuron 3 (Fig. 6.3.4l) displays preferences for stimulus correlation with constant delay and suppressed spiking activity for constant delay correlations of 10 ms.
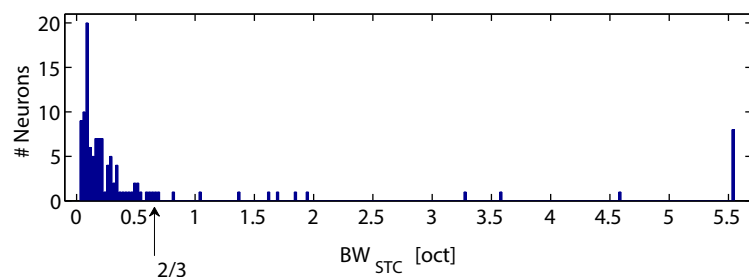

Figure 6.3.3: Distribution of STC bandwidths. The width of the coherent frequency range around the best frequency, for which the STCs are significant lies within 2/3 octave for most of the neurons ($\Theta = 1.6\sigma$, p=0.05).
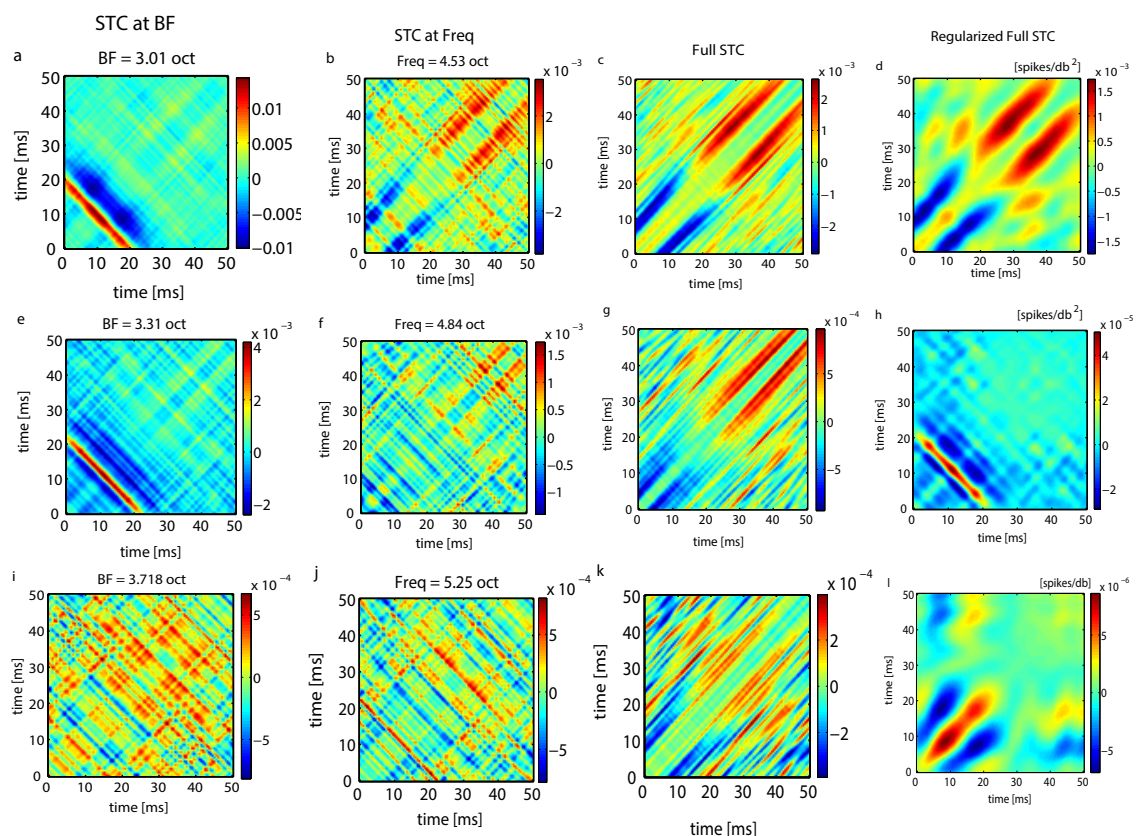
Figure 6.3.4: Examples of spike-triggered covariances for three neurons. Neuron 1: (a-d), Neuron 2: (e-h), Neuron 3: (i-l). For each neuron the spike-triggered covariance at the best frequency (a, e, i) and at a frequency channel more than one octave away (b, f, j) are displayed. The average STC across all frequency channels of one neuron is given, the 'Full STC' (c, g, k) and its regularized STC, which is constructed only from significant STCs, the 'Regularized Full STC' (d, h, l). For all $\Theta = 3\sigma$, p=0.05, except for subfigure h), $\Theta = 1.6\sigma$, p=0.001).

In some cases, the STC does not show a pronounced pattern, but yields significant singular values (Fig. 6.3.4 f,i,j). In order to investigate whether significant STCs can be predicted by other descriptions of the neurons, these were compared to the occurence of significant STCs. For three neurons, the spike-triggered covariances at the best frequency are displayed together with their significant STRFs, their derived ripple transfer functions (RTF) and conditioned response histograms (CRH) (see Sec. 3.5.4 for definitions of RTF and CRH) in Fig. 6.3.5.

The CRH counts the occurrences of spikes to a particular parameter combination. The RTF also shows preferences for parameter combinations, however, it is derived from the linear STRF, and might not show the full preference spectrum of the neuron, since the STRF might be averaging out preferences. A large deviation between the CRH and RTF could indicate nonlinear response properties. Significant STCs occurred for neurons that showed a large overlap of RTF and CRH (Neuron 1), but also for neurons for which the overlap was not very big (Neuron 2).
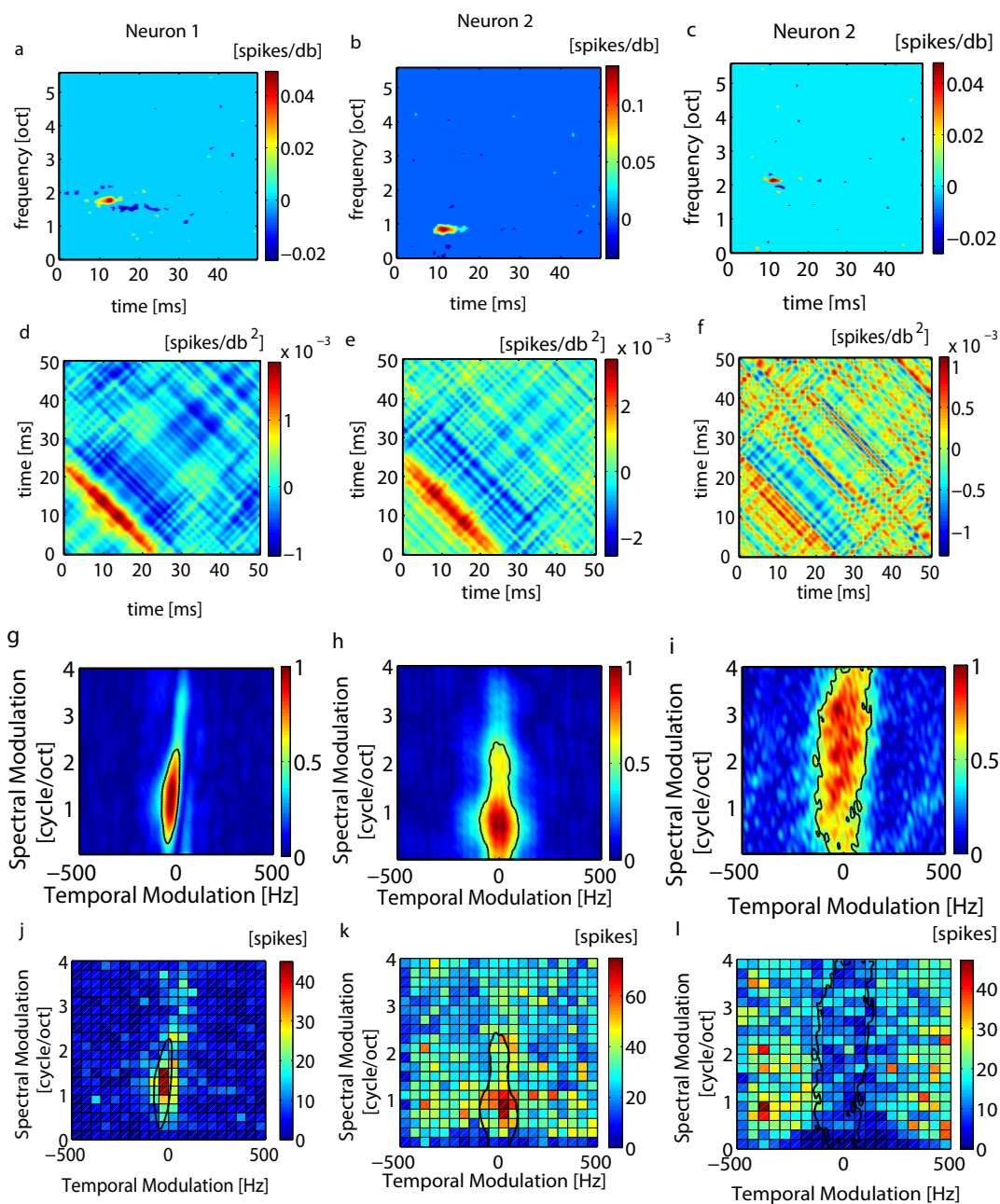
Figure 6.3.5: Comparison of STC to other descriptors of neuronal preferences. For Neuron 1, 2 and 3, the significant linear spectrotemporal receptive field STRF (a-c), the spike-triggered covariance STC at BF (d-f), the ripple transfer function RTF (g-i) and the conditioned response histogram CRH (j-l) are displayed ($\Theta = 3\sigma$, p=0.001).

In the example of Neuron 2, the CRH is very diffuse and shows no preference, however, the RTF does show bandpass tuning in the temporal domain and lowpass tuning in the spectral domain. These are the two cases, for which mostly significant STCs were encountered: good overlap of RTF and CRH, or no good overlap due to a diffusive CRH.

For neurons which yielded *no* significant STC, the RTF and CRH usually showed no overlap, either they showed a complementary behavior with no overlap in the parameter space, or the CRH was diffusive (Neuron 3).

6.3.3 Comparison of significant linear STRF and STC

In Sec. 6.3.1, it was shown that dominant frequency regions are very similar for the linear STRF and the STC. To test if this is also true in the temporal domain, the linear STRF is compared to the temporal representation of the STC in each frequency channel, the STRF-like representation (see Sec. 6.2.6). Examples of this comparison are given for two neurons, Fig. 6.3.6. For Neuron 1 the linear STRF displays a broad frequency bandwidth and an excitatory-inhibitory temporal energy pattern with a delay time of about 10 ms (Fig. 6.3.6a). The bandwidth of the STC is smaller and contained within the one of the linear STRF. In the STRF-like representation, several significant vectors around the BF display energy with an excitatory-inhibitory-excitatory pattern also centered at a delay of 10 ms. Thus, the tuning is opposite in these two receptive fields, and the nonlinear contribution from the STC could be modulatory as proposed previously for auditory neurons in the midbrain of songbirds [67].
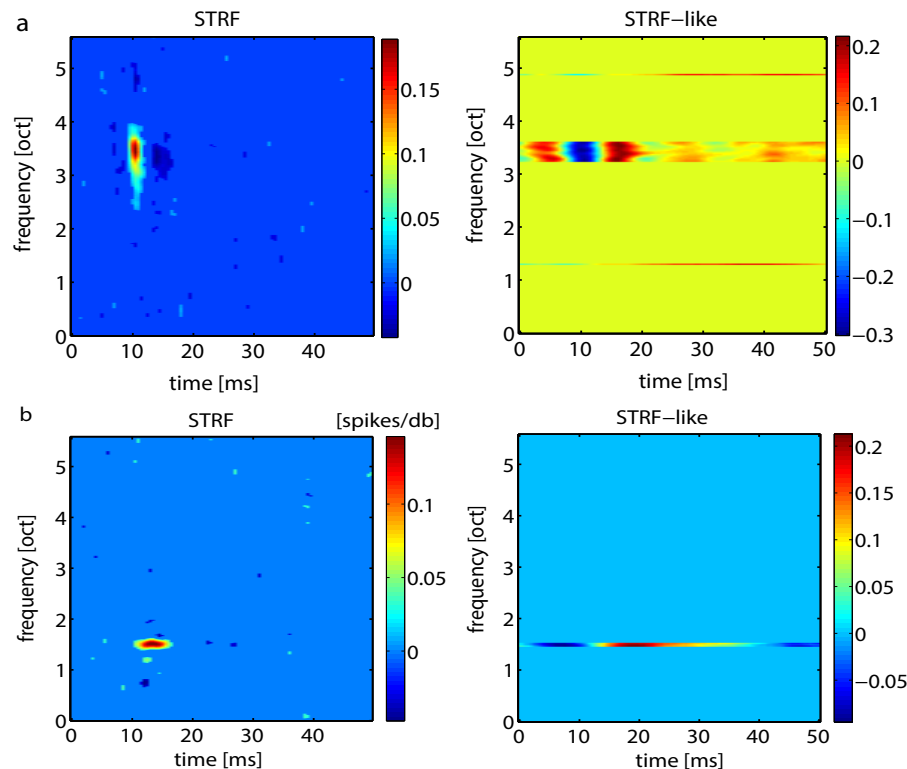

Figure 6.3.6: Comparison of significant linear STRF and STC. The significant linear STRF is compared to the first significant vectors of each frequency channel ('STRF-like' representation), for a) Neuron 1; and b) Neuron 2 ($\Theta = 3\sigma$, p=0.001).
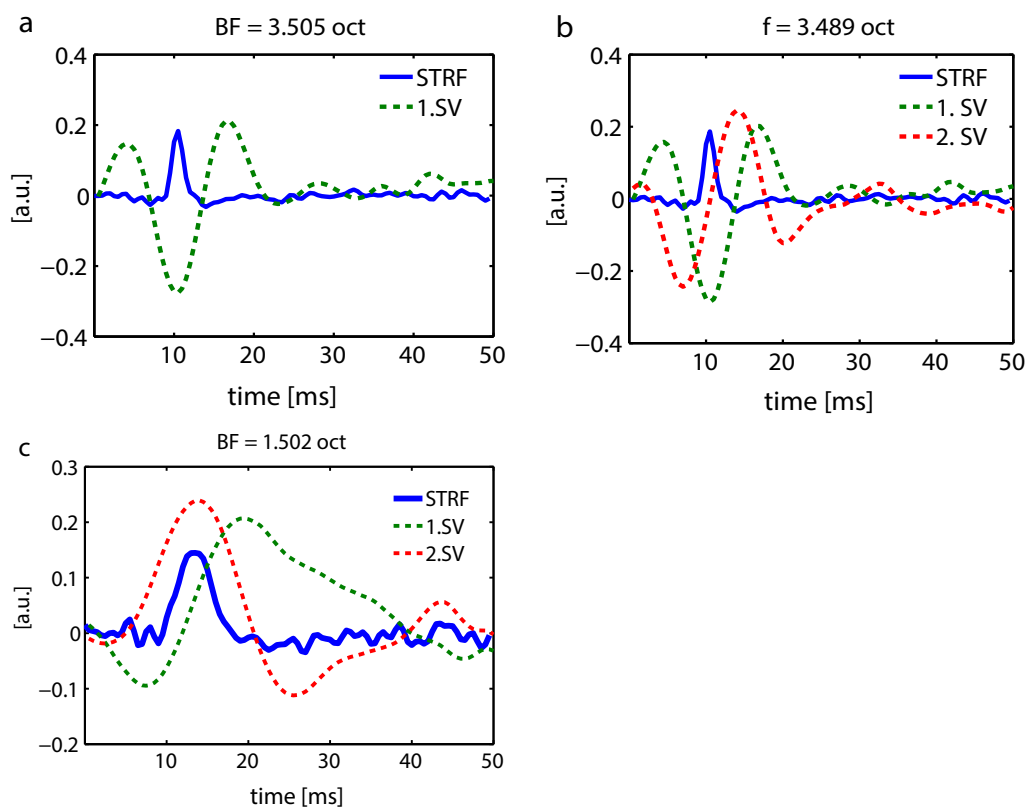
Figure 6.3.7: Comparison of temporal course of significant linear STRF and first (second) singular vectors for the two neurons displayed in Fig. 6.3.6. a) Neuron 1: temporal courses of linear STRF and first significant vector at the best frequency (no second significant singular vector is present); b) Neuron 1: temporal courses of linear STRF and first and second significant singular vector at a frequency channel near the best frequency; c) Neuron 2: temporal courses of linear STRF and first and second significant singular vector at best frequency. Neuron 2 only displays significant values at the best frequency.

In the second example only one significant vector is present and it is located at the best frequency (Fig. 6.3.6b). Neuron 2 shows an excitatory region at about 12.5 ms delay. The STC displays an inhibitory-excitatory temporal pattern with an excitatory region at a delay time of 20 ms. Thus, also in this case, the temporal modulations do not coincide, but are shifted. The linear STRF and STRF-like representation of the STC show enhanced (excitatory and suppressive) energy in similar spectral and temporal regions. Significant singular vectors and the temporal courses of the corresponding frequency channels in the linear STRF are shown in Fig. 6.3.7. The opposing tuning of the linear STRF temporal course at the best frequency and of the first significant singular vector is shown in Fig. 6.3.7a. For a neighboring frequency the tuning is very similar. The second significant singular vector is shifted by a few milliseconds relative to the first one. In the example of Neuron 2, the second vector coincides temporally with the excitatory region of the linear STRF.

The relationship between the delay time determined from the STRF and from the STC of one neuron is displayed for all neurons in Fig. 6.3.8a. The delay times are determined from the linear STRF and from the delay time of the first extremum, either minimum or maximum value of the first significant singular vector at the best frequency. For 47 % of the neurons, these two delays match (difference < 1.1 ms), for both, minima and maxima. Hence, the tuning of the singular vector is the same or opposite to the linear STRF. For almost all other neurons, the delay of the STC was higher, with differences ranging mostly between 5 and 30 ms. This was true for STCs with same and opposite tuning. For a small percentage of the neurons (10 %), delay times of the STC were about 5 ms shorter than the ones found with the linear STRF. To further investigate temporal tuning, the width between extrema in the temporal course of the singular vectors was compared to the best temporal modulation frequency of the neuron, Fig. 6.3.8b. A tendency for larger widths corresponding to low best temporal modulations frequencies (TMF), and vice versa is visible.

Previous studies have suggested that nonlinear receptive fields might contribute to an increase of the neuron's feature selectivity [56]. The highest significant value at the best frequency was taken as a measure to quantify the nonlinear processing of a neuron. This measure was compared to the degree of phase-locking (PLI), see Fig. 6.3.9a, and to the feature-selectivity index (FSI), see Fig. 6.3.9b. Between the phase-locking index and the degree of nonlinear responses no clear relationship is perceivable. A tendency of higher singular values for low feature selectivity is visible.
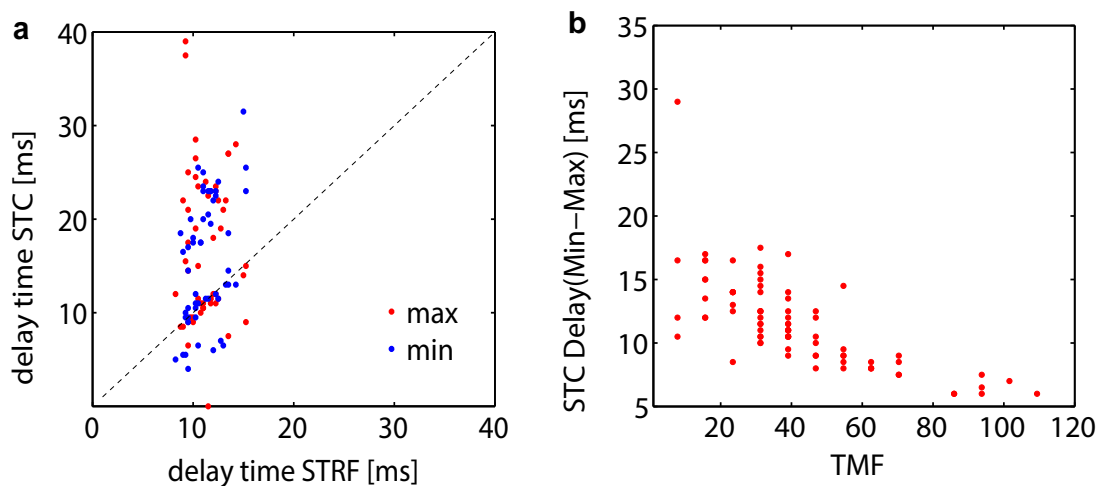

Figure 6.3.8: Temporal tuning of linear STRF and STC across all neurons. a) delay times obtained from the STC and linear STRF; b) width in tuning of STC and best temporal modulation frequency (TMF).
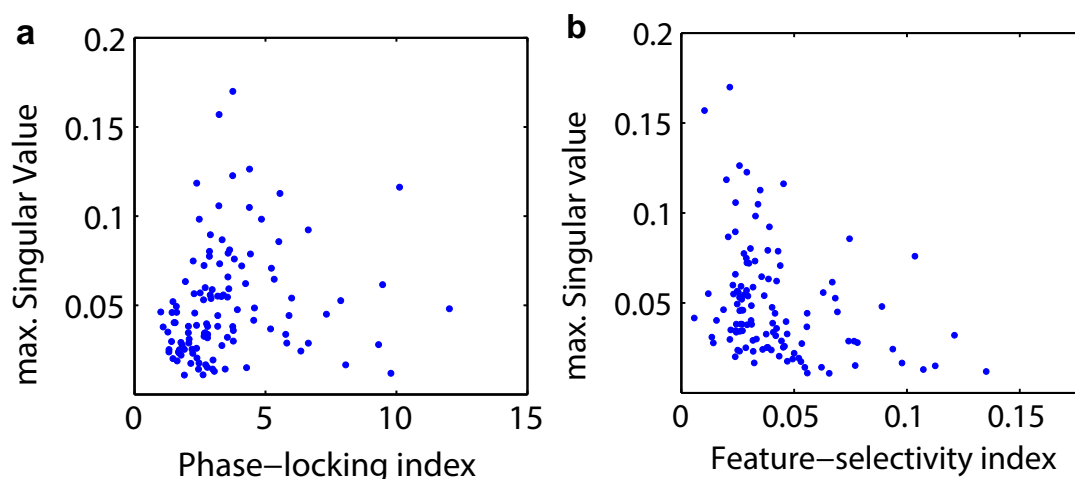
Figure 6.3.9: Relationship between selectivity and degree of nonlinearity. For all neurons the a) Phase-locking index (PLI) against the highest singular value, and; b) the feature-selectivity index (FSI) against the highest singular value are shown.

6.4 Discussion

In this chapter, using spike-triggered covariance of the envelope of DMR sound, it was shown that an important amount of neurons in the ICC display significant preferences to stimulus correlations (75 % / 57 %), respectively for significance thresholds of $\Theta = 3\sigma$ and $\Theta = 1.6\sigma$. These percentages were determined by attributing significance to a nonlinearity if at least one frequency channel displayed a significant STC. However, these neurons also showed significant linear STRFs. Thus the percentage cannot be directly compared to a previous study which found ~ 40 % of the neurons to be nonlinear [14], because in that work linear neurons were assessed if they responded in the same way to (ripple) noise and to the spectrotemporally varying DMR sound.

The stimulus correlations vary across neurons and across frequency channels. The shapes of the STCs seem to be as diverse and rich as the linear STRFs. A pronounced nonlinearity, present in more than half of the neurons, was found for stimulus correlations with maximum time delays of about 30 ms. Dominant nonlinear response properties are almost exclusively present at the best frequency of the neuron, and comparing the significant linear STRF and the first significant vector of the covariance reveals a large spectral overlap, but also reveals shifted temporal courses. About half of the neurons displayed the same tuning pattern and the other half displayed opposite tuning pattern. Same tuning could enhance the response and the opposite contribute to further tuning. The nonlinear contribution from the STC could be modulatory as proposed previously for auditory neurons in the midbrain of songbirds [67].

The degree of nonlinear responses, which was assessed with the magnitude of significant values, did not show a clear dependence on the phase-locking or feature-selectivity index, which might indicate that the singular value magnitude is not the appropriate measure for quantifying the degree of nonlinearity. Furthermore, nonlinear response properties were not linked to the match of the RTF and CRH, or to any particular type of linear STRF. Thus, the found nonlinearities do not seem to be linked to a specific type of neuron but to be inherent to ICC neurons with very diverse filtering characteristics. Using the DMR sound amplitude spectrum for the analysis, it could be shown that dominant nonlinearities were found at the best frequency of the neuron. Previous studies demonstrated a good match between the averaged linear STRF for the whole waveform and the highest significant singular vector of the STC [82]. In this work, not only similar temporal courses of the linear STRF and STC were found, but also opposing tuning. Lying within the STRF, the linear receptive field, these nonlinear responses may form a processing unit and have modulatory effects as suggested for neurons in the songbird [67]. In contrast to findings in the visual system for which the complex receptive field displaying nonlinear summation properties lies outside the classical receptive field [119, 65, 66], in this work it was shown, that at least in the spectral dimensions these two receptive fields coincide. The presented findings might differ from neural processing in awake animals, as the used recordings were taken from animals that were put under anesthesia, which was shown to have an effect on the neural activity [118]. The anesthesia might change inhibitory mechanisms which would affect the nature of the STRFs and the nonlinearities [85]. The used recording time of 20 minutes is relatively long. However, for the reverse correlation analysis more recording time would be desirable. The obtained receptive fields, the STRF and the STC, are stimulus-dependent [97, 14].

The DMR sound, despite its many advantages also poses some problems, e.g. it might not capture all features which are displayed by natural sound. Even though the neurons responded to the complete DMR waveform, the reverse correlation analysis considered only the stimulus envelope, and the fine structure might contain further crucial information. In the cortex, neuronal sensitivity to the fine structure has been shown and such a sensitivity could be present in the ICC and might be more pronounced since temporal resolution is higher in the ICC than in the cortex [85]. Therefore, the advantage of explicitly studying the envelope representation, compromises that the complete processing of the neurons might not be captured. The DMR was presented at one intensity, and even though it covers a relatively large dynamic range of 40 dB, responses to DMR might change for other intensities and yield different STRFs and STCs [124, 85].

In this work, nonlinear response properties have been analyzed for each frequency carrier. A next step in this analysis would be to further investigate preferences to stimulus correlations from different frequency carriers. This could be achieved with the spike-triggered cross-covariance of two *different* frequency carriers. This analysis would be important, since earlier studies have indicated nonlinear interactions between different sound frequencies.

Discussion

Pienkowski and Eggermont [125] showed that intra- and inter frequency interaction in the responses of neurons in the cats' primary auditory cortex exist. In their analysis they used trains of tone pips, which lack the natural sound attributes of frequency and amplitude modulations that are covered by the DMR sound. The found nonlinearities could be already present in the input to the ICC or they could originate from nonlinear processing of ICC neurons. Nonlinearities are already present in the neural responses originating in the cochlea. But how these nonlinearities are represented in the ICC is not clear yet. One could hypothesize that if the nonlinearities come from different stations preceding the ICC, then the nonlinearities would show no correlation with the linear response. Both would be measured as the neural response. If, on the other hand, the nonlinearity results from the neuron's own processing, very probably the linear and nonlinear part would display similarities, which was shown for about half of the neurons. Thus, an important part of the nonlinearities do originate in the ICC and further study of these is likely to reveal aspects of natural sound processing.

7 Discussion

7.1 Summary and Contributions

In this thesis, I investigated the representation and encoding of complex sounds by neurons in the central auditory system. Using cross-correlation, neural discrimination and reverse correlation analysis on a set of 1145 multi- and 150 single-units, the following main results were found.

The representation of behaviorally relevant vocalizations in the central inferior colliculus (ICC) of the guinea pig has been examined for groups of single neurons (Chap. 4, 5). These multi-units display correlations of their spiking responses which decrease almost linearly with spatial distance (Chap. 4). The correlations are of comparable strength along the two major spatial axes of the ICC: along the best frequency gradient and across the best modulation frequency gradient. A significant amount of the response correlations can be attributed to neural interactions between the multi-units (neural correlation).

Using neural discrimination I showed that these neural correlations do not facilitate encoding of the vocalizations (Chap. 5). Total discrimination between the 11 vocalizations based on individual multi-unit spiking responses is high (>75%) and is near-to-perfect when combining 3-5 multi-units. Individual vocalizations can be well discriminated against other vocalizations across a wide frequency range of the tonotopy. Furthermore, in the here presented work the encoding of complex sounds by neurons in the ICC was investigated with respect to neural preferences to specific acoustic properties (Chap. 6). Preferences of single neurons to temporal stimulus envelope correlations, which are also present in vocalizations, were analyzed for spiking responses to artificial dynamic moving ripple sound using reverse correlation. It could be demonstrated that the majority of ICC neurons (>57 %) show an increased spiking response to envelope correlations with a maximum delay of about 10-30 ms. These nonlinear responses are most pronounced for carrier frequencies of the stimulus envelope that correspond to the best frequency of the neuron. At the neuron's best frequency, for which linear and nonlinear responses are most pronounced, the temporal tuning of the nonlinear preferences is either opposite or similiar to that of the linear receptive field. The nonlinear response properties are present for neurons across the ICC, and no relation to other specific tuning properties such as feature selectivity or phase-locking was observed.

7.2 Critical Review and Future Work

Multi-unit neuron clusters were shown to reliably encode vocalizations across the ICC. Still, recordings of multi-units would be desirable, that allow spike-sorting and the recuperation of single-unit responses and thus the knowledge of the exact number of neurons contained within one cluster. This would allow making inferences not only about collective neural behavior but also simultaneously analyze individual single neuron behavior within the cluster. Single-unit recordings would allow to test for call-selectivity to individual vocalizations. They would also allow for a fairer comparison of the individual vocalization's spectral content and its best discrimination in dependence of the neuron's best frequency (see Sec. 5.4.1). Furthermore, single-unit recordings would allow to reliably quantify the amount of correlations due to neuronal interactions, and possibly infer structural connectivities between neurons, which is not possible for multi-units [116].

In this work, the number of neurons contained within one cluster can be roughly estimated, however, an exact number would be desirable in order to attribute the findings to a neural group of well-defined size. A single-neuron resolution within the multi-unit cluster would allow answering the question if a multi-unit responds differently to different vocalizations because it either recruits different groups of neurons for each vocalization, or, because the spike times of the multi-unit and thus of the single neurons are different for each vocalization; the same holds for averaged spike-rates. Thus, the findings on the representation of vocalization are valid on the level of multi-units without assuming that the same neurons within the multi-unit respond to different vocalizations. Multi-unit recordings are likely to display overlapping action potentials, thus distorting the waveforms which are thresholded and thus corrupting the estimate of the the spike timing and also of the averaged spike-rate. The effects are enlarged when combining spike timing responses from several multi-unit clusters, which was performed for the neural discrimination of pooled multi-units.

For the correlation and neural discrimination, the exact position of the recording electrodes, deducible from histological analysis, would have been desirable as it enables to map the vocalization encoding to specific locations within the ICC. Finally, longer recording times, thus more response trials to the vocalizations or even white noise would have allowed computing the spike-triggered covariances for the multi-units (with an added whitening procedure [106] for natural stimuli), and to compare these to the single-unit spike-triggered covariance.

Presenting the exact same version of the vocalization several times bears a bias due to pseudoreplication, and could be mitigated by presenting several exemplars of the same stimulus. Another distortion in the neural representation, which might hamper and corrupt inferences that are obtained from direct comparison of the stimulus and the neural response, result from the presentation of the calibrated vocalizations. First, the ear canal mold of *one* guinea pig was used to calibrate the vocalizations for all other guinea pigs, even though the specific shape and filtering of ear canals is unique to each individual.

Furthermore, the sound which is delivered, was not filtered by the individual pinna, and hence is distorted compared to vocalizations, which are usually filtered by the pinna and perceived by the guinea pig. Second, the calibrated vocalizations which impinged on the ear drum were not recorded. These vocalizations, however, should have been used for comparison with the neural responses. As was displayed in Chap. 4, the spectral contents and onsets do coincide in most cases with the original vocalizations. But, it is not clear if these vocalizations match the original vocalizations perfectly (Fig. 3.1.1). The calibrated and original vocalizations do show good overlap for seven cases, but also huge variations in onset and spectral content for others (see Sec. 3.1.1). It is not clear how the vocalizations impinging on the ear drum were altered. In order to obtain the best approximation of an input-output function, it is desirable to have a well-defined input and a clear, as noiseless as possible, output.

In Set-up 2, dynamic moving ripple (DMR) sound was used to probe neuronal nonlinear response properties (Chap. 6). Despite its many advantages, the DMR envelope may lack features that are present in the natural, behaviorally relevant full sound waveform. The DMR can be used for the reverse correlation analysis without further required whitening, however, short-term correlations exist and might bias the obtained nonlinear receptive fields. In general, receptive fields are stimulus-dependent [97, 14].

In the natural case, the same sound is perceived by both ears, and information from both ears influences neural responses in the ICC of either hemisphere, with a dominant part of the contralateral ear. Thus, in the case of independent DMR sound presented to both ears, ignoring the effect and sound information from the ipsilateral ear, responses recorded from one ICC might be altered with respect to presentation of the same sound to both ears. In Set-up 1, on the other hand, vocalizations were presented monaurally. Although, here, the influence from a dissimilar sound of the ipsilateral ear is absent, the effects of the ipsilateral contribution of the same sound are also missing, and thus not the full picture of ICC-processing is captured.

The recordings used in this work were taken from anesthetized animals. Although obvious advantages exist when using anesthesia, e.g. elimination of movement or the role of attention, differences for recordings in awake and anesthetized guinea pigs exist [118]. For example, it has been shown that anesthesia can suppress the neural response to a particular vocalization but may also enhance it. Since anesthesia has non-negligible effects on the neural activity [118], neural discrimination, especially for units that did not yield perfect classification when pooling, is likely to improve in awake animals. The anesthesia might change inhibitory mechanisms which would affect the nature of the receptive fields and the nonlinearities [85]. Correlation values and the obtained receptive fields are likely to be different in recordings from awake animals. Neuronal encoding strategies found in the the ICC of mammalian animals are considered to be similar to those found in the human ICC. This, however, is an idealization, and although findings give indications for processing in the human ICC, its detailed fundamental neuronal mechanisms cannot be ascertained.

Studies aiming to improve perception of hearing impaired humans can yield insight into some aspects of human hearing, such as the necessary spectral and temporal resolution of electrical stimulation in the ICC in order to ensure speech recognition.

This work was originally inspired by the attempt to improve the auditory midbrain implant. Although it developed to be a fundamental auditory neuroscience project, the finding that responses from groups of multi-unit clusters reliably encode speech-like sounds, suggests that stimulating a few groups of clusters might at least provide an optimistic starting point to test speech perception of implantees. Still, the silent underlying assumption that findings obtained from recorded, already processed activity of the ICC, can be used to effectively stimulate electrically, hence to simulate input to the ICC, is a coarse approximation. However, a second option in which the goal could be to simulate the input separately for each of the various ascending processing pathways, and electrically stimulate the ICC precisely at the terminating locations of these pathways, does not seem surgically feasible, since insertion of the stimulating electrode into the ICC poses challenges [17].

In a proposed experiment, electrode arrays (double-shank and double-tetrode) with high impedances ($\approx 2\,\text{M}\Omega$) are used to record from the ICC of guinea pigs. The higher impedance would allow obtaining multi-unit activity from which single neuron responses can be isolated, and to investigate the interplay of single neurons to produce this compound activity. It is important to examine if the presented findings for multi-units are valid at the single cell level. Previous work stated that for multi-units with up to five single neurons, no statistically significant differences exist in discrimination performance, optimal temporal resolution, and integration time to single neuron responses in songbirds [126]. However, this does not need to be true for the presented multi-units in the mammalian ICC that possibly comprise more than five single neurons.

In a previous study, correlations of neuron clusters in the auditory cortex were investigated by multi-unit activity, which was spike-sorted and then added again to eliminate contributions from thalamocortical afferents or fast spikes from interneurons [47].

The acoustic stimuli employed for the proposed experiment are vocalizations, spectrotemporally modified vocalizations, dynamic moving ripple sound, and white noise. These are presented binaurally via loudspeakers to the pinna of the guinea pigs. In this way, filtering of the pinna and ear canal of the individual animal takes place and elicits responses, which, at least in the case of vocalizations, are usually perceived by the individual. Recordings from the ICC for at least 10 min are taken, in order to ensure enough collected data for the proposed analysis. From the single neuron responses to the DMR stimuli, the spectrotemporal receptive field (STRF) and the the spike-triggered covariance (STC), as described in Chap. 6, are computed. Single neuron's responsiveness to vocalizations is directly compared to preferences found with the STRF and the STC. Previous studies in the auditory midbrain of songbirds compared the STRF to responses to natural sound but did not take into account the STC [38]. The spike-triggered covariance analysis would be extended to inter-frequency stimulus correlations.

The question whether preferences to stimulus correlations of the vocalization can be directly linked to a faithful representation of the individual vocalization by the single neuron, as ascertained by neural discrimination, is investigated. A second goal of this analysis would be to investigate the representation and encoding of variability, more specifically the representation of several exemplars of the same vocalization. Are preferences in the receptive fields present for all exemplars of the vocalization and how much do the averaged responses (PSTH) differ? How well can different exemplars of the same stimulus be discriminated based on their neural responses, and is discrimination best in the main frequency content of the vocalization or in minor spectral contents which differ for the different exemplars. Thus, the analysis performed in Sec. 5.4.1 could be performed for different exemplars of the same vocalizations. However, it would also be very interesting to apply this analysis to single neuron recordings, as it might point to selectivity of single neurons to individual vocalizations and allow for a more detailed comparison of spectral content and best discrimination. Ideally, a very large data set is collected with several insertion locations, in order to assure broad mapping of ICC activity. Histology analysis of the insertion locations enables to assign the obtained receptive fields and local circuits to specific ICC locations, in particular to a detailed location within isofrequency laminae. The animals proposed for this experiment are guinea pigs, because the vocalizations of these mammals are spectrotemporally very rich. Experiments would have to be conducted under anesthesia, since recordings in awake guinea pigs might be hard to realize. Urethane which was shown to display minor effects on the neural activity [127] could be a potential anesthesia. Mice show a variety of vocalizations, but these are shifted to an ultra-sound frequency range [128]. Alternatively, recordings could be done from songbirds (e.g. zebra finches), which display spectrotemporally rich calls [128], but are not mammals, thus inferences to humans are harder to draw. Recordings in awake animals are well feasible in rabbits or cats, which, however, do not show a rich set of vocal communication sounds.

The neural representation and encoding of complex sounds in the inferior colliculus was investigated in this thesis, on large sets of neurons. Vocalizations are represented broadly across the ICC by clusters of neurons. The neural responses to the vocalizations, resulting from the interplay of the input to and the compound receptive field of the neuron cluster, decrease in similarity with spatial distance. It was shown, that despite diverse inputs to the ICC from different ascending projections, terminating on different parts in the ICC [109], and existing response heterogeneities [21], representations of vocalizations are gradually decreasing in similarity across the best frequency and modulation gradients. On the level of multi-units, these changes in temporal response between pairs of multi-units are gradual. A clear gradient of decreasing response similarity, which can be exponentially fitted, for the representation of individual vocalizations was found.

Thus, despite the vocalizations displaying very diverse and inhomogeneous spectral contents, and the ICC having a clear logarithmic gradient in best frequency, the vocalizations exhibit a gradual linear response similarity across the ICC. However, finer differences as perceivable by single neurons might not be captured at this resolution level. The reliable encoding of speech-like sounds across the ICC, and spatial and spectral distances still yielding a high response similarity, are not only substantial to better understand processing of this nucleus, but could be important for improving medical devices. For the auditory midbrain prosthesis, these distances might give indications for the spatial separation of stimulating electrodes.

The multi-units faithfully encode the spectrotemporally rich set of behaviorally relevant vocalizations. They allow discriminating the 11 vocalizations based on their neural responses across a wide range of the tonotopy, despite overlapping spectral contents of the vocalizations. High response similarities of spiking response are confined to distances below 300 μm, and thus spiking responses differ substantially for larger distances. It was found that multi-unit clusters display mostly independent encoding, i.e. temporal correlations and noise correlations do not contribute to a more detailed encoding of the vocalizations. A similar suggestion which was made for single neurons in inferior temporal cortical neurons could hold for multi-unit clusters in the ICC: in order to avoid redundant information, connection patterns and weights are favored which keep the responses of adjacent neuron (neural groups) as independent as possible, both, for signal and noise, namely that local circuits are organized with the constraint to avoid correlated signals [129].

7.3 Conclusion

In this work it has been shown that response correlations in the mammalian ICC decrease with increasing spatial distance of the multi-units. Thus, the different inputs and gradients of response preferences across the ICC result in a gradual decrease of response correlations with distance. This decrease follows an exponential function, which has also been found for the primary auditory cortex [47]. However, with a steeper decrease which indicates smaller regions of neurons with similar responses than in the auditory cortex. It was shown that neural correlations in the mammalian ICC are substantial, thus, interactions between groups of neurons exist but they do not contribute to a more detailed encoding of vocalizations. The responses from 3-5 multi-units are sufficient to almost perfectly discriminate 11 vocalizations, but discrimination is not improved if further responses are added. Different tuning will add independent information to the discrimination, but many times adjacent multi-units will have similar tuning, and it seems puzzling that the small existing differences in the units' responses are important enough to contribute new information to the discrimination. This result, however, is supported by work of Holmstrom and colleagues who found that neural responses in the mammalian ICC are heterogeneous, and that this heterogeneity appears to be enough for an efficient encoding of vocalizations [21].

The representation of vocalizations does not follow a purely linear mapping according to the stimulus' spectral content, but deviations exist. These could be due to nonlinear response mechanisms and this needs to be investigated in future studies. It would be interesting to investigate such a mapping in detail at the single neuron level. The found nonlinear response properties agree with previous work [14], but this is the first study to investigate nonlinear response properties independently for different frequency carriers. In conclusion, multi-units provide enough information to reliably distinguish complex sounds in broad best frequency ranges. Thus, a robust representation of vocalization against smaller groups or deficient neurons exists, which might be important in the behavioral context. Multi-units in the central inferior colliculus act as independent encoders in the sense that correlations with other neurons do not facilitate encoding, as demonstrated for single retinal ganglion cells [33], and redundancy is minimized as suggested for central systems [32]. The nonlinear response properties seem to be inherent to neurons across the ICC, and also across the best frequency range.

Bibliography

[1] L.M. Aitkin and S.C. Phillips. Is the inferior colliculus an obligatory relay in the cat auditory system? *Neuroscience Letters*, 44:259–64, 1984b.

[2] M.S. Malmierca, M.A. Merchan, C.K. Henkel, and D.L. Oliver. Direct projections from cochlear nucleus to auditory thalamus in the rat. *The Journal of Neuroscience*, 22:10891–10897, 2002.

[3] J.E. Rose, D.D. Greeenwood, and J.M. Goldberg. Some discharge characteristics of single neurons in the inferior colliculus of the cat. I. Tonotopical organization, relation of spike-counts to intensity, and firing patterns of single neurons. *Journal of Neurophysiology*, 26:294–320, 1963.

[4] M.M. Merzenich and M.D. Reid. Representation of the cochlea in the inferior colliculus of the cat. *Brain Research*, 77(3):397–415, 1974.

[5] C.E. Schreiner and G. Langner. Laminar fine structure of frequency organization in auditory midbrain. *Nature*, 388:383–6, 1997.

[6] C.E. Schreiner and G. Langner. Periodicity Coding in the Inferior Colliculus of the Cat. II. Topographical organization. *Journal of Neurophysiology*, 60(6):1823–1840, 1988.

[7] G. Langner, M. Albert, and T. Briede. Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger). *Hearing Research*, 168(1-2):110–30, 2002.

[8] S. Baumann, T.D. Griffiths, L. Sun, C.I. Petkov, A. Thiele, and A. Rees. Orthogonal representation of sound dimensions in the primate midbrain. *Europe PMC Funders Group*, 14(4):423–425, 2011.

[9] E. Covey. Inputs to the Inferior Colliculus. In G.M. Shepherd G. Westheimer T.D. Albright R.H. Masland P. Dallos D. Oertel S. Firestein G.K. Beauchamp M.C. Bushnell J.H. Kaas E. Gardner A.I. Basbaum, A. Kaneko, editor, *The Senses: A Comprehensive Reference*, volume 3, pages 733–50. Academic Press, San Diego, 2008.

[10] M.S. Malmierca. The Inferior Colliculus: A Center for Convergence of Ascending and Descending Auditory Information. *Neuroembryology and Aging*, 3(4):215–229, 2004.

BIBLIOGRAPHY

[11] D.R.F. Irvine. *Physiology of the auditory brainstem. In: Handbook of Auditory Research, The Mammalian Auditory Pathway: Neurophysiology.* Springer, 1992.

[12] P.X. Joris, C.E. Schreiner, and A. Rees. Neural processing of amplitude-modulated sounds. *Physiological Reviews*, 84:541–577, 2004.

[13] S.S. Kety. Regional neurochemistry and its application to brain function. *Bulletin of the New York Academy of Medicine*, 38(12):799–812, 1962.

[14] M.A. Escabí and C.E. Schreiner. Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. *The Journal of Neuroscience*, 22(10):4114–4131, 2002.

[15] F. Rieke, D.A. Bodnar, and W. Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1365):259–265, 1995.

[16] T. Lenarz, H.H. Lim, G. Reuter, J.F. Patrick, and M. Lenarz. The auditory midbrain implant: A new auditory prosthesis for neural deafness-concept and device description. *Otol. Neurotol.*, 6:838–843, 2006.

[17] H.H. Lim, T. Lenarz, D.J. Anderson, and M. Lenarz. The auditory midbrain implant: Effects of electrode location. *Hearing Research*, 242:74–85, 2008.

[18] J. Grace, N. Amin, N.C. Singh, and F.E. Theunissen. Selectivity for conspecific song in the zebra finch auditory forebrain. *Journal of Neurophysiology*, 89(1):472–87, 2003.

[19] J.W.H. Schnupp, T.M. Hall, R.F. Kokelaar, and B. Ahmed. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *The Journal of Neuroscience*, 26(18):4785–4795, 2006.

[20] R.V. Shannon, F. Zeng, J. Kamath, V.and Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. *Science*, 270(5234):303–4, 1995.

[21] L.A. Holmstrom, L.B.M. Eeuwes, P.D. Roberts, and C.V. Portfors. Efficient encoding of vocalizations in the auditory midbrain. *The Journal of Neuroscience*, 30(3):802–819, 2010.

[22] C.V. Portfors, Z.M. Mayko, K. Jonson, G.F. Cha, and P.D. Roberts. Spatial organization of receptive fields in the auditory midbrain of awake mouse. *Neuroscience*, 193:429–439, 2011.

[23] D. Suta, E. Kvasnák, J. Popelár, and J. Syka. Representation of species-specific vocalizations in the inferior colliculus of the guinea pig. *Journal of Neurophysiology*, 90:3794–3808, 2003.

[24] D. McAlpine. Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *Journal of Neurophysiology*, 92(3):1295–311, 2004.

[25] M.A. Escabí, R. Nassiri, L.M. Miller, C.E. Schreiner, and H.L. Read. The Contribution of Spike Threshold to Acoustic Feature Selectivity, Spike Information Content, and Information Throughput. *The Journal of Neuroscience*, 41(25):9524–9534, 2005.

[26] C.V. Portfors, P.D. Roberts, and K. Jonson. Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. *Neuroscience*, 162(2):486–500, 2009.

[27] D.M. Schneider and S.M.N. Woolley. Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *Journal of Neurophysiology*, 103:3248–65, 2010.

[28] L. Wang, R. Narayan, G. Graña, M. Shamir, and K. Sen. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *The Journal of Neuroscience*, 27(3):582–9, 2007.

[29] C.K. Machens, H. Schütze, A. Franz, O. Kolesnikova, M.B. Stemmler, B. Ronacher, and A.V.M. Herz. Single auditory neurons rapidly discriminate conspecific communication signals. *Nature Neuroscience*, 6(4):341–2, 2003.

[30] C.T. Engineer, C. Perez, Y. H. Chen, R.S. Carraway, A.C. Reed, J. Shetake, V. Jakkamsetti, K.Q. Chang, and M.P. Kilgard. Cortical activity patterns predict speech discrimination ability. *Nature Neuroscience*, 11(5):603–608, 2008.

[31] H. Barlow. Possible principles underlying the transformation of sensory messages. In W.A. Rosenblith, editor, *Sensory Communication.*, pages 217–34. MIT Press, Cambridge, MA, 1961.

[32] H. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1(4):371–94, 1972.

[33] S. Nirenberg, S.M. Carcieri, A.L. Jacobs, and P.E. Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411:698–701, 2001.

[34] J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, and E.P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–9, 2008.

[35] L.F. Abbott and P. Dayan. The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11:91–101, 1999.

[36] A.S. Ecker, P. Berens, A.S. Tolias, and M. Bethge. The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31(40):14272–83, 2011.

[37] B.B. Averbeck, P.E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–66, 2006.

[38] S.M.N. Woolley, T.E. Fremouw, A. Hsu, and F.E. Theunissen. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, 8(10):1371–79, 2005.

[39] S. Andoni and G.D. Pollak. Selectivity for spectral motion as a neural computation for encoding natural communication signals in bat inferior colliculus. *The Journal of Neuroscience*, 31(46):16529–40, 2011.

[40] J. Winer and C. Schreiner, editors. *The Inferior Colliculus*. Springer, New York, 2005.

[41] T. Ito and D.L. Oliver. Origins of Glutamatergic Terminals in the Inferior Colliculus Identified by Retrograde Transport and Expression of VGLUT1 and VGLUT2 Genes. *Frontiers in Neuroanatomy*, 4:135, 2010.

[42] L.-F. Liu, A.R. Palmer, and M.N. Wallace. Phase-locked responses to pure tones in the inferior colliculus. *Journal of Neurophysiology*, 95(3):1926–1935, 2006.

[43] D. McAlpine, D. Jiang, and A.R. Palmer. A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, 4(4):396–401, 2001.

[44] C.E. Carr and M. Konishi. A circuit for detection of interaural time differences in the brain stem of the barn owl. *The Journal of Neuroscience*, 10(10):3227–46, 1990.

[45] C. Chen, F.C. Rodriguez, H.L. Read, and M.A. Escabí. Spectrotemporal sound preferences of neighboring inferior colliculus neurons: implications for local circuitry and processing. *Frontiers in Neural Circuits*, 6(62), 2012.

[46] W.J. Epping and J.J. Eggermont. Coherent neural activity in the auditory midbrain of the grassfrog. *Journal of Neurophysiology*, 57(5):1464–83, 1987.

[47] J.J. Eggermont. Properties of correlated neural activity clusters in cat auditory cortex resemble those of neural assemblies. *Journal of Neurophysiology*, 96(2):746–64, 2006.

[48] S. Wohlgemuth and B. Ronacher. Auditory discrimination of amplitude modulations based on metric distances of spike trains. *Journal of Neurophysiology*, 97(4):3082–92, 2007.

[49] J. Schnupp, I. Nelken, and A. King. *Auditory Neuroscience: Making Sense of Sound.* MIT Press, Cambridge, MA, 2011.

[50] D. Suta, J. Popelár, and J. Syka. Coding of communication calls in the subcortical and cortical structures of the auditory system. *Physiological Research*, 57:149–159, 2008.

[51] A. Calabrese, J.W. Schumacher, D.M. Schneider, L. Paninski, and S.M.N. Woolley. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PloS One*, 6(1):e16104, 2011.

[52] M. Weliky, J. Fiser, R. H. Hunt, and D.N. Wagner. Coding of natural scenes in primary visual cortex. *Neuron*, 37(4):703–718, 2003.

[53] R.A. Da Silveira and M.J. Berry II. High-fidelity coding with correlated neurons. *arXiv:1307.3591v3 [q-bio.NC]*, pages 1–48, 2013.

[54] S. Schreiber, J.M. Fellous, D. Whitmer, P. Tiesinga, and T.J. Sejnowski. A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52–54:925–931, 2003.

[55] T.O. Sharpee, C.A. Atencio, and C.E. Schreiner. Hierarchical representations in the auditory cortex. *Current Opinion in Neurobiology*, 21(5):761–7, 2011.

[56] W.E. Vinje and J.L. Gallant. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *The Journal of Neuroscience*, 22:2904–2915, 2002.

[57] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160:106–154, 1962.

[58] B.Y.J.A. Movshon, I.D. Thompson, and D.J. Tolhurst. Spatial summation of receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology*, 283:53–77, 1978.

[59] G.C. DeAngelis, I. Ohzawa, and R.D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69(4):1118–35, 1993.

[60] J.A. Movshon, I.D. Thompson, and D.J. Tolhurst. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of Physiology*, 283:79–99, 1978.

[61] B.Y.A.F Dean and D.J. Tolhurst. On the distinctness of simple and complex cells in the visual cortex of the cat. *The Journal of Physiology*, 344:305–325, 1983.

[62] R.C. Emerson, M.C. Citron, W.J. Vaughn, and S.A. Klein. Nonlinear directionally selective subunits in complex cells of cat striate cortex. *Journal of Neurophysiology*, 58:33–65, 1987.

[63] D.H. Wiesel and T.N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195:215–243, 1968.

[64] H. Schulze and G. Langner. Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: evidence for wide spectral integration. *Journal of Comparative Physiology*, 185:493–508, 1999.

[65] J. Allman, F. Miezin, and E. McGuinness. Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14:105–126, 1985.

[66] J. Allman, F. Miezin, and E. McGuinness. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Ann. Rev. Neurosci.*, 8:407–430.

[67] D.M. Schneider and S.M.N. Woolley. Extra-classical tuning predict stimulus-dependent receptive fields in auditory neurons. *The Journal of Neuroscience*, 31(33):11867–78, 2011.

[68] E. de Boer and P. Kuyper. Triggered correlation. *IEEE Trans. Biomed. Eng.*, 3:169–179, 1968.

[69] A.M.H.J. Aertsen, J.H.J. Olders, and P.I.M. Johannesma. Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biological Cybernetics*, 39(3):195–209, 1981.

[70] A.M.H.J. Aertsen and P.I.M. Johannesma. The spectro-temporal receptive-a functional characteristic of auditory neurons. *Biological Cybernetics*, 42:133–43, 1981.

[71] J.J. Eggermont, P.M. Johannesma, and A.M. Aertsen. Reverse-correlation methods in auditory research. *Q. Rev. Biophys.*, 16(3):341–414, 1983.

[72] J.J. Eggermont. Wiener and Volterra analyses applied to the auditory system. *Hearing Research*, 66(2):177–201, 1993.

[73] T. Sharpee, N.C. Rust, and W. Bialek. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Computation*, 16:223–250, 2004.

[74] J.W. Pillow and E.P. Simoncelli. Dimensionality reduction in neural models: An information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of Vision*, 6:414–428, 2006.

[75] R.J. Rowekamp and T.O. Sharpee. Analyzing multicomponent receptive fields from neural responses to natural stimuli. *Network Computation in Neural Systems*, 22(1-4):45–73, 2011.

[76] P.Z. Marmarelis and K.-I. Naka. White-Noise Analysis of a Neuron Chain: An Application of the Wiener Theory. *Science*, 175(4027):1276–8, 1972.

[77] P.Z. Marmarelis and K.-I. Naka. Identification of Mulit-Input Biological Systems. *IEEE Trans. Biomed. Eng.*, 21(2):88–101, 1974.

[78] O. Schwartz, E.J. Chichilnisky, and E.P. Simoncelli. Characterizing neural gain control using spike-triggered covariance. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 269–76. MIT Press, Cambridge, MA, 2002.

[79] W.M. Yamada and E.R. Lewis. Predicting the temporal responses of non-phase-locking bullfrog auditory units to complex acoustic waveforms. *Hearing Research*, 130:150–170, 1999.

[80] E.R. Lewis and P. van Dijk. New variations on the derivation of spectro-temporal receptive fields for primary auditory afferent axons. *Hearing Research*, 186:30–46, 2003.

[81] E.R. Lewis, K.R. Henry, and W.M. Yamada. Tuning and timing in the gerbil ear: Wiener-kernel analysis. *Hearing Research*, 174, 2002.

[82] E.R. Lewis, K.R. Henry, and W.M. Yamada. Tuning and timing of excitation and inhibition in primary auditory nerve fibers. *Hearing Research*, 171:13–31, 2002.

[83] M.G. Sneary and E.R. Lewis. Tuning properties of turtle auditory nerve fibers: Evidence for suppression and adaptation. *Hearing Research*, 228:22–30, 2007.

[84] P. van Dijk, H.P. Wit, and J.M. Segenhout. Wiener kernel analysis of inner ear function in the American bullfrog. *J. Acoust. Soc. Am.*, 95(2):956–66, 1994.

[85] C.A. Atencio, T.O. Sharpee, and C.E. Schreiner. Receptive field dimensionality increases from the auditory midbrain to cortex. *Journal of Neurophysiology*, 107(10):2594–603, 2012.

[86] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, 1985.

[87] T. Rode, T. Hartmann, P. Hubka, V. Scheper, M. Lenarz, T. Lenarz, A. Kral, and H.H. Lim. Neural representation in the auditory midbrain of the envelope of vocalizations based on a peripheral ear model. *Frontiers in Neural Circuits*, 7(166), 2013.

[88] J.C. Berryman. Guinea pig vocalizations: Their structure, causation and function. *Z. Tierpsychologie*, 41(1):80–106, 1976.

[89] F.A. Rodriguez, C. Chen, H.L. Read, and M.A. Escabí. Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. *The Journal of Neuroscience*, 30(47):15969–15980, 2010.

[90] C. Chen, H.L. Read, and M.A. Escabí. Precise feature based time scales and frequency decorrelation lead to a sparse auditory code. *The Journal of Neuroscience*, 32(39):8454–68, 2012.

[91] C.A. Atencio, T.O. Sharpee, and C.E. Schreiner. Cooperative nonlinearities in auditory cortical neurons. *Neuron*, 58(6):956–66, 2008.

[92] F.A. Rodriguez, H.L. Read, and M.A. Escabí. Spectral and Temporal Modulation Tradeoff in the Inferior Colliculus. *Journal of Neurophysiology*, 103(2):887–903, 2010.

[93] S. Greenberg. Speaking in shorthand-a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 29:159–76, 1999.

[94] M.A. Escabí, L.M. Miller, H.L. Read, and C.E. Schreiner. Naturalistic Auditory Contrast Improves Spectrotemporal Coding in the Cat Inferior Colliculus. *The Journal of Neuroscience*, 23(37):11489–504, 2003.

[95] Y.W. Lee and M. Schetzen. Measurement of the Wiener Kernels of a Non-linear System by Cross-correlation. *Int. J. Contr.*, 2(3):237–254, 1965.

[96] J.P. Rauschecker, B. Tian, and M. Hauser. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207):111–4, 1995.

[97] F.E. Theunissen, K. Sen, and A.J. Doupe. Spectro-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of Neuroscience*, 20(6):2–17, 2000.

[98] K.D. Harris, D. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, 84(1):401–14, 2000.

[99] R. Quian Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16(8):1661–87, 2004.

[100] M. Blatt, S. Wiseman, and E. Domany. Superparamagnetic clustering of data. *Phys. Rev. Lett.*, 76(18):3251–4, 1996.

[101] M. Abeles. *Local cortical circuits: An electrophysiological study.* Springer, Berlin, 1982.

[102] R. Meddis, E.A. Lopez-Poveda, A.N. Popper, and R.R. Fay, editors. *Computational Models of the Auditory System*, volume 1. Springer, 2010.

[103] P.C. Nelson and L.H. Carney. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J. Acoust. Soc. Am.*, 116(4):2173–86, 2004.

[104] M.J. Hewitt and R. Meddis. A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *J. Acoust. Soc. Am.*, 95(4):2145–59, 1994.

[105] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification.* Wiley-Interscience, Oxford, 2nd edition, 2000.

[106] W. Bialek and R.R. de Ruyter van Steveninck. Features and dimensions: Motion estimation in fly vision. *arXiv:q-bio.NC/0505003*, 2005.

[107] D.J. Klein, D.A. Depireux, J.Z. Simon, and S.A. Shamma. Robust spectrotemporal reverse correlation for the auditory system: Optimizing stimulus design. *Journal of Computational Neuroscience*, 9(1):85–111, 2000.

[108] N.K. Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–71, 2003.

[109] D.L. Oliver. Neuronal organization in the inferior colliculus. In J. Winer and C. Schreiner, editors, *The Inferior Colliculus.* Springer, New York, 2005.

[110] K.H. Pettersen, H. Linden, A.M. Dale, and G.T. Einevoll. Extracellular spikes and CSD. In R. Brette and A. Destexhe, editors, *Handbook of Neural Activity Measurement*, volume 1, pages 92–135. University Press, Cambridge, 2012.

[111] M.C.W. van Rossum. A novel spike distance. *Neural Computation*, 13:751–763, 2001.

[112] C.K. Machens, M.B. Stemmler, P. Prinz, R. Krahe, B. Ronacher, and A.V.M. Herz. Representation of acoustic communication signals by insect auditory receptor neurons. *The Journal of Neuroscience*, 21(9):3215–3227, 2001.

[113] J.O. Pickles. *An Introduction to the Physiology of Hearing.* Emerald Group, Bingley, UK, 3rd edition edition, 2006.

[114] M.S. Malmierca, M. Izquierdo, S. Cristaudo, O. Hernández, D. Pérez-González, E. Covey, and D.L. Oliver. A discontinuous tonotopic organization in the inferior colliculus of the rat. *The Journal of Neuroscience*, 28(18):4767–76, 2008.

[115] G. Langner, C. Schreiner, and M. Merzenich. Covariation of latency and temporal resolution in the inferior colliculus of the cat. *Hearing Research*, 31:197–202, 1987.

[116] W.J. Melssen and W.J.M. Epping. Detection and estimation of neural connectivity based on crosscorrelation analysis. *Biological Cybernetics*, 57:403–414, 1987.

[117] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–12, 2006.

[118] J. Astl, J. Popelár, J.E. Kvasnak, and J. Syka. Comparison of response properties of neurons in the inferior colliculus of guinea pigs under different anesthetics. *Audiology*, 35(6):335–345, 1996.

[119] D.H. Hubel and T.N. Wiesel. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, 26:994–1002, 1963.

[120] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley & Sons, Inc., USA, 1980.

[121] How linear are auditory cortical responses? volume 15, pages 109–116. MIT Press, 2003.

[122] F.E. Theunissen, S.V. David, N.C. Singh, A. Hsu, W.E. Vinje, and J.L. Gallant. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Computation in Neural Systems*, 12(3):289–316, 2001.

[123] O. Schwartz, J.W. Pillow, N.C. Rust, and E.P. Simoncelli. Spike-triggered neural characterization. *Journal of Vision*, 6(4):484–507, 2006.

[124] N.A. Lesica and B. Grothe. Efficient temporal processing of naturalistic sounds. *PLoS One*, 3:1–12, 2008.

[125] M. Pienkowski and J.J. Eggermont. Nonlinear cross-frequency interactions in primary auditory cortex spectrotemporal receptive fields: a Wiener-Volterra analysis. *Journal of Computational Neuroscience*, 28:285–303, 2010.

[126] R. Narayan, G. Graña, and K. Sen. Distinct time scales in cortical discrimination of natural sounds in songbirds. *Journal of Neurophysiology*, 96(1):252–258, 2006.

[127] C. Huetz, B. Philibert, and J.-M. Edeline. A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *The Journal of Neuroscience*, 29(2):334–50, 2009.

[128] G. Arriaga, E.P. Zhou, and E.D. Jarvis. Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PloS One*, 7(10):e46610, 2012.

[129] T.J. Gawne and B.J. Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *The Journal of Neuroscience*, 13(7):2758–2771, 1993.

Danksagung