The following full text is a publisher’s version.

For additional information about this publication click this link.
http://repository.ubn.ru.nl/handle/2066/128475

Please be advised that this information was generated on 2019-02-27 and may be subject to change.
Interaction of task-related and acoustic signals in single neurons of monkey auditory cortex

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 3 juli 2014
om 12:30 uur precies

door

Roohollah Massoudi

geboren op 9 juli 1979 te Eqlid, Iran
Promotor:
Prof. dr. A.J. van Opstal

Copromotoren:
Dr. M.M. van Wanrooij
Dr. H. Versnel (UMCU)

Manuscriptcommissie:
Prof. dr. O. Jensen
Prof. dr. E. Formisano (UM)
Dr. M.P. Zwiers

ژندگی صحنه ی یکنای هنرمندی ماست
هر کسی نغمه ی خود خواند و از صحنه رود
صحنه پیوسته به جاست
خم آن نغمه که مردم بسیارند به یاد
زاله اصفهانی
## CONTENTS

1. GENERAL introduction .............................................................................................................. 7
  1.1 The Ascending auditory pathway ....................................................................................... 8
  1.2 Monkey auditory cortex subdivisions .............................................................................. 10
  1.3 Tuning characteristics of AC subdivisions ........................................................................ 10
  1.4 Cortical vs. subcortical processing in THE auditory system ........................................... 11
  1.5 Response properties in the auditory cortex ...................................................................... 12
  1.6 Spectrotemporal properties of auditory cortical cells .................................................... 13
  1.7 Spectrotemporal characteristics of AC cells ..................................................................... 15
  1.8 The effects of learning, behavior, and motivation in auditory cortical responses ............... 16
  1.9 Top-down and bottom-up interactions in auditory cortex ................................................ 18

2. Spectrotemporal response properties of auditory cortex neurons in awake monkey ............ 23
  2.1 Introduction ...................................................................................................................... 24
  2.2 Materials and Methods ................................................................................................... 25
     2.2.1 Subjects .................................................................................................................. 25
     2.2.2 Surgical procedures ............................................................................................... 26
     2.2.3 Experimental setup ............................................................................................... 27
     2.2.4 Sound stimuli .......................................................................................................... 27
     2.2.5 Experimental paradigms ....................................................................................... 28
     2.2.6 Characterization of recording sites ........................................................................... 29
     2.2.7 Data analysis .......................................................................................................... 30
  2.3 Results ............................................................................................................................. 33
     2.3.1 General properties ................................................................................................. 33
     2.3.2 Distributions of ripple responses ............................................................................. 38
     2.3.3 Spectral-temporal separability .............................................................................. 40
     2.3.4 Phase functions ...................................................................................................... 44
     2.3.5 Linear predictions .................................................................................................. 47
  2.4 Discussion ....................................................................................................................... 50
     2.4.1 Summary ................................................................................................................. 50
     2.4.2 General characteristics ......................................................................................... 50
     2.4.3 Separability and direction selectivity ....................................................................... 51
     2.4.4 Linearity and response prediction ......................................................................... 52

3. Stable bottom-up processing during dynamic top-down modulations in monkey auditory cortex... 55
  3.1 Introduction ..................................................................................................................... 56
  3.2 Material and Methods ..................................................................................................... 57
4. Behavioral context affects temporal processing in awake monkey auditory cortex

4.1 Introduction ................................................................. 81

4.2 Materials and Methods .................................................. 82
  4.2.1 Subjects .................................................................. 84
  4.2.2 Electrophysiology .................................................... 84
  4.2.3 Sound stimuli and behavioral paradigms ...................... 85
  4.2.4 Data Analysis ......................................................... 85

4.3 Results ........................................................................... 88
  4.3.1 Effects of modulation frequency on the reaction times ..... 88
  4.3.2 Influence of behavioral context on spontaneous activity and onset responses ............................................. 90
  4.3.3 Changes in behavioral context influence temporal processing ................................................................. 93

4.4 Discussion .................................................................. 100
  4.4.1 Summary ................................................................. 100
  4.4.2 Behavioral response properties .................................... 100
  4.4.3 Effect of task engagement on spontaneous and onset responses .............................................................. 100
  4.4.4 Effect of task engagement on temporal envelope processing ................................................................. 101

5. Task-related preparatory modulations multiply with acoustic processing in monkey auditory cortex

5.1 Introduction .................................................................. 106

5.2 Materials and Methods .................................................. 107
  5.2.1 Subjects .................................................................. 107
  5.2.2 Surgical procedures .................................................. 107
5.2.3 Experimental setup ........................................................................................................... 108
5.2.4 Sound stimuli ..................................................................................................................... 108
5.2.5 Experimental paradigms .................................................................................................... 109
5.2.6 Characterization of recording site .................................................................................... 110
5.2.7 Data analysis ..................................................................................................................... 111
5.3 Results ................................................................................................................................ 112
5.3.1 Neural responses: passive and active listening ................................................................. 112
5.3.2 Behavior ........................................................................................................................... 114
5.3.3 Top-down neural signals .................................................................................................. 118
5.3.4 Model selection ............................................................................................................... 124
5.4 Discussion ........................................................................................................................... 125
5.4.1 Summary ......................................................................................................................... 125
5.4.2 Behavioral response properties ....................................................................................... 126
5.4.3 Task-related effects on AC responses ............................................................................. 126
5.4.4 Additive and multiplicative interactions ........................................................................ 128

Summaries .................................................................................................................................. 131

English summary ...................................................................................................................... 132
Chapter 2: Spectrotemporal response properties of auditory cortex neurons in awake monkey 132
Chapter 3: Stable bottom-up processing during dynamic top-down modulations in monkey auditory cortex .................................................................................................................. 133
Chapter 4: Behavioral context affects temporal processing in awake monkey auditory cortex . 133
Chapter 5: Task-related preparatory modulations multiply with acoustic processing in monkey auditory cortex .................................................................................................................. 134

Nederlandse samenvatting ....................................................................................................... 135
Hoofdstuk 2: Spectraal-temporele respons eigenschappen van auditieve cortex neurone in de wakkere makaak........................................................................................................................................ 136
Hoofdstuk 3: Stabiele bottom-up verweking tijdens dynamische top-down modulaties in de auditieve cortex van makaken ........................................................................................................... 136
Hoofdstuk 4: De gedragmatige context bïnvloedt temporele verwerking in de auditieve cortex van makaken ................................................................................................................................. 137
Hoofdstuk 5: Taak-afhankelijke modulaties vermenigvuldigen met de akoestische signaalverwerking in de auditieve cortex van makaken ......................................................................................... 138

خلاصه فارسی (Persian summary) .................................................................................................. 139
فصل دوم: مطالعه ویژگی‌های زمانی فرکانسی پاسخ‌های قشر شنوایی میمون ... 139
فصل سوم: برداشتن پایدار سیگنال‌های پایین-به-بالا (bottom-up) تحت توابع اصلی (top-down) در قشر شنوایی میمون ................................................................. 140
فصل چهارم: شرایط متفاوت رفتاری می تواند برداشتهای زمانی را در قشر شنوایی میمون تغییر دهد 140
فصل پنجم: سیگنال‌های بالا‌به پایین با سیگنال‌های شنوایی در قشر شنوایی میمون ضرب می‌شوند

Bibliography .................................................................................................................. 142
Acknowledgment ........................................................................................................... 158
Curriculum Vitae ........................................................................................................... 162
List of Publications ....................................................................................................... 163
Donders graduate school for cognitive neuroscience series ......................................... 164
1. GENERAL INTRODUCTION
Sound is a mechanical perturbation of air pressure that travels from the source that produced it through the environment with a speed of about 343 m/s. The acoustic wave contains a multitude of information cues about the sound source: its spectral and temporal properties, the sound-source direction with respect to the observer, the emotional state of the speaker, etc. Like vision, hearing is an exteroceptive sense that allows the organism to obtain detailed sensory information about objects in the environment. The auditory system is specialized to process the different features in the acoustic input in order to extract the information that allows us to communicate, to orient and navigate, to interpret the environment, and thus to survive.

In the auditory system, the different neural stages from ear to cerebral cortex, are all specialized to process (and integrate) different acoustic features, like the sound’s spectrum, its temporal modulations, the sound-source direction, but also more complex, higher-order properties, like source identity, signal and noise, meaning of the message, etc.

In this thesis, I will present my work on the neural representation of complex spectral-temporal modulated sounds in the auditory cortex of the non-human primate. In this introduction, I will first briefly outline the relevant properties of the auditory system that have played an important role in my studies.

1.1 THE ASCENDING AUDITORY PATHWAY

Hearing starts when sound waves enter the ear canal, and push against the eardrum to set the ossicles in the middle ear in motion. The middle ear transmits the sound energy to the fluid-filled cochlea, which contains the intricate inner-ear structure (the organ of Corti) that transforms the mechanical vibrations of the sound, via the inner hair cells (IHC) into the neural signals carried by the auditory nerve (AN). Because of its micromechanical properties, reflected by a systematic position-dependent variation in impedance of the basilar membrane (BM), and the systematic change in vibrational properties of the outer hair cells from base (at the oval window) to apex (at the top of the cochlea), the organ of Corti, and hence the AN, is tonotopically organized: high frequencies are represented at the cochlear base, whereas the lowest frequencies are found at the apex, with intermediate frequencies at intermediate locations. This tonotopic organisation is a prominent structural feature of the auditory system that is maintained throughout the ascending auditory pathway.

The AN fibers join as the 8th cranial nerve that projects topographically (i.e. tonotopically) to the cochlear nucleus (CN) in the brainstem. Although some CN axons project directly to the auditory midbrain Inferior Colliculus (IC), most reach the IC via the superior olivary complex (SOC) and lateral lemniscus. The pathway from CN to the IC crosses the brain’s midline, with the result that neurons in midbrain and cortex are predominantly stimulated by sounds entered into the contralateral ear. The IC provides further binaural interaction via the commissural connections between the left and
right IC. The IC axons depart toward the main auditory relay nucleus of the thalamus, the medial geniculate body (MGB). The vast majority of MGB paths travel to the auditory cortex (AC) in the temporal lobes (Figure 1.1).

The auditory cortical fields on both sides are also interconnected via commissural connections through the corpus callosum, which allows for further exchange of information between left and right hemispheres. Thus, the activity of an AC neuron would be mostly influenced by stimuli provided to either ear.

Despite some differences, the organization of the ascending auditory pathway up to, and including, the level of the MGB seems to be fundamentally similar among different mammalian species. However, at the level of the auditory cortical fields the homologies are less clear, and therefore different names are used to specify cortical subdivisions in different species.

Nevertheless, it is possible that cortical fields with different names may actually have analogous functions. For instance, both carnivores and primates have two side-by-side primary cortical areas, which receive the strongest thalamic input; but while these fields are called A1 (primary auditory cortex) and AAF (for anterior auditory field) in carnivores, they are named A1 and R in monkeys. The extent of AAF and R similarity, however, is not well known.

The cortex is evolutionary the youngest and most plastic part of the brain. This characteristic has led to the development of specialized cortical fields in some mammals, which do not exist in other species. For example, some areas in the auditory cortex of echolocating bats process echo delays and Doppler shifts, for which there is no homologue in the monkey brain.

However, what appears to be similar in the auditory cortex among all mammals is the existence of distinct primary (core) and secondary (belt) areas, which have been suggested to communicate with different parts of the brain, such as the prefrontal lobe, or the infra-temporal areas involved in object recognition (Schnupp et al., 2012).
1.2 MONKEY AUDITORY CORTEX SUBDIVISIONS

In the macaque monkey the auditory cortex (AC) is buried in the ventral bank and covered by dorsal part of the lateral sulcus. It has three main areas: core, belt and parabelt. The core consists of three subdivision fields, called A1, R and RT. These areas have strong and dense connections with the MGB in the thalamus. They have a tonotopic organization and respond well to pure tones (Merzenich and Brugge, 1973; Morel et al., 1993; Kosaki et al., 1997; Hackett et al., 1998a, 1998b, 2001). The core axons travel to the belt, which in turn projects to the parabelt (Galaburda and Pandya, 1983; Morel et al., 1993; Hackett et al., 1998b; Jones, 2006). The caudomedial area (CM), one of the belt areas, has similar architectonic characteristics as the core and belt areas (Hackett et al., 2001; de la Mothe et al., 2006). While CM’s tonotopic organization is similar to A1 (Rauschecker et al., 1997), many CM neurons are also responsive to somatic stimuli (Figure 1.1) (Schroeder et al., 2001; Fu et al., 2003).

1.3 TUNING CHARACTERISTICS OF AC SUBDIVISIONS

The areas A1 and R possess a mirror image of each other’s tonotopicities. Area A1 has its best high to low frequencies represented along the caudo-rostral axis, while it is organized along the rostro-caudal direction in area R. The tonotopic organization of core area RT is analogous to that of
area A1, and in belt area CM it is similar to core area R (Fig 1.1). Although the core areas (A1, R and RT) and belt area CM all respond to tones, they possess different tuning characteristics for other sound types. The response onset latency cannot dissociate the core areas A1 and the belt area CM in macaque monkeys, while core area R typically has a longer latency than areas A1 and CM. On the other hand, the tuning bandwidths of core areas A1 and R are identical, whereas cells in belt area CM tend to have broader bandwidths. In terms of their sound-level thresholds (i.e., the minimum sound level to produce a neural response), core areas A1 and R have lower thresholds (and are therefore more sensitive) than belt area CM (Recanzone et al., 2000).

**Figure 1.2:** Schematic view of location and internal connections of the macaque auditory cortex. The dashed line (cut) represents edges of removed dorsal part of the lateral sulcus (LS) to expose the auditory cortex, which is located in the ventral bank of LS. The dark shading area represents the core area containing A1, R, and RT subdivisions. The tonotopy of these subdivisions mirror each other, indicated by the letters H (high frequency) and L (low frequency). The belt areas are indicated in light shading, which surrounds the core area (dark shading). The area without shading contains areas RP and CP that are part of the parabelt area. There are numerous connections between belt and parabelt areas (arrows). Moreover, there are also projections within each area, and between different subdivisions (with permission, Hackett et al., 2001).

### 1.4 CORTICAL VS. SUBCORTICAL PROCESSING IN THE AUDITORY SYSTEM

A major difference of the auditory system when compared to other sensory systems is the complexity of its subcortical paths. Four major processing levels (CN, SOC, IC, and MGB) are located between the auditory nerve and the auditory cortex, compared to only one intermediate level in the visual system, and two in the somatosensory system. For example, well-known primary visual cortex (V1) properties, such as sensitivity to bar orientation and binocular disparity, and the presence of simple cells and complex cells, emerge at the level of V1 (Hirsch and Martinez, 2006). In contrast, most of the basic acoustic stimulus characteristics, such as sound frequency, duration and amplitude,
and even higher-level information, like pitch, already take place at subcortical auditory levels (Palmer, 2007). For that reason, it has been proposed that the midbrain inferior colliculus may occupy a hierarchical processing level that is equivalent to V1, in terms of its higher-order sensitivity to different stimulus properties (Nelken et al., 2003a).

The already quite advanced processing stages at subcortical auditory levels poses the question which higher-order functions may be subserved by the auditory cortex. It is not clear (albeit unlikely) whether auditory cortex has evolved to merely reprocess the subcortically processed information. To address such a question, one should apply more complex acoustic probes than pure tones or static noises, to expose the role of the AC in higher level auditory processing (e.g. its potential top-down involvement, such as in auditory attention, behavior, or memory), besides purely acoustic processing (bottom-up).

1.5 RESPONSE PROPERTIES IN THE AUDITORY CORTEX

Because the AC may occupy a position that is high up in the hierarchy of auditory processing, one way to proceed in unraveling its function would be to apply more natural stimuli, such as species-specific vocalizations. This approach has a long history (Wollberg and Newman, 1972), and studies have shown that AC neurons respond to a broad range of stimuli. However, the information about the stimulus that is carried by individual cortical neurons can be irrelevant for adjacent simultaneously recorded cells (Chechik et al., 2006). AC neurons respond typically to specific acoustic components, even in the presence of other components. Thus, it has been proposed that AC neurons might rather represent ‘auditory objects’ than different acoustic features (Nelken et al., 2003b).

The exact definition of an auditory object has raised vigorous debate among researchers (Griffiths and Warren, 2004; Bizley and Cohen, 2013). We here consider it as the auditory system’s representation of grouped acoustic events and their corresponding sources. For instance, while sitting on your couch, you hear the whistle of a boiling kettle, and the water tap that is leaking in the kitchen. Each of these two sound sources could be regarded as a distinct auditory object. The auditory system perceives the acoustic events (whistle or leaking) and their assigned sources (kettle, and tap) as two different auditory objects. Each auditory object shares certain properties, such as i) it is caused by a single physical source. 2) It has certain spectrotemporal properties that dissociate it from other auditory objects. For example, a common rise and fall of specific (often harmonically related) frequency bands (as in a human voice), a common start and stop of frequency bands, etc. These spectral and temporal properties can thus serve as ‘boundaries’ of the auditory object. 3) It is context invariant, and 4) the listener can describe it by combining different acoustic features. The auditory system constructs an auditory object by detecting the particular (learned, stored, attended) sound
features, extracts their information, and then segregates and groups them into a perceptual unit (Griffiths and Warren, 2004; Winkler et al., 2009; Bizley and Cohen, 2013).

Selectivity to auditory objects could arise from a combination of different sensitivities (Suga, 1989). This is corroborated by strong neural responses to specific combinations of stimulus features (Wang et al., 2005). This selective property seems similar to the specific selectivity of visual IT neurons to complex visual features, such as faces (Ungerleider and Mishkin, 1982), and suggests that to understand AC, one should perform experiments, in which neurons are confronted with a complex computational problem, such as auditory scene analysis.

1.6 SPECTROTEMPORAL PROPERTIES OF AUDITORY CORTICAL CELLS

A methodological approach to find an auditory neuron’s optimal sound stimulus is by using a large number of sounds that can be modulated systematically in both the temporal and spectral domains. A specific type of stimulus in this context is the so-called dynamic ripple, which is a broadband sound that consists of a superposition of narrowly spaced sinusoids, all with random phase, typically at 1/20 octave apart, and distributed over several octaves:

\[
C(t) = \sum_{n=0}^{125} \sin(2\pi f_n t + \varphi_n) \quad f_n = f_0 \cdot 2^n/20 \text{ Hz} \quad (Eq. 1.1)
\]

When the lowest frequency, \(f_0\), is 250 Hz, the highest frequency of this complex is about 19 kHz. This complex is modulated with a sinusoidal spectral profile along the logarithmic frequency axis that drifts along the linear time axis at a constant velocity:

\[
R(t) = \left(1 + \Delta M \cdot \sin(2\pi (\omega t + \Omega x))\right) \cdot C(t) \quad (Eq. 1.2)
\]

where \(\omega\) (in Hz) is the ripple velocity (temporal modulation), and \(\Omega\) (in cycles/octave) the ripple density (spectral modulation). The variable \(x\) is the spectral position of the ripple component \(f_x\) in octaves above \(f_0\), i.e., \(x = 2\log (f_x/f_0)\). The ripple’s modulation depth, \(\Delta M\), is taken between 0 and 1. Stimuli with \(\Omega=0\) resemble pure amplitude modulations (e.g., Fig. 1.3B), whereas ripples with \(\omega=0\) only contain pure spectral modulations. Figure 1.3 shows two representative examples of ripple stimuli, out of the 55 different ripples that were used in this study.

The dynamic ripple can be considered a ‘naturalistic’ sound, which, unlike real natural sounds, can be readily parameterized, and has been widely used to extract the spectral-temporal tuning properties of auditory neurons in different parts of the auditory pathway, including auditory cortex. Moreover, as these sounds are abstract constructs, based on the concepts underlying Fourier analysis, they are not recognizable as natural sounds, which prevents potential cognitive influences on a neuron’s response.
Figure 1.3: Spectrograms of a dynamic ripple, and of an AM stimulus. A) Ripple with a downward sweep direction, at a velocity $\omega = 4$ Hz, and spectral density of $\Omega = 1.2$ cyc/oct. B) Ripple with flat amplitude spectrum, amplitude-modulated at $\omega = 4$ Hz, and $\Omega = 0$ cyc/oct. In the experiments described in this thesis we typically applied 55 different ripples, for which $\omega = 8:8:40$ Hz, and $\Omega = -2.0:0.4:+2.0$ cyc/oct.

An advantage of dynamic ripples such as in Fig. 1.3 is their convenient use in estimating the spectrotemporal receptive fields (STRFs) of an auditory neuron. The STRF is a linear, time-frequency representation of a neuron’s stimulus preference, and typically exhibits excitatory and inhibitory regions in the spectrotemporal domain (e.g., Figure 1.4; Aertsen and Johannesma, 1981, Depireux et al., 2001). The two-dimensional Fourier transform of an STRF yields its so-called amplitude- and phase-modulation transfer functions (MTFs) that characterize the neuron’s preferred tuning to spectral and temporal stimulus envelopes. The STRF is thus a rich, albeit linear, descriptor of a neuron’s acoustic preference.

Dynamic ripples form a spectrotemporal basis for dynamic complex natural sounds, and thus allow for a full representation of the stimulus (like the sinusoid acts as a complete basis of Fourier analysis). Despite its limitations, the STRFs obtained from auditory cortex cells have been used with some success to reconstruct the acoustic input from measured neural responses of a population of AC cells (David et al., 2009; Mesgarani et al., 2009).

Figure 1.4: STRFs of four AC cells recorded and computed by our group. Examples of (A) narrowband, (B) complex, (C,D) broadband STRFs.
By measuring the STRFs of auditory cortex neurons different questions can be studied:

(i) To what extent do AC neurons process time and frequency independently (in which case the neural processing is termed *separable*), or in a mutually dependent way (*inseparable*)? Separable processing suggests independent channels for temporal and spectral analysis, whereas inseparable processing could suggest the grouping of complex spectrotemporal stimulus features, like FM-sweeps, or even more complex sound snippets, like vocalizations.

(ii) To what extent do AC neurons process acoustic stimuli linearly or nonlinearly? If an AC neuron behaves as a linear system, the STRF, \( S(f,t) \), can be used as a kernel to predict the neuron’s response, \( r(t) \), to arbitrary sounds, \( x(t) \):

\[
r(t) = \int_0^\infty \int_0^\infty S(f,\tau) \cdot x(f,t-\tau) \cdot d\tau \cdot df 
\]

(Eq. 1.3)

Although the STRF misses important nonlinear aspects of neurons (e.g. typical non-essential nonlinearities like saturation and rectification, but also more serious nonlinearities like stimulus-dependent properties, adaptation, and non-stationary behavior), its combined spectrotemporal representation provides useful hints about the encoding of natural sounds, which are rarely static.

(iii) What are the encoding properties of the STRFs in monkey AC? Which spectral-temporal features are represented? Are there specific preferences to particular spectral-temporal modulations?

(iv) To what extent do STRFs change, or remain invariant to the animal’s active involvement in sound processing? This question touches on an important aspect of auditory perception that will be dealt with below.

1.7 SPECTROTEMPORAL CHARACTERISTICS OF AC CELLS

So far, studies of the auditory cortex in awake monkey have not focused in much detail on its basic acoustic spectral-temporal encoding properties. As described above, the use of dynamic ripple stimuli can provide a convenient means to study the basic spectral and temporal characteristics of auditory cells. As a first step, we therefore compared the response characteristics of monkey AC cells to dynamic ripples, pure tones, a set of natural calls, and Gaussian broadband AM noises with the
characteristics that we previously obtained from the midbrain IC to the same stimuli (Versnel et al., 2009).

The results of this study are described in Chapter 2, where we quantified the response linearity of AC cells, their spectral-temporal separability, and their selectivity to ripple direction, ripple density, and ripple velocity. Our results show that AC cells, like neurons in the IC, display a wide diversity of selective responses to ripple densities and velocities, without specific direction selectivity. It also demonstrated spectral-temporal separability for the large majority of recorded neurons. The main difference that we observed between AC and IC cells appeared in the neuronal stimulus-response linearity (Eq. 1.3). We observed a much poorer linear predictability of AC responses to acoustic stimuli than had been obtained in the IC, which is indicative for the presence of more prominent nonlinearities in the response behavior of AC cells than in IC cells. We also observed a much poorer agreement between an AC neuron’s best frequency, obtained for pure tones, and the best frequency obtained from the STRF. For a linear neuron these two measures should be identical. The correspondence between best frequencies for IC cells was decent. These findings extend previous reports that suggest a hierarchical increase in nonlinear processing along the ascending auditory pathway.

1.8 THE EFFECTS OF LEARNING, BEHAVIOR, AND MOTIVATION IN AUDITORY CORTICAL RESPONSES

The auditory system is not a static system, but it is highly adaptive and plastic. The auditory cortex is no exception. Auditory cortical processing can rapidly adapt and change, depending on experience and on the behavioral and environmental demands. As a result, the representation of sound by a cortical neuron may not be static, but could change based on changes of the animal’s needs. Also input from other sensory modalities may change the acoustic tuning characteristics of auditory neurons (Brosch et al., 2005; Ghazanfar et al., 2005; Hackett et al., 2007; Kayser et al., 2007; Lakatos et al., 2007; Smiley et al., 2007; Bizley and King, 2008, 2009; Ghazanfar, 2009; Hackett and Schroeder, 2009). Such context dependent (non-acoustic) properties argue for a role of AC cells in multisensory integration.

There is considerable interest in establishing the influence of non-acoustic factors on auditory cortical activity, which could shed more light on its role in monitoring or guiding the animal’s behavior. One way to address this topic is to record neural activity from single units in awake animals while they actively perform a task.

Earlier studies have indicated that auditory-evoked AC responses are influenced by task engagement, by the value and expectation of a reward, by attention, by stimulus expectation, memory, behavioral training, sensorimotor association, behavioral choices and decision making, and motor
output. This long list of non-acoustic influences on AC responses help to better understand the importance of auditory cortex in sound processing, beyond the mere encoding of acoustic information, but also as a center that is involved in the integration of more abstract (e.g., top-down) variables that are needed for listening tasks under different behavioral and environmental conditions.

Convergent studies of different sensory modalities have demonstrated rapid, task and context-dependent changes in situations that require different attentive states of the animal. In such experiments the animal directs its attention towards a task-related sensory cue (target) that eventually (after a series of trials) leads to a selective reconfiguration of cortical connections. This reconfiguration results in receptive field changes of the involved neurons, which boosts the animal’s performance in the task (Bakin et al., 1996; Crist et al., 2001; Li et al., 2001).

Recent studies have shown that the cognitive behaviors that require targeted interactions between stimulus driven activity (bottom-up) and top-down modulations, such as attention, reward, expectation, or learning, can indeed induce changes in cortical receptive fields. Studies of AC cells of ferrets have thus reported changes in their STRFs when the animal is engaged in an auditory task that requires directed attention to a particular acoustic feature that is part of its STRF (e.g., a specific target frequency; Fig 1.5; (Fritz et al., 2003, 2005b, 2007, 2010; Elhilali et al., 2007; Atiani et al., 2009).

**Figure 1.5.** STRF plasticity in ferret primary auditory cortex. (A-D) STRFs of four example cells recorded during pre-behavior “passive STRF” in left column, and during task performance “behavior STRF” in middle column. The black arrows represent the frequency of the target tone during the detection task. More excitation (A, B) or less inhibition (C, D) is observed at target frequency location. (A) In the right column, the difference between the passive and behavior STRFs (STRF\_diff) is shown. An asterisk denotes the location of maximum change, and a circle represents the extent of total changes. (B-C) In the right column, the post behavior passive STRFs, which reverted to their original shapes as shown in the left column (with permission, Fritz et al., 2003).
1.9 TOP-DOWN AND BOTTOM-UP INTERACTIONS IN AUDITORY CORTEX

In the studies mentioned above, the receptive field modifications during auditory task performance were triggered by attention towards a specific stimulus feature, such as a target’s frequency. However, the question remains what would happen to a neuron’s tuning if the task demands are not related to a particular stimulus feature or, in other words, when there is no directed attention, preference or aversion towards any particular acoustic property? In another word, if auditory cortical neurons would continuously change their receptive fields under all behavioral conditions in changing acoustic environments, they would seem ill-suited to encode a stable and invariant representation of the acoustic environment, which could be used for perception and goal-directed behavioral responses. A central question of receptive field plasticity thus revolves around an important dichotomy: on the one hand, the need to maintain a stable representation of the environment, and on the other hand, the need to adapt to new behavioral demands. To study the former point, we introduce a behavioral auditory paradigm in this thesis, for which the behavioral or attentive demands were not tied to a particular acoustic feature.

Our paradigm consisted of three different hearing conditions for the animal, in which we ensured that the acoustics remained the same.

i) A passive hearing condition, in which the animal was passively exposed to a series of sounds that started automatically with a static noise (here, with a fixed duration of 500 ms), followed by a spectral-temporal dynamic ripple, or a Gaussian AM noise.

ii) A first behavioral condition (A500) that was acoustically identical to the passive condition, but required the animal to manually react to the change from static noise to dynamic ripple, in order to obtain a reward.

iii) A second behavioral condition (A1000), in which, unlike in the passive and A500 trials, the static epoch duration was 1000 ms. However, like in the A500 trials, it required the animal to react as fast as possible to the change from static to dynamic sound (Massoudi et al., 2013, 2014).

Passive trials were presented in a separate block, while the A500 and A1000 trials were randomly interleaved in another block of trials. As the animals were well trained for this behavioral paradigm, they could fully predict the modulation onset in A1000 trials after 500 ms static noise had passed. This aspect of the paradigm added a cognitive factor, namely prediction, to the experiment. As will become clear in this thesis, this predictive effect was clearly evidenced by the animal’s response-reaction times, and was reflected likewise in the neural responses.
In our active paradigm, the only requirement to the animal was the rapid detection of the (note: any) transition from static noise to a (again: any) dynamic sound to receive a reward. As all dynamic sounds in the experiments were equally likely and important to the animal, the paradigm did not require (or allow) directed attention towards any specific sound type, or any specific feature in the sound. However, in predictable A1000 trials the animal could adopt the strategy (after 500 ms) not to listen to the sound any more to respond to the change, but instead to rely on the learned prediction of its occurrence.

The results of our behavioral paradigm were analyzed with three potential scenarios in mind: First, the AC cells could be purely acoustically sensitive, so that engaging in the task would affect neither their firing rate, nor their receptive fields. Second, as shown by previous studies in attentive auditory tasks, both the AC neural responses and their tuning characteristics would change. If true, it would be interesting to determine how the receptive fields had changed for a task with no direct
relation to any acoustic feature. For example, the STRFs could sharpen their spectral-temporal tuning (narrower tuning widths, better signal-to-noise ratios, etc.). The third possibility could be that neural responses would change, but the tuning properties remain stable for both passive and active trials. Such a result would imply that the receptive field characteristics are invariant to the behavioral demands in the paradigm (see also Massoudi et al., 2013).

The results of these experiments are described in Chapters 3-5 of this thesis. In Chapter 3 we show that, although the responses of AC cells are profoundly different for the three behavioral conditions, the extracted STRFs of auditory cortical cells remained remarkably stable. We also demonstrated the presence of a unique task-related signal at the level of single-neurons that started to rise well in advance of the animal’s behavioral response.

These results support the third hypothesis, and therefore imply that AC plasticity may not be linked to the behavioral relevance of a sound per se, but rather requires the behavioral relevance of specific acoustic features in the sound (Massoudi et al., 2013). This characteristic allows the AC to build and maintain a stable representation of the acoustic environment that is invariant to a variety of behavioral situations, as long as the task demands do not require repeated or prolonged directed attention of a particular stimulus feature.

Applying the same behavioral paradigm, we also presented amplitude-modulated Gaussian white noises (GWN AM), which covered a much broader range of temporal modulation frequencies ($f_{AM} = 2-256$ Hz) than the spectral-temporal dynamic ripples ($\omega=8-40$ Hz), described in chapter 3. In this way we could study the effect of task engagement on temporal envelope encoding in AC cells. The results of this study are presented in chapter 4.

Our findings demonstrate that both the mean firing rate and the strength of phase locking of AC neurons to the temporal envelope modulations were influenced by the behavioral conditions. Multiplying the cell’s mean firing rate ($r$) and its vector strength ($p$), which is a measure for the strength of phase locking, yields the magnitude ($M$) of the first harmonic of its response-period histogram:

$$M(f_{AM}) = r(f_{AM}) \cdot p(f_{AM}) \quad (Eq. \ 1.4)$$

We found that $M(f_{AM})$ remained virtually unaffected by the behavioral task for a considerable range of modulation frequencies, from 5.6 to 45 Hz. This invariant band width corresponds well to the temporal range of dynamic ripples used in Chapter 3, and for which we demonstrated that the STRFs remained unaffected by the task as well. Our findings thus suggested that $M(f_{AM})$ remains stable because AC cells may employ a mixed code of rate and temporal coding for AM sounds. Furthermore, $M(f_{AM})$ could perhaps be a better measure to quantify AM encoding by AC neurons than either their firing rate, or their phase-locking, as it combines the effect of both measures.

The results from Chapter 3 and 4 suggested the possibility that the top-down task-related signals and bottom-up acoustic signals were independently processed by AC neurons. As these signals interact at the level of single cells, we wondered about the nature of this interaction. We considered
two possibilities: additive (i.e., linear) vs. multiplicative (c.q., nonlinear) interactions. In case the signals interact linearly, the neural response in active trials, $A(t)$, can be described as:

$$A(S, t) = P(S, t) + B(t + \Delta T - RT) \quad (Eq. 1.5)$$

where $P(t)$ is the neural response to the sound $S$, measured in the passive trials, and $B(x)$ is a task-related signal that is independent of the acoustics, which starts to rise from zero to some maximum value at $\Delta T$ ms before the reaction time, $RT$ (our data indicate that $\Delta T$ is about 100-150 ms; $B(t)$ is zero elsewhere).

When the two signals interact in a multiplicative way, the response in active trials is represented as a gain modulation of the passive response:

$$A(S, t) = (1 + B(t + \Delta T - RT)) \cdot P(S, t) \quad (Eq. 1.6)$$

We addressed this issue in Chapter 5. To that end, we analyzed the AC activity for both amplitude modulated Gaussian noises (AM) and spectral-temporal ripple stimuli during the same task paradigm as presented in Chapters 3 and 4, and compared the responses for the different stimuli and behavioral conditions (passive, unpredictable A500, and predictable A1000 trials). Our results revealed that the acoustic responses of the AC cells varied systematically with the different stimuli. However, when the responses to AM noises and ripple stimuli were aligned to the animal’s reaction time, they became highly similar, thus revealing a clear top-down modulation signal. This top-down signal resulted to be similar for the A500 and A1000 trials, implying that an additional prediction signal (which was obvious from the considerable changes in the reaction-time distributions) was not encoded by the AC cells. Our results further indicated that the top-down signals were invariant to the different acoustic modulations and of the stimulus type (spectral vs. temporal modulations). Finally, we demonstrated that the bottom-up and top-down interactions were best described by the multiplicative gain model of Eq. 1.6.
2. SPECTROTEMPORAL RESPONSE PROPERTIES OF AUDITORY CORTEX NEURONS IN AWAKE MONKEY

Roohollah Massoudi
Marc M. Van Wanrooij
Huib Versnel
A. John Van Opstal

under revision by PlosOne (2014)
2.1 INTRODUCTION

The core auditory cortex (AC) may be considered the final stage of acoustic processing in the ascending auditory pathway. Most acoustic properties, such as sound spectrum, sound duration, and sound level, but also perceptual qualities, like pitch or binaural disparities, are already processed at subcortical levels (Palmer, 2007). To study the presumed role of AC in higher-order auditory processing thus requires the use of complex sounds that vary in both the spectral and temporal domains. Useful naturalistic stimuli that meet these requirements are so-called dynamic ripples (Kowalski et al., 1996a; deCharms et al., 1998; Depireux et al., 2001; Linden et al., 2003; Atencio and Schreiner, 2008). Although ripples lack the higher-order statistical properties of natural sounds (Atencio and Schreiner, 2008) and the 1/f dynamics (Garcia-Lazaro et al., 2006) that cortical neurons might be sensitive to, they are particularly useful to quantify and compare neural tuning characteristics, because they can be described and varied in a straightforward, parametric way.

Ripples share many properties with natural sounds, and they provide adequate information to extract a neuron’s spectrotemporal receptive field (STRF). The STRF is a linear representation of the joint temporal and spectral sensitivity of an auditory neuron (Aertsen and Johannesma, 1981; Eggermont, 2011). An interesting question is whether a neuron treats time and spectrum as independent variables. If so, the STRF is separable, because it can be decomposed into the product of a spectral and a temporal sensitivity function. In contrast, when a neuron’s STRF is inseparable, it is most sensitive to a particular combined change in time and frequency (like in an upward or downward FM sweep). While recordings in cat and ferret AC have indicated inseparable STRFs in a significant proportion of neurons, most AC cells have separable STRFs (Depireux et al., 2001; Atencio and Schreiner, 2008). Also in the midbrain inferior colliculus (IC) of the monkey, the majority of cells have separable STRFs, with about 30% of the neurons being inseparable (Versnel et al., 2009).

As the STRF is a linear response kernel of the cell under study, it can also be used to test a cell’s linearity by predicting its responses to arbitrary sounds. Versnel et al. (2009) reported a considerable linear contribution to the neural responses in a large fraction of IC neurons recorded in alert monkeys. Although studies have also reported linear properties for AC neurons (Kowalski et al., 1996b; deCharms et al., 1998; Schnupp et al., 2001), other reports have demonstrated non-linear characteristics (Machens et al., 2004; Ahmed et al., 2006; Atencio and Schreiner, 2008, 2012). For example, Atencio and Schreiner (2008) demonstrated that so-called fast-spiking units are more separable and more linear than regular units. On the other hand, they did not obtain significant differences in separability or linearity between neurons with either broad or narrow frequency tunings (Atencio and Schreiner, 2012). Note that most studies have been performed in anesthetized animal models. Anesthesia, however, may affect crucial response properties, such as inhibitory mechanisms...
(Populin, 2005), which could in turn affect spectral-temporal separability and linearity of the cells (Andoni et al., 2007), and their sustained responsiveness (Wang et al., 2005).

Studies of awake monkey AC cells have so far focused on different aspects of neural responses, often in the context of complex sounds, or perception and overt behavior, rather than on basic acoustic spectral-temporal response properties (see, however, (deCharms et al., 1998) which addressed linearity using STRFs, and (Tian et al., 2013) which distinguished cells based on basic onset and offset responses). For example, AC cells can be modulated by non-acoustic signals (Werner-Reiss et al., 2003; Brosch et al., 2005, 2011a; Recanzone et al., 2011; Niwa et al., 2012a; Massoudi et al., 2013), or by complex sounds like con-specific vocalizations (Recanzone, 2008; Sadagopan and Wang, 2009; Miller et al., 2010), or pitch (Bendor and Wang, 2010; Bendor et al., 2012). However, the properties of STRFs and linearity of awake monkey AC cells have so far not been studied in detail.

We therefore examined the STRFs of AC cells from two awake monkeys using dynamic ripples, and quantified selectivity to ripple density, ripple velocity and direction, as well as their spectral-temporal separability. By comparing the best frequencies for tone responses to those of ripple responses, and the predictability of responses to a set of natural stimuli, we also assessed a neuron’s response linearity. By comparing our earlier results from monkey IC (Versnel et al., 2009) with the current AC recordings, we aimed to clarify the IC-to-AC spectrotemporal processing of sound.

### 2.2 MATERIALS AND METHODS

-----------------------------------------------------

#### 2.2.1 SUBJECTS

The rhesus monkeys (Macaca mulatta, Monkey J; 7-9 kg, and Monkey T; 8-10 kg) were obtained from the national primate research center (BPRC) in Rijswijk, the Netherlands. Animals were housed in the Central Animal Facility (CDL) of the Radboud University, and participated in the recording sessions for about two years. Experiments were conducted in accordance with the European Communities Parliament and Council Directive (September 22, 2010, 2010/63/EU). All experimental protocols were approved by the local Ethics Committee on Animal Research of the Radboud University Nijmegen (RU-DEC, ‘Radboud University Dier Experimenten Commissie’).
Monkeys were pair-housed to facilitate normal interactive behavior, including grooming. Their joined cages measured 1.6 x 2.4 x 2.0 m (hxwxd), and cage enrichment was provided in the form of a swing, plastic 3D puzzles, a mirror, and tools. The room, in which four such paired cages were placed, was further enriched with background pop-music. To promote foraging behavior, small seeds were dispersed across the floorbed on a daily basis. All animals in the room received a fixed amount of dry food daily, and when outside the experimental sessions they each had daily access to a bottle containing 400 ml of water.

About 24 hours before the start of an experimental session, water intake of the monkey was limited to 20 ml/kg. In the experiment, the animal earned a small water reward of 0.2 ml per successful trial. We ensured that the animals earned at least the minimum of 20 ml/kg on an experimental day. After an experimental session, water was supplemented to the required minimum amount, if needed, and the animal received additional pieces of fruit. In weekends, the animals’ fluid intake was increased to 400 ml daily.

To monitor the animal’s health status, we kept quantitative records of body weight, and water and food intake. Expert veterinarian assistance was available on site. The facility’s expert animal technician performed the surgeries on the animals, described below. Quarterly testing of hematocrit values ensured that the animal’s kidney function remained within the normal physiological range. Our procedures followed the water-restriction protocol of the Animal Use and Care Administrative Advisory Committee of the University of California at Davis (UC Davis, AUCAAC, 2001). Whenever an animal showed signs of discomfort, or illness, experiments were stopped and the animal was treated until the problem was solved.

The animals were sacrificed at the end of the study by an i.v. injection of 1 ml of heparine, followed by an overdose of pentobarbital. The animals were then perfused, and their brains were removed for further anatomical study.

2.2.2 SURGICAL PROCEDURES

After completing training in order to respond to spectrotemporal ripple stimuli at a performance level of at least 80% (see details in Massoudi et al., 2013), the animal underwent surgery under full anesthesia and sterile conditions. Anesthesia was maintained by artificial respiration (0.5% isoflurane and N2O), and additional pentobarbital (IV; 3 mg/kg/hour), ketamine (IM; 0.1 ml/kg), and fentanyl (IV; 20 μg/kg/hour) were administered. A stainless steel recording chamber (12 mm diameter) was placed over a trepanned hole in the skull (10 mm diameter). The orientation and coordinates of the chamber were directed to the auditory cortex, as determined on the basis of MRI images. The chamber allowed a vertical approach of the left AC. A stainless-steel bolt, embedded in dental cement on the skull, allowed firm fixation of the head during recording sessions.
2.2.3 EXPERIMENTAL SETUP

The head-restrained monkey sat in a primate chair within a completely dark and sound-attenuated room ($2.45 \times 2.45 \times 3.5$ m), while a glass coated tungsten microelectrode (impedance 1-2 MΩ; Alpha Omega, Ubstadt-Weiher, Germany) was carefully positioned and lowered into the brain through a stainless steel guide tube by an electronically-driven stepping motor (National Aperture Inc. MM-3M-F-1). Electrode signals were grounded to a contact mounted in the skull. The analog electrode signal was amplified (BAK Electronics; model A-1), band-pass filtered between 0.1 and 12.5 kHz (custom-built 8th order LP Butterworth filter in series with Krohn-Hite, model 3343, 100 Hz HP cut-off), and monitored through a speaker and oscilloscope. The raw signal was then digitized (at 25 kHz, A/D convertor, TDT2 system; module AD-1; Tucker-Davis Technologies). An automated algorithm that was controlled by the Brainware program (BrainWare, V 9.07 for TDT, under Windows 98; DELL PC) detected individual action potentials. Data analysis and spike sorting was performed offline in MATLAB (version 7.14.0.739, R2012a, Natick, MA, USA).

2.2.4 SOUND STIMULI

Sound stimuli were digitally generated at a sampling rate of 100 kHz and delivered via the Brainware software package and TDT2 hardware. A trigger, provided by a TG6 module, started sound presentation (DA1, low-pass filtered at 20 kHz through a TDT2-FT6 module), and spike data acquisition. A speaker (Blaupunkt PCxg352, flat frequency characteristics within 5 dB between 0.2 and 20 kHz), positioned at the frontal central position at a distance of 1.0 m from the monkey, presented the stimuli; sound levels were set by two programmable attenuators (PA4).

The ambient background acoustic noise level was about 30 dBA. Acoustic foam that was mounted on the walls, floor, ceiling, and every large object in the room effectively absorbed reflections above 500 Hz. In this study, we presented two types of acoustic stimuli: (1) pure tones, and (2) frozen static noise, followed by a spectrotemporal ripple.

*Tones.* Pure tones lasted for 150 ms and were presented over a frequency range from 250 to 16000 Hz at 4 different sound levels (10, 30, 50, and 70 dB sound pressure level (SPL)). Trials were presented at least four times in a randomized manner. The frequency-tuning curve of a neuron was determined from the average firing rate for each tone across all sound level presentations. The best frequency ($B_{\text{tone}}$) of each neuron was taken at the maximum of the tuning curve. Characteristic frequency (CF) was defined as the frequency that produced a response higher than mean plus 2 SDs of the baseline activity at the lowest intensity. The cell’s response onset latency was defined as the moment after the pure tone onset at which the firing rate exceeded the mean baseline activity plus twice its standard deviation (SD) for the first time for at least 10 milliseconds. The peak latency was the time at which the firing rate reached its maximum.
**Ripples.** The ripple stimuli consisted of a broadband complex of 126 spectral components, equally distributed (20/octave) from 250 Hz to 20 kHz (Depireux et al., 2001; Versnel et al., 2009). All components had random phase. The ripple envelopes were sinusoidally modulated in the spectrotemporal domain. The amplitude of each component was described as follows:

\[ S(t,x) = 1 + \sin(2\pi \omega t + 2\pi \Omega x) \]  

(Eq. 2.1)

with \( t \) time (in S), \( x \) position of the spectral component in octaves above the lowest frequency (250 Hz), \( \omega \) ripple velocity (Hz; temporal modulation), \( \Omega \) ripple density [cycles/octave, or c/o; spectral modulation]. The set of 55 different ripples used in our study consisted of all combinations of 11 different ripple densities, \( \Omega \) in \([-2.0: 0.4: +2.0]\) cyc/oct, and 5 different velocities, \( \omega \) in \([8: 8: 40]\) Hz.

A negative density corresponds to an upward direction of the spectral envelopes, a positive density to a downward direction, and \( \Omega = 0 \) means a pure amplitude-modulated (AM) spectrally flat complex (Versnel et al., 2009). The ripples with \( \Omega \neq 0 \) are referred to as moving ripples, and the ripples with \( \Omega = 0 \) as AM complexes. For all ripples the modulation depth was 100%. The sound level was 60 dB SPL, and the duration was 1000 ms.

**Vocalizations.** Three different macaque calls and three different bird calls were presented to both monkeys. The bird calls and macaque vocalizations are described in detail in Versnel et al. (2009). Briefly, all vocalizations had been resampled to 50 kHz. Artifacts and background noises were removed by cutting and high-pass filtering. The calls were presented at three sound levels (40, 50, and 60 dB SPL), their durations varied from 300 to 1600 ms, and the number of repetitions in each recording was at least 10.

---

**2.2.5 EXPERIMENTAL PARADIGMS**

Neural responses were measured while monkeys were exposed to the spectral-temporal ripples. The data presented here were collected during recording sessions in which also the AC responses were recorded during a behavioral task (see Massoudi et al., 2013). Each stimulus started (Fig 3.1, sound onset) with a static epoch (\( \omega=0 \) Hz, \( \Omega=0 \) c/o), which smoothly changed into a dynamic pseudo-randomly selected ripple that lasted for 1000 ms (Eq. 2.1). The static noise was frozen within each block of trials, it had the same sound level (60 dB SPL) as the dynamic ripple, and its duration was 500 ms.

Each trial started automatically with the static noise. Data collection started 300 ms before sound onset, and ended 700 ms after sound offset, yielding a total recording duration of 2500 ms. The number of trial repetitions was between 4 and 10.
2.2.6 CHARACTERIZATION OF RECORDING SITES

Although we cannot with certainty identify the exact AC subdivision(s) in which we encountered individual neurons, we are confident that we recorded from the AC core (primary auditory cortex A1 and its immediate rostral part, area R) and the caudomedial field (CM) for the following reasons: 1) MRI scans were used for stereotaxic placement of the recording chamber; the subsequent coordinates of the successful recording sites within the chamber corresponded closely to the stereotaxic coordinates of A1 as provided by the atlas of the Rhesus monkey brain by (Paxinos et al., 2000); 2) before reaching an AC recording site there was a physiologically silent period, corresponding to the gap between upper and lower parts of the lateral sulcus (Kaas and Hackett, 2000); 3) tone-onset latency of the recorded sites was 22.6 ± 5.9 and 23.6 ± 5.6 ms for monkey J and T, respectively; 4) almost all neurons (96%) responded well to pure tones (BF: 250 – 16000 Hz); 5) The pure-tone tuning bandwidths for monkey J and T were 1.5 ± 1.2 and 1.5 ± 1.3 octaves, respectively; 6) The pure-tone thresholds for monkey J was 21 ± 13 dB SPL, and for monkey T was 23 ± 12 dB SPL. These tuning characteristics all fall in the same ranges as reported by Recanzone et al. (2000) for behaving monkeys in AC areas A1, R and CM (Massoudi et al., 2013). 7) Finally, we reconstructed tonotopic maps of the recording sites in both animals. The maps demonstrated a systematic increase in CF from anterior to posterior locations over a spatial extent that corresponded well to earlier studies (e.g., (Bendor and Wang, 2008; Niwa et al., 2012a; Dong et al., 2013). Figure 2.1 shows the tonotopic map obtained for monkey T, in which the CF increases along the medial-lateral axis, indicative for an area between A1 and R (Bendor and Wang, 2008). The results for monkey J were similar, albeit that the CF was more stable along the medial-lateral axis. This indicates that recordings in this animal were most likely taken from A1 (not shown).

Our database consisted of 426 cells from which we successfully recorded during presentation of the entire set of 55 ripples (monkey J: n=178; monkey T: n=248).

Figure 2.1. Reconstructed tonotopic map for the cells of monkey T. Each bin represents the average best frequency obtained at the recording location. Characteristic frequency (color coded) increases systematically from anterior to posterior recording sites, and along the medial-lateral direction, indicative for primary auditory cortex (A1) and area R.
2.2.7 DATA ANALYSIS

2.2.7.1 SPIKE-DENSITY FUNCTIONS

To convert each spike-raster plot into a continuous spike-density function, we first binned the recorded spike times into a digital sequence with a time resolution of one millisecond, and then convolved each spike with a Gaussian kernel with a standard deviation of 5 ms.

2.2.7.2 THE SPECTROTEMPORAL RECEPTIVE FIELD (STRF)

We estimated a cell’s STRF from ripple stimuli by using the same off-line method as described by Versnel et al. (2009) and Massoudi et al. (2013). First, detected spikes were sorted and binned into peri-stimulus time histograms. We then wrapped the 900-ms response window (the ripple duration, excluding the first 100 ms to exclude transient onset responses) into 32-bin period histograms, in which the ripple velocity determined the period as 1/ω. We subsequently performed a fast Fourier transformation on the period histograms. The magnitude $A(\omega, \Omega)$ (spikes/s) and phase $\Phi(\omega, \Omega)$ (rad) of the period histograms were derived from the first harmonic of the resulting Fourier spectrum to generate the spectrotemporal transfer function:

$$T(\omega, \Omega) = A(\omega, \Omega) \cdot \exp(i \cdot \phi(\omega, \Omega))$$ (Eq. 2.2)

The 2D inverse Fourier transformation of $T(\omega, \Omega)$ then produces the spectrotemporal response field, or STRF, of the cell:

$$\text{STRF}(t, x) = \text{FFT}^{-1}[T(\omega, \Omega)]$$ (Eq. 2.3)

with $x$ the frequency in octaves (between 0 and 2.5 in 0.25 octave steps), and $t$ running from 0 to 125 ms (at 12.5 ms resolution). The spectral dimension (abscissa) of the STRF reflects the frequency tuning, and the temporal direction (ordinate) reflects the cell’s linear impulse response. Note that the frequency range of the STRF is determined by the step size of the ripple densities employed in the experiment: $[\text{range } x] = 1/ [\text{step size } \Omega]$. Typically, the step size was 0.4 cycles/octave, resulting in a frequency range of 2.5 octaves. Likewise, the temporal range is determined by the resolution in applied ripple velocities: $[\text{range } t] = 1/ [\text{step size } \omega]$, leading to 125 ms. The position of the frequency range (in Hz) is ambiguous, as the lower frequency of the STRF could be either 250 Hz (the lowest component in the ripple stimuli), or multiples of 2.5 octaves above 250 Hz (i.e., at 1414 or 8000 Hz). The pure-tone responses were used to resolve this ambiguity.
2.2.7.3 PHASE LOCKING TO AMPLITUDE MODULATIONS

A measure, \( q \), to quantify the strength of phase locking of a cell to a particular ripple stimulus was computed as follows:

\[
q = \frac{A_i}{\sqrt{\sum_{i=1}^{16} |A_i|^2}} \quad \text{(Eq. 2.4)}
\]

with \( A_i \) the amplitude of the \( i \)-th harmonic of the ripple’s period (1/\( \omega \)). The parameter \( q \) reflects the quality for the best sinusoidal fit with the period histogram. If \( q=1 \) the period histogram corresponds to a pure sinusoid. If \( q=0 \) the cell has no phase-locking at all. We applied \( q \) as an index to judge the sensitivity of the neuron to the ripple stimulus. Thus, the neuron yielded 55 \( q \)-values: \( q_1 \) to \( q_{25} \) corresponds to upward ripples (\( \Omega < 0 \)), \( q_{26} \) to \( q_{30} \) to the AM complexes (\( \Omega = 0 \)), and \( q_{31} \) to \( q_{55} \) to downward ripples (\( \Omega > 0 \)). The sensitivities to all 50 moving ripples (\( \Omega \neq 0 \)) and 5 AM complexes (\( \Omega = 0 \)) were compared separately with a statistical criterion that was set at \( p = 0.01 \). Simulating random responses then resulted in criterion levels of the 25th percentile \( q_{\text{crit}} = 0.387 \) for \( \Omega \neq 0 \), and the median \( q_{\text{med}} = 0.376 \) for \( \Omega = 0 \). Thus, for instance, it is unlikely (\( p < 0.01 \)) that \( q_{\text{crit}} > 0.387 \) if a neuron responds in a random way to moving ripples.

2.2.7.4 BEST RIPPLE DENSITY AND VELOCITY AND DIRECTION SELECTIVITY

Three response parameters were derived from the transfer function \( T \) (Eq. 2.2): the best ripple velocity \( \omega_B \) of a cell, its best ripple density \( \Omega_B \), and its direction selectivity \( D \). The parameters \( \omega_B \) and \( \Omega_B \) were determined as the \( \Omega \) and \( \omega \) at which we obtained the largest response amplitude (max MTF). We computed the direction selectivity, \( D \), by (Mendelson and Cynader, 1985):

\[
D = \frac{R_{\text{up}} - R_{\text{down}}}{R_{\text{up}} + R_{\text{down}}} \quad \text{(Eq. 2.5)}
\]

with \( R_{\text{up}} \) and \( R_{\text{down}} \) the sum of the response amplitudes to the upward (\( \Omega < 0 \)) and downward (\( \Omega > 0 \)) ripples, respectively. We subsequently determined the ripple/AM ratio to indicate the preference of a neuron for moving ripples versus AM complexes. To that end, the responses to the best \( \Omega \neq 0 \) were summed for all presented velocities and divided by the sum of the responses to \( \Omega = 0 \) at all velocities.

Finally, a best frequency (BF_{strf}) and latency were derived as the frequency and latency at which the STRF had its maximum response.

2.2.7.5 SEPARABILITY

We examined whether the two-dimensional ripple transfer function \( T(w, \Omega) \) can be obtained by the product of two separate, one-dimensional, transfer functions, the temporal function \( F(w) \) and the spectral function \( G(\Omega) \). The separability analysis was performed for the complete transfer function.
\(-2 \leq \Omega \leq 2\) cyc/oct, as well as for each of the two quadrants of the transfer function \((\Omega < 0\) and \(\Omega > 0\)). The former measure targets full separability, whereas the latter quantifies the so-called quadrant separability (Depireux et al., 2001; Versnel et al., 2009). We applied two different methods to examine the extent of separability.

First, we performed singular value decomposition (SVD) of the ripple transfer function (Depireux et al., 2001). For a fully separable function, the following equation applies:

\[
T(w, \Omega) = F(w) \times G(\Omega) \quad \text{(Eq. 2.6)}
\]

Briefly, the SVD method decomposes \(T(w, \Omega)\) as follows: \(T(w, \Omega) = F(w) \times S \times G(\Omega)\), with \(S\) a diagonal matrix with eigenvalues \(\lambda_i\) with \(\lambda_1 \geq \lambda_2 \geq \lambda_3\) and so forth. For a separable function, \(\lambda_1 = 1\), and the other eigenvalues are zero. As a measure to quantify separability, we defined the inseparability index, \(\alpha\), as used by (Depireux et al., 2001; Linden et al., 2003; Versnel et al., 2009):

\[
\alpha = 1 - \frac{\lambda_1^2}{\sum_{i=1}^{n} \lambda_i^2} \quad \text{(Eq. 2.7)}
\]

with \(\lambda_i\) the eigenvalues. The transfer function is inseparable when \(\alpha = 1\) and fully separable for \(\alpha = 0\).

Second, a temporal transfer function, \(F_B(\omega)\), at \(\Omega = \Omega_B\), and a spectral function, \(G_B(\Omega)\), at \(\omega = \omega_B\) were selected from the recorded complete transfer function \(T(\omega, \Omega)\). We then computed an estimate \(T_{est}(\omega, \Omega)\) according to Eq. 2.6, \(T_{est}(\omega, \Omega) = F_B(\omega) \times G_B(\Omega)\), and correlated the resulting STRF\(_{est}\) (Eq. 2.3) with the original STRF. The correlation coefficient, \(\rho\), was used as an additional measure of separability.

### 2.2.7.6 PHASE FUNCTIONS

For a majority of transfer functions, the phase functions \(\Phi(\omega, \Omega)\) appeared to be linear, and could thus be described for each ripple direction as follows (Depireux et al., 2001; Versnel et al., 2009):

\[
\Phi(\omega, \Omega) = -2\pi\omega \tau + 2\pi\Omega x + \chi \quad \text{(Eq. 2.8)}
\]

The phase intercepts for downward and upward sweep directions, \(\chi_{down}\) and \(\chi_{up}\), contain temporal and spectral components, denoted by \(\theta\) and \(\varphi\), respectively. This can be written as follows (Depireux et al., 2001; Versnel et al., 2009):

\[
\chi_{down} = -\theta + \varphi
\]

and

\[
\chi_{up} = \theta + \varphi \quad \text{(Eq. 2.9)}
\]

The separated phase functions that resulted from the SVD analysis were used to derive the parameters \(\tau\), \(x\), \(\theta\), and \(\varphi\). The slope \(\tau\) corresponds to the temporal position (group delay), and the slope \(x\) corresponds to the spectral position (reflecting BF); these parameters can be determined for both ripple directions. The phase constants \(\theta\) and \(\varphi\) are related to asymmetry of the STRF around BF. The parameter \(\theta\) reflects inhibition before and/or after excitation at BF: \(\theta < 0\) reflects excitation at
onset followed by inhibition, and $\theta > 0$ reflects inhibition at onset followed by excitation. The parameter $\varphi$ reflects the extent of sideband inhibition above and/or below BF: $\varphi < 0$ indicates inhibition dominant below BF, and $\varphi > 0$ indicates inhibition dominant above BF (see Depireux et al., 2001, for details).

### 2.2.7.7 RESPONSE PREDICTIONS

From the STRF, a linear estimate of the temporal response pattern of a cell, $R(t)$, to any sound stimulus can be predicted by time convolution and spectral integration with the stimulus spectrogram, $S(x, t)$:

$$R(t) = \int_{F_{\text{MIN}}}^{F_{\text{MAX}}} dx \int_{0}^{\infty} dv \cdot \text{STRF}(x, v) \cdot S(x, t - v) \quad \text{(Eq. 2.10)}$$

with $F_{\text{MIN}}$ and $F_{\text{MAX}}$ the minimum and maximum frequencies of the spectral range of the STRF, respectively. Applying this equation, we predicted responses to the six animal vocalizations and compared the predictions with the recorded responses. In case of STRFs with a 2.5 octave spectral range (the default recording), we applied a spectral shift of the STRF to align its maximum with the center of the spectral range, and aligned the stimulus spectrograms accordingly (see Fig. 2.13A). Convolving the temporal domain for each spectral bin resulted in a predicted response spectrogram (see Fig. 2.13B). Finally, the response spectrogram was summed along the spectral domain, which yielded the predicted response (see Fig. 2.13C). The bin size, $(df, dt)$, of $S(x, t)$ in both the spectral and temporal dimensions was set according to the bin size of the STRF, typically $df = 0.25$ octave, and $dt = 12.5$ ms.

### 2.3 RESULTS

#### 2.3.1 GENERAL PROPERTIES

We recorded neural activity during ripple presentation from 426 AC neurons (248 in monkey T, 178 in monkey J), which showed at least an acoustically elicited onset or offset response. Further systematic responsiveness of neurons to moving ripples ($\Omega \neq 0$) and AM complexes ($\Omega = 0$) was quantified by the phase-locking index $q$ (Eq. 2.4; see Methods; summarized in Table 2.1). Forty-one cells (~10%) responded to both moving ripples and AM complexes. Three cells responded only to moving ripples, while 115 cells (27%) only to AM complexes. More than half of the neurons (63%, $N = 267$) demonstrated no phase-locking response to any ripple, and were therefore excluded for further analysis.

Tone responses were elicited for 246 cells out of 255 cells recorded during tone stimulus presentation (Table 1; using the same response criteria as in Recanzone et al., 2000; see Methods).
Table 2.1: Number of neurons sensitive to moving ripples (Ω ≠ 0) and AM complexes (Ω = 0) according to their q-value (q>0.387 for Ω ≠ 0 and q>0.376 for Ω = 0). The numbers in parentheses indicate the number of the cells that were also responsive to pure-tone stimuli. Bold numbers indicate the cells that were selected for further analysis (N=159 for ripples, and N=120 to both tones and ripples). The number between brackets represents the number of cells for which also tone responses were recorded (total of 255).

<table>
<thead>
<tr>
<th>N = 426 [255]</th>
<th>Ω ≠ 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>responsive</td>
</tr>
<tr>
<td>Ω = 0</td>
<td>41 (39)</td>
</tr>
<tr>
<td>non-responsive</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Total</td>
<td>44 (42)</td>
</tr>
</tbody>
</table>

Figure 2.2 shows the pure-tone responses and resulting tuning curves of four example cells. The raster plots (left-hand columns) show each neuron’s spiking activity, which significantly increased above background during tone presentation (from 300-450 ms; blue lines). The four cells showed a clear short-latency onset peak, whereas cell T77 also had a clear offset response (Fig. 2.2D). The tuning curves (right-hand columns) represent the mean response (averaged over the tone duration) as function of sound level and frequency. Each of the cells was tuned to a narrow range of frequencies, and their tuning widths typically broadened at higher sound levels. Two cells responded to their preferred frequency at all presented sound levels (10-70 dB SPL; Figs. 2.2A and D), while the other two only responded at the higher sound levels (Figs. 2.2B and C). We encountered both monotonic (Figs. 2.2B and 2D) and non-monotonic (Fig. 2.2A) rate-level characteristics in our sample of cells. In general, the AC cells in our study showed high sensitivity to pure tones in terms of a low sound level threshold, short response latency, and/or narrow frequency tuning.
Figure 2.2. Example tone responses. The raster plot (left) and tuning-curve (right) of four example cells: A. J112, B. J113, C. T81, and D. T77. Note the onset responses in the raster plots. The tone duration (150 ms) is indicated by the blue lines. The responses are sorted first by frequency (0.25 – 16 kHz), and second by the sound levels (10-70 dB SPL). The tuning curves represent the responses for the various sound level – frequency combinations. The excitatory response bandwidths generally increase by sound level.

The Best frequencies (BFs) ranged from 0.25 to 16 kHz, with the majority of encountered BFs below 2 kHz (76%; Fig. 2.3).
Figure 2.3. Best frequency (BF) range. From 255 cells responsive to either moving ripples or AM complexes, 246 cells were also responsive to pure tones. The BFs spread across the presented tone stimuli, however, the majority preferred the lower frequencies.

In Figure 2.4 we present the magnitude transfer functions (MTFs) and corresponding STRFs for the four example cells shown in Figure 2.2. Because the spectral range of the STRFs extended to 2.5 octaves (see Methods), the spectral position of the STRFs is ambiguous. To solve this ambiguity, we used the tone-evoked tuning curves (right-hand column), to determine the appropriate spectral range for each cell (see Material and Methods). An important feature revealed by the transfer functions is each cell’s selectivity to a limited range of ripple densities, \( \Omega \), and velocities, \( \omega \), which strongly varied among the neurons. We encountered neurons that preferred low (4B-C) or high ripple densities (Fig. 2.4A), and neurons that preferred low (Fig. 2.4A, D), or high ripple velocities (Fig. 2.4B). Also, direction selectivity varied, as neurons could strongly prefer either down- or upward moving ripples (Figs. 2.4C-D, in these examples upward), or responses to both directions were equally strong (Figs. 2.4A-B). Two of the neurons responded significantly better to dynamic ripples than to AM complexes (Figs. 2.4A, D), while the two other neurons responded to both moving ripples and AM complexes (Figs. 2.4B, C). The STRFs generally showed excitation for a narrow range of frequencies with latencies between 10 and 30 ms, so that a BF\(_{strf} \) could easily be obtained (Methods, and see below). Inhibition occurred around the time of or after excitation (Figs. 2.4A-D).
Interestingly, the frequency tuning reflected by the STRF did not agree well with the neuron’s tuning to pure tones. The difference between best frequencies could be as large as 1.25 octaves for cell J112 (Fig. 2.4A), one octave for cell J113 (Fig. 2.4B), 0.5 octave for cell T81 (Fig. 2.4C), or 0.75 octave for cell T77 (Fig. 2.4D). Such a discrepancy implies nonlinearity in the response behavior of AC cells (see also below).

Figure 2.4. Magnitude transfer functions, corresponding STRFs, and tone responses of the cells shown in Figure 2.2. A. J112, example of low best ripple velocity ($\omega_B = 8$ Hz), and high best ripple density (\(\Omega_B = -2\) cyc/oct); furthermore it prefers only moving ripples. Note the large difference (1.25 octaves) between BF\_tone and BF\_strf. B. J113, example of high best ripple velocity ($\omega_B = 32$ Hz) and low best ripple density (\(\Omega_B = 0\) cyc/oct). It is responsive to both moving ripples and AM complexes. The difference between BF\_tone and BF\_strf is one octave. C. T81, example of asymmetric transfer function with preference for upward direction (\(\Omega < 0\)) with \(\Omega_B = -0.4\) cyc/oct and $\omega_B = 32$ Hz. The difference between BF\_tone and BF\_strf is 0.5 octaves. D. T77, another example of preference for low velocity moving ripples ($\omega_B = 8$ Hz; \(\Omega_B = -0.8\) cyc/oct).

To quantify this point in more detail for the population of cells, we compared the BFs for all neurons that were stimulated with tones as well as ripples (N=120, Fig. 2.5). Only 23 of the 120 neurons had good agreement between their respective BFs (inserted table), with differences less than 0.25 octaves (i.e., at or below the spectral resolution of the STRFs). For 35 cells the BFs fell between 0.25 and 0.5 octaves of each other, while for a majority of 62 neurons (51%) the correspondence between BFs was quite poor (differences >0.5 octave). These results applied to both animals.
Figure 2.5. BF\textsubscript{tone} and BF\textsubscript{strf} similarity. BF obtained from the ripple-based STRF against the BF to tones for 120 AC neurons that responded both to tones and ripples. The solid line represents the unity line (y=x), the dotted lines correspond to the BF differences at half octaves (y=x±0.5 octave), and the dashed lines correspond to the BF differences at half octaves (y=x±1.25 octave). For only 23 cells (19%) the differences fell within the boundaries of ±0.25 octaves. Inset table: distribution of difference frequencies in four bins.

2.3.2 DISTRIBUTIONS OF RIPPLE RESPONSES

For each neuron responding to ripples and/or AM complexes (159 neurons; 83 from monkey T and 76 from monkey J) we determined the best ripple velocity, \(\omega_B\), best ripple density, \(\Omega_B\), and the cell’s direction selectivity, \(D\), from the ripple-response transfer function. The distributions of the extracted parameters, shown in Figure 2.6A-C, display a wide range of ripple selectivities in monkey AC. The distribution of \(\Omega_B\) was skewed toward low densities, as a large majority of neurons (79%) preferred the two lowest ripple densities (0–0.4 cyc/oct; Fig. 2.6A). As shown in Figure 2.3, most of the recorded AC neurons had relatively low pure-tone BFs, while neurons with a low BF tend to have broader tuning curves, i.e. low preferred densities. To test for a potential relation between BF\textsubscript{tone} and preferred density, we plotted the tone-evoked BF vs. \(\Omega_B\) for the 120 AC cells for which we also recorded tone responses. Figure 2.6D shows the absence of a significant correlation (t\textsubscript{119} = -15, p <
0.0001; \( r = -0.07, p = 0.43 \), implying that AC neurons with broad tuning properties (low \( \Omega_B \)) do not necessarily have low BF\(_{tonic}\).

The distribution of best ripple velocities, \( \omega_B \), in the range routinely tested in our experiments (8–40 Hz) was roughly dominated by lower velocities, with 34% preferring the lowest applied velocity in this study (8 Hz; Fig 2.6B). We observed no correlation between \( \omega_B \) and BF\(_{tonic}\) (\( r = 0.08, p = 0.37 \); Fig. 2.6E).

Direction selectivity \( D \) of the neurons was symmetrically distributed around zero (median \( D = 0.002 \)), and a majority of neurons (62%) had similar preference for upward and downward moving ripples (\(-0.1 < D < 0.1 \); Fig. 2.6C). We obtained a strong direction preference (\( D < -0.2 \) or \( D > 0.2 \)) for 13 neurons (8%). The joint distribution of \( \omega_B \) and \( \Omega_B \) indicated that ripple selectivity did not cover a wide range of spectral-temporal combinations as few cells were responsive to high \( \omega_B \) and \( \Omega_B \); furthermore, selectivity to \( \Omega \) did not correlate with selectivity to \( \omega \) (Fig. 2.6D). Moreover, direction selectivity was uncorrelated with either \( \omega_B \) (\( t_{158} = 23, p < 0.0001 \)) or \( \Omega_B \) (\( t_{158} = 9, p < 0.0001 \); not shown).

**Figure 2.6.** Ripple parameters derived from the magnitude transfer function. Distribution of the A. best ripple density (\( \Omega_B \)), B. best ripple velocity (\( \omega_B \)) and C. direction selectivity (D). D and E. The relation between best tone frequency versus \( \Omega_B \) and \( \omega_B \) respectively. F. combined distribution of \( \Omega_B \) versus \( \omega_B \). The size of the data points is proportional to the number of cells. For A-C and F the number of cells was 159, and for D and E was 120 for which we had tone responses.
The scatter plot in Figure 2.7 (left) shows the ripple/AM ratio, which quantifies ripple selectivity of the neurons, as a function of best ripple density. As expected, neurons with $\Omega_B = 0$ tended to have a ratio below one, while neurons with $\Omega_B > 0$ had a ratio above one. Note that exceptions also occurred, as $\Omega_B$ was obtained at a single velocity, while the ripple/AM ratio was computed across all velocities. The ratio did not depend on $\Omega_B > 0$, which means that, on average, neurons with a low $\Omega_B$ responded equally strong to AM complexes as neurons with a high $\Omega_B$. The distribution of the ripple/AM ratio (right-hand side of Fig. 2.7) was unimodal and skewed towards low ratios, with the peak at a ratio of 0.5 meaning that most neurons (60%) preferred AM complexes rather than moving ripples. Most of the cells preferring ripples had their ratio close to one (median = 1.5) implying that they responded nearly as well to spectrally flat AM complexes as to moving ripples (cf. Table 1).

![Figure 2.7](image)

**Figure 2.7.** Moving ripple and AM complexes ratio. Ratio of responses to moving ripple and AM complexes as a function of $\Omega_B$ (left), and its distribution (right). Bin size, bs, is chosen according to $bs = \text{range} / \sqrt{n}$ with range computed over data.

### 2.3.3 SPECTRAL-TEMPORAL SEPARABILITY

For a separable transfer function the spectral transfer functions taken at fixed, different velocities ($w$) should be scaled versions of one another. The same holds for the temporal transfer functions taken at different densities, $\Omega$. We first quantified separability with the SVD method (see Materials and Methods). Figure 2.8 shows the result of the SVD analysis on the transfer function of neuron J140 for downward ripple directions ($\Omega > 0$). At velocities 16, 32, and 40 Hz the magnitude
curves peaked at ~0.8 cyc/oct, and for velocities 8 and 24 it peaked at ~1.2 cyc/oct. The phase curves were all very similar, running parallel to each other (Fig. 2.8A). The same property holds for upward ripple directions (not shown). These similarities imply a high degree of separability, which is indeed reflected by a low inseparability index (Eq. 2.7, Materials and Methods) for this neuron (α = 0.05). The responses can therefore be well described by the mean separated spectral and temporal tuning curves (corresponding to λ₁), obtained from SVD (Fig. 2.8B). Indeed, the STRF reconstructed by multiplying the temporal and spectral functions for the magnitude and phase characteristics of both directions and AM complexes (Eq. 2.6) were very similar to the original STRF (r = 0.98) (Fig. 2.8C).

Figure 2.8. SVD analysis results. SVD analysis of ripple transfer functions (magnitude and phase), shown for downward ripple direction (Ω>0) of an example cell. For this example cell, the eigenvalues (normalized with respect to λ₁) are λ₁=1, λ₂=0.18, λ₃=0.1, λ₄=0.06, λ₅=0.004, and the total inseparability index α=0.027. A. Transfer function of downward ripples in Cell J140. B. Separated spectral and temporal transfer functions corresponding to λ₁. C. The original STRF (left) and a reconstructed STRF (right) by multiplying separated transfer functions in section B according to Eq. 2.6. The correlation between the STRFs was 0.98.

Figure 2.9 shows the separated transfer functions (cf. with format Fig. 2.8B) of the four neurons of Figure 2.4. Three of these neurons responded to a sufficiently wide range of ripples to allow for a meaningful SVD analysis (median q over total transfer function >0.5; Figs. 9A, C, and D). Neuron J113 responded to only a limited range of ripples, mostly at zero density (q = 0.36; Figs. 2.4B
and 2.9B). In the quadrant-separated representation of the data, the shape of the transfer functions for the two ripple directions (solid and dashed curves, respectively) appeared quite similar. Note also that the phase functions could typically be well approximated by straight lines (in the range with substantial responses), as predicted by Eq 2.8, where only the intercepts of the phase curves could differ for the two directions. The slopes of the phase curves, which reflect the spectral position and group delay (see Eq. 2.8), were similar (note that slopes for Ω < 0 are shown inverted, for clarity). As all four neurons responded well to AM complexes, their AM transfer functions are plotted as dotted lines in the temporal graphs for comparison. Note that also the AM curves were similar to the separated ripple transfer functions, differing only in magnitude.

![Figure 2.9](image-url)

**Figure 2.9.** Separated spectral and temporal transfer functions of the neurons shown in Figures 2.2 and 2.4. Magnitude (spikes per seconds) and phase (radians) versus ripple density and ripple velocity. Solid curves represent functions for downward sweep direction (Ω > 0), and dashed curves represent functions for upward direction (Ω < 0). Dotted curves represent AM transfer functions (Ω = 0). For comparison purposes, the spectral data for Ω < 0 are shown along the positive Ω axis. The total and quadrant inseparability indices (α) are mentioned in the spectral phase plots.

We performed the SVD analysis (Eq. 2.7) for all cells in our population that had good phase locking over a wide range of ripples (median q > 0.4; N = 75 cells). Figure 2.10A shows that for a
large majority of selected neurons the transfer functions in both directions were separable ($\alpha_{\text{up and down}} < 0.2$ for 72% of the neurons). For 12 neurons (16%) good separability was obtained for one direction only ($\alpha_{\text{up or down}} < 0.2$). The direction for which the lowest $\alpha$ was obtained typically coincided with the preferred direction, D. A statistical simulation indicated that the probability of finding a value $\alpha < 0.3$ for a random 5×5 transfer function is <0.01. Using $\alpha = 0.3$ as a selection criterion for (in)separability, we found that six of the single-direction transfer functions were inseparable, and that only two neurons had inseparable transfer functions for both directions. The distributions per animal were very similar (gray and black dots and bars in Figure 2.10 represent each monkey). In conclusion, AC neurons show a significant degree of spectral–temporal separability for a single direction, a feature that has become known as quadrant separability (Depireux et al., 2001; Versnel et al., 2009).

The value of $\alpha$ for the complete transfer function normally exceeds the lowest index obtained separately for the two directions. Assuming quadrant separability, we also performed a statistical simulation on complete 11×5 transfer matrices that consisted of three components: two different quadrant-separable 5×5 matrices representing transfer functions to both directions, and a 1×5 matrix representing the AM noise transfer function. When the three components differed randomly, we obtained a probability <0.01 of finding a value $\alpha < 0.3$. Using $\alpha = 0.3$ as a criterion, we found that only 7% (5 cells) of the complete transfer functions were inseparable (Figure 2.9B). The inseparability indices for these transfer functions were confined to an intermediate range of values ($0.3 < \alpha < 0.5$).

![Figure 2.10. Inseparability indices. A. Distribution of quadrant separabilities, $\alpha_{\text{down}}$ vs. $\alpha_{\text{up}}$. B. Distribution of total separability index ($\alpha_{\text{total}}$). The number of cells, for which the $q > 0.4$, was 75.](image)

Because the data indicated quadrant separability for the AC neurons that responded well to moving ripples (see above), the full transfer function $T_{\text{sep}}(w, \Omega)$ can in principle be obtained by
determining only the temporal and spectral transfer functions, $F_B(w)$ and $G_B(\Omega)$, and subsequently applying Eq 2.6: $T_{sep}(w, \Omega) = F_B(w) \times G_B(\Omega)$ (cf. Fig. 2.8C). To test whether this alternative approach would yield a good STRF estimate, temporal and spectral slices $F(w, \Omega = \Omega_B)$ and $G(\Omega, w = \omega_B)$ were taken from the recorded transfer function, $T(w, \Omega)$, at their optimal velocity $\omega_B$ and density $\Omega_B$. The correlation $\rho$ between $T(w, \Omega)$ and $T_{sep}(w, \Omega)$ was computed as a measure of separability, with $\rho = 1$ indicating perfect separability (cf. $\alpha = 0$). We verified that $\rho$ was indeed significantly related to the inseparability index $\alpha$ ($r^2 = 0.95$; Fig. 2.11A). The median $\rho$ was 0.94, confirming the high degree of quadrant separability and validity of this approach to estimate the STRF of a separable cell (Fig. 2.11A). The correlation coefficients $\rho$ were also estimated separately for the amplitude and phase of the transfer function, resulting in a median $\rho$ of 0.88 for amplitude and a median $\rho$ of 0.67 for phase. This indicated that the separability of amplitude was higher than for phase.

![Figure 2.11](image)

**Figure 2.11.** Correlation between original and reconstructed STRFs ($\rho$). A. $\rho$ as a function of total inseparability index ($\alpha_{total}$). B. Distribution of $\rho$.  

### 2.3.4 PHASE FUNCTIONS

The phase functions, such as shown in Figures 2.8 and 2.9, could be approximated by straight lines. This feature, which results from phase locking to the ripple envelope, is further quantified in Figure 2.12A by fitting linear regression lines through the temporal and spectral phase data for upward (left) and downward (right) ripples (same neuron as in Fig. 2.8). The slopes and intercepts of the regression lines yielded meaningful parameters (see Eq. 2.8): the slopes can be associated to group delay and spectral position, whereas the spectral-temporal parameters derived from the intercepts (see
Eq. 2.9) correspond to STRF asymmetries in excitatory and inhibitory sidebands in the spectral and temporal domains.

Figure 2.12B shows that the group delays obtained for the population of 75 cells fell in a range of 15–40 ms. The group delays for upward and downward ripple directions agreed within 3 ms for a majority of neurons (72%). The group delays were longer than the onset latencies to tone stimuli by approximately 5 ms, which was significant (t_{67} = -4.7, p < 0.001). The group delays did not correlate with the tone-evoked latencies (r^2 = 0.02; p = 0.23), and the latency differences did not correlate with direction selectivity D either (r^2 < 0.001; p > 0.9). The spectral positions found for the two directions had almost identical values for a large majority of neurons (Fig. 2.12C). Spectral position differences also did not correlate with preferred ripple direction (r^2 = 0.004; p = 0.56).

Figure 2.12D shows that the temporal phase constant θ varied between −180° and 0° for most neurons (64%), which implies an onset excitation at BF_{STRF} (see Materials and Methods). We obtained a positive θ for a substantial minority of neurons (36%). This feature hints at an inhibitory onset response for the STRF. The spectral phase constant ϕ (Fig. 2.12E) was almost normally distributed around 0, which means that dominant inhibition above (ϕ > 0) and below (ϕ < 0) a cell’s BF was found in a similar number of neurons. A large majority of neurons has a ϕ near 0 indicating symmetry of sideband inhibition above and below BF.
Figure 2.12. Phase properties. A. Phase functions of neuron shown in Figure 2.8 (140). Phase, derived from SVD (see Fig. 2.8B), plotted as a function of velocity and density. For both directions, slopes correspond to temporal position (group delay) and spectral position (BF), see Eq. 2.8. B. Temporal positions, or group delays, for upward versus those for downward sweep direction. C. Spectral positions for upward versus those for downward sweep direction. The spectral positions are between -1.25 and 1.25 octave. When the difference between the two positions is > 1.25 octave, a 2.5 octave correction is made. D. Distribution of temporal phase constant $\theta$ (-180$^\circ$<0<180$^\circ$). F. Distribution of spectral phase constant $\phi$ (-90$^\circ$<$\phi$<90$^\circ$).
2.3.5 LINEAR PREDICTIONS

For 75 cells with good phase locking to a wide range of ripples (median q > 0.4), we also analyzed the responses to six different natural vocalizations. We predicted the responses to the vocalizations by applying the linear STRF model of Eq. 2.10 (details in Materials and Methods). The result of such an analysis for one of the neurons (J114) is illustrated by the black curve in Figure 2.13C for the monkey grunt sound. The actual average response (gray curve) is shown for comparison. Note that the predicted response was not rectified; negative values thus indicate a predicted response inhibition, while positive values can be interpreted as the firing rate of the cell. Although the largest peak of the actual response was successfully predicted, the linear model failed to make a decent prediction for the rest of the trial, as evidenced by the low correlation coefficient between the two curves (r = 0.12). Importantly, the cell had a consistent response to the vocalization as judged by inter-trial repeatability.

Figure 2.13. Prediction response of an example cell (J114). A. STRF and spectrogram of the stimulus (monkey vocalization, grunt). The STRF is spectrally centered around its BF. The spectrogram was taken at the same frequency range and the same spectral-temporal resolution. B. Prediction spectrogram yielded by convolution along the temporal dimension of STRF and stimulus spectrogram. C. Predicted response (black) obtained by summation along the spectral dimension of predicted spectrogram. The gray curve represents the actual response for comparison. The average spontaneous spike rate (here, 58 spikes/s) is subtracted from the actual response. The amplitude of the predicted response is scaled to the actual response.
Figure 2.14 shows the prediction for four of the other vocalizations presented to different neurons of the two monkeys. Although most acoustic energy of these vocalizations overlapped considerably with each neuron’s STRF, the linear predictions were still poor.

**Figure 2.14.** Prediction of response to four different vocalizations compared to actual response for four AC cells. A. Cell T79 with high BF (6.7 kHz) responding to bird call (Meadowlark). B. Cell T81 with low BF (0.8 kHz) and a broad excitation which followed by inhibition, responding to Blackbird call. C. T84, an inhibitory onset cell with high BF (10 kHz) responding to undulating vocalization. D. J145, another inhibitory onset cell with broad inhibition-excitation and medium BF (2kHz). The predicted (black) and actual (gray) responses are shown for each cell. In general the prediction was poor.
To summarize the overall performance of the linear predictive power of the STRF for all vocalizations across the 75 neurons, Figure 2.15 shows the response correlations as a function of each neuron’s prediction strength. We defined the latter as the root-mean-square of the predicted response amplitude for each vocalization, to avoid trivially low correlations for stimuli for which the expected response would be close to zero. In our previous study of monkey IC neurons, the highest linear response correlations were typically obtained for the strongest responses (Versnel et al., 2009). In contrast, however, we found no such relationship for the population of AC cells. In general, the correlations between actual and predicted responses were low for each of the six vocalizations, as shown in the table. The best predictions were obtained for the monkey grunt sound (r=0.14, on average), although they were still very poor. Rectification of the predicted responses had no influence on these results.

**Figure 2.15.** Predictability and prediction strength. Correlation between actual and predicted responses to 6 vocalizations as function of the prediction strength for 75 AC cells. The table shows the median correlation of recorded response versus predicted and rectified predicted responses, respectively. Rectification did not improve the prediction.
2.4 DISCUSSION

2.4.1 SUMMARY

Our study allows for a direct comparison of the spectral-temporal tuning properties of AC cells with those obtained from IC cells in awake macaque monkeys, as they were tested with the same experimental protocols and sound stimuli (Versnel et al., 2009). Our results indicate that both structures share a wide diversity of selective responses to ripple densities and velocities (Figs. 2.4 and 2.6), and that both demonstrate quadrant and full spectral-temporal separability in the majority of recorded neurons (Fig. 2.10). In terms of neuronal stimulus-response linearity, however, we observed marked differences between the IC and AC cells. First, while the correspondence between best frequencies for tones and dynamic ripples remained within 0.25 octaves for most IC cells, it was quite poor, with often more than half an octave difference, for the population of AC cells (Fig. 2.5). Second, the linear prediction for AC responses to the set of natural stimuli on the basis of the neuron’s STRF was typically poor (Figs. 2.13-15), whereas for IC neurons the prediction was fair (correlations 0.4-0.6) for stimuli yielding high firing rates. These differences argue for stronger nonlinear responsiveness for AC cells than for IC cells.

2.4.2 GENERAL CHARACTERISTICS

Spectral tuning. The AC population showed a predominant preference for low ripple densities ($\Omega_B \leq 0.4$ cyc/oct; Fig. 2.6B). This result is similar to monkey IC (Versnel et al., 2009), and to data obtained from thalamus and A1 in anesthetized animals (Shamma et al., 1995; Kowalski et al., 1996a; Versnel and Shamma, 1998; Escabi and Schreiner, 2002; Miller et al., 2002; Ahmed et al., 2006). This suggests that ripple-density selectivity might be faithfully transmitted from IC to AC, and that anesthesia may have little influence on spectral tuning characteristics. A similar preference for low ripple densities was also reported for human auditory cortex (Langers et al., 2003), and for human psychophysics (Chi et al., 1999). This prominent preference for low ripple densities might be confounded by the high percentage (76%) of recorded AC cells with a low $BF_{\text{tone}}$ (<2 kHz), considering that low $BF_{\text{tone}}$ cells typically have broader tuning characteristics (e.g. Kowalski et al., 1995; Versnel et al., 1995). However, the absence of a correlation between $BF_{\text{tone}}$ and $\Omega_B$ and the low $\Omega_B$ values for high-$BF_{\text{tone}}$ neurons observed in this non-human primate study, indicate that neurons across monkey AC have low $\Omega_B$ (Fig. 2.6D).

Temporal tuning. Previous studies have suggested that the range of preferred ripple velocities decreases from IC to cortex (Escabi and Schreiner, 2002; Miller et al., 2002). In line with these studies, the peak at low $\omega_B$ in our study was similar to that of monkey IC ($\omega_B = 8$ Hz for 34% of AC cells, vs. 31% for IC cells; Versnel et al., 2009), but preference for the two highest measured
velocities has decreased substantially from IC to AC ($\omega_B = 32$-40 Hz in 36% of IC cells vs. 24% of AC cells; Fig. 2.6A; Kowalski et al., 1996a; Miller et al., 2002; Ahmed et al., 2006; Versnel et al., 2009). Also, the joint distribution of optimal velocities and densities for the AC population did not cover the entire 2D spectral-temporal space as for IC neurons (Versnel et al., 2009). For example, cells tuned to higher $\omega_B$ (> 24 Hz), tended to be associated with a smaller range of best densities than cells tuned to low $\omega_B$ (Fig. 2.6F). This suggests that AC cells can process sound features over a range that is spectrally narrower than IC cells (Versnel et al., 2009).

**Phase functions.** The phase functions of the spectral-temporal transfer function were approximately linear in both quadrants (Fig. 2.12A), which was also reported for ferret A1 neurons (Kowalski et al., 1996a; Depireux et al., 2001) and for monkey IC (Versnel et al., 2009). These straight-line relationships indicate good phase locking to the ripple envelopes. The regression coefficients derived from these straight-line phase relations are informative descriptors of the STRF: the slope $\tau$ of the temporal phase function determines the group delay (BF$_{strf}$ latency), whereas the slope $\chi$ of the spectral phase function determines its spectral position (BF$_{strf}$). Like reported for IC neurons, these parameters were symmetrically distributed for both upward and downward ripple directions for AC cells as well (Figs. 2.12B, C). However, unlike IC neurons, there was no significant relation between the up-down differences of direction selectivities (D) and the up-down group-delay differences (Versnel et al., 2009).

The spectral ($\phi$) and temporal ($\theta$) components of the intercepts of the straight-line phase functions refer to STRF asymmetries in excitatory and inhibitory bands around the cell’s BF. Like for ferret A1 (Kowalski et al., 1996a; Depireux et al., 2001) and monkey IC (Versnel et al., 2009), we found similar distributions for these intercepts. The majority of AC neurons have an onset excitation at their BFs ($\theta < 0$; Fig. 2.12D). Inhibitory side bands above ($\phi > 0$) and below ($\phi < 0$) the BF$_{strf}$ were found for the same fractions of AC neurons (Fig. 2.12E).

---

### 2.4.3 SEPARABILITY AND DIRECTION SELECTIVITY

In anesthetized preparations a substantial proportion of A1 cells were characterized by inseparable STRFs (ferret, Depireux et al., 2001; Simon et al., 2007; cat, Miller et al., 2002; mouse, Linden et al., 2003). Most often, however, inseparability was due to an asymmetry between ripple directions ($\Omega>0$ vs. $\Omega<0$), as the transfer functions for inseparable STRFs were typically quadrant separable. Versnel et al. (2009) reported that the majority of monkey IC neurons (>70%) had fully separable STRFs, and IC cells with inseparable STRFs were typically quadrant separable. We here found that STRFs of AC cells were even more separable than IC neurons, since 93% of the cells were both fully separable and quadrant separable. Our results may be quite surprising since direction selectivity is typically found to increase from IC to cortex (rat, Felsheim and Ostwald, 1996) and separability is found less often in awake than in anaesthetized conditions (ferret, Simon et al., 2007).
We suggest that species differences underlie the apparent differences in (in)separability between previous studies and our results from awake monkey. The extent of separability may arise from the species-specific relevance of environmental sounds (e.g., vocalizations in bat have with strong spectrotemporal direction dominance leading to high inseparability of STRFs (Andoni et al., 2007)).

---

2.4.4 LINEARITY AND RESPONSE PREDICTION

---

As a simple measure for a potential nonlinearity in AC responsiveness we took the difference between the best frequencies found for pure tones (BF_{tone}) and for the STRF derived from dynamic broadband ripples (BF_{strf}). Our results revealed poor similarity between the two BFs (Fig. 2.5). Previous studies demonstrated a strong correspondence between the BFs in A1 cells of anesthetized ferrets (Shamma et al., 1995; Kowalski et al., 1996a; Versnel and Shamma, 1998), which could suggest that anesthesia might linearize potentially nonlinear response behavior of AC neurons. Recordings taken from IC showed a much better agreement between BF_{tone} and BF_{strf} in awake bats (Andoni et al., 2007), as well as in awake monkeys (Versnel et al., 2009). This might indicate that the contribution of nonlinear processing to the neural responses increases along the ascending auditory pathway.

The nonlinear response behavior implied for AC cells in the current study was further supported by the general failure to faithfully predict responses to the set of vocalizations on the basis of the STRF (Figs. 2.13-15). Previous studies, applying the STRF method, demonstrated that most AC neurons responded linearly to broadband sounds in rats (Machens et al., 2004), to ripples in anesthetized ferrets (Kowalski et al., 1996b; Versnel and Shamma, 1998; Klein et al., 2006), and to virtual acoustic space stimuli in anesthetized ferrets and cats (Jenison et al., 2001; Schnupp et al., 2001). This could again suggest that cortical responses may be more linear in anesthetized preparations than in awake animals. If so, the apparent discrepancy between results may be attributed to the state of alertness of the animal (Ryan and Miller, 1977; Populin, 2005). Note that other studies have shown that differences in an animal’s alertness do not systematically change the shape of the STRF (Britvina and Eggermont, 2008; Massoudi et al., 2013), but rather seem to affect the response strength (Eggermont, 2011). This could indicate that the presumed nonlinearity in cortical neural responses is not due to systematic changes in the STRFs, but to other factors.

When the neural response behavior is nonlinear, the shape of a neuron’s STRF, and hence its potential usefulness to predict responses to other stimuli, may critically depend on the stimuli used to extract the STRF, as well as on their resemblance with test stimuli used to compute the linear prediction, or response approximation (Theunissen et al., 2000; Blake and Merzenich, 2002; Woolley et al., 2005; Nagel and Doupe, 2008; David et al., 2009; Gourévitch et al., 2009). Indeed, broadband sustained stimuli, such as the spectral-temporal ripples used in this study, led to a different estimate of the STRF for cortical neurons than narrow-band stimuli, like natural vocalizations (Klein et al., 2006).
In line with this, some studies have reported that the linear STRF prediction of AC cells improves with the similarity between test stimuli and the sounds used to derive the STRFs (in anesthetized ferrets Versnel and Shamma, 1998; in anesthetized zebra finches, Theunissen et al., 2000). We therefore suspect that part of the failure to predict AC responses to the set of vocalizations in Figures 2.13-15 may be explained by this nonlinear aspect, as the STRFs were derived from responses to ripples, which differed substantially from the natural vocalizations. We here briefly note that the linear prediction of best ripple responses with the ripple-derived STRFs for the AC cells tended to go well for the first few hundred ms of the ripple, but after that a non-stationary adaptation kicks in, which causes a temporal drift of the response periodicity. This causes the overall correlations between prediction and response to be low (data not shown). Further quantitative details regarding the nonlinear behavior of AC cells in our experiments is beyond the scope of the present study, and will be a topic of a follow-up paper.

Single-unit recordings from the IC of awake bats (Andoni et al., 2007) and monkeys (Versnel et al., 2009) showed a fair linear predictability of the responses. Those results, however, also indicated that IC responses could not be fully explained by a linear kernel such as the STRF. This could imply an increased contribution from nonlinear processing along the ascending auditory pathway from IC to AC, which is in line with the observations of Atencio et al. (2012). The data of Yeshurun et al. (Yeshurun et al., 1985, 1989) showing that natural sound responses were better predicted by medial geniculate body (MGB) neurons than by cortical neurons, indicate the strong nonlinearities likely arise at the level of the AC.

Further note that the STRF cannot capture the typical nonlinear response behavior of neurons to either changes in sound level (Valentine and Eggermont, 2004; Lesica and Grothe, 2008), or to higher-order correlations that are present in natural sounds, like vocalizations (Atencio et al., 2008).

Finally, nonlinear response behavior of AC cells is further corroborated by studies reporting multiplicative encoding of different sound features in ferret AC (Walker et al., 2011), and multiplicative interactions between bottom-up (acoustic) and top-down (task-related) signals at the level of single units in monkey AC (Massoudi et al., 2013, 2014).

Our findings add to previous findings of a hierarchical increase in nonlinear processing along the auditory system. Through comparison with other studies, inter-species and vigilance effects on encoding complex sounds in single-unit AC neurons are highlighted. Moreover, this study provides a standard comparison of acoustic properties between IC and AC neurons, as similar acoustic stimuli, behavioral conditions, and species were used as in an earlier IC study (Versnel et al., 2009).
3. STABLE BOTTOM-UP PROCESSING DURING DYNAMIC TOP-DOWN MODULATIONS IN MONKEY AUDITORY CORTEX

Roohollah Massoudi
Marc M. Van Wanrooij
Sigrid M.C.I. Van Wetter
Huib Versnel
A. John Van Opstal

Published in *European Journal of Neuroscience* (2013), 37(11):1830-42
3.1 INTRODUCTION

To establish perceptual invariance, the auditory system faces the dichotomy to either adapt its acoustic representations when perceptual errors call for learning and plasticity, or to preserve a stable acoustic representation despite changes of the behavioral context. Much of the low-level (bottom-up) acoustic processing seems to take place at subcortical stages (Nelken et al., 2003; Palmer, 2007), with evidence for higher-order, non-acoustic (top-down) cognitive (Griffiths et al., 2004; Gutschalk et al., 2005; Snyder et al., 2006; Nelken and Bar-Yosef, 2008) and multisensory (Fu et al., 2003; Lakatos et al., 2007; Ghazanfar et al., 2008; Kayser et al., 2008a, 2010) processing in auditory cortex (AC).

Since the auditory cortex occupies a central position within the acoustic and non-acoustic processing pathways (Aertsen et al., 1981; Edeline et al., 2001; Edeline, 2003; Fritz et al., 2003, 2005a, 2005b; Schroeder and Foxe, 2005; Brosch et al., 2005; Polley et al., 2006; Lakatos et al., 2007; Riecke et al., 2007; Elhilali et al., 2007; King et al., 2007; Atiani et al., 2009), it is expected to play a role in stable perception of the auditory environment, but evidence is lacking. Recent studies have indicated that consistent attention to a specific acoustic feature of a target sound (e.g. a cued frequency) can induce changes in the spectro-temporal tuning characteristics of AC neurons toward the attended feature (Fritz et al., 2003, 2005a, 2005b; Elhilali et al., 2007; Atiani et al., 2009). However, if such changes are contingent upon general auditory task performance, the question arises how the auditory system maintains a stable representation of the acoustic environment under different behavioral conditions.

To address this, we investigated whether AC cells can preserve the bottom-up encoding of a sound’s spectral-temporal acoustics, despite prominent top-down signals from non-acoustic, context-dependent sources. We analyzed single-unit activity in monkey AC for three different situations, one in a passive condition, and two in an active listening condition. In the passive condition, we merely exposed animals to sounds that started automatically with 500 ms static noise, followed by a spectral-temporal dynamic ripple, and determined the spectral-temporal receptive field (STRF) from the ripple-evoked responses (Depireux et al., 2001; Eggermont, 2011). In the active listening condition the animal had to manually react to the change from static noise to dynamic ripple to obtain a reward. In half of the trials static noise lasted for 500 ms (A500), while in the other half it lasted for 1000 ms (A1000). Since both trial types were randomly interleaved, the change occurred unexpectedly in A500 trials. However, animals could fully predict ripple onset in A1000 trials after more than 500 ms static noise had passed, which added a cognitive factor to the active paradigm. This predictive effect was clearly evidenced by the animal’s response-reaction times and neural responses. Our recordings demonstrate that in active trials neural activity to static noise and dynamic ripples differed markedly from passively evoked responses. We show that the STRFs nonetheless remained identical for all
paradigms, in line with perceptual stability, and that the neural modulations in the active paradigms reflected a true task-related top-down signal.

3.2 MATERIAL AND METHODS

3.2.1 SUBJECTS

Neurophysiological recordings were performed in the left auditory cortex of two adult male rhesus monkeys (*Macaca mulatta*, Monkey J; 7-9 kg, and monkey T; 8-10 kg). Each monkey participated in the recording sessions for about two years. They were trained to respond to the onset of spectral-temporal modulations of a sound to receive a drop of water as a reward. Experiments were conducted in accordance with the European Communities Parliament and Council Directive (September 22, 2010, 2010/63/EU). All experimental protocols were approved by the Ethics Committee on Animal Research of the Radboud University Nijmegen (RU-DEC, ‘Radboud University Dier Experimenten Commissie’). Monkeys were pair-housed to promote normal interactive behavior. Our procedures follow the waterrestriction protocol of the Animal Use and Care Administrative Advisory Committee of the University of California at Davis (UC Davis, AUCAAC, 2001). About 24 hours before the start of an experimental session, water intake was limited to 20 ml/kg. In the experiment, the monkey earned a small water reward of 0.2 ml per successful trial. We ensured that monkeys earned at least the minimum of 20 ml/kg on an experimental day. After an experimental session, water was supplemented to the required minimum amount, if needed, and the animal received additional pieces of fruit. In weekends, the animals’ fluid intake was increased to 400 ml daily. To monitor the animal’s health status, we kept records of body weight, and water and food intake. Expert veterinarian assistance was available on site. Quarterly testing of hematocrit values ensured that the animal’s kidney function remained within the normal physiological range. Whenever an animal showed signs of discomfort, or illness, experiments were stopped and the animal was treated until the problem was solved.

3.2.2 SURGICAL PROCEDURES

When the initial sound-change detection training was completed (day-to-day performance level became 80%, or better), surgery was performed under full anesthesia and sterile conditions. Anesthesia was maintained by artificial respiration (0.5% isoflurane and N₂O), and additional pentobarbital (IV), ketamine (IM), and fentanyl (IV) were administered. A stainless steel recording chamber (12 mm diameter) was placed over a trepaned hole in the skull (10 mm diameter), the coordinates of which had been determined on the basis of MRI images. The chamber allowed a
vertical approach of the AC on the left side. A stainless-steel bolt, embedded in dental cement on the skull, allowed firm fixation of the head during recording sessions.

3.2.3 EXPERIMENTAL SETUP

The head-restrained monkey sat in a primate chair within a completely dark and sound-attenuated room (2.45 × 2.45 × 3.5 m), while a glass coated tungsten microelectrode (impedance 1-2 MΩ; Alpha Omega, Ubstadt-Weiher, Germany) was carefully positioned and lowered into the brain through a stainless steel guide tube by an electronically-driven stepping motor (National Aperture Inc. MM-3M-F-1). Electrode signals were grounded to a contact in the recording chamber. The analog electrode signal was amplified (BAK Electronics; model A-1), band-pass filtered between 0.1 and 15 kHz (custom-built 8\textsuperscript{th} order LP butterworth filter; HP with Krohn-Hite, model 3343), and monitored through a speaker and oscilloscope. The raw signal was then digitized (at 25 kHz, A/D convertor, TDT2 system; module AD-1; Tucker-Davis Technologies). An automated spike-detection algorithm (BrainWare, V 9.07 for TDT), run on a PC (Windows 98; DELL) isolated action-potential waveforms and determined their moments of occurrence. Data analysis and spike sorting was performed offline in MATLAB (version 7.9.0, R2009b, Natick, MA, USA).

3.2.4 SOUND STIMULI

Sound stimuli were digitally generated at a sampling rate of 100 kHz and delivered via Brainware-software and TDT2 hardware. A trigger provided by a TG6 module started sound presentation (DA1, low-pass filtered at 20 kHz through an FT6 module), and spike data acquisition. The sounds were presented in the free field from the frontal central position at a distance of 100 cm from the monkey by a speaker (Blaupunkt PCxg352, flat frequency characteristics within 5 dB between 0.2 and 20 kHz) at a fixed sound level of 60 dB (A-weighted (dBA), set by two programmable attenuators, PA4, and measured with a calibrated sound amplifier and microphone, BK2610/BK4134, Bruel and Kjaer).

Ambient background acoustic noise level was ~30 dBA. Reflections above 500 Hz were effectively absorbed by acoustic foam that was mounted on the walls, floor, ceiling, and every large object present. Ripple stimuli were used to determine the spectrotemporal properties of the recorded cells, and to investigate how detection of a change from a static noise to a dynamic ripple influenced the neuron’s responses. They consisted of a broadband complex of 126 components equally distributed (20/octave) from 250 Hz to 20 kHz (Depireux et al., 2001; Versnel et al., 2009). All components had random phase. The amplitude of each component was described as follows:

\[ S(t, x) = 1 + \sin(2\pi wt + 2\pi f t) \]  
(Eq. 3.1)
with t time (s), x position of the spectral component in octaves above the lowest frequency (250 Hz), \( \omega \) ripple velocity (Hz), \( \Omega \) ripple density (cycles/octave, or c/o). The ripple stimuli followed static noise of either 500 or 1000 ms duration, which consisted of the same components as the ripple but all with equal amplitude.

The set of 55 different ripples used in our study consisted of all combinations of 11 different ripple densities, \( \Omega \) in \([-2.0:0.4:+2.0]\) c/o, and 5 different velocities, \( \omega \) in \([8:8:40]\) Hz. A negative density corresponds to an upward direction of the spectral envelopes, a positive density to a downward direction, and \( \Omega = 0 \) means a pure amplitude modulated (AM) sound (see Figure 3.1 in Versnel et al., 2009). The modulation depth was fixed at 100%. The ripple stimuli were used to compute the spectrotemporal receptive field (STRF) of each cell, according to the procedure described below and illustrated in Figure 3.3. The sound level of the static noise and ripple was 60 dB sound pressure level (SPL).

Pure tones were were presented passively in a separate block at the start of a cell’s recording, and were used exclusively to extract the frequency-tuning properties of the recorded cells (Recanzone et al., 2000). They lasted for 150 ms and were presented at 13 frequencies from 250 to 16000 Hz at 4 different sound levels (10, 30, 50, and 70 dB SPL), set by the programmable PA4 attenuators. Trials were presented at least four times in a randomized manner. The spontaneous firing-rate was determined during a 300-ms prestimulus period across all (\( \geq 208 \)) presentations. Driven responses were defined as the average firing rate of all (\( \geq 4 \)) presentations of each stimulus that was greater than the mean plus two standard deviations of the spontaneous activity. The best frequency (BF) of each neuron was defined as the frequency that produced a driven response at the lowest intensity (threshold). The cell’s response onset latency was defined as the moment after pure-tone onset at which there was a driven response for at least three consecutive 2 ms time bins. The tuning bandwidth (BW) was measured as the difference in octaves between the low- and high-frequency edges of the driven responses, linearly interpolated at 10 dB above the threshold.

3.2.5 EXPERIMENTAL PARADIGMS

Single-unit responses to ripples were measured in two different paradigms (Fig. 3.1): i) passive sound exposure, in which the monkey was exposed to the rippled sounds without a task, and the straight-ahead fixation light was off, ii) active listening: upon presentation of a red fixation light at straight ahead, the monkey initiated a trial by pressing a bar. It had to attend to the same sounds as presented during passive exposure, as it had to respond to the ripple-modulation onset within 100-600 ms in order to receive a drop of water as reward. Trials in which the monkey did not detect the ripple onset were repeated at a randomly selected trial within the same recording block.

The passive and active conditions were presented in two different blocks, typically starting with the passive condition. In some experiments, the passive condition was repeated after the active
condition to check the cell’s response and STRF stability (not shown). In the passive condition the trial started automatically with the static noise (Fig. 3.1, sound onset). After 500 ms (Fig. 3.1, ripple onset), the ripple was presented for 1000 ms (Fig. 3.1, sound offset). Recording duration was 2500 ms, which started 300 ms before the sound onset, and ended 700 ms after sound offset (Fig. 3.1, black). The number of ripple repetitions was between 4 and 10.

The active listening block contained A500 and A1000 trials (based on the duration of the static noise epochs, at 500 ms and 1000 ms, respectively) that were randomly interleaved in the experiment at equal probability. The static noise was presented 300 ms after the monkey pressed the bar, and then changed into a ripple (at a pseudorandomly selected velocity and density) that lasted for 1000 ms. We used A500 and A1000 trials for two reasons: 1) to prevent the monkeys from learning that the ripple started after a constant time, which might have prompted them to start counting instead of listening; 2) to add a cognitive (i.e., predictive) aspect to the behavioral trials: the well-trained monkeys could anticipate a change at 1000 ms in A1000 trials with probability one.

The recording duration was 2500 ms in both A500 and A1000 trials, and the number of repetitions depended on the monkey’s performance (usually no repetition). The monkeys’ performance scores were ~ 80% (A500 = 80%, A1000 = 81%); monkey J had slightly better performance (A500 = 81%, A1000 = 85%) than monkey T (A500 = 78%, A1000 = 76%).
Figure 3.1. Experimental paradigms. Tasks: passive sound exposure (black), and active (A500 [red] and A1000 [blue]) trials, with relevant events in the trial indicated. Encircled numbers identify recording epochs during presentation of static noise (1-4) and dynamic ripples (5-7) for the different trial types, used for the correlation analyses throughout this paper. For illustrative purposes, the ripple is presented as an amplitude-modulated waveform. Histograms: reaction-time distributions for active conditions show faster responses in the A1000 trials than the A500 trials (light shading: monkey T; dark shading: monkey J).

3.2.6 DATA SELECTION

We recorded neural responses to the dynamic ripples in both the passive and active conditions from 128 cells in the auditory cortex (AC) of the two monkeys (J: n=65, T: n=63). For 23 out of 128 cells, we could not collect one full repetition for all ripples in either, or both active trial types. Thus, we had, in total, 105 cells with complete recordings (J: n=52, T: n=53). For most analyses in this study we further selected cells on the basis of an STRF-based quality criterion that excluded neurons with noisy responsiveness in the active condition. We therefore calculated the signal-to-noise ratio (SNR) in the STRF for the pooled active trials by dividing the maximum of the entire active STRF by
its standard deviation ($\sigma$) between 75 and 125 ms (note that the STRFs of AC cells typically do not contain acoustic tuning information after 75 ms; see also Results):

$$SNR = \frac{\text{max}(\text{STRF})}{\sigma}$$  \hspace{1cm} (Eq. 3.2)

The distribution of log (SNR) values for the active STRFs (pooled for A1000 and A500 trials) across the entire population of cells is shown in Fig 3.2. The SNR values in our recorded population show up as a bimodal distribution (low vs. high SNR values); the border between the two distributions was at about SNR=3.3. We included 51 cells (J: n=31; T: n=20) with SNRs above 3.3 for both the individual and population-based analyses. In the results we also show the population-based analysis for the 54 cells with SNRs below 3.3. Low SNR values in active trials could result because the number of presentations for each ripple in the active tasks had to be limited (typically to one), as otherwise the monkey would become satiated before the end of the experiment.

![Figure 3.2](image)

**Figure 3.2:** The logarithmic distribution of SNR values across the entire population of cells. SNRs can be divided in low vs. high SNR values; the border between the two distributions was set at SNR =3.3 (dashed line).

3.2.7 CHARACTERIZATION OF RECORDING SITES

Although we cannot with certainty identify the exact AC subdivision(s) in which we encountered individual neurons, we are confident that we recorded from the AC core (primary auditory cortex A1 and area R) and caudomedial field (CM) for the following reasons: 1) MRI scans
were used for stereotaxic placement of the recording chamber; the subsequent coordinates of the successful recording sites within the chamber corresponded closely to the stereotaxic coordinates of A1 as provided by the atlas of the Rhesus monkey brain by Paxinos et al. (2000); 2) before reaching an AC recording site there was a physiologically silent period, corresponding to the gap between upper and lower parts of the lateral sulcus (Kaas and Hackett, 2000); 3) tone-onset latency of the recorded sites was 22.6 ± 5.9 and 23.6 ± 6.6 ms for monkey J and T, respectively; 4) all neurons responded well to pure tones (BF: 250 – 16000 Hz); 5) The pure-tone tuning bandwidths for monkey J and T were 1.5 ± 1.2 and 1.5 ± 1.3 octaves, respectively; 6) The pure-tone thresholds for monkey J was 21 ± 13 dB SPL, and for monkey T was 23 ± 12 dB SPL. All above-mentioned tuning characteristics fall in the same ranges as reported by Recanzone et al. (2000) (Recanzone et al., 2000) for behaving monkeys in AC areas A1, R and CM.

Recording stability was verified by spike-waveform variability during the different behavioral tasks. When available, the passive STRFs could be obtained from pooling data recorded both before and after the active listening condition. Due to recording stability, this often led to a better estimate of the STRF.

3.2.8 DATA ANALYSIS

3.2.8.1 SPIKE-DENSITY FUNCTION

To convert each raster plot into a continuous spike-density function, we first binned the recorded spike times into a digital sequence at a time resolution of one ms, and then convolved the data with a Gaussian kernel with a standard deviation of 5 ms.

3.2.8.2 SPECTROTEMPORAL RECEPTIVE FIELD (STRF)

We estimated a cell’s STRF from ripple stimuli by using the same off-line method as described by Versnel et al. (2009). First, detected spikes (Fig. 3.3A) were sorted and binned into peri-stimulus time histograms. We then wrapped the 900-ms response window (the ripple duration, excluding the first 100 ms to exclude transient onset responses) into 32-bin period histograms, in which the ripple velocity determined the period as 1/ω (Fig. 3.3B). We subsequently performed a fast Fourier transform on the period histograms. The magnitude $A(\omega, \Omega)$ (spikes/s) and phase $\Phi(\omega, \Omega)$ (rad; Fig. 3.3C and 3.2D) of the period histograms were derived from the first harmonic of the resulting Fourier spectrum to generate the spectrotemporal transfer function:

$$T(\omega, \Omega) = A(\omega, \Omega) \exp(i \Phi(\omega, \Omega))$$  \hspace{1cm} (Eq. 3.3)

The 2D inverse Fourier transformation of $T(\omega, \Omega)$ then produces the spectrotemporal response field, or STRF (Fig. 3.3E), of the cell:

$$STRF(t, x) = FFT^{-1}[T(\omega, \Omega)]$$  \hspace{1cm} (Eq. 3.4)
with \( x \) the frequency in octaves (between 0 and 2.5 in 0.25 octave steps), and \( t \) running from 0 to 125 ms (at 12.5 ms resolution). The spectral dimension (abscissa) of the STRF reflects the frequency tuning, and the temporal direction (ordinate) reflects the cell’s linear impulse response. Note that the frequency range of the STRF is determined by the step size of the ripple densities employed in the experiment: \[ \text{range } x = 1/ [\text{step size } \Omega]. \] Typically, the step size was 0.4 cycles/octave, resulting in a frequency range of 2.5 octaves. Likewise, the temporal range is determined by the resolution in applied ripple velocities: \[ \text{range } t = 1/ [\text{step size } \omega], \] leading to 125 ms. Note that the position of the frequency range (in Hz) is ambiguous, as the lower frequency of the STRF could be either 250 Hz (the lowest component in the ripple stimuli), or multiples of 2.5 octaves above 250 Hz (i.e., at 1414 or 8000 Hz). The pure-tone responses were used to resolve this ambiguity. We shifted the STRF one octave higher or lower if the excitatory or inhibitory response areas were located near the edges of the STRF frequency domain, to relocate the center of the STRF close to the middle of the frequency axis.
Figure 3.3: Procedure to derive the STRF from neural responses to the 55 ripples that varied over the full velocity – density range. (A) Sorted spike trains of AC neuron T84 to ripples with densities between -2.0 and +2.0 cycles/octave, and velocities between 8 and 40 Hz. Each row represents one trial and each dot is a spike. Note the clear phase locking to ripple velocity. (B) Period histograms (32 bins) for all 55 stimuli. The period is defined by the inverse of ripple velocity. (C) Magnitude (in spikes/s) and phase (in degrees) were calculated from the best-fit cosine to the period histogram (here for the highlighted ripple at 0.8 cycles/octave and 32 Hz in B). (D) Magnitude and phase characteristics of the total ripple transfer function. (E) Spectrotemporal receptive field (STRF) obtained from 2-dimensional inverse Fourier transformation of the complex transfer function.

3.2.8.3 ANALYSIS OF STRF CHANGES

We compared the active STRFs (STRFₐ) with the passive STRFs (STRFₚ) to see whether the spectrotemporal tuning changes systematically in the different listening conditions. To that end, we first aligned the frequency bin (x) at which the maximum of the passive STRF was obtained, with the center of the spectral range (x’), and shifted the active STRFs by the same amount (Δx = x’-x). Then,
we performed a linear regression analysis for each cell \( k \) by comparing the active and passive STRFs within ± 0.25 octaves around the center of the spectral range, over a time window within 12.5 – 37.5 ms (called the local STRF):

\[
STRF_{Ak}(t, x') = g_k STRF_{pk}(t, x') + b_k \quad \text{(Eq. 3.5)}
\]

Here, \( g_k \) is a fixed gain that reflects the change in contrast between excitation and inhibition during active listening, while \( b_k \) is a constant bias (either increase in overall excitation, or inhibition) of the local STRF, and \( x' \) is the shifted spectral coordinate. We then normalized the entire (non-local) active STRF by:

\[
\hat{STRF}_{Ak}(t, x') = \frac{STRF_{Ak}(t, x') - b_k}{g_k} \quad \text{(Eq. 3.6)}
\]

Finally, the difference STRF for cell \( k \) was determined by:

\[
\Delta STRF_k(t, x') = STRF_{Ak}(t, x') - STRF_{pk}(t, x') \quad \text{(Eq. 3.7)}
\]

which is the residual of the linear regression. Any systematic pattern in \( \Delta STRF_k(t, x') \) now reflects a change in spectrotemporal tuning of the neuron, irrespective of any gain change. Finally, the difference STRF was normalized by its root mean square (rms) value, to allow averaging across neurons with very different firing rates:

\[
\overline{\Delta STRF}(t, x') = \frac{1}{N} \sum_{k=1}^{N} \frac{\Delta STRF_k(t, x')}{rms_k} \quad \text{(Eq. 3.8)}
\]

3.2.8.4 CORRELATION ANALYSIS

To determine whether the temporal firing patterns of neurons differed for the three types of listening trials, we calculated the correlation coefficient of the spike-density functions (at 1 ms time resolution). For these calculations, we skipped the first 100 ms of each epoch to avoid artificially high correlations due to a persistent transient onset response.

To quantify the similarity between raw STRFs derived from different behavioral trials, we reshaped each recorded STRF (a matrix, consisting of 10x10 entries) into a vector (100x1 elements), and calculated Pearson’s correlation coefficients for the three possible pairs of two vectors.

3.3 RESULTS

3.3.1 TASK

As described in Methods (and Fig. 3.1) the different durations employed in the static epochs (and hence the timing of the ripple onset to initiate bar release) added a cognitive (i.e., predictive) aspect to the behavioral trials. During the first 500 ms of A500 and A1000 trials, the probability that the change into a dynamic ripple would occur at 500 ms was 0.5 for both trial types. However, the well-trained monkeys could anticipate the change at 1000 ms with probability one in an A1000 trial, if
it did not occur at 500 ms. This aspect of the task is indeed strongly reflected in the animals’ behavior. In A500 trials the median manual reaction times (monkey J: 390 ms, monkey T: 376 ms; Fig. 3.1, red histograms) were substantially longer than in A1000 trials (monkey J: 218 ms, monkey T: 141 ms; Fig. 3.1, blue histograms), in which clear predictive responses also occurred (i.e., reaction times below 200 ms, sometimes even before ripple onset). We verified that the reaction times did not systematically change over (and during) recording sessions, and that A1000 trials always elicited faster median reaction times than A500 trials. Therefore, the experiment did not induce perceptual learning, which would have led to faster responses as time progressed, and reaction times could be pooled across sessions.

To assess whether the differences in reaction times for A500 and A1000 trials, as well as the considerable reaction-time variability (see distributions in Fig. 3.1) could somehow be attributed to certain acoustic features of the different ripples in the two active trial types, we analyzed and compared the reaction times for the individual ripples. If animals would always respond to a particular acoustic feature in the ripple stimuli, the pattern of reaction times should be similar for A500 and A1000 trials, albeit perhaps shifted by a constant amount toward lower reaction times for A1000 trials. Because the reaction-time distributions of the two animals were very similar (Fig. 3.1), we pooled their reaction-time data. Fig. 3.4 shows the mean manual reaction times for each ripple of the monkeys (color coded) obtained during the recordings of all 51 cells. In A500 trials (Fig. 3.4A), animals detected the ripple onsets on the basis of both spectral and temporal modulations, as a clear spectral-temporal pattern of reaction times emerged from this analysis. Along the spectral dimension (ripple density), reaction times increased with increasing (positive and negative) density, which indicates that higher spectral modulations were consistently harder to perceive. At the high-density ripples on both ends, high velocities were harder to detect as lower ripple velocities; for these high-density/high-velocity ripples we consistently obtained the longest reaction times. In summary, even though the modulation depth was 100% for all stimuli, the animals used both spectral and temporal acoustic modulations to detect the ripple onsets in the A500 trials.

In contrast, for A1000 trials (Fig. 3.4B) the clear acoustic pattern observed in Figure 3.4A is virtually lost. This strongly suggests that in this case both animals mainly responded to the ripple on the basis of prediction, rather than on the ripple’s acoustic characteristics.
Figure 3.4. Mean reaction time of each ripple during A500 (A) and A1000 (B) trials. Because of the resemblance of the two monkeys’ results, we pooled the results. There is a clear pattern of reaction time in A500 from lower to higher modulations, which is absent in A1000 trials.

Table 1 summarizes the performance of the two monkeys in the active trials. Trials were counted ‘correct’ (and hence rewarded) when the manual reaction times fell between 100 and 700 ms (see Methods). Overall, the correct trial percentages were similar (~80%) for both active trials. In about 4% of trials the animals did not respond with a manual reaction (‘misses’); these trials were excluded from the database for further analysis. Note that while in 15.3% of the A500 trials animals reacted later than 700 ms, and hardly ever responded too early, for the A1000 trials we obtained the opposite result: 14.5% of the trials had reaction times shorter than 100 ms, but animals hardly ever responded too late. These results further corroborate our interpretation that for A500 trials the animals’ reactions were based on the sound characteristics, and since some ripples (high-velocity/high-density) were harder to detect than others, reaction times could be long; in contrast, for A1000 trials animals often predicted the upcoming change, leading to very short (sometimes even negative; see Fig. 3.1) reaction times.

<table>
<thead>
<tr>
<th>Active trial</th>
<th>Total number</th>
<th>correct</th>
<th>missed</th>
<th>late</th>
<th>early</th>
</tr>
</thead>
<tbody>
<tr>
<td>A500</td>
<td>4866</td>
<td>79.60%</td>
<td>4.10%</td>
<td>15.30%</td>
<td>1.00%</td>
</tr>
<tr>
<td>A1000</td>
<td>4819</td>
<td>80.90%</td>
<td>4.00%</td>
<td>0.60%</td>
<td>14.50%</td>
</tr>
</tbody>
</table>

Table 1. Behavioral performance. Data from both monkeys is pooled. Reaction times were between 100-700 ms for correct trials, < 100 ms for early trials, and > 700 ms in late trials. In the missed trials, the monkeys did not respond.

Our experimental design contained seven different epochs (Fig. 3.1) with varying acoustic (static epochs 1-4, versus dynamic epochs 5-7) and/or behavioral (passive versus active- unpredictive versus active-predictive) states. The rationale of our paradigms is summarized in Figure 3.5, which shows the predictions of a pairwise correlational analysis between the different epochs, given that neural responses would be modulated by the animal’s behavioral state in the active conditions. We reasoned that (i) by comparing neural activity in these epochs for the different behavioral conditions,
we could identify potential top-down modulations, and (ii) by comparing the STRFs obtained from the dynamic epochs, we could determine the effect of top-down factors on the cell’s acoustic tuning. Thus, if top-down factors were to systematically influence neural activity of AC cells, responses between passive and active conditions should differ (yielding low correlations, red). Even in the simple detection task these top-down factors might include: attention, sound-change prediction (in epoch 4), motor preparation (handle bar release in epochs 6 and 7), and reward prediction (epochs 6 and 7). Note that the comparison of static epochs 2 and 3 (Fig. 3.5A) is a crucial control, because the acoustics and the monkeys’ behavioral states in these epochs are equal: the same sound is presented (500 ms of static noise), the monkey performs a task, and the animal is equally uncertain about the ripple onset. Therefore, despite the influence of top-down factors on AC activity, responses in epochs 2 and 3 should still correlate well (green). Depending on whether or not the bottom-up tuning characteristics change under the influence of task performance, an a priori prediction of the correlations between STRFs cannot be made (Fig. 3.5C, yellow).

Figure 3.5. Rationale of the experiments. If top-down signals modulate AC activity, correlations (A and B) between corresponding activity patterns are reduced for all but two epochs: the first 500 ms of static noise of A500 (2) and A1000 (3). In these epochs the animal attends to the sound, but is equally uncertain about the upcoming change. Bh and Pre indicate potential top-down signals in the active epochs: Bh - signals related to nonspecific behavior (e.g. vigilance, attention), Pre – modulations specific to change prediction, reward prediction or motor preparation. Whether and how STRFs would change (C) due to top-down modulations is not known a priori. Expected correlation values are indicated by the color bar.

3.3.2 TASK MODULATION OF AC ACTIVITY

Our correlation analysis is based on the recordings from 51 cells (monkey J: N=31, monkey T: N= 20). All selected cells possessed well-defined STRFs for the passive exposure condition, were selectively tuned to pure tones (best frequencies ranged from 250 Hz to 16 kHz, with the large majority ≤ 2 kHz), and recordings were stable during the entire session (see Experimental Procedures for selection criteria).

Task participation substantially changed neuronal response behavior, as exemplified by the spike-raster plots and the associated average spike densities of three representative AC cells during
the dynamic epochs (Fig. 3.6). For both active listening conditions (Fig. 3.6A, epochs 6 and 7), the mean spike density was typically higher than for passive sound exposure (Fig. 3.6A, epoch 5; mean firing rate gain for three cells: A500/P = 1.30, A1000/P = 1.31, A1000/A500 = 1.10). However, the increase in firing rate was not merely due to a simple magnitude scaling of the passive response, since the temporal patterns also differed substantially. This resulted in poor correlations between the spike-density functions of these dynamic epochs (Fig. 3.6B, cf. Fig. 3.5B, center).

**Figure 3.6.** (A) Spike-raster plots, sorted according to, first, ascending ripple velocity, and then by ascending ripple density, and trial-averaged spike density functions for three representative cells (J42, T98 and T75) during dynamic ripples (duration = 1000 ms) in the passive (epoch 5, black), A500 (epoch 6, red) and A1000 (epoch 7, blue) trials. Each dot in the raster plot represents one spike. Small tick marks at the top left-hand panel correspond to the applied ripple densities, the values of which are provided at the central left-hand panel. (B) Correlations (color-coded; numbers in boxes) between the spike-density functions of epochs 5-7.

The observed changes in the spike trains during the acoustic detection task suggest an influence of non-acoustic (top-down) factors. To fully quantify this influence on the neuronal responses for each recorded cell and for every response epoch, we analyzed the temporal firing patterns, and the acoustic spectrotemporal tuning characteristics as determined by the STRFs for the Passive, A500 and A1000 trials.
We first describe this analysis for example neuron J67 (Fig. 3.7). The spike rasters for the 55 ripples and the average spike-density functions for the Passive (Fig. 3.7A, left), A500 (Fig. 3.7A, center) and A1000 (Fig. 3.7A, right) trials demonstrate that task participation had a considerable influence on the neural response patterns during the entire trial. During passive exposure the neuron exhibited low baseline activity during silence (t<0 ms), followed by a strong phasic response (about 100 spikes/s), starting ~17 ms after sound onset (at t=0) and some additional response peaks during static-noise epoch 1. In contrast, during the silent period of the two active listening trials the neuron ramped up its firing rate already before sound onset. This prelude activity was followed by a high phasic response (about 100 spikes/s), followed by sustained activity during the remainder of static epochs 2-4.

Passive exposure during dynamic epoch 5 (t = 500-1500 ms) showed a fine temporal structure in the spike rasters that was locked to the ripples’ repetition rates (1/velocity). At sound offset (t = 1500 ms), activity quickly dropped back to baseline. For the two active-listening trials (dynamic epochs 6 and 7) the neuron produced a much stronger response than in the passive condition, and after ripple offset (at 1500 ms and 2000 ms, respectively) it persisted with sustained firing.

The correlations comparing the firing patterns of the various epochs were low or insignificant (Fig. 3.7D, left and center, cf. Fig. 3.5), except when comparing the first 500 ms of static noise for the active-listening trials (epochs 2 and 3). The high correlation between epochs 2 and 3 (r = 0.93) implies that the changes in response firing rates in all epochs were specific to the same acoustical, mental, and behavioral state of the monkey, and were not simply due to random response variations.

Next, we derived the cell’s STRF (Fig. 3.7C; see Materials and Methods, Fig. 3.3), from the cell’s first-order phase and magnitude responses in the dynamic epochs (Eqs. 3.3 and 3.4). Interestingly, the phase of the cell’s ripple transfer function (PTFs; Fig. 3.7B) was highly similar for the three conditions (r = 0.97 – 0.98). This indicates that the cell phase-locked to the acoustic modulations of the ripple envelope in a similar fashion for all three conditions, despite the substantial differences in spike rates. The magnitude transfer functions (MTFs) differed to some extent (r = 0.50 - 0.57), because in the active-listening paradigms the MTFs contained stronger responses (red colors) for several, but not all, ripples than when the monkey passively heard these stimuli (Fig. 3.7B; mean/max MTF: Passive = 10/17 spikes/s, A500 = 24/42 spikes/s, and A1000 = 22/41 spikes/s).

Because the MTFs varied across conditions, the resulting STRFs (Fig. 3.7C) differed slightly in their peak and trough magnitudes for this cell (max/min of STRF: Passive = 7/4 spikes/s, A500 = 15/-7 spikes/s, and A1000 = 14/-7 spikes/s; dark-red/blue colors, respectively, Fig. 3.7C). However, as a result of the robust phase-locking behavior, the patterns of excitation and inhibition in the STRFs were highly similar (r = 0.85 - 0.90 for the three comparisons; Fig. 3.7D, right).
Figure 3.7. Full quantitative analysis of example cell J67. (A) Raster plots (top) and spike density functions (bottom) for the three listening conditions during the trial. Vertical lines denote the different recording epochs (cf. Fig. 3.1). The trials are sorted according to, first, ascending ripple velocities, and second, to ascending ripple densities (see tick marks). Averaged spike-density functions are normalized to the peak passive response (set to 100%). Note higher activity in active trials than in passive trials (mean A500/P = 5.7, A1000/P = 5.4, A1000/A500 = 0.94). (B) Magnitude (left) and phase (right) ripple transfer functions and corresponding STRFs (C) for the three listening conditions. The similar phase transfer functions in different conditions indicate robust phase locking to ripple envelope modulations. (D) Correlations between spike density functions for static (left) and dynamic (center) epochs, and STRFs (right). A1000A denotes the first 500 ms of the static noise in A1000 trials, and corresponds to epoch 3 in Figure 3.1; A1000B represents the second 500 ms of the static noise (epoch 4 in Fig. 3.1).

3.3.3 PRESERVATION OF ACOUSTIC TUNING

Similarly, the cells shown in Figure 3.6 preserved their acoustic encoding features during active listening (Fig. 3.8A). Their STRFs were virtually unaffected by task performance, despite the marked changes in their firing behavior (Fig. 3.5). The cells showed a clear excitation for a narrow frequency range, with latencies between 10-40 ms, typically surrounded by a complex of spectral-temporal inhibitory sidebands. The excitatory (red) and inhibitory (blue) segments were located at exactly the
same time-frequency bins for the three listening conditions. Also the other parts of the STRF remained unaffected, yielding high mutual STRF correlations for these cells (Fig. 3.8B).

**Figure 3.8.** Spectral-temporal receptive fields. (A) STRFs of the three cells shown in Figure 3.6 for the passive (epoch 5), A500 (epoch 6) and A1000 (epoch 7) trials (see Fig. 3.1). (B) Correlation coefficients for the STRFs for each pair of epochs. The STRFs for the three listening conditions are highly similar for each of the three cells.

Despite the high resemblance of the STRFs for the different hearing conditions, subtle but systematic changes might potentially be missed by the correlation analysis. To check for any such consistent spectral-temporal changes in the active STRFs in comparison to the passive ones, we applied the linear regression analysis of Eqs. 3.5-3.8 (see Materials and Methods). Figure 3.9A represents the average passive-evoked STRF of all 51 cells, showing an excitatory region followed by inhibition. Panels 8B-D show the average normalized difference STRFs (Eq. 3.8) between A500-P, A1000-P, and A1000-A500 trials, respectively. Systematic patterns in the difference STRFs are absent, from which we conclude that there were no discernible shape changes in the STRFs during active listening. The distribution of the gain changes computed for all cells was unimodal and on average about one (mean ± standard deviation of the gains: A500/P = 1.11 ± 0.85, A1000/P = 1.06 ± 0.71, A1000/A500 = 0.81 ± 0.33). The bias values were around zero (mean ± standard deviation of biases: A500/P = 0.09 ± 0.37, A1000/P = 0.15 ± 0.46, A1000/A500 = 0.11 ± 0.37).
We also applied the population regression analysis to the 54 cells with low SNR, which is shown in Figure 3.9E-H. Again, no systematic changes were observed for the STRFs obtained in active trials. The results for cells yielding low SNRs and high SNRs were therefore quite similar.

**Figure 3.9.** STRF shape-change analysis for high-SNR cells (A-D) and low-SNR cells (E-H). (A) Shifted passive STRF, centered at the spectral octave range and averaged across all high-SNR cells (N=51). (B) Normalized average difference STRF (Eq. 3.8) between passive and A500 listening. Note absence of a systematic response, indicating no systematic change in the STRF of the neurons. (C) Normalized difference between passive and A1000 listening. (D) Normalized difference between A500 and A1000 listening. Color in panels A-C is scaled to the averaged passive STRF. Color in panel D is scaled to the averaged A500 STRF. (E-H) Results for the low-SNR cells are qualitatively similar as the high-SNR data shown in (A-D). Note differences in scale (color bars) for low and high SNR cells, which roughly correspond to the average firing rates for the selected populations.

For the vast majority of cells (Fig. 3.10C), the STRF correlations between the three behavioral conditions were high (median $r = 0.74 – 0.81$). Low correlation values as found for a few cells could be ascribed to a relatively low signal-to-noise ratio. Thus, in spite of the large differences in firing behavior, as evidenced from the low correlations between the spike-density functions during the dynamic epochs (Fig. 3.10B), and higher responses in the active condition (mean ± standard deviation of firing rate gain for all 51 cells: A500/P = 1.85 ± 0.93, A1000/P = 3.05 ± 1.14, A1000/A500 = 1.21 ± 0.2), AC cells faithfully preserved their bottom-up acoustic-tuning characteristics.

During the static-noise epochs (1-4) the stimulus acoustics were the same, but the behavioral states of the animal differed (except for epochs 2 and 3). As a result, the correlations between the mean spike densities for passive and active listening in these epochs were very low for five out of six comparisons (Fig. 3.10A). In line with the rationale of our paradigm (Fig. 3.5A), evoked responses for the large majority of neurons during the first 500 ms of A500 (epoch 2) and A1000 trials (epoch 3) were highly correlated (median $r = 0.81$; Fig. 3.10A; c.f. Fig. 3.7D, left). The high correlations...
between epochs 2 and 3 demonstrate that the activity changes across the other epochs did not result from random noise. Thus, the decorrelation of acoustically identical, yet behaviorally different epochs indicates involvement of a top-down signal.

**Figure 3.10. Population correlation analysis.** (A) Correlation coefficients between the spike density functions of the static epochs (1-4) for the three listening conditions of all 51 neurons, color coded according to the median value (color bar). (B) Correlation coefficients between the spike density functions of the dynamic epochs (5-7). (C) Correlations between the STRFs obtained from the three behavioral situations. Correlations are high only for the STRF comparisons and the epoch 2-3 comparison.

### 3.3.4 TOP-DOWN SIGNALS ARE TIME-LOCKED TO BEHAVIOR

During the dynamic epoch, neural responses for passive sound exposure and the two active trial types differed profoundly, as evidenced by a low correlation between the average spike-density functions (Fig. 3.10B). Upon detecting the ripple in the active listening task, the monkey prepares a manual response as fast as possible to obtain a reward. Therefore, an important factor in the active-listening paradigm, so far ignored in the analysis, is the considerable inter-trial variability of reaction times, within and between the two active trials (Figs. 3.1 and 3.4). This variability could potentially conceal any systematic motor-related or predictive component in the cell’s firing rate. Here we test whether this component was present in the activity of AC cells by re-aligning the spike rasters to the animal’s reaction time.

The resulting average spike-density functions of example cell J67 (same cell as in Fig. 3.7) show that the correlations between the dynamic A500 and A1000 trials increased markedly from 0.13, when aligned to ripple onset (Fig. 3.11A), to 0.64, when aligned to reaction time (Fig. 3.11B). Indeed, this correlation improved for 47 out of 51 cells ($T_{50} = -9.5, p < 10^{-6}$) (Fig. 3.11E).

This result also generalizes to the population averages of the spike-density functions, which show a marked improvement in correlation when the AC responses are aligned to reaction time (Fig.
3.11D), vs. ripple onset (Fig. 3.11C). Note that the population average started to increase well before (approximately 400 ms) bar release, thus representing a top-down signal that could be potentially related to change-prediction, motor preparation, or to the prediction of the upcoming reward. The present paradigm cannot dissociate these different factors, as reward delivery immediately followed bar release.

In summary, the absence of a correlation between the neural activity patterns during active listening (Fig. 3.10B) was not due to random noise in the spike timings. Instead, the changes in firing rate prior to bar release in both animals restored the neural response correlations, and can therefore be attributed to top-down signals related to movement preparation and/or reward prediction signals.

**Figure 3.11. Top-down signals.** (A, C) Average spike density functions for all A500 (red) and A1000 (blue) trials aligned to ripple onset, and (B,D) aligned to bar release for (A,B) cell J67 and (C,D) the population averages. Mean and standard error are denoted by solid lines and transparent patches, respectively. The firing activity was normalized by dividing the firing rates by the response range (maximum – minimum) for each cell. (E) Correlation coefficients between spike density functions of dynamic epochs 6 and 7 of A500 and A1000 trials, when aligned to reaction time (ordinate) versus when aligned to ripple onset (abscissa). The red dot corresponds to example cell J67.
3.4 DISCUSSION

This study is the first to quantify robust spectrotemporal tuning of AC neurons despite strong task-related modulations in their firing rates. Our results show that AC responses differed systematically between active and passive listening conditions, demonstrating clear non-acoustic top-down processes in monkey AC.

The data indicate that during passive exposure AC responses were primarily driven by bottom-up, stimulus driven, acoustic signals, resulting in low pre- and post-stimulus background activity, short onset responses (e.g., Fig. 3.7A), and often strong phase locking to the ripple envelope (e.g., Fig. 3.7B). In the active listening condition, the first-order phase locking remained unaltered (e.g., Fig. 3.7B). As a consequence, the shapes of the spectrotemporal receptive fields were virtually identical for the active and passive conditions (Figs. 3.7C, 3.8 and 3.10C), as there was no systematic shift and/or broadening in the active STRF when compared to the passive one (Fig. 3.9).

In contrast, the firing patterns of the different dynamic stimulus epochs became decorrelated (Figs. 3.6, 3.7A and 3.10B). These differences did not simply reflect random noise, e.g. sounds produced by animal movement, licking behavior, or changes in the animal’s vigilance. First, the modulations were highly systematic when responses were aligned to the reaction time (e.g. Fig. 3.11); second, in line with the rationale of our paradigms (Fig. 3.5), although the static noise-evoked firing patterns for passive and active conditions were uncorrelated, spike-density functions for the two active-listening trials during the first 500 ms epochs (2 and 3) were highly similar (Figs. 3.7D and 3.10A). In this situation, both the stimulus acoustics (randomly generated noise) and the monkey’s behavioral state (engagement in, and attention to the task and the sound, and equal uncertainty about the upcoming ripple) were identical. Although the acoustics did not change in the second half (epoch 4) of the A1000 static noise either, neural responses did change. We attribute this to differences in the monkey’s predictive state that led to decreased reaction times for A1000 trials (Figs. 3.1 and 3.4). A comparison of each ripple’s reaction time during A500 and A1000 trials showed that the monkeys’ decisions were not triggered by the same acoustic features (Fig. 3.4). In A500 trials, the animal’s reaction was based on both spectral and temporal modulations (Fig. 3.4A), whereas in A1000 trials change detection was predominantly based on prediction, rather than on acoustic ripple modulations (Fig. 3.4B). This predictive state led to apparently different spike-density functions for the A500 and A1000 dynamic epochs (Figs. 3.6 and 3.10B). The correlation patterns shown in Fig. 3.10 did not change when trials with negative reaction times in the A1000 trials were excluded from the analysis (not shown). Thus, the low correlations between static-noise epochs 3 and 4 did not result from a signal related to motor execution at variable reaction times during the static noise in epoch 4. Realigning the neural responses to reaction time, however, revealed the same preparatory signal and a concomitant increase in response correlation, for the A500 and A1000 trials in both animals (Figs.
A similar preparatory signal was reported for monkey auditory cortex by Brosch et al. (2005), and Niwa et al. (2012b), and for rat auditory cortex by Jaramillo and Zador (2011) (Brosch et al., 2005; Jaramillo and Zador, 2011; Niwa et al., 2012a).

Our sound-detection paradigm identified several top-down modulations in the active responses of AC neurons. Typically, neural baseline activity was higher for active listening during the entire trial (Figs. 3.6A and 3.7A) and persisted even during the silent epochs, i.e. before sound onset and after sound offset (Fig. 3.7A), which was also demonstrated recently by Niwa et al. (2012a) (Niwa et al., 2012b). This was also apparent from the slight increase in the magnitude transfer characteristics (e.g. Fig. 3.7B). This overall increase could be due to a fixed, static attentional signal that was present only when the monkeys were actively engaged in an auditory task. Most studies have reported that attentional demands, including arousal and vigilance, increase cortical responses (Edeline et al., 2001; Steriade et al., 2001; Niwa et al., 2012b). Our results are consistent with these findings (see, however, Otazu et al., 2009 (Otazu et al., 2009) who reported a decreasing AC response due to task engagement in rats). Other top-down modulations were highly dynamic, changing on a millisecond time scale from a task-preparation, to a stimulus-onset prediction signal that started before sound onset (Fig. 3.7A) (in line with Selenzeva et al., 2006), and finally to a motor-preparation or reward-prediction signal during the ripple (Figs 3.11B and 3.11D). These results confirm and extend previous reports on response enhancement in AC neurons prior to motor action (Brosch et al., 2011b; Niwa et al., 2012a), and in relation to reward expectation (Brosch et al., 2011a). Given the different interpretations of the non-acoustic modulations in AC cells, we conjecture that they arose from different cortical sources that could involve auditory belt and parabelt regions (acoustic task) (Durif et al., 2003; Brosch et al., 2005), parietal and frontal cortical areas (prediction, reward) (Romanski et al., 1999; Fritz et al., 2010), as well as premotor cortex (motor preparation) (Lemus et al., 2009).

While we demonstrated no changes in STRFs in task-related listening conditions (Figs 3.7C, 3.8, 3.9 and 3.10C), numerous studies have shown changes in STRFs resulting from various behavioral conditions. A tone-detection task modified STRFs of A1 in trained ferrets by enhancing neural responses at the attended target frequency (Fritz et al., 2003). Moreover, tone-discrimination and variation of signal-to-noise ratio in these tasks could all induce changes in the STRF by selectively enhancing sensitivity to the target frequency, and selectively suppressing tuning to the reference frequency (Fritz et al., 2005a, 2005b; Atiani et al., 2009). Consistent focused attention to a particular acoustic feature (the target tone) may have been the key trigger to initiate the receptive field changes (Fritz et al., 2007). Factors other than attention could also induce receptive field plasticity in A1, such as learning (Ohl and Scheich, 1997, 2005; Kilgard et al., 2001, 2002; Kilgard and Merzenich, 2002), conditioning (Bakin et al., 1996; Ji et al., 2001), expectation (Jaramillo and Zador, 2011), reward (David et al., 2012), sound localization (Lee and Middlebrooks, 2011), and electrical microstimulation (Suga et al., 2002; Suga and Ma, 2003). These findings suggest that a variety of non-acoustic factors can affect spectrotemporal tuning of A1 cells. However, if auditory task
performance by itself were to modify STRFs, AC would seem ill suited to encode the acoustic environment. In our experiments monkeys did not attend to, or were learning, a specific acoustic feature of any given target ripple, as all ripples were equally likely and important. Instead, the animals were merely awaiting (in A500; Fig. 3.4A) or predicting (in A1000; Fig. 3.4B) their appearance. Although after the change detection animals were not required to listen to the ripples anymore, AC cells still followed their spectral-temporal dynamics with high fidelity. This suggests that AC plasticity may not be linked to behavioral relevance per se, but rather requires behavioral relevance of specific acoustic features.

Model studies have suggested that multiplexing different signals at the population level of narrowly tuned cells allows the target structures of the population to selectively extract the relevant variable(s) for further processing (Pouget and Sejnowski, 1995; van Opstal and Hepp, 1995). Such a mechanism provides an efficient and flexible encoding scheme to deal with multiple task constraints and representations. Multiplicative encoding (‘gain fields’) has been reported for various sensorimotor processing stages in the monkey brain. For example, changes in static eye position gain-modulate visual responses in the lateral intraparietal sulcus without affecting visual receptive fields (Andersen et al., 1985). Similarly, multiplicative eye-position modulations were reported for saccadic eye-movement responses in the midbrain superior colliculus (Van Opstal et al., 1995), and for the acoustic responses of a subpopulation of cells in the inferior colliculus (Groh et al., 2001; Werner-Reiss et al., 2003; Zwiers et al., 2004). A recent study also indicated a multiplicative mechanism for bottom-up encoding of sound attributes in A1 cells of anesthetized ferrets (Walker et al., 2011). While attentional signals have been shown to gain-modulate visual responses in primate (McAdams and Reid, 2005) and cat (Kara et al., 2002) V1, to our knowledge, a multiplicative interaction of acoustic and behavioral signals in AC has not been reported before. Our results imply that primate AC neurons can efficiently represent rapidly changing bottom-up and top-down signals in a multiplexed fashion. This allows for a stable representation and perceptual invariance of acoustic signals in the presence of strong non-acoustic top-down modulations.
4. BEHAVIORAL CONTEXT AFFECTS TEMPORAL PROCESSING IN AWAKE MONKEY AUDITORY CORTEX

Roohollah Massoudi
Marc M. Van Wanrooij
Huib Versnel
John Van Opstal
4.1 INTRODUCTION

An important function of the central auditory system is temporal envelope processing of complex sounds. This capacity plays an important role in speech intelligibility. For example, despite limited spectral information, many users of cochlear prostheses are able to understand speech, denoting the importance of envelope information to the understanding of complex communication signals (Shannon et al., 1995).

Responsiveness of neurons to amplitude-modulated sounds is a major characteristic of the core auditory cortex (Lu et al., 2001; Liang et al., 2002; Joris et al., 2004; Bartlett and Wang, 2007; Yin et al., 2011; Niwa et al., 2012a, 2012b; Malone et al., 2013; Massoudi et al., 2013, 2014). Auditory cortex cells encode AM signals by a temporal code (synchronized, or phase-locked, to the envelope modulations), by a rate-code (non-synchronized to the modulations), or by a combination of both mechanisms. It has been suggested that a gradual transformation takes place from a pure temporal code at lower levels within the auditory processing stream, to a rate coding mechanism at higher levels (Lu et al., 2001; Liang et al., 2002; Joris et al., 2004; Bartlett and Wang, 2007; Yin et al., 2011).

Envelope processing could be modified by different factors. The carrier of the envelope, i.e. the spectral content of the sound, has been reported to affect temporal processing in monkey auditory cortex (Malone et al., 2013). It has also been shown that active discrimination of amplitude modulation frequencies can affect envelope processing in auditory cortex (Lemus et al., 2009; Dong et al., 2011; Niwa et al., 2012a, 2012b).

In this study, we determined whether and how neuronal envelope processing in monkey auditory cortex is affected by performing in a simple auditory detection task, in which the animal had to react as fast as possible with a manual response after detecting the onset of an amplitude modulation in an ongoing sound. In 50% of the trials, the animal could predict the onset of the upcoming amplitude modulation. We employed this paradigm in previous studies, where we demonstrated a clear sensitivity of the reaction times to acoustic modulations only when the animal could not predict the upcoming sound modulation (Massoudi et al., 2013, 2014). To better dissociate how temporal and rate coding may be affected by the behavioral demands, we here focused on the pure amplitude modulations (Gaussian AM noises) over a broad frequency range (2-256 Hz), and analyzed different tuning characteristics of the neurons during the trial in relation to the animal’s reaction time patterns.

We analyzed the cell’s spontaneous firing rate, its sound-onset response, as well as its sustained response during the static and modulated periods of the acoustic stimuli (cf. Wang et al., 2005), as function of the animal’s reaction time, and sound-modulation frequency.

How the reaction times and the different properties of neuronal responses vary with the sound-modulation frequency is quantified by modulation transfer functions (MTFs). We explicitly
tested whether neuronal MTFs remained invariant when the monkey’s behavior changed from awake, yet idle, to task engagement, with or without predictability of the sound modulation. In our analyses, we determined different metrics to quantify and dissociate temporal vs. rate-coding mechanisms.

Our results show that during task performance the mean firing rate of AC cells strongly increases, while their phase locking decreases. However, the magnitude of the first harmonic of the period histogram, which combines both factors, was unaffected by the task between AMF ~ 5-45 Hz. This implies that a tradeoff between rate and temporal encoding may have yielded invariant neural responses for a considerable range of sound modulations.
4.2 MATERIALS AND METHODS

The experimental paradigms described in this study were the same as in our previous reports (Massoudi et al., 2013, 2014). Detailed descriptions of the surgical procedures, experimental setup, and characterization of recording sites are available there. Recordings to amplitude-modulated noises were collected in the same session as separated stimulus-response blocks. In other blocks, responses to other types of sounds (pure tones, dynamic spectral-temporal ripples, natural vocalizations of macaques and and birds; see Massoudi et al., 2013, 2014) were also recorded.

4.2.1 SUBJECTS

Single-unit recordings were performed from the left auditory cortex of two adult male rhesus monkeys (Macaca mulatta; Monkey J, 7–9 kg; Monkey T, 8–10 kg). Each monkey participated in the recording sessions for about 2 years. Animals were trained to respond to the onset of spectral–temporal modulations of a sound to receive a drop of water as a reward. Experiments were conducted in accordance with the European Communities Parliament and Council Directive (September 22, 2010, 2010/63/EU). All experimental protocols were approved by the Ethics Committee on Animal Research of the Radboud University Nijmegen (Radboud University Dier Experimenten Commissie). Monkeys were pair-housed to promote normal interactive behavior. Our procedures followed the water-restriction protocol of the Animal Use and Care Administrative Advisory Committee of the University of California at Davis (2001). At about 24 h before the start of an experimental session, water intake was limited to 20 ml/kg. In the experiment, the monkey earned a small water reward of 0.2 ml per successful trial. We ensured that monkeys earned at least the minimum of 20 ml/kg on an experimental day. After an experimental session, water was supplemented to the required minimum amount, if needed, and the animal received additional pieces of fruit. At weekends, the animal’s fluid intake was increased to 400 ml daily. To monitor the animal’s health status, we kept records of body weight, and water and food intake. Expert veterinarian assistance was available on site. Quarterly testing of hematocrit values ensured that the animal’s kidney function remained within the normal physiological range. Whenever an animal showed signs of discomfort, or illness, experiments were stopped and the animal was treated until the problem was solved.

4.2.2 ELECTROPHYSIOLOGY

We report on single-unit recordings from 148 auditory cortex cells (monkey J, N = 97; monkey T, N = 51), while monkeys were either passively exposed to, or actively detecting, a sound envelope modulation. To verify the stability of the recordings, we confirmed that the average spike-waveforms were similar across the active and idle blocks (note the overlap of the average waveforms of passive, A500 and A1000 trials of an example cell in the Fig. 4.2E inset). The responses of 100
cells from this population have also used for other purposes in our previous study (Massoudi et al., 2014) Here, we include these neurons within a larger database, and with a focus of analysis of different response characteristics than in our previous study.

**4.2.3 SOUND STIMULI AND BEHAVIORAL PARADIGMS**

All sounds started with freshly generated static broadband Gaussian white noise, which was followed by an amplitude modulation. In passive trials, animals were merely exposed to the sounds without a required behavioral response. In the active trials, the monkey pressed a bar to initiate a trial. Based on the static noise duration, this experiment contained two different conditions: i) A500 trials the static noise lasted 500 ms, and ii) A1000 trials, in which it lasted for 1000 ms. To obtain a reward the animal had to react to the perceived sound change from static noise to AM by releasing the bar from 100 to 700 ms after the sound change. Since both trial types were randomly interleaved at equal probabilities, the change occurred unexpectedly (probability 50%) in the A500 trials, but animals could in principle anticipate the sound onset in A1000 trials (with probability 100%), as soon as 500 ms static noise had passed.

**4.2.4 DATA ANALYSIS**

All data analyses were performed in Matlab (version 2009b; Mathworks, Natick, MA).

**4.2.4.1 SPIKE-DENSITY FUNCTIONS**

To convert each spike-raster plot into a continuous spike-density function (a measure for the instantaneous mean firing rate across trials), we first binned the recorded spike times into a digital sequence with a time resolution of one millisecond, and then convolved each spike with a Gaussian kernel with a standard deviation of 10 ms. Finally, the signals were added across trials, and we normalized the resulting function by the number of trials.

By default the data were aligned to sound onset (e.g., Figs 4.2D and E). To relate the top-down signals to reaction time, we re-aligned all neural responses to the bar-release time by shifting the spike timings in each trial by the associated reaction time (e.g., Figs 4.2F).

**4.2.4.2 MODULATION AND RESPONSE-ONSET ANALYSIS**

For each neuron, we estimated the modulation transfer functions (MTFs) in a similar fashion as described in detail for spectral-temporal ripples in our previous reports (Versnel et al., 2009; Massoudi et al., 2013). The MTF describes a neuron’s sensitivity to the temporal envelope modulation, (AMF). The sensitivity was quantified by several parameters extracted from the spike trains, e.g. the mean firing rate, or the strength of phase locking, giving rise to different MTFs. First, spike timings were binned into single-cycle period histograms (divided in 32 bins, evenly distributed
over the modulation period $T_{AM} = 1/AMF$). We subsequently performed a fast Fourier transform on the period histograms to derive the temporal and rate MTFs (mMTFs and rMTFs), respectively: the synchronized rate, $\tau(AMF)$ (spikes/s), for each AM stimulus of the mMTF was derived from the amplitude of the first harmonic of the resulting Fourier spectrum, while the unsynchronized, average firing rate, $\rho(AMF)$ (Hz), was taken as the DC. The mMTF reflects the strength of the neuron’s temporal modulation-following response, while the rMTF indicates how much the average firing rate changes with the modulation frequency, $AMF$.

In addition to the mMTF, we also determined the response vector strength, $R(AMF)$ (Goldberg and Brown, 1969), for the $n$ recorded spikes for each sound modulation. The vector strength quantifies the strength of phase-locking of the neuron to the particular sound modulation, and is defined as:

$$R(AMF) = \sqrt{\frac{\left(\sum_{i=1}^{n} \cos \phi_i\right)^2 + \left(\sum_{i=1}^{n} \sin \phi_i\right)^2}{n}}$$  \hspace{1cm} (Eq. 4.1)

where each spike $i$ is treated as a unit vector with a phase $\phi_i$ between 0 and $2\pi$, measured as the spike time modulo the modulation period, $T_{AM}$. Equivalently, Eq. 4.1 can also be expressed by:

$$R(AMF) = \frac{\tau(AMF)}{\rho(AMF)}$$  \hspace{1cm} (Eq. 4.2)

with $\tau(AMF)$, the magnitude of the first harmonic, and the normalization $\rho(AMF)$, equals the DC component (mean firing rate during the modulation; see above). When vector strength $R=1$, all spikes occur at exactly the same modulation phase, and when $R=0$ the spikes are distributed uniformly along the modulation cycle.

4.2.4.3 ONSET RESPONSES

We analyzed the cell’s onset response to the sound in a similar way, by determining the strength of time-locking of all spikes (applying Eq. 4.1) within the 40 ms time window following sound onset. In this case, when $R=1$ all spikes occur at the same moment after sound onset, and when $R=0$ the spikes are distributed uniformly over the window (absence of an onset peak).

4.2.4.4 PHASE LOCKING AND TRIAL SIMILARITY

Different values of the vector strength, $R(AMF)$, the first-harmonic magnitude, $\tau(AMF)$, and the average firing rate, $\rho(AMF)$, each imply a particular shape of the period (or onset) histogram. A high $R$ indicates a peak at a specific modulation phase (or onset latency), a high $\tau$ corresponds to a sinusoidal histogram, and a high value for $\rho$ hints at a uniform spike distribution, respectively. To check for systematic changes in a neuron’s response, while allowing for any arbitrary shape of the period histogram, we also determined a trial-similarity metric (Malone et al., 2007, 2013). This was achieved by dividing the trials in pseudo-random order for each AM stimulus into 2 equal-sized sets.
For each set we then generated the corresponding period histograms and computed Pearson’s correlation coefficient between them. By bootstrapping, we performed this procedure 100 times to obtain a stable estimate for the correlation coefficient for each AMF, leading to the tsMTF. The trial similarity metric is a measure of reproducibility of the period histogram, independent of the dispersion of spikes along the modulation cycle.

To determine the best modulation frequency (BMF) for the cell, we refer to the peaks in the different MTFs as mBMF, rBMF, vsBMF and tsBMF, for the magnitude, rate, vector strength, and trial similarity metrics, respectively.

4.2.4.5 REACTION-TIME ANALYSIS

The reaction time was defined as the moment of bar release relative to AM sound onset. We calculated the median reaction times for all modulation frequencies for each of the two active trial types (A500 and A1000). We fitted a simple regression model to the data that quantified where in the modulation cycle (period $T_{AM}$) the monkey detected the sound-modulation onset (threshold, $\lambda_\phi$), and how much time the animal needed to prepare the manual response ($\lambda_0$) (Massoudi et al., 2014):

$$RT(T_{AM}) = \lambda_\phi \cdot T_{AM} + \lambda_0 \quad (\text{Eq. 4.3})$$

In Massoudi et al. (2014) we noted that the reaction times for the A1000 trials appeared to reflect different strategies of the animal: fast A1000 responses seemed to be made irrespective of the acoustic modulations, whereas the slowest A1000 responses reflected also sensitivity to the sound modulations. To better quantify this issue, we here fitted Eq. 4.3 to the different percentiles of the RT distributions (from 10% - 90% in 10% increments) of both the A500 and A1000 reaction times on the data sets pooled across all recording sessions (see Fig. 4.1).

4.2.4.6 STATISTICAL SIGNIFICANCE

For the magnitude MTF, we determined a measure to quantify the significance of the strength of locking of the cell response to a particular sound-modulation frequency, AMF, by first determining from the Fourier Transform of the period histogram (Versnel et al., 2009):

$$q = \frac{A_1}{\sqrt{\sum_{i=1}^{16} A_i^2}} \quad (\text{Eq. 4.4})$$

with $A_i$ the amplitude of the $i$-th harmonic of the modulation period ($T_{AM}$). The parameter $q$ reflects the quality for the best sinusoidal fit to the period histogram. If $q=1$ the period histogram corresponds to a pure sinusoid. If $q=0$ the cell has no sinusoidal envelope-following response. We applied $q$ as a significance measure to judge the sensitivity of the neuron to the modulation frequency under study. To assess the level of significance for $q$, we simulated random spike trains, constructed period histograms, and determined the associated $q$-value. From these simulations we obtained a threshold value of $q_\theta=0.12$, which yielded a $p$-value of 0.001 (i.e., 99.9% of all simulated responses yielded lower $q$-values).
For rate MTFs, the significance of a difference between mean firing rates, \( \rho(AMF) \), across the different stimulus modulation frequencies was determined by comparing the spike rate distributions across trials according to a Wilcoxon rank-sum test (\( p<0.001 \)).

For the vector strength MTF we used the Rayleigh statistic \( 2nR^2 \) (equivalently: \( 2T^2/\rho \)) to test for significant phase locking (>13.8, \( p<0.001 \), Mardia and Jupp, 1999).

For the trial similarity MTF, we simulated trials with randomly distributed spike trains, compared their similarity, and constructed a distribution of the trial similarity metric. Generally, a value of 0.8 yielded a significance of \( p<0.001 \) for 32-bin period histograms.

### 4.2.4.7 SIGNIFICANT VS. NONSIGNIFICANT CELLS

In this paper we applied the same nomenclature as Malone et al. (2013) to denote the significance of a cell’s sensitivity regarding a given stimulus parameter (firing rate, vector strength, trial similarity and first harmonic amplitude). A cell was termed ‘significant’ if for at least one modulation frequency and under at least one of the three listening conditions the parameter was significant (according to the criteria described in the previous section). Otherwise the cell was termed ‘insignificant’. Note that also the insignificant cells had clear auditory responses.

### 4.2.4.8 CHANGES IN MODULATION TRANSFER FUNCTION

We quantified the differences between MTFs, obtained for the different behavioral conditions (passive, A500 an A1000), with a Similarity Index (SI; Malone et al., 2013). This index was defined as the vector norm (containing the 15 AM frequencies) of the difference vector between the two MTFs, divided by the sum of their respective vector norms:

\[
SI = 1 - \frac{||MTF_1 - MTF_2||}{||MTF_1|| + ||MTF_2||}
\]

(Eq. 4.6)

To compare the differences in MTFs due to behavioral conditions across the population, we used a nonparametric Wilcoxon signed rank test (\( p<0.05 \), e.g., Fig. 4.7).

### 4.3 RESULTS

#### 4.3.1 EFFECTS OF MODULATION FREQUENCY ON THE REACTION TIMES

As reported in our previous study (Massoudi et al., 2013, 2014), the different demands imposed by the unpredictable A500 and predictable A1000 trials yielded different reaction-time distributions for both monkeys. The prediction caused 17.6% of the reaction times in A1000 trials to fall before the sound change, while almost all A500 reaction times (99.45%) fell after the sound change.
For the unpredictable A500 trials the reaction times also depended systematically on the modulation frequency of AM sounds (ANOVA, $F_{df=14}=9.3$, $p=0.009$; Fig. 4.1A, red triangles). The red lines in Figure 4.1A correspond to the 25th, 50th, and 75th percentiles of the fit, respectively (Eq. 4.3). As in our previous study, for the predictable modulation onsets (A1000 trials; Fig. 4.1A, blue triangles), the reaction times were virtually independent of the acoustic modulations.

Note that the 25th percentile of the reaction times resulted in a similar frequency-independent pattern as the mean (Fig. 4.1A, lower blue line), while the 75% percentile of the responses (Fig. 4.1A, top blue line) started to resemble the A500 reaction-time pattern. To better quantify that the slower reaction times in the A1000 trials might rely in a systematic way also on the acoustic modulations, Fig. 4.1B, shows the period dependence ($\lambda_0$, see Materials and Methods, slope in Eq. 4.3) as function of the latency offset ($\lambda_0$) for the different percentiles of the pooled data set. It demonstrates that the 60% fastest A1000 responses (<60th percentile, blue line) remained invariant with respect to the sound-modulation frequencies, while the 40% longest reaction times (>60th percentile, blue line) depended systematically on the sound modulation frequency. This indicates that fast A1000 responses relied heavily on the prediction of the upcoming sound change in the trials. Interestingly, the slower A1000 responses depended on the sound modulations (AMF) in a similar way as the A500 responses (Fig. 4.1B, red line). For example, the fastest A500 responses (10th percentile) relied on AMF in almost the same way as the slowest A1000 responses (90th percentile). Yet, the acoustic dependence of the A500 responses appeared to level off, and even started to decline again, for the slowest reactions (>70th percentiles). This could suggest that when the animal is slow (i.e., reaction times >400 ms) its sensitivity to acoustic modulations decreased, possibly due to a decreased vigilance.

It is important to note that the percentiles of the reaction times were taken from the pooled data base, which was collected from over a hundred of recording sessions. Therefore, despite the considerable day-to-day variability of the mean reaction times for the A500 and A1000 trials, the overall patterns of the reaction times, in terms of their dependence or invariance to acoustic modulations, remained stable over time. This is further corroborated by Fig. 4.1C, which illustrates the median reaction times obtained from 128 recording sessions for both monkeys (grey and black dots). Besides the clearly longer reaction times for A500 trials in all recording sessions, there is also a clear correlation between the median A500 and A1000 reaction times ($r = 0.45$, $p < 0.0001$). This implies that when in a given session the animal was slow, or fast, his reaction times for the A500 and A1000 trials both increased, or decreased, respectively. As a result, part of the variability in the reaction times, which remains often unexplained, contained a clear signature as to whether the animal would rely more (slow responses) or less (fast responses) on the acoustic amplitude modulations.
Figure 4.1. Reaction-time analysis. (A) Reaction time as a function of sound-modulation frequency for A500 (red) and A1000 (blue) trials. Solid lines correspond to the optimal regression of Eq. 4.3 through the 25th, 50th and 75th percentiles of the reaction times, pooled across 128 recording days and animals. (B) Regression slope (period dependence; Eq. 4.3) as a function of regression offset (latency offset) for the [10th, 20th, …, 90th] percentiles. Symbols in (B) (blue: A1000, red: A500 trials) indicate a significant regression slope (p<0.05). (C) Median reaction times of each recording session for A1000 vs. A500 trials of monkey J (grey dots) and monkey T (black dots) are strongly correlated.

4.3.2 INFLUENCE OF BEHAVIORAL CONTEXT ON SPONTANEOUS ACTIVITY AND ONSET RESPONSES

Figure 4.2 shows the results of our analysis of the spontaneous activity, and of the sound-onset responses, for example cell T84. The top row presents the raster plots for passive (green, Fig. 4.2A), A500 (red, Fig. 4.2B), and A1000 (blue, Fig. 4.2C) trials. Although the cell responded vigorously in the passive condition (Fig. 4.2A), engaging in a behavioral task evoked even more spikes for both the A500 an A1000 trials (Figs. 4.2B and C). Despite the high firing rate, the cell’s spike trains show a visible temporal structure for the passive condition (Fig. 4.2A), which is also present, but less obvious, for the A500 (Fig. 4.2B) and A1000 (Fig. 4.2C) trials (see also below). The spontaneous activity in passive trials was significantly lower than for the active trials (Fig. 4.2D; mean ± SD: passive = 89 ± 7, A500 = 134 ± 12, A1000 = 136 ± 13 spikes/s). Furthermore, the passive and active trials lacked a significant rise in the firing rates from about 100 ms prior to sound onset (at t = 0; Fig. 4.2D).

Figure 4.2E compares the onset responses for the three behavioral conditions. The onset peak had virtually indistinguishable magnitudes for the three trial types (Fig. 4.2E; maximum peak: passive = 302, A500 = 311, A1000 = 290 spikes/s). The inset shows the stability of the average spike waveforms collected during the passive, A500, and A1000 trials, which all have similar shape.

When the responses in active trials were aligned to the reaction times, the cell showed a steady increase in activity that started well before the reaction time, both for A500 (Fig. 4.2F, red) and A1000 trials (Fig. 4.2F, blue). This is denoted by the bold lines representing best regression fits to the last 100 ms of the spike-density functions (see also Massoudi et al., 2014).
**Figure 4.2.** Analysis of different neural response epochs to AM noises. Spike rasters in panels (A), (B) and (C) are aligned to sound-modulation onset at time $t = 0$. Data from neuron T84 for passive (A), A500 (B) and A1000 (C) listening conditions, respectively. Trials are ordered to modulation frequency, AMF. Note clear neural modulation responses in synchrony with sound-amplitude modulations, especially in passive trials. (D) Analysis of the first 300 ms of the recording (before sound onset at -500 ms in (A,B), and -1000 ms in (C)) to characterize the cell’s spontaneous firing rate. The spontaneous activity in active trials was higher than for passive listening. The spike-density functions (thin colored lines) were obtained by averaging across trials. Bold straight lines denote best-fit regression lines to the last 100 ms. (E) average spike-density onset peaks of the neuron were similar across hearing conditions. The inset illustrates the average spike waveforms (on different time- and amplitude scales) obtained for the three hearing conditions. (F) average spike-density functions for a500 (red) and a1000 (blue) trials, aligned to handle-bar release. Bold straight lines denote best-fit regression lines to the last 100 ms. Note steady increase of neural activity toward bar release.

In Figure 4.3 we present the results, based on the same analysis as in Figure 4.2, for the population of cells ($N = 148$). We pooled the spontaneous pre-stimulus firing, and the onset responses for A500 and A1000 trials, as they were randomly interleaved within a block of trials, and the animal had no prior knowledge about the static duration of the noise. We also pooled the top-down modulatory signals aligned to the reaction time, because in our previous study we have shown that these were the same for A500 and A1000 trials ((Massoudi et al., 2014); see also, e.g. Fig. 4.2F).
We determined the proportion of cells in the population (N = 148) that systematically increased their spontaneous activities during the last 100 ms prior to sound onset. We then plotted the slope (β) of the linear regression fits to the last 100 ms of the spontaneous activities (-100 < t < 0; see e.g., bold lines in Fig. 4.2D) for active vs. passive trials. The result of this analysis is shown in Figure 4.3A. The cells with a significant active slope (p<0.001; n=74) are shown as grey circles; non-significant cells (n=74) by open circles. Although many cells had their slopes near zero, the largest slopes were positive, and obtained for the active trials (t_{df=147} = -2.2, p = 0.03). This positive slope is indicative for an additional cortical signal that could possibly be attributed to sound-onset anticipation in the active trials, in which the monkey initiated the trial (and hence the sound) himself by pressing the handle bar (see Materials and Methods).

Figure 4.3B shows that the majority (72%) of AC neurons had a higher onset peak response in the passive condition than in the active trials (t_{df=147} = 7.2, p < 0.0001). We also analyzed the temporal locking of the spikes to sound onset through the synchronization index (see Materials and Methods; Eq. 4.1). Figure 4.3C shows that the onset synchronization is very high for both passive and active onset responses (typically > 94%). Yet, the spikes of the passively evoked onset peak are even better time locked than the spikes for the active trials (Fig. 4.3C; t_{df=147} = 8.3, p < 0.0001). This result shows that involvement in a behavioral acoustic task affects the precision of the onset peak response, which by itself could be considered as a bottom-up acoustic response with little acoustic tuning specificity.

Although many auditory cortex cells carried a clear modulatory top-down signal that is aligned to the reaction time (e.g., Fig. 4.2F), others did not. In our previous study (Massoudi et al., 2014), we reported that 47% of the included auditory cortex cells (n=100) carried this signal. Here, we fitted simple linear regression lines to the last 100 ms before the reaction time (cf. bold lines in Fig. 4.2F), and plotted the resulting slopes (β) for A1000 vs. A500 trials. Fig. 4.3D shows that there is good agreement between the slope values for both trial types (r^2 = 0.78, p < 0.0001) across the larger population of 148 cells. Moreover, 70 cells (47%) yielded significant slopes for both A500 and A1000 trials (grey circles in Fig. 4.3D).
Figure 4.3. Population analysis of activity epochs for passive vs. active listening. (A) Slopes of the regression fit to the last 100 ms before sound onset for pooled active vs. passive trials. Grey circles: cells with at least one significant slope; open circles: non-significant slopes. Slopes for active trials are typically larger than for passive trials (B) Peak of the onset responses for the pooled A500 and A1000 vs. passive trials. Note that the onset peak is higher in passive trials for the majority of cells. (C) The onset synchronization index (SI, Methods) for passive against pooled A500 and A1000 trials. The passive onset spikes were better synchronized for the majority of cells. (D) Slope of the top-down behavioral signal prior to the reaction time for A500 (abscissa) and A1000 (ordinate) trials. Grey circles: cells with a significant slope; open circles: non-significant slopes.

4.3.3 CHANGES IN BEHAVIORAL CONTEXT INFLUENCE TEMPORAL PROCESSING

Figure 4.4 shows the four different MTF measurements, described in Methods, for cell T84. The magnitude MTF (mMTF) in Figure 4.4A corresponds to the temporal modulation filter of this auditory cortex cell (Miller et al., 2002). The mMTF values follow qualitatively similar patterns for passive (green), A500 (red), and A1000 (blue) trials, with a clear peak between 16<AMF<90 Hz. Overall the mMTF values are lower for the passive than for the active trials (Fig. 4.4A). The active experiments also elicited higher mean firing rates (rMTF) for A500 (Fig. 4.3B, red) and A1000 trials (Fig. 4.4B, blue), than for passive trials (Fig. 4.4B, green). Again, all three trial types yielded qualitatively similar patterns across AMFs. While the active responses increased more profoundly for the highest AMF (at 181 and 256 Hz; filled triangles, Fig. 4.4B), the rMTFs were slightly suppressed
for modulation frequencies between 22 to 128 Hz (Fig. 4.4B), which roughly corresponds to the location of the peak in the vsMTF (Fig. 4.4C). This characteristic suggests that this cell decreased its high firing rate to better encode its temporally preferred AMFs (higher vector strength). The vector strengths in the vsMTFs follow a similar pattern across AMFs for the three hearing conditions, although the MTF again rises earlier (AMFs > 11 Hz) for passive hearing (Fig. 4.4C).

Even at the high firing rates (Fig. 4.4B), the cell still shows a high trial similarity (tsMTF>0.8) for AMFs < 64 Hz in the passive hearing condition (green, Fig. 4.4D). In general, the active trials yielded lower tsMTFs, but most values were significant for AMFs < 64 Hz (red and blue triangles, Fig. 4.4D). Based on the different characteristics in panels 4.4B and 4.4C, one could characterize cell T84 as a significant rate-coding cell.

Figure 4.4. Task engagement influences sound-envelope processing. (A-D) The four panels contain the different modulation transfer functions (MTFs): temporal-synchronized firing rate (mMTF, in A), average firing rate (rMTF, DC, in B), vector strength (vsMTF, in C) and trial similarity (tsMTF, in D). Values correspond to passive (green), A500 (red) and A1000 (blue) conditions, respectively. Markers indicate statistical significance of the metric (at p<0.001; Methods). Vertical ticks denote the standard error. Numbers in each panel denote the Similarity Indices (SI) for the MTFs (Eq. 4.5, Methods).

Figure 4.5 illustrates our MTF analyses for a different cell (J21). In this case, the mMTF curves are virtually identical for the three behavioral trial types, with a peak at AMF = 64 Hz (Fig. 4.5A). The rMTF and vsMTF patterns both have a slightly broader tuning characteristic for AMFs >
11 Hz, but both peak also around 64 Hz (Figs. 4.5B and 5C), and fall off for the highest modulation frequencies. While in general the rMTF was higher for the active trials (Fig. 4.5B), the passive responses were better phase locked to AMFs, as indicated by the higher values of the vsMTF (Fig. 4.5C). The tsMTFs started to rise earlier (AMFs > 5.6 Hz) for the passive condition. To summarize these characteristics, cell J21 may be regarded as a predominantly temporal, or phase-locking, neuron.

Figure 4.5. MTF analysis for neuron J21. This neuron has strong phase-locking for AMFs in the mid-frequency range, whereas its rate coding capacity drops at higher frequencies. Same conventions as in Fig. 4.4.

In Figures 4.4 and 4.5 we illustrated the effects of the behavioral context on the temporal and rate response properties for a range of sound-modulation frequencies for two example neurons. While the detailed effects on the temporal and rate-coding characteristics could vary considerably from neuron to neuron, we now attempt to characterize and quantify the effects for temporal and rate coding at the population level. To that end, we divided the cells into two groups for the mMTF and rMTF measures, respectively: a significant and a non-significant group (see Materials and Methods; see also Malone et al., 2013). The firing rates for the mMTF and rMTF were first normalized for each cell by dividing each response by the maximum firing rate obtained at its best AMF in the passive hearing condition. We then converted the resulting normalized values to dB, in order to moderate the effects of the high-firing rate neurons on the population averages. Note that because of the averaging
across cells, and because cells have their optimal AMFs distributed across frequencies, the mean values of the normalized firing rates will typically lie well below zero dB.

Figure 4.6A shows the population averages for the significant mMTFs (see Materials and Methods). The A500 (red, Fig. 4.6A) and A1000 (blue, Fig. 4.6A) mMTFs were virtually identical, but the passive mMTFs coincided with the active results only for AMF between 5.6 and 45 Hz. Also for the non-significant MTFs, the mMTFs of the A500 and A1000 trials remained identical. In contrast, the passive mMTFs were lower than the active results for all AMFs. Overall, both the significant and nonsignificant mMTFs display a low-pass characteristic for all listening conditions. From the significant mMTFs we obtain an estimated cut-off frequency for temporal envelope following at about 45-64 Hz, which is remarkably high.

To determine the actual contribution of sound and/or behavior to the changes in neural firing, we subtracted for each neuron’s response during AM presentation its mean spontaneous activity for each behavioral condition (taken over the first 200 ms of the recording, prior to sound presentation; see Materials and Methods). The results revealed that, although, the rMTF appeared higher for the active trials (Figs. 4.4B and 5B), the spontaneous-adjusted driven rMTFs of both passive and A1000 trials were identical for the significant cells (blue and green, Fig. 4.6C). The A500 rMTFs, however, showed a steeper increase between AMFs = 2 - 11 Hz (red, Fig. 4.6C). This was not observed for the A1000 responses (blue). The cause for this apparent different behavior relates to the differences in reaction times for A1000 and A500 trials as function of AMF (Fig. 4.1; see Discussion).

Overall, in contrast to the mMTF, the rMTFs showed a high-pass characteristic across the significant population, for the three hearing conditions. The results suggest a high-pass cutoff at about 11 Hz. A different behavior resulted for the non-significant cells, for which the responses seemed predominantly flat, and the overall passive spontaneous-adjusted driven rMTF was significantly higher than for both active trials (Fig. 4.6D). In summary, task-related behavior could have a moderate influence on the temporal encoding of AM noises at low and high AMFs, and led to changes in the firing rates for rate-coding cells at the low modulation frequencies.

By performing a similar population analysis on the vector strengths across the population, the resulting average vsMTF was dominated by low values (<0.4) for all listening conditions, irrespective of the significance or non-significance of the cells (not shown). This may seem at odds with the example cell of Fig. 4.4C, where the vector strength resulted be quite high for a range of AMFs. The low average values are due to two factors: first, there is a wide distribution of best frequencies for vector strength, and second, the vsMTFs tend to have a limited bandwidth. As a result, the average vector strength for each AMF appears to be low (i.e. lower than the maximum vector strengths obtained at that frequency). The same logic applies to the trial similarity analysis for the population, which also led to low values (not shown), despite the fact that individual cells could have quite high trial-similarity values for a particular range of AMFs (Figs. 4.4D and 4.5D). The averaging across cells had a limited effect on the population characteristics of the mMTF and rMTF because temporal
encoding of AC cells dominated for AMFs $< 64$ Hz, while rate encoding dominated for AMFs $> 11$ Hz. Thus, the overall low-pass and high-pass characteristics for these respective cell groups survived averaging across cells.

**Figure 4.6.** Average population MTFs. The average MTF for (A, B) first-harmonic magnitude, and (C, D) the mean firing rate corrected for spontaneous activity, separated for cells with at least one significant value in any of the three hearing conditions (left), and cells without any significant values (right). Firing rates for mMTF and rMTF were converted to dB relative to the maximum firing rate in the passive condition (see text). Color indicates behavioral condition (green: passive; red: A500; blue: A1000). Markers denote that the active metric differed significantly from the passive metric (Wilcoxon signed rank test, $p<0.001$; see Methods). Vertical ticks indicate standard error (see Methods). N: number of cells contributing to the graph.

Figure 4.7A illustrates the distribution of maximum vsMTFs for passive (green), A500 (red), and A1000 (blue) trials for the 148 cells. Although the distributions of passive and A500 trials did not differ ($t_{df=147} = -0.14$, $p = 0.89$), the A1000 values differed from the other two ($t_{A500,A1000, df=147} = 2.6$, $p = 0.009$; $t_{A500,A1000, df=147} = 4.6$, $p < 0.0001$), mostly due to differences in the tail (large CS values) of the distributions. While a substantial minority of cells had relatively high vsMTF values (> 0.6) for passive (N = 33; 22%) and A500 (N = 30; 20%) trials, this was true for only 14 cells (10%) in the A1000 trials. The lower numbers found in the latter condition are again related to the fast reaction times in these trials at the lower AMFs, where vector strength is typically highest.
Figure 4.7B shows the distributions of best trial similarity values. The passive trials yielded slightly higher similarity values, and differed significantly from the A500 ($t_{p-A500, df=147} = 4.3, p < 0.0001$) and A1000 ($t_{p-A1000, df=147} = 4, p < 0.0001$) trials. The values obtained for the two active trial types were virtually identical ($t_{A500-A1000, df=147} = -1, p = 0.34$). Although a considerable number of cells had relatively high trial similarity values, the majority of tsMTFs values fell below 0.4, indicating an inconsistency in the spike trains of these AC cells for the different behavioral conditions.

![Figure 4.7.](image)

**Figure 4.7.** Maximum vsMTF and tsMTF values distributions. (A) The vector strength (vsMTF) maximum values distributions of 148 AC neurons for passive (green), A500 (red), and A1000 (blue) trials. (B) The distribution of maximum trial similarity values (tsMTF) for the three behavioral conditions. The maximum values were selected, irrespective of their AMF.

### 4.3.3.1 Best Modulation Frequencies

From the rate and vector strength modulation transfer functions (rMTF, vsMTF) we determined the best AMF for each of the three behavioral conditions for each cell (see Materials and Methods). Figure 4.8A,B shows the joint distributions of rBMF (Figs. 4.8A,B) and vsBMF (Figs. 4.8C,D) for passive listening (ordinate) against the A500 (Figs. 4.8A, C) and A1000 (Figs. 4.8B,D) hearing conditions. Note that a purely acoustic cell would not be affected by the behavioral state, and the marginal distributions (histograms provided on the top and left of the matrices) should be identical. Moreover, there should be a robust diagonal structure in the joint distributions, indicating a high correspondence with the two BMFs.

Instead, Figures 4.8A and B show that the marginal distributions for rBMFs were slightly shifted towards higher AMFs for the passive condition: the median rBMF was 32 Hz for passive, 22.6 Hz for A500, and 16 Hz for A1000 trials. The rBMF distributions were broad for the three hearing conditions, and not significantly different between passive against A500 ($p = 0.99$; two-sample Kolmogorov–Smirnov (KS) test), or A1000 ($p = 0.07$, KS test). Moreover, the agreement between
rBMFs for passive against A500 and A1000 was fair for AMFs > 64 Hz, as judged from the diagonal structure in the joint distribution of rBMFs.

The marginal distributions of vsBMFs are shown in Figures 4.8C and D. These distributions are skewed towards the low AMFs. While the distributions of BMFs were indistinguishable for passive and A1000 (p = 0.70, KS test), they were slightly different for passive vs. A500 (p = 0.036, KS test). However, because a temporal code, quantified by the vsMTF, typically belongs to low AMFs (see Joris et al., 2004), we also tested the vsBMF distributions for AMFs <= 16 Hz. In that case the BMF distributions remained indistinguishable (p_{passive-A500} = 0.13, p_{passive-A1000} = 0.40, KS test). This is further supported by the agreement between vsBMFs for the low AMFs as seen along the diagonal axis in the joint distributions (Figs 4.8C and D).

Taken together, these results imply that despite the overall behavior-induced changes in the MTFs (Figs. 4.4-7), the preferred AMF remained the same.

**Figure 4.8.** Analysis of best modulation frequencies for passive vs. active listening. Panels show joint distributions of each neuron’s best modulation frequencies derived from the cell’s rMTF (A,B) and vsMTF (C,D), respectively. Histograms denote the marginal cell-count distributions (normalized to their maximum) as function of AMF for active (red, blue) and passive (green) conditions. (A,C) Passive vs. A500 data. (B,D) Passive vs. A1000 data. Color saturation in each matrix corresponds to cell count from zero (white) to maximum (dark red/blue). Data for all 148 cells. Note that vector strength predominantly peaks at the low AMF range (~8 Hz), whereas rate tuning tends to prefer high AMFs as well. The listening condition affects a cell’s best AMF, because the data are not concentrated along the main diagonal.
4.4 DISCUSSION

4.4.1 SUMMARY

To our knowledge, this is the first study applying the magnitude of the first harmonic from the Fourier spectrum of the period histograms (mMTF) as a measure for encoding AM modulations. Although the overall increase in the mean firing rate (rMTF) in the detection tasks was mostly due to a higher spontaneous activity in active trials (e.g., Fig. 4.2D), the rMTFs and the strength of phase locking (vsMTF) of AC neurons were both influenced by the task in different AM-frequency ranges (Figs. 4.4-7). Interestingly, the mMTFs remained virtually unaffected by the task for amplitude modulations between 5 to 45 Hz (Fig. 4.6A). Within this same range of modulations the spectrotemporal receptive fields of AC neurons were shown to be stable during the same task conditions (Massoudi et al., 2013). An invariant mMTF during different listening conditions could suggest that AC cells employ a mixed code of rate and temporal coding.

4.4.2 BEHAVIORAL RESPONSE PROPERTIES

In our previous study we showed that monkeys rely predominantly on acoustic modulations to initiate their response in unpredictable trials (A500), and not in predictable trials (A1000; Jaramillo and Zador, 2011; Massoudi et al., 2013, 2014). The analysis in Fig. 4.1 extends the earlier results by looking at the different bins of the reaction-time distributions, taken from recordings that were collected over more than a year. Despite the large inter-session variability of median reaction times, the A1000 trials were always faster than A500 trials (Fig. 4.1C), and highly correlated. We observed no systematic trend with time in the mean or median reaction times, which indicates that the animals did not improve by perceptual or procedural learning. The collected reaction times could thus be pooled without the need for normalization, leading to a consistent pattern for the entire data set (Fig. 4.1B). The data indicate that when animals reacted faster or slower on a given day in A500 trials, they would also react faster or slower for the A1000 trials on that day by about the same amount. Interestingly, the absolute reaction times were predictive for the animal’s sensitivity to sound modulations (quantified by the slope of Eq. 4.3; Fig. 4.1B), regardless when they were measured. These data show that reaction time is a reliable and sensitive parameter to assess auditory perception.

4.4.3 EFFECT OF TASK ENGAGEMENT ON SPONTANEOUS AND ONSET RESPONSES

The different behavioral states affected the spontaneous and the evoked firing rates in auditory cortex cells (Fig. 4.2). Several studies have reported different effects on these quantities,
ranging from suppression (Otazu et al., 2009; Jaramillo and Zador, 2011), to no change at all (Benson and Hienz, 1978), and to response enhancement (Scott et al., 2007; Niwa et al., 2012b; Massoudi et al., 2013, 2014). These discrepancies may perhaps be due to subtle differences in behavioral demands (see, e.g., Sutter and Shamma, 2011), which could yield drastically different responses in cortical areas (Groh et al., 1996; Boudreau et al., 2006). In our behavioral paradigms the spontaneous activity was typically higher for active trials when compared to the passive trials (Fig. 4.2D). This overall enhancement might be attributed to an attention signal that was present only when animals were engaged in the auditory task. Moreover, in about 50% of AC cells the spontaneous firing rate in active trials ramped up about 100 ms prior to sound onset (Fig. 4.2D and 4.3A). This additional signal could reflect sound-onset anticipation, as the sound would always start 300 ms after the monkey pressed the handle bar (see, e.g., Brosch et al., 2005; Jaramillo and Zador, 2011; Massoudi et al., 2013, 2014).

Despite the lower spontaneous activity for passive listening, the cells’ onset responses in passive trials were typically higher than in active trials for the majority of cells, even without correcting for the increased baseline (Fig. 4.3B). This suggests that active engagement in an auditory task may have a suppressive effect on a cell’s bottom-up onset activity. A similar effect was reported by Beaton and Miller (1975) in a frequency-discrimination task.

Our results further indicate that the spikes associated with the sound onset were strongly time-locked to the sound for all behavioral conditions (synchronization index > 94%; Fig. 4.3C), but systematically better for passive hearing. Taken together, our results imply that engaging in an auditory task moderately weakens the precision of onset tuning properties of AC cells.

---

**4.4.4 EFFECT OF TASK ENGAGEMENT ON TEMPORAL ENVELOPE PROCESSING**

---

We employed four different MTF characteristics to test whether task involvement could alter the sound-envelope processing capacities of AC cells (Figs. 4.4-7). When looking at individual cells, the mMTFs for passive and active hearing could vary from identical mMTFs across the tested modulation frequencies (Fig. 4.5A), to a partial overlap for only a range of AMFs (Fig. 4.4A), to even entirely different response curves (not shown). At the population level (N=49; significant cells) the mMTFs were identical across hearing conditions for the range between 5.6 to 45 Hz, but differed between passive and active for both lower and higher modulation frequencies (Fig. 4.6A). The mMTFs for the A500 and A1000 trials were indistinguishable. As the mMTF is derived from the first harmonic of a cell’s period histogram, it equals the magnitude transfer function obtained from responses to spectral-temporal dynamic ripples (Massoudi et al., 2013). The current findings on amplitude modulated GWN over a much large ranger of modulation frequencies (2-256 Hz) extend the observations from Massoudi et al. (2013), which were restricted to ripple velocities between 8-40 Hz. That study showed that the spectral-temporal receptive fields of AC cells were unaffected by
auditory task performance, despite the considerable changes in a neuron’s firing behavior. Here we demonstrate that task invariance of the neural responses may in fact be band-limited to the 5 – 45 Hz range of sound-amplitude modulations.

The reason for this band-limited task invariance is not clear. Possibly, it reflects the different encoding mechanisms for amplitude modulations employed by AC cells. Low-frequency modulations appear to be predominantly encoded by temporal locking, while rate coding is the main mechanism for high AMF (Joris et al., 2004). The cut-off frequencies associated with these two mechanisms depend on the auditory stage (cortical vs. sub-cortical), and may also reflect species differences (Lu et al., 2001; Liang et al., 2002; Joris et al., 2004; Bartlett and Wang, 2007; Yin et al., 2011). Our AC data suggest that temporal codes dominate for AMFs < 5.6 Hz, and rate codes for AMFs > 45 Hz. Both mechanisms could play a role in the range between 5.6-45 Hz. The latter is indeed reflected by the mMTF measure (Eq. 4.2). Thus, if task engagement would hamper one encoding mechanism (specifically, temporal at the higher modulation frequencies), the other mechanism could compensate, resulting in invariant tuning.

We observed that the mean firing rate in active trials tended to be higher than in passive trials (e.g., Figs. 4.4B and 5B; Massoudi et al., 2013, 2014). Two factors could contribute to these higher average firing rates (rMTF): a higher overall spontaneous activity in active trials (Figs. 4.2D and 4.3A), and an additional top-down signal related to the preparation of bar release (or reward expectation) during the static noise and AM sound (Figs. 4.2F, and D; Massoudi et al., 2014).

Recently, Niwa et al. (2012b), showed that after normalizing active responses relative to the spontaneous, the passive driven sound-evoked activity was actually higher (Niwa et al., 2012b). Our results, however, indicate that the spontaneous-adjusted firing rates of A500/A1000 and passive trials were virtually identical for the higher modulation frequencies (Fig. 4.6C). This also implies that the amplitude of the top-down signal that is related to the reaction time, and which is proportional to the acoustically driven activity through a multiplicative interaction (see Massoudi et al., 2014), is too small to drastically affect the average sound evoked firing rate. These findings are in line with studies that reported no effect of task engagement in AC (Hoehman et al., 1976, 1981; Gilat and Perlman, 1984). Many studies, however, have demonstrated an increase in the CA responses in auditory task performance (e.g., Miller et al., 1972; Gottlieb et al., 1989; Scott et al., 2007), while others showed decreased responses in rats (Otazu et al., 2009), ferrets (Atiani et al., 2009), and an overall inhibition during behavior in the somatosensory cortex of mice (Crochet and Petersen, 2006; Poulet and Petersen, 2008). Possibly, these differences in results might be due to task requirements, stimulus specifics, and/or species differences. Indeed, we observed differences in baseline corrected firing rates for the lower modulation frequencies (Fig. 4.6C).

In the A500 trials the (baseline corrected) firing rates fell well below those of the passive trials for the low-frequency modulations between AMFs = 2 - 11 Hz (Fig. 4.6C). Unexpectedly, we obtained a different behavior of the firing rates for the A1000 trials. This is unexpected, as animals
had no prior knowledge as to whether they would be engaged in an A500 or in an A1000 trial. This was indeed apparent from the baseline firing patterns and sound onset responses, which were indistinguishable for the two types of trials (Figs. 4.2 and 4.3). Moreover, neural responses during the first 500 ms of static noise were identical as well (Massoudi et al., 2013). The only acoustic difference between the two trial types is the additional 500 ms of static noise in A1000 trials. If this acoustic difference would underlie the overall differences in firing rates during the amplitude modulations, the rMTFs should have been different for all modulation frequencies. We have shown that the top-down modulatory signal related to the reaction time is also identical for the two trial types (Massoudi et al., 2014), so that the different firing rates cannot be attributed to large differences in top-down modulations either.

The only remaining explanation for the observed differences between A500 and A1000 rMTFs is the systematic difference in reaction times. The A1000 reaction times were in large part based on prediction (> 60% of the trials), rather than on stimulus acoustics, whereas for A500 trials reaction times depended on modulation frequency (Fig. 4.1). In Figure 4.6C the A500 rMTFs up to 11 Hz fall well below the passive and A1000 trials. The reaction times for A1000 trials were short for all modulation frequencies, and on average almost 150 ms faster than the shortest A500 trials. However, after the monkey releases the bar, the firing rate stays at a higher level until the next trial (Fig. 4.2F; also, Massoudi et al., 2014). This additional activity is not related to acoustic or behavioral events. As a result, trials with very short reaction times produce higher average firing rates than trials with long reaction times, and this is indeed what we observed for the low-frequency A1000 trials. To obtain a purer measure for the rMTF, it would be preferred to calculate the firing rate up to about 100 ms before bar release (at the onset of the top-down modulation; Fig. 4.2F). Unfortunately, for A1000 trials most responses to low-frequency modulations had to be discarded, as too few acoustic responses remained for a reliable statistic. However, given the consistent similarities in response behaviors of AC cells during A500 and A1000 trials up to the moment of bar release, we conjecture that the A1000 rMTFs are expected to be very similar to the A500 rMTFs, and hence different from the passive responses up to about 11 Hz. The same reasoning applies to the apparent A500-A1000 differences seen in the distribution of best vsMTFs (Fig. 4.7A), since the highest values for vector strength are obtained for low modulation frequencies.

Although some auditory cortex cells had a relatively high vsMTF (Fig. 4.5C), for the majority of cells they resulted to be low (Figs. 4.4C and 4.7A). For 33/148 cells (22%) we obtained significant vsMTF values for at least one behavioral condition. This contrasts with the results of Malone et al. (2013) who reported a significant percentage of 82%. Note that this difference is already apparent for the passive responses (Fig. 4.7A), so that it is not due to our behavioral paradigms. Perhaps the difference betrays a species differences (squirrel monkey and macaque) or differences in recording sites within AC. A low vsMTF does not necessarily mean absence of phase locking in a cell’s spiking activity, but rather indicates the absence of a clear peak at a specific phase of the frequency of interest.
(Joris et al., 2004). Another potential cause for low vsMTF values could be the very high firing rates observed in monkey AC cells (Joris et al., 2004). Our findings suggest that both vector strength and trial similarity measures are by themselves not adequate to quantify temporal envelope processing in AC cells during a behavioral paradigm.

The best modulation frequency for each of the four MTF characteristics remained similar for the different hearing conditions (Fig. 4.8; non-significant KS-statistics). This result is in line with a recent study by Malone et al. (2013) showing that different carriers are less effective on the BMFs. Because of the observed differences in MTF curves (Fig. 4.6), our results suggest that especially the non-BMF modulation frequencies are vulnerable to changes imposed by auditory task performance.
5. TASK-RELATED PREPARATORY MODULATIONS MULTIPLY WITH ACOUSTIC PROCESSING IN MONKEY AUDITORY CORTEX

Roohollah Massoudi
Marc M. Van Wanrooij
Sigrid M.C.I. Van Wetter
Huib Versnel
John Van Opstal

Accepted by *European Journal of Neuroscience*, online since March 20, 2014.
5.1 INTRODUCTION

The mammalian auditory cortex (AC) not only encodes the spectral-temporal acoustical properties of sounds, but also processes a variety of other signals. For example, AC neurons has been shown to be modulated by multisensory integration (Schroeder et al., 2001; Foxe et al., 2002; Fu et al., 2003; Brosch et al., 2005; Ghazanfar et al., 2005; Kayser et al., 2007, 2008b, 2009, 2010; Lakatos et al., 2007), by attentive behavior (Hubel et al., 1959; Fritz et al., 2003, 2005a, 2007; Sussman et al., 2005, 2007; Brechmann and Scheich, 2005; Scheich et al., 2007; Yin et al., 2008; Atiani et al., 2009; Niwa et al., 2012b, 2012a; Bizley et al., 2013; Dong et al., 2013; Massoudi et al., 2013), by auditory learning and conditioning (Bakin et al., 1996; Ohl and Scheich, 1997, 2005; Ji et al., 2001; Kilgard et al., 2001, 2002; Kilgard and Merzenich, 2002), by expected reward (Jaramillo and Zador, 2011; David et al., 2012), by the sound’s location (Recanzone, 2000; Lee and Middlebrooks, 2011), and by changes in eye position (Werner-Reiss et al., 2003; Maier and Groh, 2010).

Non-acoustic top-down signals could interact with acoustic signal processing in a variety of ways. For a purely additive signal interaction, the neural response in a given trial under active listening is a linear combination of the response to the acoustic signal response and a non-acoustic top-down signal. In case of a purely multiplicative interaction, the latter combines with the acoustically evoked response through a nonlinear, multiplicative gain modulation. These different modes of interaction are usually not dissociated (Maier and Groh, 2010).

Here, we determined the type of interaction at the single-neuron level in AC during a simple auditory reaction-time task. In such a task the acoustic tuning properties of AC neurons are unaffected by task performance, although evoked neural spike trains during active listening differ substantially from those recorded during passive sound exposure (Massoudi et al., 2013). A stable acoustic representation by AC cells could facilitate the formation of an invariant percept of the acoustic environment when changing between passive and active listening conditions. While the systematic differences in neural activity patterns suggested the presence of a top-down signal, the precise nature of this signal remained elusive (Massoudi et al., 2013).

In this study, we characterized this signal by analyzing the type of interaction between sound-evoked and top-down signals in AC responses, as well as the potential contribution of acoustic and non-acoustic factors to this interaction. To that end, we compared single-unit activity to a wide range of acoustic stimuli (dynamic spectral-temporal ripples, and amplitude-modulated [AM] Gaussian noises), presented during either passive sound exposure, or during an auditory reaction-time task. As the latter contained unexpected (in non-predictable trials) as well as predictable onsets of spectral and/or temporal sound modulations that were randomly interleaved, it added a cognitive component to the paradigm (the presence or absence of prediction). We analyzed the monkeys’ reaction times as a function of the spectral and temporal modulation rates and of modulated sound-onset predictability,
and identified the neural signature of these behavioral components on the single-unit activity patterns. Finally, we quantitatively dissociated additive and multiplicative signal-interaction modes in the neural response patterns.

5.2 MATERIALS AND METHODS

5.2.1 SUBJECTS

We performed neurophysiological recordings in the left auditory cortex of two adult male rhesus monkeys (Macaca mulatta, Monkey J; 7-9 kg, and monkey T; 8-10 kg). Animals participated in the recording sessions for about two years. They were trained to manually respond to an acoustic change from a static broadband sound into a temporal and/or spectral modulation to receive a drop of water as a reward for each successful trial. Experiments were conducted in accordance with the European Communities Parliament and Council Directive (September 22, 2010, 2010/63/EU). All experimental protocols were approved by the local Ethics Committee on Animal Research of the Radboud University Nijmegen (RU-DEC, ‘Radboud University Dier Experimenten Commissie’). Monkeys were pair-housed to facilitate normal interactive behavior. About 24 hours before the start of an experimental session, water intake was limited to 20 ml/kg. In the experiment, the monkey earned a small water reward of 0.2 ml per successful trial. We ensured that the animals earned at least the minimum of 20 ml/kg on an experimental day. After an experimental session, water was supplemented to the required minimum amount, if needed, and the animal received additional pieces of fruit. In weekends, the animals’ fluid intake was increased to 400 ml daily. To monitor the animal’s health status, we kept records of body weight, and water and food intake. Expert veterinarian assistance was available on site. Quarterly testing of hematocrit values ensured that the animal’s kidney function remained within the normal physiological range. Our procedures followed the water-restriction protocol of the Animal Use and Care Administrative Advisory Committee of the University of California at Davis (UC Davis, AUCAAC, 2001). Whenever an animal showed signs of discomfort, or illness, experiments were stopped and the animal was treated until the problem was solved.

5.2.2 SURGICAL PROCEDURES

After completing the initial training (when the monkey reached a day-to-day performance level of at least 80%), the animal underwent surgery under full anesthesia and sterile conditions. Anesthesia was maintained by artificial respiration (0.5% isoflurane and N2O), and additional pentobarbital (IV; 3 mg/kg/hour), ketamine (IM; 0.1 mL/kg), and fentanyl (IV; 20 μg/kg/hour) were
administered. A stainless steel recording chamber (12 mm diameter) was placed over a trepanned hole in the skull (10 mm diameter). The orientation and coordinates of the chamber were directed to the auditory cortex, as determined on the basis of MRI images. The chamber allowed a vertical approach of the left AC. A stainless-steel bolt, embedded in dental cement on the skull, allowed firm fixation of the head during recording sessions.

5.2.3 EXPERIMENTAL SETUP

The head-restrained monkey sat in a primate chair within a completely dark and sound-attenuated room (2.45×2.45×3.5 m), while a glass coated tungsten microelectrode (impedance 1-2 MΩ; Alpha Omega, Ubstadt-Weiher, Germany) was carefully positioned and lowered into the brain through a stainless steel guide tube by an electronically-driven stepping motor (National Aperture Inc. MM-3M-F-1). Electrode signals were grounded to a contact mounted in the skull. The analog electrode signal was amplified (BAK Electronics; model A-1), band-pass filtered between 0.1 and 15 kHz (custom-built 8th order Butterworth filter; Krohn-Hite, model 3343); 100 Hz HP cut-off), and monitored through a speaker and oscilloscope. The raw signal was then digitized (at 25 kHz, A/D convertor, TDT2 system; module AD-1; Tucker-Davis Technologies). An automated algorithm detected individual action potentials (BrainWare, V 9.07 for TDT, running on a PC, Windows 98; DELL). Data analysis and spike sorting was performed offline in MATLAB (version 7.14.0.739, R2012a, Natick, MA, USA).

During the experimental and training sessions the behavior and wakefulness of the animals were observed via an infra-red camera.

5.2.4 SOUND STIMULI

Sound stimuli were digitally generated at a sampling rate of 100 kHz and delivered via Brainware-software and TDT2 hardware. A trigger, provided by a TG6 module, started sound presentation (DA1, low-pass filtered at 20 kHz through a TDT2-FT6 module), and spike data acquisition. A speaker (Blaupunkt PCxg352, flat frequency characteristics within 5 dB between 0.2 and 20 kHz), positioned at the frontal central position at a distance of 1.0 m from the monkey, presented the stimuli at a fixed sound level of 60 dBA (set by two programmable attenuators, PA4).

The ambient background acoustic noise level was about 30 dBA. Acoustic foam that was mounted on the walls, floor, ceiling, and every large object in the room effectively absorbed reflections above 500 Hz. In this study, we presented two different types of acoustic stimuli: (1) frozen static noise, followed by a spectrotemporal ripple, and (2) freshly generated static Gaussian white noise (GWN), followed by amplitude modulated (AM) noise.
**Tones.** Pure tones lasted for 150 ms and were presented over a frequency range from 250 to 16000 Hz at 4 different sound levels (10, 30, 50, and 70 dB sound pressure level (SPL)). Trials were presented at least four times in a randomized manner. The frequency-tuning curve of a neuron was determined from the average firing rate for each tone across all sound level presentations. The best frequency (BF_tone) of each neuron was taken at the maximum of the tuning curve. Characteristic frequency (CF) was defined as the frequency that produced a response higher than the mean plus 2 SDs of the baseline activity at the lowest intensity. The cell’s response-onset latency was defined as the moment after the pure-tone onset at which the firing rate exceeded the mean baseline activity plus twice its standard deviation (SD) for the first time for at least 10 milliseconds. The peak latency was the time at which the firing rate reached its maximum.

**Ripples.** The ripple stimuli consisted of a broadband complex of 126 spectral components, equally distributed (20/octave) from 250 Hz to 20 kHz (Depireux et al., 2001; Versnel et al., 2009). All components had random phase. The envelopes of the complex were sinusoidally modulated in the spectrotemporal domain. The amplitude of each component was described as follows:

\[
S(t, x) = 1 + \sin(2\pi \omega t + \Omega x) \tag{Eq. 5.1}
\]

with \( t \) time (in S), \( x \) position of the spectral component in octaves above the lowest frequency (250 Hz), \( \omega \) ripple velocity (Hz; temporal modulation), \( \Omega \) ripple density [cycles/octave, or c/o; spectral modulation].

The set of 55 different ripples used in our study consisted of all combinations of 11 different ripple densities, \( \Omega \) in \([-2.0:0.4:+2.0]\) c/o, and 5 different velocities, \( \omega \) in \([8:8:40]\) Hz. A negative density corresponds to an upward direction of the spectral envelopes, a positive density to a downward direction, and \( \Omega = 0 \) means a pure amplitude-modulated (AM) sound (e.g. Versnel et al., 2009). For all ripples the modulation depth was 100%. The static noise was frozen within each block of trials, and was generated setting the ripple modulation depth to 0, i.e., all spectral components had the same amplitude. The sound level of the static noise and ripple was 60 dB SPL.

**AM noise.** The AM noise stimuli were temporally modulated in 15 steps, logarithmically spaced from 2 to 256 Hz. Like the dynamic ripples, the modulation depth was 100%. In contrast to the ripples, the broadband noise was generated fresh for each stimulus. The average sound level was 60 dB SPL.

**5.2.5 EXPERIMENTAL PARADIGMS**

Neural responses were measured while monkeys were exposed to the spectral-temporal ripples and AM noises, each presented in different blocks of trials. Each stimulus started (Fig. 5.2, sound onset) with a static epoch (\( \omega=0 \) Hz, \( \Omega=0 \) c/o), which smoothly changed (Fig. 5.2, dynamic onset) into a dynamic epoch that was either a pseudo-randomly selected ripple that lasted for 1000 ms, or an AM noise, which lasted for 700 ms (Fig. 5.2, sound offset).
Sounds were presented to the animal in two different behavioral paradigms: i) passive listening (Fig. 5.2, top), in which the monkey was merely exposed to the sounds without a task, and the fixation light was off; ii) active listening: upon presentation of a red fixation light at straight ahead, the monkey had to initiate a trial by pressing a light-weight bar. It heard the same sounds as during passive exposure, but now had to respond to the sound-modulation onset within 100-600 ms to receive a drop of water as reward. Trials in which the monkey did not detect the ripple onset correctly were repeated at a randomly selected trial within the same recording block. The passive and active paradigms were presented in different blocks, typically starting with the passive sound exposure.

In the passive experiment the trial started automatically with the static noise (Fig. 5.2, sound onset). Data collection started 300 ms before sound onset. For ripples the trial ended 700 ms after sound offset, yielding a total recording duration of 2500 ms. For AM noise stimuli the data acquisition ended 500 ms after AM offset, giving a recording duration of 2000 ms per trial (Fig. 5.2, top). The number of trial repetitions was between 4 and 10; the inter-stimulus interval was 2.5 s.

The active-listening paradigm contained A500 (Fig. 5.2, center) and A1000 (Fig. 5.2, bottom) trials that were randomly interleaved in the experiment at equal probability. The static noise was presented 300 ms after the monkey pressed the bar, and lasted either 500 ms (A500 trial), or 1000 ms (A1000 trial), upon which it changed into a spectral-temporal ripple (or AM noise) that lasted for 1000 ms (or 700 ms). All trials contained both static and dynamic epochs; there were no catch trials without a dynamic modulation. For each trial the data recording started at the moment of bar press, and finished 2500 ms later. The number of repetitions depended on the monkey’s performance (typically no repetition for ripples, and 5 repetitions for AM).

5.2.6 CHARACTERIZATION OF RECORDING SITE

Although we cannot with certainty identify the exact AC subdivision(s) in which we encountered individual neurons, we are confident that we recorded from the AC core (primary auditory cortex A1 and its immediate rostral part, area R) for the following reasons: 1) MRI scans were used for stereotaxic placement of the recording chamber; the subsequent coordinates of the successful recording sites within the chamber corresponded closely to the stereotaxic coordinates of A1 as provided by the atlas of the Rhesus monkey brain by Paxinos et al., (2000); 2) For both monkeys we reconstructed the tonotopic organization of the recorded neurons, which demonstrated a systematic increase in characteristic frequency (CF) from anterior to posterior locations over a spatial range that corresponds well to other reports (e.g., Niwa et al., 2012a; Dong et al., 2013). This implies that the majority of recordings were likely taken from A1. As an illustration, Figure 5.1 shows the tonotopic map for the recorded cells of monkey J. For monkey T, the CF also increased over both the anterior-posterior and medial-lateral axes, which implies that recordings were taken from an area between R and A1 (not shown; Bendor and Wang, 2008). 3) before reaching an AC recording site
there was a physiologically silent period, corresponding to the gap between upper and lower parts of the lateral sulcus (Kaas and Hackett, 2000); 4) tone-onset latency of the recorded sites was 22.6 ± 5.9 and 23.6 ± 5.6 ms for monkey J and T, respectively; 5) all neurons responded well to pure tones (BF: 250 – 16000 Hz); 6) The pure-tone tuning bandwidths for monkey J and T were 1.5 ± 1.2 and 1.5 ± 1.3 octaves, respectively; 7) The pure-tone threshold for monkey J was 21 ± 13 dB SPL, and for monkey T it was 23 ± 12 dB SPL. These tuning characteristics all fall in the same ranges as reported by Recanzone et al. (2000) for behaving monkeys in AC areas A1, R and CM.

Recording stability was verified by analyzing spike-waveform variability during the different behavioral tasks. We included cells from which we obtained stable single-unit recordings with limited spike variability during presentation of both the ripples and AM noises in the active and passive paradigms. As a result, we obtained 100 cells from the two monkeys (J: n=57, T: n=43).

Figure 5.1. Reconstructed tonotopic map for monkey J. Characteristic frequency (color coded) increases from anterior to posterior recording sites, which is indicative for primary auditory cortex, A1. Each pixel represents the average best frequency found at that location.

5.2.7. DATA ANALYSIS

5.2.7.1 SPIKE-DENSITY FUNCTION

To convert each spike-raster plot into a continuous spike-density function (a measure for the instantaneous mean firing rate across trials), we first binned the recorded spike times into a digital sequence with a time resolution of one millisecond, and then convolved each spike with a Gaussian kernel with a standard deviation of 5 ms. Finally, the signals were added across trials, and we normalized the resulting function by the number of trials.
By default the data were aligned to sound onset (e.g., Figs. 5.5A and C). To relate the top-down signals to reaction time, we re-aligned all neural responses to the bar-release time by shifting the spike timings in each trial by the associated reaction time (e.g., Figs. 5.5B and D).

5.3 RESULTS

5.3.1 NEURAL RESPONSES: PASSIVE AND ACTIVE LISTENING

Figure 5.2 illustrates the experimental paradigms, and shows the neural responses of a representative neuron (J67), exposed to the AM stimuli (AM frequencies [AMFs] between 2 and 256 Hz). All sounds started with freshly generated static broadband GWN, which was followed by an amplitude modulation (Fig. 5.2, sound). In passive hearing (Fig. 5.2, top), animals were merely exposed to the sounds without a required behavioral response. In that case the neuron was well tuned to the acoustics of the stimuli, as evidenced from selective, phase-locked, responses to the different AMFs (see central group of trials in Fig. 5.2, top). Similar selective response behavior was obtained for the population of 100 neurons (data not shown here). This was also true for the neural responses to dynamic spectral-temporal ripples (see Massoudi et al., 2013), for which the acoustic modulations varied in the temporal (velocity between 8 and 40 Hz) and spectral (density between -2 and +2 cyc/oct) domains (see Methods).

In the active hearing trials (Fig. 5.2, center and bottom), the monkey pressed a bar to initiate a trial. This experiment contained two different conditions: in A500 trials (Fig. 5.2, center) the static noise lasted 500 ms, whereas in A1000 trials (Fig. 5.2, bottom) it lasted for 1000 ms. To obtain a reward the animal had to react, by releasing the bar within 600 ms, to the perceived sound change from static noise to dynamic modulations. Since both trial types were randomly interleaved at equal probabilities, the change occurred unexpectedly (probability 50%) in the A500 trials, but animals could in principle anticipate the sound onset in A1000 trials (with probability 100%), as soon as 500 ms static noise had passed (highlighted by the light-gray epoch in Fig. 5.2, bottom).

Being in a task drastically changed the neural responses, which was readily obvious in changes of the mean overall firing rates (e.g. prestimulus firing rate increased for active trials: ripple mean ± SD, passive: 18 ± 17, A500: 32 ± 31, A1000: 31 ± 30 spikes/s, AM mean ± SD, passive: 23 ± 23, A500: 35 ± 36, A1000: 35 ± 36 spikes/s). Here, we will not focus on these general differences, which might be due to a nonspecific attentional signal (like alertness), but focus instead on the substantial changes in the neural firing patterns related to task execution. During the dynamic response epochs of active trials (Fig. 5.2, center and bottom) the phase locking to acoustic modulations seemed less obvious in the spike rasters than for the passive trials (Fig. 5.2, top). This is due to two factors: i) a higher firing rate introduces additional noise to the firing patterns because of
the increased and variable baseline, and ii) additional activity that is locked to the monkey’s reaction time varies considerably from trial to trial, and therefore blurs the phase locking when spike rasters are aligned to sound onset, rather than to bar release. Note, however, that our previous study showed that despite these additional signals, the underlying phase locking of AC cells to ripples, and hence their spectral-temporal tuning characteristics, remained fully intact for the active trials (Massoudi et al., 2013). Below, we will demonstrate that the changes in firing patterns betray the presence of a dynamic task-related signal that interacts with the neuron’s acoustic tuning response, but varies systematically from trial to trial with the animal’s behavior. To better appreciate the different task-related factors of the experiment, we first quantify the behavioral effects of the acoustic and non-acoustic aspects of our paradigm.
Responses of cell J67 for AM noise during the passive (top), A500 (center) and A1000 (bottom) paradigms. Each panel shows the different epochs in the stimuli (shown below the spike rasters), indicated by dark-gray vertical lines (at sound onset, dynamic onset, sound offset). Spike rasters are sorted according to AMF (bottom: 2 Hz, top: 256 Hz). Each dot corresponds to a spike. Red lines show the trial-averaged spike-density curves (filtered with $\sigma=5$ ms kernel). The passive paradigm elicited clear phase locking to temporal sound modulations in the mid-AMF range. Center and bottom panels: active paradigm to the identical sounds for A500 and A1000 trials, respectively (randomly interleaved in the experiment). The phase locking to sound modulations seems lost. Note higher firing rates before stimulus onset, and a lower transient peak response. Panels also indicate the behavioral paradigm: bar press and release, and reward delivery. Gray histograms on the bar lines represent the reaction-time distributions of monkey J for this experiment. In A500 trials reaction times fall around the third stimulus modulation. In the A1000 paradigm, reaction times shift to much shorter values, around the first stimulus modulation, and sometimes even before modulation onset. Gray shading in the static stimulus epoch highlights the prediction period in which the monkey can anticipate the modulation onset.

5.3.2 BEHAVIOR

The A500 and A1000 trials led to clearly different behavioral responses of the monkeys (Figs. 5.2 and 5.3). For example, during the recording session of cell J67 (Fig. 5.2, grey histograms), the reaction times for AM A1000 trials were on average 198 ms shorter than for AM A500 trials ($t_{df=14} =$
13.5, P<0.0001; A1000: median ± SE: 170 ± 11 ms, A500: median ± SE: 368 ± 14 ms). This result was obtained for all recording sessions, both stimulus types, and both animals. In the unpredictable A500 trials the reaction times across sessions was 384 ± 21 ms for ripples, and 357 ± 21 ms for the AM noises, with reaction times typically falling after the sound change (e.g. Fig. 5.2, grey histogram center). These reaction times are in line with a sensory-evoked hand-movement response (Rogal et al., 1985). The predictability of sound-change onset in the A1000 trials invariably led to a substantial reduction of the reaction times to 188 ± 16 ms for ripples, and 158 ± 16 ms for AM noises, with 17.5% of the responses falling even before the stimulus change (e.g. Fig. 5.2, grey histogram, bottom). Importantly, although median reaction times varied substantially from day to day (as evidenced by the considerable standard deviations), there was no systematic trend in these values over time, or in the difference between non-predictive and predictive trials (data not shown). Thus, animals did not display evidence of procedural learning during the course of the experiments.

Table 5.1 represents the combined performance of the two monkeys in the active trials for AM and ripple stimuli. Trials were counted ‘correct’ (and hence rewarded) when the manual reaction times happened between 100 and 700 ms (see Materials and methods). In total, the correct trial percentages were around 90% for non-predictable A500 trials and 60-65% for the predictable A1000 trials. In about 4% of trials the animals did not respond with a manual reaction (‘misses’); these trials were excluded from the database for further analysis. Note that, whereas in 8.50% of the ripple A500 trials animals reacted later than 700 ms, it decreased to 0.04% for AM A500 trials, which implies that the AM stimuli were easier to detect than the dynamic ripples. Only about 1% of A500 trials were early. While late responses for A1000 trials were rare, more than 30% of them were early responses. This indicates that animals often released the bar on the basis of sound-modulation onset prediction in A1000 trials.

<table>
<thead>
<tr>
<th>Active trials</th>
<th>Total</th>
<th>correct</th>
<th>missed</th>
<th>late</th>
<th>early</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ripple A500</td>
<td>6602</td>
<td>87.00%</td>
<td>3.20%</td>
<td>8.50%</td>
<td>1.00%</td>
</tr>
<tr>
<td>Ripple A1000</td>
<td>6805</td>
<td>65.50%</td>
<td>2.00%</td>
<td>0.40%</td>
<td>32.00%</td>
</tr>
<tr>
<td>AM A500</td>
<td>7535</td>
<td>95.40%</td>
<td>3.60%</td>
<td>0.04%</td>
<td>0.90%</td>
</tr>
<tr>
<td>AM A1000</td>
<td>8875</td>
<td>61.40%</td>
<td>2.30%</td>
<td>0.00%</td>
<td>36.10%</td>
</tr>
</tbody>
</table>

Table 5.1. Behavioral performance. Data from both monkeys is pooled. Reaction times were between 100-700 ms for correct trials, < 100 ms for early trials, and > 700 ms in late trials. In the missed trials, the monkeys did not respond.

5.3.2.1 REACTION TIMES AND ACOUSTICS

To verify that the predictive and non-predictive differences reflected a cognitive component in the behavior (i.e., uncertainty and predictability of sound-modulation onset), we analyzed the
reaction times of both animals as a function of the acoustic modulation parameters for the 70 different stimuli (i.e., for 55 different spectral-temporal ripples, and 15 AM stimuli), and the two response conditions (A500 and A1000; data pooled and averaged across animals and cells).

When the modulation onset was unpredictable (i.e. a probability of 0.5 in A500 trials), the monkeys’ reaction times strongly depended on both spectral (Fig. 5.3A; ANOVA: $F_{df=5}=91, p<0.0001$) and temporal (Fig. 5.3A, C; ANOVA, ripples: $F_{df=4}=7.5, p<0.0001$; AM: $F_{df=14}=121, p<0.0001$) acoustic modulations. Ripple densities of 0 cyc/oct elicited the shortest reaction times of $337 \pm 2.5$ ms (mean $\pm$ SE; Fig. 5.3A, green), which gradually increased to $458 \pm 12$ ms for the highest ripple densities (-2 and +2 cyc/oct, Fig. 5.3A, dark red). Moreover, for the ripples the animals’ reaction times also reflected a significant interaction between spectral and temporal modulations (ANOVA: $F_{df=20}=1.88, p = 0.001$) with the combination of higher velocities and densities yielding an additional increase in reaction times.

In the AM detection paradigms we employed a much larger range (2-256 Hz) of temporal modulations than in the ripple experiments (8-40 Hz). The modulation frequency clearly affected the monkeys’ behavioral responses for the AM noises (Fig. 5.3C, cyan dots), and in a similar fashion for the purely amplitude-modulated ripples at zero density (Fig. 5.3C, red dots). The monkeys reacted slowest for the lowest modulation frequency at 2 Hz (median $\pm$ SE: $483 \pm 4$ ms), and fastest for the highest modulation frequency at 256 Hz (median $\pm$ SE: $315 \pm 4$ ms). This finding seems at odds with reports suggesting that lower modulation frequencies are easier to perceive and detect than high modulation frequencies (e.g. Chi et al., 1999). However, a possible explanation for this reaction time effect could be that it simply takes more time for slow acoustic amplitude modulations to reach a perceptual detection threshold than for faster amplitude modulations (e.g. Heil and Neubauer, 2003), even though in general the detection threshold is lower for the lowest modulation frequencies. In line with this idea, we could fit the following simple linear regression model (with respect to the modulation period) to the AM reaction times:

$$RT_{AM500} = \frac{130}{w} + 359 \text{ ms} \quad \text{(Eq. 5.2)}$$

with $w$ the modulation frequency (in Hz) and RT the reaction time (in ms). This result ($F_{df=7241}=1632, p < 0.0001$) indicates that the perceptual detection threshold for AM stimuli (presented at 100% modulation depth) is reached at about 13% of the period of the modulation, and that an additional 359 ms are needed to prepare the manual motor response (Fig. 5.3C; cyan line).

In contrast, for the predictable modulation onsets (A1000 trials; Fig. 5.3B, C), reaction times were virtually independent of the acoustic modulation parameters (ANOVA, ripple spectral modulations: $F_{df=5} = 3.4, p=0.1$; ripple temporal modulations: $F_{df=4}=1, p=0.41$; AM: $F_{df=14} = 1.5, p=0.1$), which becomes especially evident when applying the linear model of Eq. 5.2:

$$RT_{AM1000} = \frac{0}{w} + 162 \text{ ms} \quad \text{(Eq. 5.3)}$$
The optimal fit to the data ($F_{df=8282}=11.1$, $p=0.0008$) indicates that the manual responses did not depend on the modulation frequency ($w$), and required only 162 ms of preparation (Fig. 5.3C; green line). Again, the data obtained for the amplitude-modulated (zero density) ripples nicely coincide with the AM responses (Fig. 5.3C, blue triangles). The predictive aspect of the A1000 task was therefore clearly reflected in a considerable reduction of manual reaction times of about 220 ms for the ripples (Figs. 5.3A,B) and about 200 ms for the AM noises (Fig. 5.3C, see also above).

![Figure 5.3](image)

**Figure 5.3.** Median reaction times (color-coded) for each of the 55 dynamic ripples during A500 (A) and A1000 (B) trials. Note the clear pattern of faster and slower reaction times for A500 trials as function of temporal (velocity) and spectral (density) modulations, which is absent for the A1000 trials. (C) Median reaction times of each AM noise during A500 (cyan) and A1000 (green) trials. Cyan curve: Eq. 5.1; green line: Eq. 5.2. Data points for zero-density ripples in A500 (red) and A1000 (dark blue) are superimposed.

Although the reaction-time distributions for A500 and A1000 trials differed substantially, they nonetheless overlapped, which means that a sizeable fraction of the A1000 responses could be considered as relatively slow (e.g. Fig. 5.2, grey histograms). To test whether these slower A1000 responses depended on spectral-temporal modulations, like the A500 responses, we divided the behavioral responses of all recorded A1000 trials into fast (<225 ms) and slow (>225 ms) reaction times, and analyzed them as function of AMF (for AM noises), and of ripple density. Through this definition, 40% of the reaction times to ripples, and 33% to AM noises were categorized as slow responses for A1000 trials. Figures 5.4A and 5.4B show the reaction times (relative to the mean A1000 reaction time) for slow (blue), and fast (red) A1000 trials, and for all A500 trials (grey). Interestingly, the slower A1000 responses were controlled by modulation-onset prediction (like fast A1000 responses), as well as by the acoustic modulations (like A500 responses).

Taken together, these findings demonstrate that the animals used all available information in the trial to obtain a reward: for A500 trials responses were exclusively guided by acoustic information. In A1000 trials, however, both acoustic (when the animal had a long reaction time) and non-acoustic (prediction, when the reaction was fast) signals determined the animal’s responses.
5.3.3 TOP-DOWN NEURAL SIGNALS

The neural responses in active trials differed substantially from the passive-evoked acoustic responses to the same sounds (e.g., Fig. 5.2; Massoudi et al., 2013). This suggests the presence of an additional task-related signal, which interacts with the acoustically evoked response. The behavioral results shown in Figures 5.3 and 5.4 indicate that the manual reactions were guided by at least two factors: acoustic spectral-temporal modulations in the unpredictable A500 trials as well as in the late A1000 trials, and a non-acoustic predictive signal in all A1000 trials (both early and late). Thus, to identify the top-down neural signals in the active trials, different possibilities should be considered.

First, if the top-down signal is a general attentive signal (e.g., reflecting nonspecific task engagement, or vigilance), it could induce an overall change in the firing rate of the active trials, which would relate to neither the acoustic events in the trial, nor to the motor events of the monkey’s reaction. Second, if the top-down signal would be purely acoustic, the change in activity should be locked to the stimulus events (e.g., as a gain-amplitude modulation). Thirdly, if the top-down modulation would represent a purely behavioral signal (e.g. the preparation of the manual release response) it should be locked to the manual reaction time and be independent of the sound type. Note that in principle the top-down signal could reflect contributions from all these different factors, in which case there will be no single variable that fully explains the extracted neural modulation. To investigate these possibilities, we aligned each trial’s response to either the sound-modulation onset, or to the manual reaction time.

The results of this analysis for one of the neurons (T103) are shown in Figure 5.5. We determined the average spike-density functions across trials (Methods) for passive (black), unpredictable A500 (red) and predictable A1000 (blue) trial types, and for ripples (top, A-B) and AM...
noises (bottom, C-D). In the left-hand column the data were aligned to sound onset. Although the average passive firing rate remained almost flat for the two sound types (ripples and AM noises), the A500 and A1000 neural responses changed considerably around the dynamic sound onset (indicated by the red and blue vertical lines, respectively), with roughly similar firing patterns for both sound types. For A500 trials the change in the activity started a few hundred milliseconds after dynamic sound onset, while for the A1000 trials the change started even before the sound-modulation onset. As the activity patterns do not align for the A500 and A1000 responses (neither when aligned to sound onset [Figs. 5.5A, C] or to modulation onset [not shown]), the top-down signals were clearly not primarily driven by the sound acoustics. Interestingly, the neural responses appeared to follow roughly similar patterns as the underlying reaction time distributions for the two trial types (filled histograms at the bottom of panels A and C). This suggests that the neural responses may be better related to the animal’s reaction times.

Figures 5.5B and 5.5D show the average active neural responses when trials were aligned to the reaction time. The firing rates show a steady increase that start about 150 ms before the reaction time, with patterns that are highly similar for both unpredictable (A500) and predictable (A1000) trial types. This indicates that the top-down signal could be attributed to a motor preparation and/or a reward expectation signal (Figs. 5.4B and D). The peak at about 20 ms after bar-release has a similar shape and onset latency as the sound-onset peak, and can be attributed to the soft sound produced by the rapid bar-release.
Figure 5.5. Spike-density analysis for example cell T103. (A, B) Responses to ripples. (C, D) Responses to AM noises. (A, C) Spike-density functions for all passive (black), A500 (red) and A1000 (blue) trials aligned to sound onset. Mean and SE are denoted by solid lines and transparent patches, respectively. Filled histograms represent reaction-time probability (%) distributions. Note the resemblance between the reaction times and the active response modulations. The red and blue vertical lines at 500 ms and 1000 ms denote modulation onset for passive and A500 trials and for A1000 trials, respectively. (B, D) Spike-density functions aligned to bar release. The curves fully superimpose. The vertical line at 0 ms denotes handle-bar release.

Figure 5.6 shows the results of this analysis for the population of 100 cells. The observations for neuron T103 persist for the total population: the neural responses in the active trials were much better aligned to the monkey’s reaction time than to the acoustic modulations, and demonstrate that the top-down signals reflect a dynamic, non-acoustic signal at the level of single AC cells. The A1000 trials with early-responses (~30% of the A1000 trials; Table 5.1) had a similar top-down signal (data not shown; there were too few early-response A500 trials to reliably obtain instantaneous firing rate).
In the analysis so far we averaged neural responses across the different stimuli in the experiment. Although the top-down signals were not locked to the acoustic modulations (Figs. 5.6A,C), it is conceivable that the signals themselves could somehow depend on the spectral and temporal modulation rates. For example, since the behavioral data (Figs. 5.3 and 5.4) clearly indicated that the reaction times depended quite reliably on the stimulus acoustics and behavioral context (predictive and non-predictive trials, or fast and slow responses), the amplitude of the top-down signal could depend on the reaction time itself. Alternatively, because the top-down signal rides on top of the sound-evoked activity, it could be stronger for preferred stimuli for each cell.

To investigate the influence of acoustics on the top-down signal, we selected the trials along the spectral and temporal stimulus modulation dimensions, and behavioral contexts (predictive/non-predictive), and calculated the sound-specific top-down signals for the neural population. Figures 5.7A and 5.7B separate the ripple trials for all cells and reaction times into the different ripple densities (pooled across the five ripple velocities, while pooling $\Omega>0$ with $\Omega<0$; Fig. 5.7A) and different ripple velocities (pooled across the eleven ripple densities, Fig. 5.7B). Figure 5.7C shows the
data for the fifteen different AMFs of the AM stimuli. The left-hand panels of Figure 5.7 correspond to the A500 condition, the right-hand panels to the A1000 condition. The individual curves are remarkably similar within each modulation range (ANOVA over the epoch [-200 to 0] ms: ripple: spectral modulations: A500: $F_{df=10} = 0.31$, $p=0.98$; A1000: $F_{df=10} = 0.16$, $p=1$; ripple temporal modulations: A500: $F_{df=4} = 0.06$, $p=0.99$; A1000: $F_{df=4} = 0.01$, $p=1$; AM stimuli: A500: $F_{df=14} = 0.2$, $p=1$; A1000: $F_{df=14} = 0.13$, $p=1$), and between different modulation types. Thus, spectral-temporal modulation rates, or predictive and non-predictive modulation onsets, had no influence on the size and shape of the top-down neural modulation.
Figure 5.7. Population averaged spike-density functions separated for the different spectral-temporal sound modulations. Responses aligned to reaction time, and averaged across all cells. (A) Responses for ripples at different absolute densities (pooled for 5 velocities). (B) Responses for ripple velocities (pooled across 11 densities. (C) Responses for 15 different AMFs of AM GWN stimuli. Left-hand column: A500 data; right-hand column: A1000 data. Note similarity of individual response curves across stimulus properties and hearing conditions.
5.3.4 MODEL SELECTION

We here consider two potential models for the neural interaction of acoustic and behavioral signals on single AC neurons: linear summation, and nonlinear gain modulation. For the summation model, the neural response for a trial in the active reaction-time task can be described by:

\[ R_{\text{ADD}}(S; t) = A_{\text{ACT}}(S; t) + B(t-t_R) \quad \text{(Eq. 5.4)} \]

with \( A_{\text{ACT}}(S; t) \) the acoustic-evoked response component to sound S in the active listening condition, and \( B(t-t_R) \) the behavioral top-down signal, which is referenced to the reaction time, \( t_R \) (according to Figs 5.6B,D), and taken independent of the acoustics of sound S (demonstrated in Fig. 5.7). In contrast, the multiplicative gain-modulation model for that trial reads:

\[ R_{\text{MUL}}(S; t) = (1 + B(t-t_R)) \cdot A_{\text{ACT}}(S; t) \quad \text{(Eq. 5.5)} \]

We employ the property that the acoustic component of the response during active listening is the same as for passive listening, as recently reported by Massoudi et al. (2013). In other words, \( A_{\text{ACT}}(S; t) = A_{\text{PASS}}(S; t) \). Further, in the absence of a behavioral signal \((B(t-t_R)=0)\) both models reduce to the passive-evoked response. To dissociate the two models, we use the observation that the mean firing rates of AC cells for the static and dynamic components of the stimuli during passive listening varied considerably from cell to cell (ranging from only a few spikes/s to over 100 spikes/s), but, when averaged across sound modulations, remained roughly constant during the trial (see Figs. 5.5 and 5.6, black lines). As a measure for this passive sound-evoked activity we calculated, for each cell, the mean firing rate obtained during the passive static-noise epoch, separately for ripples and for AM noises: \( \mu_{\text{PASS}} = \text{mean}(A_{\text{PASS}}(S; t)) \). To avoid a biased contribution from the transient sound-onset peaks we took \( 200<t<500 \text{ ms} \). We then subtracted this value from the mean active response during the acoustic modulation epoch for each active trial, and aligned the resulting modulatory activity to the animal’s reaction time for that trial. We subsequently determined the spike-density function, now pooled across all ripples and AM noises, to obtain the mean dynamic change in activity for each neuron resulting from the top-down signals:

\[ \Delta R_{\text{MOD}}(t-t_R) = R(t-t_R) - \mu_{\text{PASS}} \quad \text{(Eq. 5.6)} \]

Applying Eq. 5.6 to the two interaction schemes (Eqs. 5.4 and 5.5) predicts for the additive model that:

\[ \Delta R_{\text{ADD}}(t-t_R) = B(t-t_R) \quad \text{(Eq. 5.7)} \]

and for the multiplicative model:

\[ \Delta R_{\text{MUL}}(t-t_R) = \mu_{\text{PASS}} \cdot B(t-t_R) \quad \text{(Eq. 5.8)} \]

In other words, when plotting the modulatory firing rate versus the mean passive firing rate, \( \mu_{\text{PASS}} \), if would yield a flat line if the additive model is valid, with an offset at the mean top-down modulation-firing rate. In contrast, the multiplicative model predicts a linear relation with a slope that is determined by the average top-down modulation strength. A combination of additive and multiplicative signals on the neurons would give an additional offset to Eq. 5.8. Figure 5.8A shows
the results of this analysis for the population of 100 cells (pooled for the two monkeys, and across ripples and AM noises). Data points were colored according to their mean passive firing rate (see color bar). The correlation coefficient of \( r=0.62 \) is highly significant \( (F_{df=98}=61.3, \ p<0.001) \). The linear regression line through the data points has a slope of 0.61 \( (t_{df=98}=7.83, \ p<0.001) \), and an offset that is indistinguishable from zero \( (\text{offset} = -1.4 \text{ spikes/s}, \ t_{df=98} = -0.44, \ p=0.65) \). Thus, the data do not support the presence of an additive component, and clearly favor the prediction of the multiplicative model (Eqs. 5.5 and 5.8).

In Figure 5.8B we show the instantaneous modulatory firing rates (Eq. 5.6) aligned to bar-release, after sorting the cells for their mean passive firing rates (color bar). The figure illustrates that the higher the cells’ mean firing rate, the stronger the change in modulatory activity across the measured epoch (here 150 ms), as predicted by Eq. 5.8.

Figure 5.8. Evidence for multiplicative signal interaction at AC cells. (A) Change in mean modulatory firing rate as function of the mean passive firing rate in the static epoch (based on Eq. 5.6). For each cell \( (n = 100) \), data were pooled for ripples and AM noises, and for A500 and A1000 trials. Note linear relation with zero offset \( (r = 0.62) \). (B) The modulatory signal aligned to bar release. For higher passive static firing rates, the total modulatory signal is stronger too. Different colors denote different mean passive firing rates (color bar).

5.4 DISCUSSION

5.4.1 SUMMARY

This study is the first to compare task-related top-down signals in monkey AC for different acoustic stimuli during the same behavioral demands. Our results revealed clear top-down components during active hearing that could be extracted from the animal’s behavior (Figs. 5.3 and 5.4), which contained acoustic (spectral-temporal) and non-acoustic (cognitive) factors. We also identified a systematic top-down signal at the level of single-unit activity that was locked to the animal’s behavioral response but invariant to the acoustic modulations (Figs. 5.5-5.7). We
demonstrated that the neural top-down component is processed independently from the acoustic signals by AC neurons in a multiplicative, separable, way (gain modulation).

5.4.2 BEHAVIORAL RESPONSE PROPERTIES

The behavioral results demonstrated that the monkeys’ manual reaction times depended on the spectral-temporal sound characteristics in unpredictable trials (Figs. 5.3A and C) and, interestingly, also in the slower predictable trials (Fig. 5.4). In all A1000 trials the reaction times were substantially shortened by prediction of the upcoming sound change (Figs. 5.2 and 5.3; Jaramillo and Zador, 2011; Massoudi et al., 2013). These results suggest that the animals adopted a single strategy for both trial types, which led them to use all available information to earn a reward as fast as possible. In A500 trials the monkey only had acoustic information to achieve this goal, since all sounds and trial types were randomly selected, and equally likely and important. The results for the ripples and AM noises reflected this acoustic dependence quite clearly. Reaction times to unpredictable AM noises and zero-density ripples were only determined by the AM frequencies, and could be fitted well with the simple cycle-detection regression model of Eq. 5.1 (Fig. 5.3C). The manual responses to the unpredictable moving ripples (density ≠ 0) relied on both spectral and temporal modulations, showing a significant interaction between these acoustic features (Fig. 5.3A).

In the A1000 trials, however, the monkey acquired target-onset predictability as an additional source of information. This non-acoustic cognitive top-down signal appeared to override the dependence on purely acoustic information: reaction times not only shortened by about 200 ms on average, their reliance on acoustic modulation rates had also virtually disappeared (Figs. 5.3B and C). However, when the animal withheld its immediate response in A1000 trials, the longer reaction times (>225 ms) depended on the acoustic modulations as well. These results demonstrate a clear qualitative difference between the predictable and unpredictable trial types, which was reflected in the single-unit responses of AC neurons (Figs. 5.2, 5.5A,C and 5.6A,C), but not in the strength of the top-down signals (Figs. 5.5B,D, 5.6B,D and 5.7). In other words, the neural response does not reveal an additional prediction signal.

5.4.3 TASK-RELATED EFFECTS ON AC RESPONSES

Executing a sound-detection task considerably changed the neural firing patterns for all AC cells, when compared to the passive condition (Fig. 5.2). For both sound types, active engagement in the task led to an overall increase of the spontaneous (background) firing rate prior to trial initiation (Fig. 5.2), and during the acoustic evoked firing rates, when compared to passive hearing (Figs. 5.5A,C and 5.6A,C; see also Massoudi et al., 2013). However, the pure stimulus-evoked mean activity was hardly affected, as is visible in the responses during the static noise (Figs. 5.5A, C and 5.6A, C). The overall increase in mean firing rate during the dynamic epochs is therefore attributable to the
behavioral top-down modulation. Most studies have reported that attentional demands, including non-specific effects due to arousal and vigilance, increase cortical firing rates (Edeline et al., 2001; Steriade et al., 2001; Scott et al., 2007; Niwa et al., 2012a, 2012b), whereas some studies reported decreased AC responses during task performance in rats (Otazu et al., 2009), and ferrets (Atiani et al., 2009), and overall inhibition during behavior in somatosensory cortex of mice (Crochet and Petersen, 2006; Poulet and Petersen, 2008). In contrast to these reports, our data did not reveal strong non-specific effects on the firing rates of AC neurons, which was further supported by the zero offset in the regression analysis of Figure 5.8A.

Note that in our paradigms the animals did not, and could not, attend to a particular pre-cued acoustic feature of the upcoming target sound, as they were not instructed to do so, and were merely anticipating any detectable spectral-temporal modulation while listening to the static noise. Animals could not predict any particular acoustic modulation in advance either, as stimuli were all equally likely and important for obtaining a reward. The only trial-specific feature that mattered in the experiment was the duration of the static interval. Our data clearly demonstrate that the animals indeed used this information, and the neural activity patterns reflected this non-acoustic aspect of the task.

Realigning the active responses of each trial to its associated reaction time revealed a signal starting about 150 ms prior the bar release. This signal was clearly present at the single cell and population level (Figs. 5.5B, D and 5.6B, D), and found to be invariant to the acoustic modulations (Figs. 5.5-5.7). The signal may be attributed to either motor preparation for bar release, or to reward expectation. These two interpretations could not be dissociated as in our experiments reward delivery fell immediately after bar release. Previous studies have also shown the presence of motor-preparation (Brosch et al., 2005, 2011a; Selezneva et al., 2006; Yin et al., 2008; Niwa et al., 2012a), and reward expectancy (Yin et al., 2008; Brosch et al., 2011a) signals at the level of AC neurons, although the relative proportions of cells carrying such a signal differs between studies from less than 15% (e.g., Yin et al., 2008) to more than 60% (e.g., Brosch et al., 2005). A possible explanation for this apparent discrepancy could be related to the results on the multiplicative effects in Figure 5.8. Although many recorded AC cells carry the task-related signal (47% of cells; Wilcoxon signed rank test on the change in modulatory firing rate), its strength, and hence its detectability, is directly proportional to the mean acoustic-evoked firing rate. Variation in these firing rates between studies could thus underlie the different proportions of cells seen to carry multiplicative non-acoustic modulations. These task-related signals could originate from different cortical areas such as parietal and frontal cortical areas (reward) (Romanski et al., 1999; Fritz et al., 2010), or from pre-motor cortex (motor preparation) (Lemus et al., 2009).

Although for the fast A1000 trials (Fig. 5.4, red) the animals responded mainly on the basis of sound-change prediction, we observed no additional neural prediction signal in AC cells, as the A500 and A1000 curves did not systematically differ. This suggests that the prediction signal, possibly
originating from prefrontal cortex (Nobre et al., 1999), induces a temporal shift (by about 200 ms) of the top-down modulation on AC cells.

---

**5.4.4 ADDITIVE AND MULTIPLICATIVE INTERACTIONS**

Multiplicative interactions between sensory signals, top-down signals, and motor-related signals have been described for a variety of brain areas and sensory-motor systems. For example, reports have described eye-position gain fields for visually responsive cells in monkey posterior parietal cortex (Andersen et al., 1985), and for eye-movement related visual-motor cells in the monkey midbrain superior colliculus (Van Opstal et al., 1995). Similarly, in the auditory system eye-position gain field tuning, as well as multiplicative spatial sensitivity, has been reported for auditory sensitive cells in the monkey inferior colliculus (Groh et al., 2001; Zwiers et al., 2004). It has also been shown that attention may influence the firing rate of visual (McAdams and Maunsell, 1999; Treue and Martínez Trujillo, 1999; Reynolds et al., 2000; Maunsell and Treue, 2006), and somatosensory cortex cells (Sripati and Johnson, 2006) in a multiplicative fashion. This indicates that attention can change the gain of neuronal responses without changing their underlying response properties, or tuning curve. Different gain mechanisms could be invoked by attention, such as enhancing the attended stimulus strength, changing the neuronal gain of the behaviorally salient target, or sharpening the sensory selectivity of the neuron. Polack et al. (2013) (Polack et al., 2013) recently identified a noradrenergic mechanism underlying gain modulations in mice V1 cells, which led to a stronger, but less variable, change in the cells’ membrane potential during a locomotion task.

Other studies have proposed additive interactions between sensory and non-sensory signals at the level of AC neurons. For example, Maier and Groh (2010) (Maier and Groh, 2010) suggested an additive interaction of eye-position signals with auditory-evoked activity in behaving monkey AC, and Schnupp et al. (2001) (Jenison et al., 2001) demonstrated that spatial selectivity of AC neurons in anesthetized ferret is well predicted by a linear integration model of sound levels within each frequency band and ear. Additive interactions have also been proposed for attentional top-down signals in the visual cortex (Luck et al., 1997; Williford and Maunsell, 2006; Buracas and Boynton, 2007; Thiele et al., 2009).

The analysis in Figure 5.8 demonstrates that it is possible to dissociate additive and multiplicative interactions on AC cells by exploiting the cell-to-cell variability of mean acoustic responses, together with the idea that the sound-evoked responses are unaffected by task performance. We recently verified this latter requirement for the spectral-temporal receptive fields of AC neurons (Massoudi et al., 2013). Indeed, a relatively slow modulation on top of a rapid acoustic tuning response will not affect the spectrum of the latter. This requirement on the top-down signals is typically fulfilled: the monotonic rising phase of the top-down signal is confined to the low-frequency
spectrum (from dc to a few Hz; Fig. 5.8B), when compared to the rapid phase-locked response profiles of AC cells to acoustic amplitude modulations (up to 256 Hz in our sample of cells).

Our population analysis revealed that the non-acoustic top-down modulation and the sound-evoked activity combine in a separable way through multiplicative gain modulation (Eqs. 5.5 and 5.8; Fig. 5.8). Such multiplexing of different signal types is beneficial for an efficient and reliable transmission of multiple sources of information, which express different types of variables, through the same channel. A relatively simple read-out mechanism, based on linear weighting of the cell contributions in the recruited population, could subsequently decode the individual signals for further processing. Quantitative models for such read-out mechanisms have been described for eye-position, sound level and sound-location gain fields of cells in the auditory midbrain inferior colliculus (Zwiers et al., 2004), and for eye-position gain fields in the visuomotor midbrain superior colliculus (van Opstal and Hepp, 1995).
SUMMARIES
The aim of this thesis was to better understand the response properties and function of the auditory cortex (AC). To that end we performed single-unit recordings from cells in the core auditory cortex of two macaque monkeys that were trained to respond as fast as possible to a perceived modulation of complex sounds. We addressed the basic spectral-temporal tuning characteristics of AC neurons (chapter 2), the effect of the behavioral state of the animal on spectral-temporal (chapter 3) and temporal envelope (chapter 4) processing, and the interaction between acoustic information and top-down task-related signals at the single unit level of core AC (chapter 5). Our findings revealed that although AC shares many spectral-temporal properties with the auditory midbrain inferior colliculus (IC), its response characteristics appeared to be more nonlinear than in the IC. As a result we obtained a poor correspondence between the best frequencies derived from the spectrotemporal receptive fields (STRF) of AC neurons, and those extracted from pure-tone responses. In addition, the linear prediction of AC responses to sounds on the basis of the STRF was poor (chapter 2). We next showed that despite prominent differences in the evoked neural activity, the extracted STRFs remained remarkably stable during three different behavioral conditions: (i) passive exposure to sounds (the classical STRF recording, as in chapter 2), vs. active engagement in an (ii) unpredictable and (iii) highly predictable sound change-detection task (chapter 3). By applying a broader range of temporal modulation frequencies in the same task paradigm, we showed that the different behavioral conditions affected the temporal processing capabilities of AC cells for very low (below 8 Hz) and very high (above 64 Hz) modulation frequencies. However, the tuning strength remained virtually identical for the considerable range between about 10-50 Hz (chapter 4). Finally, we studied the nature of the neural interactions at the level of AC cells to acoustic and non-acoustic task-related signals. Our findings suggested a non-linear multiplicative interaction between the cell’s tuning characteristic and the top-down behavioural modulation, which further corroborates the results of chapter 2 on the AC cells’ non-linear characteristics (chapter 5).

---

CHAPTER 2: SPECTROTEMPORAL RESPONSE PROPERTIES OF AUDITORY CORTEX NEURONS IN AWAKE MONKEY

In this chapter, we studied the basic spectrotemporal characteristics of AC cells such as their best temporal and spectral modulation frequencies, separability of spectral and temporal responses, and their linearity. So far, most studies of the core auditory cortex have characterized the spectral and temporal tuning properties of cells in non-aware, anesthetized preparations. As experiments in awake animals are scarce, we here used dynamic spectral-temporal broadband ripples to study the properties of the spectrotemporal receptive fields (STRFs) of AC cells in awake monkeys. We show that AC
neurons were typically most sensitive to low ripple densities and velocities, and that most cells were not selective for a particular ripple direction. A substantial proportion of neurons preferred amplitude-modulated sounds (at zero ripple density) to dynamic ripples (at non-zero densities). The vast majority (>93%) of amplitude modulation transfer functions were separable with respect to spectral and temporal modulations, indicating that time and spectrum are independently processed in AC neurons. We assessed linearity vs. nonlinearity of AC responses by comparing a neuron’s best frequency for tones with its best frequency within the STRF. We also compared neural responses to a set of natural vocalizations to the linear STRF-based predictions of these responses. Both measures suggest a prominent nonlinearity at the level of AC cells. We discuss our findings in the light of results obtained from the monkey midbrain inferior colliculus.

CHAPTER 3: STABLE BOTTOM-UP PROCESSING DURING DYNAMIC TOP-DOWN MODULATIONS IN MONKEY AUDITORY CORTEX

This chapter tries to address whether task-related top-down processing in auditory cortex interferes with its spectrotemporal analysis of sound. Recent studies indicated non-acoustic modulations of AC responses, and that attention changes a neuron’s spectrotemporal tuning. As a result, AC would seem ill suited to represent a stable acoustic environment, which is deemed crucial for auditory perception. To assess whether top-down signals influence acoustic tuning in tasks without directed attention, we compared monkey single-unit AC responses to dynamic spectrotemporal sounds under different behavioral conditions. Recordings were mostly made from neurons located in primary fields (A1 and R) that were well tuned to pure tones, with short onset latencies. We demonstrate that responses in AC were substantially modulated during an auditory detection task and that these modulations were systematically related to top-down processes. Importantly, despite these significant modulations, the spectrotemporal receptive fields of all neurons remained remarkably stable. Our results suggest multiplexed encoding of bottom-up acoustic and top-down task-related signals at single AC neurons. This mechanism preserves a stable representation of the acoustic environment despite strong non-acoustic modulations.

CHAPTER 4: BEHAVIORAL CONTEXT AFFECTS TEMPORAL PROCESSING IN AWAKE MONKEY AUDITORY CORTEX

This chapter is devoted to temporal envelope processing of complex sounds, which is a major function of auditory cortex, and crucial for speech (c.q. vocalization) intelligibility. Although neural sensitivity to amplitude-modulated noises (AM) has been studied in AC, and some studies have reported on its role in AM discrimination tasks, little is known about how different behavioral states
of the listener would influence AM sensitivity. Here, we analyzed the spontaneous activity, the sound-onset response, and the sustained response of monkey AC cells, when animals could be in different behavioral states, varying from passive sound exposure, to engagement in a non-predictive and predictive sound-change detection task. We analyzed the monkeys’ reaction times as function of the AM modulations and predictability in the task, and found a systematic relationship between reaction time and acoustic sensitivity. We determined modulation transfer functions (MTFs) to quantify the behavioral and single-unit neuronal responses for 15 amplitude modulation frequencies (AMF) between 2 and 256 Hz. From the Fourier spectrum of the period histograms of evoked spikes, we extracted the magnitude of the first harmonic for the different AM frequencies (mMTF). We found that mMTF was task-independent for AMFs between 5.6 to 45 Hz, but changed between passive and active listening conditions for both the lower and higher modulation frequencies. Task involvement also altered the strength of envelope phase locking of cells, their mean firing rate, and trial-to-trial variability. Our findings suggested that the mMTF could be a better measure to quantify temporal envelope encoding of auditory cortex neurons, as it depends on both the firing rate and vector strength.

CHAPTER 5: TASK-RELATED PREPARATORY MODULATIONS
MULTIPLY WITH ACOUSTIC PROCESSING IN MONKEY AUDITORY CORTEX

This chapter concerns quantifying the task-related top-down signals and probing the mechanism of interaction between them and acoustic information. Therefore, we characterized task-related top-down signals in monkey auditory cortex cells, by comparing single-unit activity during passive sound exposure to neuronal activity during a predictable and unpredictable reaction-time task for a variety of spectral-temporal modulated broadband sounds. Although animals were not trained to attend to particular spectral or temporal sound modulations, their reaction times demonstrated clear acoustic spectral-temporal sensitivity for unpredictable modulation onsets. Interestingly, this sensitivity was absent for predictable trials with fast manual responses, but re-emerged for the slower reactions in these trials. Our analysis of neural activity patterns revealed a task-related dynamic modulation of AC neurons that was locked to the animal’s reaction time, but invariant to the spectral and temporal acoustic modulations. This finding suggests dissociation between acoustic and behavioral signals at the single-unit level. We further demonstrate that single-unit activity during task execution can be described by a multiplicative gain modulation of acoustic-evoked activity and a task-related top-down signal, rather than by linear summation of these signals.
NEDERLANDSE SAMENVATTING

Dit proefschrift stelde zich ten doel om beter inzicht te verkrijgen in de respons eigenschappen en functionele betrokkenheid van de auditieve cortex (AC). Om dat doel te bereiken zijn neurale afleidingen gedaan aan geïsoleerde neuronen in de centrale auditieve cortex van twee resus makaken die waren getraind om zo snel mogelijk te reageren op een waargenomen modulatie in complexe geluiden. We bepaalden de basale spectrale en temporele gevoeligheden van AC neuronen (Hoofdstuk 2), het effect van de gedragstoestand van het dier op deze spectraltemporele gevoeligheden (Hoofdstuk 3) en op de verwerking van temporele envelop informatie (Hoofdstuk 4), en op de interactie tussen akoestische informatie en taak-gerelateerde signalen op cel-niveau in de centrale AC (Hoofdstuk 5). Onze resultaten laten zien dat de centrale AC veel gemeen heeft met de spectraltemporele gevoeligheid van cellen in de auditieve middenhersenen (inferior colliculus; IC). Echter, de respons karakteristiek van AC neuronen bleek aanzienlijk sterker niet-lineair te zijn dan eerder gerapporteerd voor de IC. Het gevolg hiervan was een slechte overeenkomst tussen de beste frequenties die konden worden afgeleid uit de spectraltemporele receptieve velden (STRF) van AC neuronen, en die van responsies op pure tonen. Voor een lineaire verwerking zouden deze beste frequenties nagenoeg gelijk dienen te zijn. Daarnaast was ook de voorspelling van de lineaire neurale respons met een lineair model voor geluiden op basis van de STRF veel slechter dan is gevonden voor de IC (Hoofdstuk 2). Vervolgens toonden we aan dat ondanks de grote verschillen in neurale responsies, de afgeleide STRFs opmerkelijk stabiel bleken te zijn tijdens de drie gedrags condities in onze experimenten: (i) passieve stimulatie met breedbandige spectraltemporele ripple geluiden (het klassieke STRF paradigma, zoals beschreven in Hoofdstuk 2), vs. actief gedrag t.a.v. deze geluiden binnen een (ii) onvoorspelbare, en (iii) volledig voorspelbare geluidsverandering in de taak (Hoofdstuk 3). Door vervolgens de responsies te meten over een veel breder scala van temporele modulatiefrequenties (2 – 256 Hz) tijdens dezelfde drie taken, konden we laten zien dat de verschillende gedragscondities de temporele gevoeligheid van AC neuronen konden beïnvloeden voor zeer lage (onder de 8 Hz) en zeer hoge (boven de 64 Hz) modulatiefrequenties. Echter, de temporele gevoeligheid was onaangetast voor een aanzienlijk frekwentiebereik tussen 10-50 Hz (Hoofdstuk 4). Tenslotte bestudeerden we de wijze waarop verschillende typen signalen (akoestische en gedragsgerelateerde) met elkaar interageerden op het cel niveau in de AC. Onze resultaten en analyses suggereren een niet-lineaire, multiplicatieve interactie tussen de akoestische respons van AC cellen en de top-down gedragsmodulatie. Dit resultaat ondersteunt de bevindingen uit Hoofdstuk 2 m.b.t. de niet-lineaire karakteristieken van AC cellen (Hoofdstuk 5).
In dit hoofdstuk bestudeerden we de spectrale-temporele gevoeligheid van AC cellen, zoals hun beste temporele en spectrale modulatiefrequenties, de scheidbaarheid (separabiliteit) van temporele en spectrale responsies, en hun lineariteit. Verreweg de meeste studies in de centrale AC tot nu toe hebben de spectrale en temporele gevoeligheden van cellen beschreven voor niet-wakkere, geanesthezeerde preparaties. Omdat experimenten in wakkere dieren tot op heden zeldzaam zijn, hebben we de eigenschappen van spectraal-temporele receptieve velden (STRFs) voor AC cellen in wakkere makaken in kaart gebracht, door gebruik te maken van breedbandige, dynamische spectraal-temporele ripple geluiden. De resultaten lieten zien dat AC cellen typisch het gevoeligst waren voor lage ripple dichtheden (spectraal) en ripple snelheden (temporeel), en dat de meeste cellen niet speciek gevoelig waren voor de richting van de ripple modulaties. Een aanzienlijk deel van de neuronen prefereerde amplitude-gemoduleerde geluiden (ripple dichtheid nul) boven dynamische ripples (ripple dichtheid ongelijk nul). Meer dan 93% van de amplitude overdrachtskarakteristieken van de ripple responsies waren separabel t.a.v. spectrale en temporele modulaties. Dit suggereert dat tijd en spectrum onafhankelijk worden verwerkt op het niveau van AC neuronen. We bepaalden de mate van lineariteit vs. niet-lineariteit in de AC responsies door de beste frequenties van de STRFs en die voor pure tonen met elkaar te vergelijken. We vergeleken ook de neurale responsies op natuurlijke vocalisaties met de lineaire voorspellingen op grond van de STRFs. Beide vergelijkingen suggereren een hoge mate van niet-lineair gedrag van AC cellen. Onze bevindingen worden bediscussieerd aan de hand van eerder verkregen resultaten uit de auditieve middenhersen van makaken (inferior colliculus) onder dezelfde akoestische omstandigheden.

In dit hoofdstuk bestudeerden we in hoeverre de top-down verwerking van taak-gerelateerde signalen interferereerde met de spectrale-temporele analyse van geluid in de auditieve cortex (AC). Recente studies hebben laten zien dat er niet-akoestische modulaties plaatsvinden op het niveau van AC cellen, en suggereerden dat attentie kan leiden tot veranderingen in de spectrale-temporele gevoeligheid van AC neuronen. Dit zou kunnen betekenen dat de AC niet goed is uitgerust om een stabiele akoestische omgeving te representeren, hetgeen cruciaal is voor de auditieve perceptie. Om
vast te stellen of top-down signalen de akoestische gevoeligheid van AC neuronen beïnvloeden zonder specifieke gerichte attentie, vergeleken we de responsies van AC cellen in getrainde makaken voor dynamische spectraal-temporele geluiden onder verschillende gedragscondities. De afleidingen werden verricht aan neuronen in de primaire auditieve cortexvelden (A1 en R), waar cellen goed reageren op pure tonen met korte latenties. We laten zien dat de AC responsies aanzienlijk worden gemoduleerd tijdens een auditieve detectietoekomst, en dat deze modulaties systematisch samenhangen met een top-down signaal. Opmerkelijk was dat, ondanks deze sterke modulaties, de STRFs van de neuronen stabiel bleven. Deze resultaten suggereren een gemultiplekte codering van bottom-up (akoestische) en top-down (taak) signalen op het niveau van individuele AC cellen. Een dergelijk mechanisme behoudt een stabiele akoestische representatie van de omgeving, ondanks sterke niet-akoestische modulaties.

---

**HOOFDSTUK 4: DE GEDRAGSMATIGE CONTEXT BEÎNVLOEDT TEMPORELE VERWERKING IN DE AUDITIEVE CORTEX VAN MAKAKEN**

----

Dit hoofdstuk bestudeert de temporele envelop verwerking van complexe geluiden, wat een belangrijke functie is voor de auditieve cortex, en cruciaal voor het verstaan van spraak (en vocalisaties in het algemeen). Hoewel de neurale gevoeligheid voor amplitude-gemoduleerde ruisstimuli (AM) al eerder bestudeerd is in de AC, en er al eerder is gerapporteerd over de rol van de AC m.b.t. AM discriminatietoekomst, is er nog weinig bekend over hoe de verschillende gedragstoestanden van de luisteraar de AM gevoeligheid van cellen beïnvloeden. Hier analyseerden we de spontane neurale activiteit, de geluids-onset respons, en de aanhoudende geluidsrespons van AC cellen in wakkere makaken, waarbij de dieren in verschillende gedragssituaties konden verkeren, variërend van het passief ondergaan van geluid, tot het actief deelnemen aan een niet-voorspelbare en volledig voorspelbare geluidstaak. We bepaalden de reactietijden van de dieren als functie van de AM modulaties, en van de mate van voorspelbaarheid in de taak, en vonden een systematische relatie tussen de reactietijd en de akoestische gevoeligheid. We bepaalden modulatie-overdrachts functies (MTFs) om de gedrags- en neurale responsies voor de 15 gebruikte amplitude moduatiefrequenties (AMF tussen 2-256 Hz) te kwantificeren. Uit het Fourier spectrum van de neurale periode histogrammen, namen we de amplitude van de eerste harmonische voor de verschillende AMF (mMTF). We vonden dat de mMTF invariant, dus taak-onafhankelijk, was voor AMFs tussen 5.6 en 45 Hz, maar verschilde tussen passief en actief luisteren voor zowel lagere als hogere AMFs. Betrokkenheid in een taak veranderde ook de Vector Sterkte van de cellen, die bepaalt hoe sterk het fase volggedrag van de neuronen is t.a.v. de temporele omhullende van het geluid, alsmede de gemiddelde vuurfrekwentie en de neurale variabiliteit tussen trials. Onze resultaten suggereren dat de
mMTF een betere maat kan zijn om de codering van de temporele omhullende van complexe geluiden door AC cellen mee te beschrijven, omdat deze maat zowel afhangt van de vuurfrequenties als van de Vector Sterkte.

HOOFDSTUK 5: TAAK-AFHANKELIJKE MODULATIES VERMENIGVULDIGEN MET DE AKOESTISCHE SIGNAALVERWERKING IN DE AUDITIEVE CORTEX VAN MAKAKEN

In dit hoofdstuk kwantificeren we de taak-gerelateerde top-down signalen op de responsies van individuele auditieve cortex neuronen, en trachten het mechanisme te achterhalen waarmee deze signalen interageren met de akoestische informatie. We vergeleken daartoe de neurale activiteit van AC cellen tijdens passieve geluidsstimulatie met de activiteit gemeten voor dezelfde breedbandige spectraal-temporeel gemoduleerde geluiden tijdens een onvoorspelbare en voorspelbare reactietak. Hoewel de dieren niet specifiek waren getraind om aandacht te schenken aan een speciale spectrale of temporele modulatie, toonden hun reactietijden voor de onvoorspelbare stimuli aan dat de dieren een duidelijke spectraal-temporele voorkeur c.q. gevoeligheid hebben. Interessant was dat deze gevoeligheid op de reactietijden verdween voor de voorspelbare stimuli met snelle reactietijden. Echter, zelfs in de voorspelbare taak was de akoestische gevoeligheid zichtbaar op de langzamere reactietijden. Onze analyse van de neurale vuurpatronen liet een taakgerelateerde dynamische modulatie zien op de responsies die was gekoppeld aan de reactietijd van het dier, maar invariant was voor de verschillende spectrale en temporele modulaties. Deze bevinding suggereert een onafhankelijke verwerking van akoestische en top-down taaksignalen op het niveau van individuele AC neuronen. We laten zien dat de de neurale activiteit van een AC neuron beter kan worden beschreven door een multiplicatieve gain-modulatie van de akoestische neurale respons door het taakgerelateerde signaal, dan door een lineaire sommatie van deze twee signalen.
خلاصه فارسی (PERSIAN SUMMARY)

هدف از این پژوهش درک بهتر کارکرد قشر شنوایی میمون در زمینه پردازش اصواتی است که در دو بعد زمان و فرکانس نوسان دارد و همچنین به بررسی اثرات شرایط متغیر متغیر متفاوت رفتاری بر این پردازش پدیده می‌شود. در فصل دوم این پایان‌نامه، ما به بررسی کلیکولوس تحتانی 1 و سلول های کلمیکولوس تحتانی 2 (دو پایگاه اصلی پردازش اطلاعات شنوایی) و بیشتری در مورد سلول های قشر شنوایی و سلول های کلمیکولوس تحتانی 3 (دومین پایگاه اصلی پردازش اطلاعات شنوایی) سلول های قشر شنوایی و سلول های کلمیکولوس تحتانی 4 (دومین پایگاه اصلی پردازش اطلاعات شنوایی) کمک می‌کنیم.

1. ویژگی‌های زمانی-فرکانسی نورون‌های قشر شنوایی میمون

در فصل دوم این پژوهش، ما به بررسی ویژگی‌های زمانی-فرکانسی و سلول های کلمیکولوس تحتانی و سلول های کلمیکولوس تحتانی مشابه دارند. همچنین، به بررسی ویژگی‌های زمانی-فرکانسی در سلول های قشر شنوایی و سلول های کلمیکولوس تحتانی اشاره می‌کنیم.

2. تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا

در فصل سوم این پژوهش، ما به بررسی تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا می‌پردازیم. تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در هر دو بعد زمان و فرکانس نوسان دارد.

3. تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های قشر شنوایی

در فصل چهارم این پژوهش، ما به بررسی تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های قشر شنوایی می‌پردازیم. تاثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های قشر شنوایی در دو بعد زمان و فرکانس نوسان دارد.

4. تأثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های سلول های قشر شنوایی

در فصل پنجم این پژوهش، ما به بررسی تأثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های سلول های قشر شنوایی می‌پردازیم. تأثیرات شرایط مختلف رفتاری بر پردازش اطلاعات زمانی-فرکانسی صدا در سطح تکنورون‌های سلول های قشر شنوایی در دو بعد زمان و فرکانس نوسان دارد.
که هیچ نوسان فرکانسی نداشتند. پاسخ‌های زمانی و فرکانسی اکثریت غالب نورون‌ها (93 درصد) جادوگری بودند که این نشان می‌دهد که فرکانس و زمان به طور جداگانه در نورون‌های قشر شنوایی پردازش می‌شوند. ما میزان خطا بودن پاسخ‌های قشر شنوایی با استفاده از STRF مطلوبه به دست آمده از محکه‌های Tone مقایسه کردیم. دوم، ما پاسخ‌های پیش‌بینی شده با استفاده از یک فیلتر خطی مانند STRF (با محکه‌های طبیعی مانند صدا پرندگان و میمون، را با پاسخ‌های واقعی قشر شنوایی مقایسه کردیم. هر دو مقایسه بر وجود ویژگی‌های غیرخطی نورون‌های قشر شنوایی گواهی دادند. ما همچنین این نتایج را با پژوهش‌های خودمان در کولیکولوس تحتانی مقایسه و تفسیر کردیم که نشان از وجود ویژگی‌های غیرخطی بیشتر قشر شنوایی در مقایسه با کولیکولوس تحتانی داد.

فصل سوم: پردازش پایدار سیگنال‌های پایین-به-بالا (BOTTOM-UP) در قشر شنوایی میمون

در این فصل تلاش شده است تا اثرات سیگنال‌های بالا-به-پایین ضروری فرکانسی و زمانی فرکانسی در شرایط مختلف رفتاری میمون را بررسی کنیم. پژوهش‌های اخیر نشان داده که سیگنال‌های بالا-به-پایین می‌توانند به قشر شنوایی اثر بگذارند. در نتیجه، نورون‌های قشر شنوایی نسبت به کردن بکارگیرنده صوتی تغییراتی به نظر می‌رسد. به همین دلیل، برای تاثیرات سیگنال‌های بالا-به-پایین در شرایط مختلف رفتاری، نورون‌های قشر شنوایی میمون را در پاسخ به محکه Tone در شرایط مختلف متغیر و بدون نیاز به توجه مستقیم، مطالعه کردیم.

فصل چهارم: شرایط متفاوت رفتاری می‌تواند پردازش زمانی را در قشر شنوایی میمون تغییر دهد

این فصل به بررسی اثرات سیگنال‌های پایین-به-بالا در شرایط مختلف رفتاری میمون و در نتیجه تغییراتی در پردازش زمانی می‌پردازد که برای درک گفتمان نیز حیاتی است. در این فصل، به بررسی و تحلیل اصل قشر شنوایی اصلی به‌عنوان یک مرحله اولیه در پردازش اطلاعات زمانی می‌پردازیم که برای درک گفتمان نیز حیاتی است.

# 5 Amplitude modulated noise
# 6 Spontaneous activity
# 7 Onset response
متفاوت رفتاری از قرار گرفتن به صورت پسیو در برابر صوت تا واکنش اک تیو در شرایط رفتاری قابل پیش بینی و غیرقابل پیش بینی را بررسی کردیم. ما زمان واکنش میمونها بر اساس نوسانات زمانی صدا و قابلیت پیش بینی تغییر در صوت تحلیل کردیم. برای تعیین یک میزان قرار گرفتن پسیو-رازی یا اک تیو-واکنشی، به عنوان یک واحد اندازه‌گیری پاسخ‌های نورونی برگزیده شد. به عنوان نمونه به تابعی به نام mMTF (modulation transfer function) نشان دادیم که برای نوسانات از ۵.۶ تا ۶۵ هرتز در شرایط اک تیو و پسیو پایدار باقی مانده، ولی mMTF برای نوسانات از ۵.۴ تا ۲۵ هرتز در شرایط اک تیو و پسیو متغیر بود. همچنین در گروهی از فرکانس‌های بالاتر و پایینتر میان شرایط پسیو و اک تیو متغیر بود. برای مشخص کردن فازبندی و میانگین پاسخ‌های نورونی در فازبندی سلول‌ها نیز می‌تواند میزان میزان اثرات مسالمه‌ای زمانی اصول باشد. می‌تواند میزان و میانگین پاسخ را باعث ناپایداری شود. نتایج ما نشان می‌دهد که mMTF می‌تواند معیار بهتری برای مطالعه زمانی اشارات باشد.

فصل پنجم: سیگنال‌های بالا-به پایین با سیگنال‌های شنوایی در قشر شنوایی میمون ضرب می‌شوند.

در این فصل تلاش شده است تا ماهیت سیگنال‌های بالا-به پایین مشخص شود و سپس به مکانیسم برهم‌کنش میان آنها و اطلاعات شنوایی بپردازیم. بنا بر این، با مقایسه فعالیت سلول‌های قشر شنوایی در حین اکتیویت و شرایط رفتاری مانند گوش دادن به صدا، واکنشاتی کننده به صدا در شرایط اکتیویت و همچنین به صدا در شرایط واکنشاتی را باید با یکدیگر مقایسه و تفسیر کنیم. برای بازنگری این موضوع، می‌توان به مفهوم فازبندی واکنشاتی در شرایط مختلف بازنگری کرد. در این مدل، فازبندی واکنشاتی به عنوان یک پارامتر معنی‌دار در تفسیر واکنشاتی شناخته شد. در نتیجه، می‌توان گفت که mMTF می‌تواند یک معیار بهتری برای مطالعه زمانی اشارات باشد.

8 Magnitude
9 Harmonic
10 Phase locking
BIBLIOGRAPHY


ACKNOWLEDGMENT

John, I have learned a lot from you and I’m happy that I still have the opportunity to learn more. You have been more than a supervisor to me. I remember my first week in the Netherlands, you told that you are not my boss and we are all colleagues working at the same level. You have never had a bossy attitude towards me or anyone else. You are the best teacher that I have ever had; I realized this when you explained the difficult theoretical aspects of my project, otherwise it would be very difficult for a person with biology background, like me, to understand such complicated issues. There is no doubt what a great scientists you are, but you are an even greater human being with a big heart and ready to help everyone as much as possible. I will never forget your help with my arrival at Nijmegen and the way you treated me and Elnaz, making our new life in Nijmegen more cheerful and comfortable. I cannot possibly be more grateful for all you have done for me. Simply put: Thank you! Moreover, I would also like to thank Marijke, your wife, who has been always very friendly and nice to me and Elnaz in a way that made us feel like family.

Marc, I spent most of my PhD time with you. At the start of my PhD, I had no idea about data analysis using MATLAB. You very patiently taught me all these things. It would have been impossible to finish my thesis without your help. Moreover, it has always been fun and fruitful to go to conferences with you as we had good scientific discussions and fun evenings. Thank you so much for all your support and help. You also have been a good friend outside of work, and we have shared lots of memorable moments, especially in Efteling, with you and Marley. I wish both of you and Fin a long, happy, and successful life.

Huib, although you joined my project as co-promoter later, I had been using the setup built by you since the beginning of my PhD. Besides your scientific help, I learned to be always patient, precise, and cautious about scientific results. I have seen you double and triple checking all details obsessively, which seemed weird to me at first, but later I realized that this is an essential part of good scientific work. It has been always fun to sit with you and discuss my results. Thank you so much for all your help.

Sigrid, you did a great job with monkey handling and recordings. Most of the data presented in this thesis were collected by you. Thank you so much for your good job and I am happy that we are still colleagues.

Nothing would work without managers and their assistants. Stan, thank you so much for what you have done as the head of Biophysics department and now the faculty of science. Judith, I am grateful for all your help. You have been always ready to help whenever I knocked your door. Cerien, I appreciate what you are doing for our department. You are always a bomb of energy, cheerful and
happy, spreading energy around. I would like also to thank Irene, Arjan, Annet, Gea, and Marie Louise for their help on numerous occasions.

I would like to show my appreciation towards Richard and Jeroen, for their help, advice and comments during monkey experiments and writing this thesis. Furthermore, I am thankful to Armaz for his advice and support. I wish you good luck with your scientific career.

This thesis could never been produced without help of technicians. Hereby, I must acknowledge Gunter, Hans, Stijn, Dick, and Perry, who always provided good logistical support for my project. It was nice to work with you guys, and I wish you all good luck in the rest of your careers. I should also thank CDL staff especially Alex, Michael, Jeroen, and Wilma for looking after our monkeys.

Ali and Katrien, one of the nicest events during my PhD was getting to know you two guys. It has always been a great pleasure sitting and chatting with you two, especially since your cute and adorable son, Ono, has joined us. Thank you so much for our good moments together and your unconditional help and support.

Hereby, I should express my gratitude to my old colleagues. Rob, you were a huge help during the first two years of my PhD project especially in monkey handling. I also greatly appreciated your assistance in my first two house moves. I wish you lots of success in your current job. Peter, you are one of the most serious and hardworking people that I have ever seen. However, you are a good friend with a kind heart. I am sure that you are going to be a great scientist, just be slightly more positive. By the way, it is good that we are going to work together again. Tom, we were working in a same room for more than two years, which made a lot of good memories and it was always fun chatting with you. Now you are in USA, good luck! I wish you a happy and prosperous life with Maaike [What a coincidence! I received an email from you right now, while I was writing about you]. Denise, you were always full of energy and talking to you always cheered me up. I wish you good luck with your new job as an audiologist and a happy life together with Vincent. Jurrian, you were the only one to whom I could talk about football. I really enjoyed talking to you and hope we get more opportunities to talk about and watch football together. I wish you and Esther good luck in your new careers. David Magezi, we always had good conversations and I learned a lot. I have missed your laughs since you left. I wish you happy life with Juhanna and your two children. Julian, I have always been amazed by your planning system. You are a good example of an organized person, good luck! Artem, David, and Joke, thank you guys for your help in monkey handling.

Here, I would like to thank my current colleagues. Rachel, although sometimes you have weird ideas, it was still really nice working with you in a same group. Thanks for your advice, which helped a lot to wrap up this thesis. Lidwien, you have always been good company in drinks and social events. Moreover, you are a very good listener. Martijn, thank you for the energy you bring to our office. I
never forget our trip to California, it was lots of fun. Bahadir, thank you for our good time together especially during the Dixit game. Anja, we have been colleagues only for a few months, but I am sure that we are going to be good colleagues and friends. Ahmed, we have always had good epical conversations during the lunch breaks. Good luck with your soon-to-be-started PhD. Laurence, although you spend most of your time in Prime doing experiments, you are a quiet, nice colleague when you are in our office. I wish you good luck with your PhD project. Jose, it is only a few weeks since you have joined our group, but you seem to be a nice and funny guy, good luck. Bahram and Elahe, you could be categorized also in my Iranian friend group, but since you both have officially started your PhDs, I mentioned you two here. It is always good to have someone around to whom you can speak in your mother tongue. You provided this opportunity by being here. Thank you and good luck with your PhDs! I wish you lots of highly ranked publications. More importantly, I wish you both a happy life together. At the end of this section, I would like to thank others, to whom I had and/or still have nice chats and good moments such as Kevin, Vicence, Alberto, Mohammad, Sep, Marije, Tara, Marijn, and Attaphone.

I hereby would like to thank my Iranian friends who helped me to feel as if I were at home home. Ali and Sevda, it is almost one and half years since we met for the first time, but we got along very well since then and became best friends. We have had great times together and we always enjoy your company, Let’s keep it up. Amin and Manzar, you brought a lot of fun and cheerfulness to our moments together. I wish you two and Anahita the best. Ali and Faranak, you guys were two of the few Iranians whom I got to know when I came to the Netherlands. Thank you for your helpful information. I wish you both successes in your careers and good times together with Aida. Ebrahim and Pegah, you have been always very helpful to us and I am always grateful for that and for the fun times together especially during BBQs. I wish you both good luck and happy times with Ava. Parisa, Ali, and Sharina, thanks so much for our great times together and all your support during hard days. Although we are missing you in Nijmegen, you are still our good friends. Ehsan and Roya, we used to hangout very often, sadly we have had not the opportunity lately. I wish you two good luck and success in your scientific career. Payam, you have very good knowledge about many things out of your research field and It has been always useful for me chatting and drinking with you. Yaser, you have been also for a short time in Nijmegen, but I have enjoyed our friendship. Thank you and good luck. Alireza, It is always nice spending time with you, especially when you are in good mood and you do not have headache. Houshang, thanks for our enjoyable times together and wish you success in your work and happy life with Helia and Arian. Hamed, you have a unique personality among all my friends, especially when you are drunk. It is always fun sitting with you and chatting. Hamid, you are one of the lunch and table football buddies together with Amin, Ebrahim, and Bahram. Our time together is full of humor and cheerfulness. I really enjoy it especially after a tough day. Thank you so much guys. Here, I should also express my gratitude to other Iranian friends who have always been
there when I needed them including: Mehran, Shahla, Shaghayegh, Hamid, Sanaz, Mr Mousavi and his wife (Zahereh), Saeid, Morteza, Rasoul, Buschra, Yaghoub, Omid, Arash, and Hojat.

Elnaz, with you, I feel happier every day, everything seems more beautiful. Without you, it would be impossible to overcome the problems and get over the disappointments during my life and my PhD. Besides being my wife, you have been my best friend on whom I could always rely. I feel lucky to have you in my life. Thank you so much for everything. I love you.

At the end, I would like to thank those whose names were not mentioned here but directly or indirectly helped me with my PhD. Special thanks to the people of the Netherlands, with whom so far I have spent more than five happy and prosperous years. They treated me very well and accepted me as one of them. I am grateful for this and appreciate it.
Roohollah Massoudi was born on July 9th, 1979, in Eqlid, Iran. He entered primary school in 1985 and graduated from high school with a major on natural sciences in 1997. One year later, he was admitted to the bachelor program on plant biology at Shahid Bahonar University, Kerman, Iran, and graduated in 2002. He continued his education as a master student by studying animal physiology at the University of Tehran, Iran, from 2002 to 2005. His project was about the role of rat hippocampal GABA\textsubscript{B} receptors in morphine addiction. After doing the mandatory military service, in November 2008, he joined the Auditory system, Gaze Control & Executive Function group in the Biophysics department in of the Radboud University Nijmegen (RUNN), The Netherlands, under supervision of prof. dr. John van Opstal and drs Marc van Wanrooij and Huib Versnel. Since then his main focus has been on spectrotemporal properties of monkey auditory cortical neurons and the interaction between acoustic and task related signals.
LIST OF PUBLICATIONS


11. Lommertzen, J. (2009). *Visuomotor coupling at different levels of complexity.* Radboud University Nijmegen, Nijmegen, the Netherlands.


32. Van Dijk, J.P. (2010). On the Number of Motor Units. Radboud University Nijmegen, Nijmegen, the Netherlands.

38. Grootens, K.P. (2010). *Cognitive dysfunction and effects of antipsychotics in schizophrenia and borderline personality disorder*. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


57. van der Linden, M.H. (2011). *Experience-based cortical plasticity in object category representation.* Radboud University Nijmegen, Nijmegen, the Netherlands.


61. Van Leeuwen, T.M. (2011). *‘How one can see what is not there’: Neural mechanisms of grapheme-colour synaesthesia.* Radboud University Nijmegen, Nijmegen, the Netherlands.


64. Voermans, N. (2011). *Neuromuscular features of Ehlers-Danlos syndrome and Marfan syndrome; expanding the phenotype of inherited connective tissue disorders and investigating the role of the extracellular matrix in muscle.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


78. van Barneveld, D.C.P.B.M. (2012). Integration of exteroceptive and interoceptive cues in spatial localization. Radboud University Nijmegen, Nijmegen, the Netherlands.


86. Verhagen, L. (2012). How to grasp a ripe tomato. Utrecht University, Utrecht, the Netherlands.
87. Nonkes, L.J.P. (2012). *Serotonin transporter gene variance causes individual differences in rat behaviour: For better and for worse.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


89. Xiang, HD. (2012). *The language networks of the brain.* Radboud University Nijmegen, Nijmegen, the Netherlands.

90. Snijders, A.H. (2012). *Tackling freezing of gait in Parkinson’s disease.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


107. Van Eijndhoven, P. (2012). State and trait characteristics of early course major depressive disorder. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


172. Lagro, J. (2013). *Cardiovascular and cerebrovascular physiological measurements in clinical practice and prognostics in geriatric patients.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


