Figure-ground processing in the auditory cortex of the macaque monkey

Felix Schneider

A thesis submitted for the degree of Doctor of Philosophy

Bioscience Institute, Newcastle University

January 2020

Abstract

Natural sensory scenes are often very complex, with a multitude of overlapping objects in space and time. In order to direct behaviour, a critical aspect of everyday perception is the segregation and grouping of relevant features from those scenes, known as figure-ground segregation. The neurobiological basis of auditory figure-ground processing is poorly understood. To gain insights into different aspects of this process, I have investigated the behavioural, systemic and neuronal mechanisms the brain uses to segregate and group temporally coherent elements from a complex acoustic scene in macaque monkeys.

This thesis presents the result of this research in five chapters: Chapter 1 reviews the fundamental basics of auditory scene analysis and the auditory system. Chapter 2, 3 and 4 present experimental work and cover figure detection behaviour (Chapter 2), systemic organisation of figure-ground analysis (Chapter 3) and the underlying neuronal mechanisms (Chapter 4). Finally, Chapter 5 discusses and interprets the results in the context of previous research.

In summary, this work establishes that macaques are an excellent animal model for auditory scene analysis and provides new evidence of the cortical response mechanisms during auditory figure-ground segregation. I show that macaques have not only similar detection performance to humans but that the areal organisation measured with fMRI is comparable. Furthermore, I demonstrate robust effects on neuronal firing rates in response to auditory figures across the cortical hierarchy. Lastly, this thesis establishes neuronal differences in figure processing between anterior and posterior auditory cortical fields.

Acknowledgement

I would like to thank my supervisors, Tim, Alex and Yuki, for their constant support and guidance. Whenever I needed something, you were there and helped me to overcome any difficulties. I have learned a lot from you and I'm very grateful for the time we have spent together.

Alwin and Jochem, without you this work would still be in its infancy. Thank you for sharing code, helping with analyses, discussing my data, checking my setup, providing critical feedback and inspiration and just being there whenever I needed a helping hand. Having you around really made this PhD fun!

Molly, thank you for the countless times you have taken care of my animals. I hope I can repay you one day.

Fabien, Pradeep and Michael, thank you for your invaluable help with the MRI and for the support with the data analysis. I enjoyed our time in the basement.

A special thank you to Claudia Distler for performing the post-mortem histology.

Chris, thank you for your guidance over the past three years. Your input helped me to steer this project in the right direction. Also, thank you for providing a lab.

Ryan, thank you for spending so much time setting up the lab with me.

A big thank you to the remaining Griffiths lab, especially to Alex and Sukhbinder, for many discussion and helpful ideas.

None of this work would have been possible without the help and support of the CBC staff. Thank you for taking such good care of the monkeys and for the everyday help you provide.

Thank you to Pawel Kusmierek for sharing code.

Thanks to Ross, Jochem and Michael for unforgettable conference seasons. I've had a blast!

Finally, an enormous thank you to my family: Tabea, for giving me the opportunity to pursue this PhD, while your dreams had to wait. Liam and Mika, for taking my mind off my work. I'm grateful that you are a part of my life!

Table of Contents

List of Abbreviation

A1	Primary auditory cortex	**HI**	Hits
AC	Auditory cortex	**HIr**	Hit rate
AL	Anterolateral belt	**Hz/kHz**	Hertz/Kilohertz
ANOVA	Analysis of Variance	**IPS**	Intraparietal sulcus
ANT	Anterior	**ITPC**	Inter-trial phase coherence
AUROC	Area under the receiver operating characteristic	**LFP**	Local field potential
BOLD	Blood oxygen level dependent	**LGN**	Lateral geniculate nucleus
BPN	Band-pass noise	**ls**	Lateral sulcus
CL	Caudolateral belt	**MEG**	Magnetoencephalography
cla	Claustrum	**MGN**	Medial geniculate nucleus
CLK	Click trains	**ML**	Middle lateral belt
CM	Caudomedial belt	**MM**	Middle medial belt
CN	Cochlear nucleus	**MRI**	Magnetic resonance imaging
Coh	Coherence level	**MUA**	Multi-unit activity
CPB	Caudal Parabelt	**POS**	Posterior
CR	Correct rejections	**PT**	Pure tones
CV	Coefficient of variation	**pu**	Putamen
dB	Decibels	**PV**	Parvalbumin-positive interneurons
deg	Degree	**R**	Rostral core
EEG	Electroencephalography	**RF**	Receptive Field
EPI	Echo-Planar Imaging	**RM**	Rostromedial belt
FA	False alarms	**RPB**	Rostral Parabelt
FAr	False alarm rate	**RT**	Rostrotemporal core
FFT	Fast Fourier transform	**RTL**	Rostrotemporal-lateral belt
FMI	Figure modulation index	**RTM**	Rostrotemporal-medial belt
fMRI	Functional magnetic resonance imaging	**RTp**	Rostrotemporal-polar area
FR	Firing rate	**s/ms**	Seconds/Milliseconds
Freq	Frequency	**SFG**	Stochastic figure-ground
FWHM	Full width at half maximum	**SNR**	Signal-to-Noise ratio
GLM	General linear model	**SOC**	Superior olivary complex

SOM	Somatostatin-positive interneurons	**T1**	longitudinal relaxation time
Spks	Spikes	**T2**	transverse relaxation time
SPL	Sound pressure level	**TE**	Time to Echo
STG	Superior temporal gyrus	**TGdg**	Granular part of the dorsal temporal pole
STGr	Rostral superior temporal gyrus	**TR**	Repetition time
sts	Superior temporal sulcus	**VIP**	vasopressin-positive interneurons
T	Tesla	**WN**	White noise

List of Figures

List of Tables

1 General Introduction

1.1 Summary

The natural acoustic world comprises a huge variety of complex and dynamic auditory scenes, where different sound sources overlap in time and frequency space. From rain forests to cocktail parties, making sense of sound scenes requires the brain to disentangle sound elements that belong to the same source. This process is called auditory scene analysis. Sound objects (also called auditory figures) from the same spatial location have to be dynamically decoded using spectrotemporal features that are especially difficult to segregate in a noisy environment. The segregation of these sound objects is then the fundamental basis for attentional selection and behavioural interaction with the world. This introductory chapter covers basic principles of auditory perception, auditory scene analysis, perceptual grouping and figure-ground processing. It also particularises the aims and significance of the work presented in this thesis.

1.2 Auditory scene analysis

Everyday auditory scenes comprise a multitude of sound sources most of which are irrelevant and can be ignored. Maintaining conversations in such noisy environments can be difficult for people with hearing impairments. However, a significant portion of patients (estimates vary between 5-10%) that seek clinical help because of problems with speech-in-noise perception have normal hearing thresholds (Hind et al., 2011; Kumar et al., 2007). Those patients do not display any signs of peripheral damage but still find it difficult to follow or maintain a conversation in a crowded, noisy environment (The “cocktail-party-problem”, Cherry, 1953). In such cases, damage to the synapses between the cochlea and auditory nerve (cochlear synaptopathy) has been suggested to play a role (Kujawa and Liberman, 2009; Oxenham, 2016). However, proposed electrophysiological measures (e.g. auditory brainstem responses and envelope following responses) do not correlate well with life-time noise exposure or subjective speech-in-noise perception (Guest et al., 2018).

These findings suggest that other, more central (cognitive) processes might contribute to problems with speech-in-noise perception (Pienkowski, 2017). Previous work has linked difficulties with speech-in-noise perception to worse perceptual organisation (Holmes and Griffiths, 2019), a process that relies on the central nervous system. Thus, sensory processing might be intact, but the perceptual representation of sound might be affected in people with poor speech-in-noise perception.

In order to extract speech, or in fact any sound of interest, from the barrage of auditory input, the brain needs to detect sound elements that belong to one source, segregate them from the background and group them into one percept (Bizley and Cohen, 2013). The segmentation and perceptual organisation of the sensory scene into perceptual units is called auditory scene analysis (Bregman, 1990). Different terms are used to describe auditory perceptual grouping: Stream segregation refers to the process of segmenting the incoming sensory information into one or more perceptually segregated streams (Moore and Gockel, 2012). This term is often used to describe the separation of simple, artificial tone sequences. Even though related to the process of stream segregation, the processing of complex sound scenes into foreground and background is referred to as figure-ground segregation (Teki et al., 2013). Here, a stable perceptual object occurs amidst irrelevant background information. Perceptual organisation of any kind relies on the smallest perceptual entities: auditory objects. The following paragraph defines auditory object properties.

1.2.1 Auditory objects

The fundamental basis of our ability to perceive the world, interact with it and communicate about it is the segregation of a complex scene into distinct perceptual entities called objects. Independent of the sensory modality, all objects share some common characteristics and follow Gestalt cues like similarity, continuity, proximity, and common fate (Darwin, 1997; Wagemans et al., 2012). In the visual domain, objects can be defined as complete and coherent components of the subjective interpretation of the visual input (Feldman, 2003). In audition, this definition applies as well, however, the concept of auditory objects is not as intuitive. Objects in the auditory modality originate from a source, which is a physical entity that gives rise to sound, for example, a guitar being played (Moore and Gockel, 2012). However, at each point in time only one sound waveform that comprises all auditory information of the environment hits

the tympanic membrane. The brain needs to extract and bind information that belong to the same source in order for an object percept to arise.

Auditory objects have the following defining characteristics: Firstly, they correspond to things in the sensory world and can be emitted with (e.g. instruments, speech) or without (e.g. environmental sounds) clear intention. Secondly, they can be separated from the sensory world based on their spectrotemporal properties. Thirdly, they can span multiple acoustic events which unfold over time. Hence an auditory object can be a grouped temporal sequence of many events (e.g. footsteps). Lastly, auditory objects can be generalised and their representations is hence invariant to changes in their spectrotemporal properties. (Bizley and Cohen, 2013; Griffiths and Warren, 2004). Taken together, auditory objects can be described as perceptual units that are the computational result of the auditory system's ability to detect, extract, segregate and group spectrotemporal regularities in the acoustic environment. These perceptual units are abstract representation which combine sensory information with internal representations (e.g. memories, semantic information) about the world (Bizley and Cohen, 2013). Since auditory percepts have an inherent temporal nature, a group of successive or simultaneously occurring objects is called an auditory stream (Moore and Gockel, 2012). Perceptual representations can just contain a single stream (e.g. footsteps) or comprise multiple streams from different sources as it is the case for music or choirs. However, ultimately both streams and entire percepts still rely on the segregation of single sound objects.

1.2.2 Perceptual grouping

Perceptual grouping is a biologically widespread and relevant phenomenon that can be found across species such as non-human primates (Christison-Lagay and Cohen, 2014; Fishman et al., 2001, 2017; Knyazeva et al., 2018), birds (Bee and Klump, 2004; Itatani and Klump, 2014) and humans (Higgins et al., 2020; Moore and Gockel, 2012; Teki et al., 2011, 2013). Auditory objects are the basis for many interactions with, and reactions to, the environment. However, before behaviour can occur, objects must be converted into streams or percepts. Unlike the analysis of visual scenes, auditory scene

segmentation crucially relies on the temporal structure of component object dimensions (Bizley and Cohen, 2013). For instance, a rapid sequence of sound objects can be perceived as either one or more streams (Deike et al., 2012; Micheyl et al., 2005). Coherent elements that are masked by noise induce figure-ground effects (O'Sullivan et al., 2015a; Teki et al., 2013; Tóth et al., 2016). But how exactly does the auditory system recognise and combine related elements in a sound scene into a specific percepts? The following section describes factors that contribute to perceptual grouping and as a result cause stream segregation or figure-ground effects.

1.2.2.1 Stream segregation

Simple sound signals that are only comprised of two alternating sinusoidal tone bursts (denoted as A and B) are a great tool to investigate the most basic factors contributing to stream segregation. Due to their design (triplet vs. continuous presentation), these stimuli are called ABA (van Noorden, 1975) or ABAB-signals (Miller and Heise, 1950). Figure 1.1 and Figure 1.2 show schematic examples of these stimuli. When listening to this type of signal, human listeners report varying percepts, depending on the frequency separation between the A and B tones. The larger the frequency difference between the tones, the more likely a segregated percept is reported. For instance, listeners report they hear two streams of different frequency (Figure 1.1 a). However, with only a small difference between the tones, fusion of the elements almost always occurs (Figure 1.1 c), causing listeners to report a galloping ABAB rhythm (Miller and Heise, 1950; Moore and Gockel, 2012; van Noorden, 1975). Intermediate frequency leads to a bistable percept that switches between fusion and fission in an irregular manner (Higgins et al., 2020; Moore and Gockel, 2012; Pressnitzer and Hupé, 2006). The reason for this phenomenon has been defined in previous work (van Noorden, 1975): If the frequency separation exceeds a certain value (temporal coherence boundary), fission usually occurs. On the other hand, if tonal separation is below a critical value (fission boundary, smaller than temporal coherence boundary), a single stream is heard most of the time. In between the two boundaries, bistable percepts can occur.

Stream segregation has been shown to build up gradually over time, with the tendency of fission increasing as function of exposure time (Bregman, 1978; Micheyl

et al., 2005; Moore and Gockel, 2012; Pressnitzer et al., 2008). This suggests that the auditory system seems to operate with the bias to assume a single stream until enough evidence has been gathered to support several sources, i.e. two streams. However, those studies assume a one-stream percept before the first perceptual decision is made. By normalising the probability of a two-stream percept by the probability that a perceptual decision has been made at all, other experimental work shows no build-up of stream segregation at all for large frequency separation (Deike et al., 2012). Instead, only low and intermediate semitone differences cause a moderate built-up in streaming over time. This suggests that the initial percept depends strongly on the frequency separation between A and B tone and that there is no default perceptual bias.

Changes in sound intensity, spatial location or regularity can reset streaming and lead to an integrated percept. Sudden changes have a more drastic effect on segregation compared to gradual alterations (Haywood and Roberts, 2010; Rogers and Bregman, 1993, 1998). Gaps in the tone sequence can partially, or completely, reset the build-up of streaming (Beauvois, 1997; Cusack et al., 2004). Another factor that contributes to grouping of successive tones is the presentation rate of single elements, with low rates being more likely to induce fusion (Moore and Gockel, 2012). Temporally coherent tone elements never cause stream segregation independent of the frequency separation between both tones (Elhilali et al., 2009a, Figure 1.1 b).

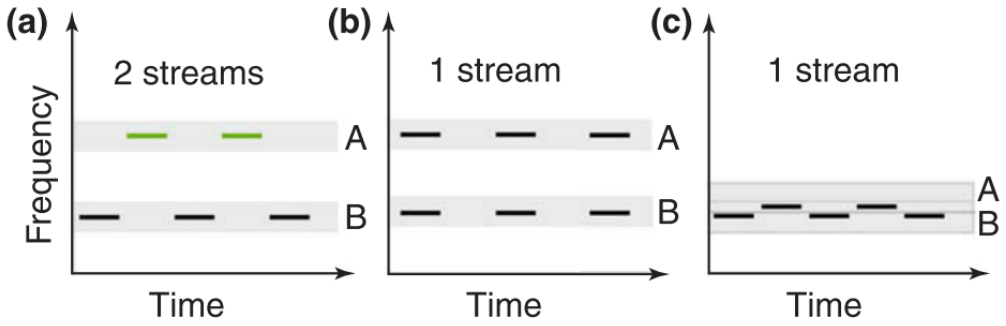

Figure 1.1: Schematic depiction of ABAB stimulus paradigms used to investigate stream segregation.

Frequency elements are indicated by green or black lines. (**a**) Alternating tone presentation with large frequency separation. (**b**) Synchronous tone presentation with large frequency separation. (**c**) Alternating tone presentation with small frequency separation. Perceptual result is indicated in the above text. Figure adapted from Shamma et al., 2011.

1.2.2.2 Figure-ground segregation

Auditory figure-ground segregation refers to the perception of a foreground sound object, the figure, against a background containing irrelevant information. It is closely related to the phenomenon of stream segregation as the presence of an auditory object needs to be streamed independently of the background. The link between stream segregation and figure-ground segregation becomes apparent when considering a previous study that investigated perceptual awareness of a tone sequence using an informational masking paradigm (Gutschalk et al., 2008a). Here, a stochastic tone cloud, consisting of randomly generated tones, masks a stream of repeating tones (multi-tone masker stimulus, Figure 1.2). Listeners were not able to detect the repeating frequency elements from the beginning of the sequence. Over time, however, the detection probability of subjects increased, reflecting perceptual grouping of random masker elements versus the coherent tone sequence.

How auditory objects are perceived in complex acoustic scenes has been tested in human subjects using Stochastic Figure-ground (SFG) stimuli. In contrast to the multi-tone masker signals, these stimuli model natural acoustic scenes, as object and masker elements spectrally overlap. Thus, integration of perceptual evidence over the time-frequency domain is required for perceptual grouping. SFG stimuli are based on randomly generated chords, where a foreground object can arise through the grouping of temporally coherent frequency elements that have to be present in successive chords (Teki et al., 2011). Figure 1.2 shows a schematic depiction of a SFG stimulus. Mechanisms based on the temporal coherence of stimulus features have been suggested to play a role in the segregation of complex acoustic scenes like in multi-talker environments (Shamma et al., 2011).

The detection performance of human listeners strongly relies on the coherence level of the figure (O'Sullivan et al., 2015a; Teki et al., 2011, 2013; Tóth et al., 2016) which is defined by the number of frequency channels that constitute the figure. Higher figure coherence generally leads to better behavioural detection performance. In addition, the duration of the figure, i.e. the number of chords that contain coherent frequency elements, impacts detection performance as well. Longer figure durations improve behavioural detection performance (Teki et al., 2013). A series of behavioural experiments established that the resulting percept of the foreground object is very stable. Human detection behaviour seems to be unaffected by chord duration changes,

interruption by noise burst or ramping of the figure in frequency space (Teki et al., 2013). However, high visual load has been reported to impair figure perception, further indicating that perceptual grouping of correlated sound elements is a cognitive process (Molloy et al., 2018).

Work presented in this thesis exclusively used SFG stimuli to investigate central mechanisms of auditory figure-ground processing.

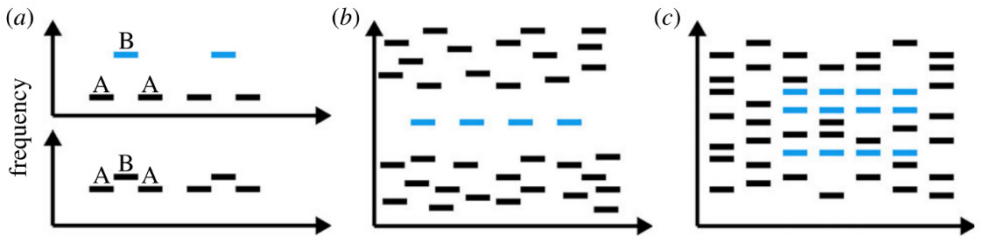

Figure 1.2: Schematic depiction of different stimulus classes used to investigate perceptual grouping.

(**a**) ABA/van Noorden paradigm. Triplets of tones with different frequency are presented. Similar to ABAB stimuli shown in Figure 1.1, percepts can vary from one or two perceived streams. (**b**) Multi-tone masker stimulus used to investigate the effects of informational masking. The sequence of repeated tones is only sometimes perceived (**c**) Stochastic figure ground stimulus that can lead to a perceptual pop-out effect of an auditory object. The stimulus is organised in chords (vertical columns), defined by the summation of randomly selected frequency elements (black lines). Temporally coherent elements that form a figure are depicted in blue. Figure reproduced from Dykstra et al., 2017.

1.3 Macaques as an animal model for figure-ground analysis

Human behavioural and imaging experiments (Molloy et al., 2018; O'Sullivan et al., 2015a; Teki et al., 2011, 2013, 2016; Tóth et al., 2016) have shown clear correlates of auditory figure-ground segmentation. However, the underlying neuronal mechanisms are still poorly understood. Systematic invasive recordings in the human brain are currently not feasible. We therefore sought to establish a non-human primate model to investigate neuronal mechanisms of figure ground processing.

Animal models are a crucial tool in order to better understand fundamental aspects of brain functionality. They allow invasive procedures such as extracellular single cell recordings, optogenetics and cortical lesioning, which

are not possible in humans. These procedures give extensive insights into the basic processing strategies of the brain and are helpful to understand the interplay between cognition, perception and action.

Mice and rats are the most widely used animal species in neuroscience research. In auditory neuroscience, the ferret is a common animal model that has been used to study cortical plasticity (David et al., 2012; Fritz et al., 2003, 2007a; Lu et al., 2017), basic neuronal coding properties (Elgueda et al., 2019; Lu et al., 2019) and scene analysis (Atilgan et al., 2018; Town et al., 2017; Wood et al., 2017a). Ferrets have a comparable cortical organisation (Bizley et al., 2005), hearing range (Kavanagh and Kelly, 1988) and are more closely related to humans than rodents are. However, there are still marked difference between species, for example less gyrification and fewer cortical fields with a tilted tonotopic axis compared to primates (Bizley et al., 2005). In contrast, macaque monkeys are the closest primate model that is available for invasive, experimental work. Macaques are able to work on challenging cognitive tasks that allow exploration of the complex nature of primate perception and cognition. The potential insights of research results are likely transferable to human auditory processing as both primates share:

- similar audiograms (Dylla et al., 2013; Jackson et al., 1999; Pfingst et al., 1975a),
- similar detection performance of tones in quiet (Heffner and Heffner, 1986) and noise (Dylla et al., 2013)
- similar pitch perception (Joly et al., 2014a)
- similar streaming abilities for two-tone (ABAB) sequences (Christison-Lagay and Cohen, 2014)
- homologous organisation of the auditory cortex (Baumann et al., 2013; Joly et al., 2014b).

Macaque monkeys and humans also use their sensory systems in a similar way which is not surprising given that they have evolved in similar environments. Both species rely primarily on visual information. This is not the case for other lab animals like rodents that rely much more on somatosensation (Diamond et al., 2008). Of course, the choice of an animal model always depends on the system under

investigation. However, in this case, a primate model of auditory scene analysis would yield valuable insights that could illuminate long unanswered scientific questions, for example how humans use auditory information to solve the cocktail party problem.

1.4 The auditory system of macaque monkeys

The auditory system of mammals is morphologically similar, however, different species have evolved different functional specialisations (Malmierca, 2013). Work described in this thesis focussed on the behavioural and neural correlates of auditory figure-ground perception in non-human primates. Hence, the primate auditory system is introduced below. However, some paragraphs describe overarching principles that are found across mammals and include references from studies on non-primate species.

The following two sections (1.4.1 and 1.4.2) are summarised based on chapter 6 (“The Subcortical Nuclei”) from the ‘Physiology of Hearing’ textbook (Pickles, 2012), chapter 2 (“The Ear”) from the ‘Auditory Neuroscience’ textbook (Schnupp et al., 2011) and chapter 30 (“Hearing”) from the ‘Principles of Neuroscience’ textbook (Kandel et al., 2000).

1.4.1 From sound to action potential

In essence, sound is vibration. Under normal circumstances, sound leads to small change in air pressure which in turn cause the tympanic membrane to vibrate. This physical oscillation is mechanically transmitted via tiny bones in the middle ear, the malleus, incus and stapes, to the fluid inside the cochlea’s compartments. However, sound can completely bypass the tympanic membrane without major changes to the resulting percept. This is evident by the fact that bone-conducting headphones exist, or that one can hear its own voice when blocking both ears. The conversion, from a mechanical signal into an electrical one, happens in the cochlea. In one of the cochlea’s compartments, the scala media, the Organ of Corti is located, which sits on top of the basilar membrane. It has stereocilia projecting towards the tectorial membrane. Deflections of the basilar membrane physically opens the ion channels of those hair cells which causes changes in membrane potential that can lead to

glutamate release at an excitatory synapse and ultimately an action potential in the auditory nerve. Due to the changing mechanical properties of the basilar membrane, high frequencies are transduced at the base and low frequencies near the apex. Action potentials of auditory ganglion cells can already encode frequency and intensity.

1.4.2 The ascending auditory pathway

The analysis of sound is done in separate, parallel streams, where critical features are progressively extracted and passed on to the next stage.

After transduction from vibration to voltage, the signal is conveyed via the auditory nerve (VIII cranial or vestibulocochlear nerve) to the cochlear nucleus (CN). The spatial arrangement of nerve fibers maintains a tonotopic organisation that is maintained all the way to the auditory cortex. Auditory nerve fibres branch as they enter the brainstem. At the CN, the first auditory relay station of the brain after the cochlea, the sound abstraction process begins. The nucleus has three subdivision (dorsal, anteroventral and posteroventral), with ascending branches terminating in the anteroventral subdivison. The ventral part of the CN processes the temporal structure of sound whereas the dorsal part cares more about spectral contrasts. Different cell types (Bushy-, Stellate-, Octopus-, Fusiform cells) in each division encode different stimulus features and send their outputs to different parts of the ascending auditory pathway. Hence, the separation of information streams commences at this point. From the CN, information gets projected via two main streams:

1) A dorsal stream, running through both, acoustic and intermediate stria, projecting directly to the inferior colliculus (IC). This stream mainly serves to support sound identification. Mostly dorsal CN cells transmit information via that route (Figure 1.3).
2) Information from anteroventral CN cells project via the ventral stream that runs through the trapezoid body, projecting to three nuclei of the superior olivary complex (SOC). Here, timing and intensity information from right and left ear converge, critically involving this structure in spatial hearing (Figure 1.3).

Subsequently, axons from CN and SOC run together towards the (predominantly contralateral) IC, where most information will eventually converge. This tract constitutes the lateral lemniscus. Some side branches may or may not intervene

different nuclei of the lateral lemniscus (NLL), where neurons with complex, multipeaked tuning curves further process complex cross-frequency interactions of sounds. Furthermore, the dorsal NLL increases the accuracy, contrast and dynamic range of localisation information.

Afterwards, the IC is the main receiving station and primary site of convergence for sound identification and localisation information. Most of the auditory input arrives in the tonotopically organised central nucleus of the IC. Neuronal responses become more complex at this stage, with stimulus features being combined, which is relevant for object analysis. Thus, IC represents a combination of auditory features that are relevant to real-world objects. In contrast to earlier stages, like the auditory nerve, IC neurons can encode broadband stimuli.

From the IC, information gets passed to the medial geniculate nucleus. This auditory relay station of the thalamus is subdivided into three divisions (ventral, dorsal and medial). Each division projects to different aspects of the auditory cortex. The ventral division sharpens the frequency resolution and projects strongly to the primary auditory cortex. However, it is reciprocally connected with the cortex and has to be seen as a functional unit. Medial and dorsal divisions form part of the non-specific projections to other auditory cortical areas. They include cells with multimodal response properties and are linked to learning and emotional responses. Similar to the IC, the MGN is organised in a laminar fashion, with different functional properties.

Taken together, the incoming sound signal has been extensively processed in a number of brain nuclei that have extracted sound features, location and multimodal signals before the information has even reached the cortex. Information about object properties have already been present from the IC onwards. Cortical response properties are determined by upstream subcortical nuclei, which differ between primary and non-primary areas.

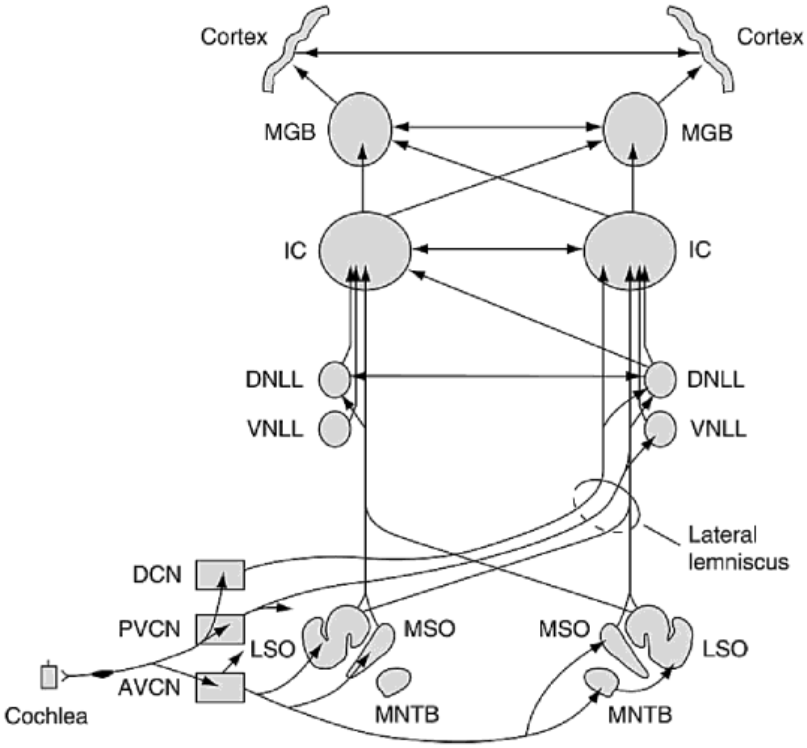

Figure 1.3: Schematic overview of the ascending auditory pathway.

Many minor and inhibitory pathways are not shown. Abbreviations: AVCN, Anteroventral cochlear; nucleus; DCN, Dorsal cochlear nucleus; DNLL, Dorsal nucleus of lateral lemniscus; IC, Inferior colliculus; LSO, Lateral superior olivary nucleus; MGB, Medial geniculate body; MNTB, Medial nucleus of trapezoid body; MSO, Medial superior olivary nucleus; PVCN, Posteroventeal cochlear nucleus; VNLL, Ventral nucleus of lateral lemniscus. Figure reproduced from Pickles, 2012.

1.4.3 The auditory cortex

In mammals, the auditory cortex is a group of adjoining cortical areas in the temporal region of the cerebral cortex that receives significant thalamic input from the medial geniculate body. The number of areas, their arrangements and feature processing abilities vary widely across species. Auditory-related areas are directly connected to the auditory cortex and can be found in every lobe of the brain (Hackett, 2011; Hackett et al., 1998a). For practical reasons, this paragraph focusses on the auditory cortex of macaque monkeys.

In macaques, most of the auditory fields can be found on the superior temporal plane (the dorsal surface of the temporal lobe) and along the superior temporal gyrus (Figure 1.4). Based on thalamic input (Hackett et al., 1998b; Kaas and Hackett, 2000; Scott et al., 2017), cytoarchitecture (Hackett et al., 1998b, 2001; Joly et al., 2014b) and corticocortical connectivity (Hackett et al., 1999, 2014; Kaas et al., 1999; Scott et al., 2015), the auditory cortex has been subdivided into three distinct hierarchical levels:

1) Primary “core” koniocortex, characterised by a prominent granular layer 4, in the centre of the auditory field,
2) non-primary “belt” regions that encompasses the core,
3) even higher order parabelt fields that are located lateral to the belt (Hackett et al., 1998b).

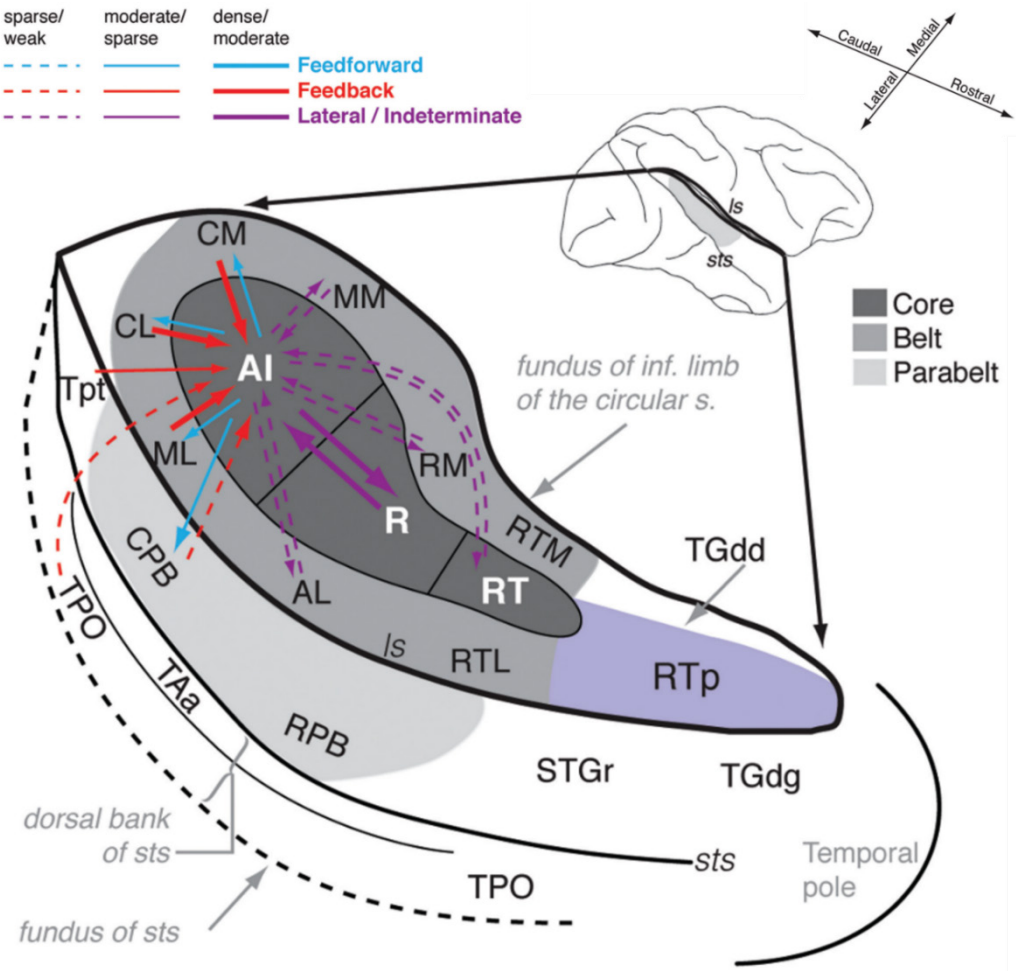

Figure 1.4: Schematic diagram of the superior temporal plane of macaque monkeys with known auditory and auditory-related areas.

Colour arrows indicate hierarchical relationship between A1 and linked fields. Line thickness indicates connection strength. Abbreviations – Structures: ls, lateral sulcus; sts, superior temporal sulcus; Core: A1, Primary auditory cortex; R, Rostral core; RT, Rostrotemporal core; Belt: AL, anterolateral belt; CL, Caudolateral belt; CM, Caudomedial belt; ML, Middle lateral belt; MM, Middle medial belt; RM, Rostromedial belt; RTL, Rostrotemporal-lateral belt; RTM, Rostrotemporal-medial belt; Parabelt: CPB, Caudal Parabelt; RPB, Rostral parabelt; Auditory-related: RTp, Rostrotemporal-polar; STGr, Rostral superior temporal gyrus; TAa, STS dorsal bank areas; TGdd, Dysgranular part of the dorsal temporal pole; TGdg, Granular part of the dorsal temporal pole; TPO, Sts dorsal bank area; Tpt, Temporo-parietal area; Figure reproduced from Scott et al., 2015.

The three core (A1, R, RT) and eight belt regions (CM, CL, ML, AL, RTL, RTM, RM, MM) are exclusively situated deep inside the lateral sulcus (Figure 1.4). Only the parabelt region stretches along the convexity of the superior temporal gyrus. The location of core areas can be highly variable with respect to anatomical features in

chimpanzees and humans. Macaque monkeys exhibit less variability regarding the location of primary regions (Hackett et al., 2001; Marie et al., 2015).

The primary areas A1 and R receive more than 80% of their ascending projections from the ventral subdivision of the medial geniculate nucleus (Scott et al., 2017). Most neurons in A1 are responsive to stimulation through either ear, however, with differing sensitivity. The cortex is subdivided into alternating summation and suppression columns. In summation columns, neurons are excited by stimulation of both ears, with stronger drive from the contralateral ear. In suppression columns, neurons are driven by unilateral input that is suppressed by stimulation of the opposite ear. These columns are situated in a right angle to the tonotopic gradient (Kandel et al., 2000). Other orthogonally directed functional columns seem to exist for amplitude modulation (Baumann et al., 2015).

Information flows in a mostly serial fashion along the caudorostral and mediolateral axes with corticocortical projections mainly originating from the primary auditory cortex (Hackett, 2011; Kikuchi et al., 2010; Scott et al., 2015, 2017). Each cortical area is strongly and reciprocally connected with its immediate neighbours but long-range connection to other auditory fields have been found as well (Hackett et al., 2014; Scott et al., 2015). Figure 1.4 illustrates how the primary auditory cortex, A1, is linked to other auditory fields. Strict hierarchical sensory processing cannot be inferred based on this connectivity pattern. Instead, the data suggest recurrent inter-areal processing. In addition to the primarily stepwise information transfer, the auditory cortex receives parallel thalamocortical projections from different thalamic nuclei (Scott et al., 2017). The composition of thalamic input changes dramatically between auditory and auditory-related areas (Figure 1.5). Multisensory thalamic and cortical information converge in posteromedial parts of the auditory cortex (Hackett et al., 2007; Smiley et al., 2007) and anteriorly to the core, all of which are strongly connected to the auditory core (Scott et al., 2015, 2017).

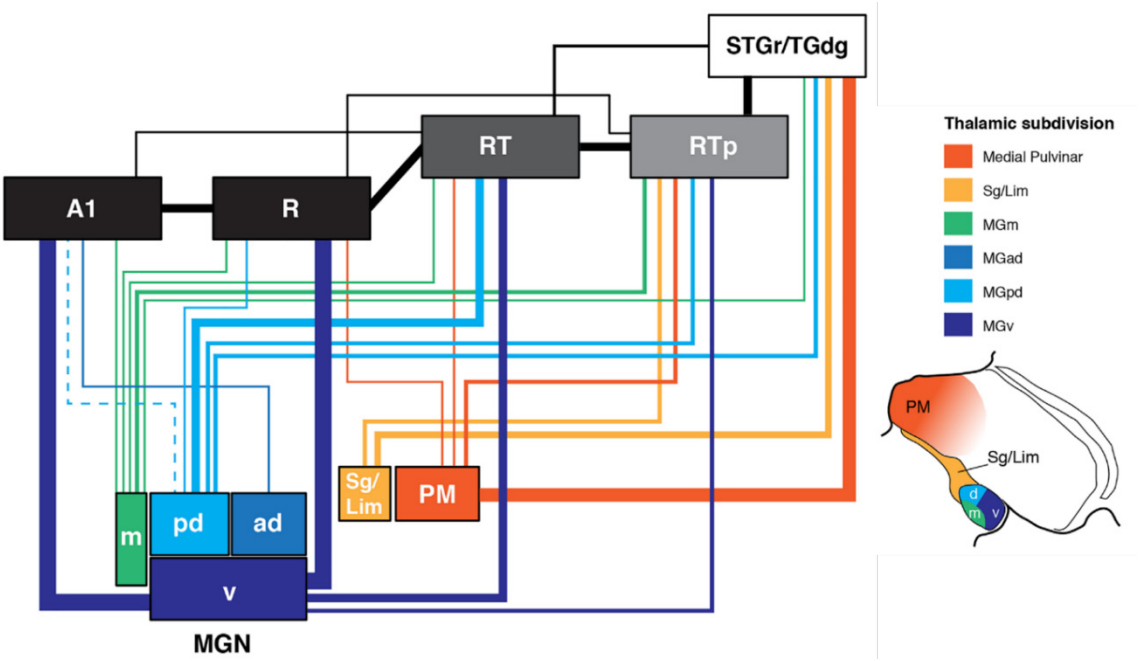

Figure 1.5: Flowchart diagram of thalamic input to auditory core and auditory-related regions in the anterior superior temporal plane.

Colour code indicates thalamic origin of input. Line thickness represents connection strength. Hierarchical, feedforward relationship between areas shown by vertical offset. Right: Schematic diagram of the thalamus shown with selected subdivisions. Abbreviations – Structures: Lim, Limitans nucleus; MGN, Medial geniculate nucleus (with anterior dorsal (ad), posterior dorsal (pd), medial (m), ventral (v) subdivision); PM, Medial pulvinar; Sg, Suprageniculate nucleus; Core: A1, Primary auditory cortex; R, Rostral core; RT, Rostrotemporal core; Auditory-related: RTp, Rostrotemporal-polar; STGr, Rostral superior temporal gyrus; TGdg, Granular part of the dorsal temporal pole; Figure reproduced from Scott et al., 2017.

The auditory cortex of mammals is characterised by a spatial/tonotopic representation of frequency selectivity of cortical neurons. Tonotopic gradient reversals mark the division boundaries between cortical fields (Bandyopadhyay et al., 2010; Besle et al., 2019; Guo et al., 2012; Hackett et al., 2011; Joly et al., 2014b; Kikuchi et al., 2019; Kuśmierek and Rauschecker, 2009a; Malmierca, 2013; Nelken et al., 2008; Recanzone et al., 2000; Saenz and Langers, 2014). On average, tonotopic organisation is well-defined in core and belt areas (Joly et al., 2014b; Kikuchi et al., 2019; Kuśmierek and Rauschecker, 2009a, 2014; Recanzone et al., 2000) but less pronounced in the parabelt (Kajikawa et al., 2015). However, when considering the characteristic frequencies of single neurons, correlated variability with the complexity of the neurons frequency selectivity has been found (Gaucher et al., 2019). In

macaques, A1 completely represents the cochlear frequency space (Merzenich and Brugge, 1973). The tonotopic gradient progresses in a high-low-high fashion. A1 represents higher frequencies caudomedially and low frequencies rostrolaterally. Area R displays a mirror-inverted frequency response profile (Baumann et al., 2015; Joly et al., 2014b; Morel et al., 1993). Tonotopic organisation of belt fields is similar to the adjacent subdivision of the auditory core (Hackett et al., 1998a; Morel et al., 1993). Functionally, core areas A1 and R have similar frequency tuning properties, response thresholds and strength of activation. However, the temporal integration of auditory input varies between core fields with longer integration windows in rostral fields (Scott et al., 2011).

Mammalian primary cortical areas are not just simple feature analyser. A1 responses are highly modulated by task demands (Bagur et al., 2018; Lu et al., 2017; Scheich et al., 2007), perceptual judgements (Bizley et al., 2013; Niwa et al., 2012; Tsunada et al., 2016), categorical decision making (Selezneva et al., 2006; Tsunada et al., 2011) and exhibit strong involvement in cognitive functions such as memory, learning and attention (Weinberger, 2004, 2012).

The following section describes the structural organisation of the mammalian neocortex in more depth.

1.4.3.1 Cortical microcircuits

Even though there are major differences in the architecture of different cortical areas, the overall microcircuit and flow of information remains similar (Douglas and Martin, 2004). The neocortex is horizontally organised into six cortical layers, which can be subdivided (from superficial to deep) into supragranular, granular and infragranular layers, based on their location with respect to the granular layer 4 (L4). In addition, sensory areas can be vertically organised into microcolumns (Diamond et al., 2008; Hubel and Wiesel, 1968; Linden, 2003). The similarity of this anatomical structure across brain areas and mammalian species has led to the development of a canonical cortical microcircuit. Generally, there are two major classes of cortical neurons:

1) Principal cell that are usually excitatory, glutamatergic pyramidal neurons. This class provides approximately 80% of cortical neurons.

2) Inhibitory GABAergic interneurons, about 20% of cortical neurons (Harris and Mrsic-Flogel, 2013).

These classes can be further subdivided into different cells types, with different projections patterns and functional response properties. In addition, neurons across the brain can be modulated by different neurotransmitters, like acetylcholine, dopamine and serotonin (Kandel et al., 2000).

The vast majority of sensory information arrives from the thalamus and targets principal cells most densely in L4 and at the L5/6 border (Figure 1.4, Huang and Winer, 2000; Smith *et al.*, 2012; Constantinople and Bruno, 2013). However, thalamic input can arrive in all cortical layers (Harris and Mrsic-Flogel, 2013; Winer and Lee, 2007). L4 principal cells then project to all other layers, but most densely to L2/3. The cortical wiring suggests a primary flow of information from granular to supragranular to infragranular layers (Blasdel et al., 1985) but cortical layers are highly interconnected and form a multiple circuits between different types of cells (Callaway, 1998; Harris and Mrsic-Flogel, 2013).

Cortico-cortical connection arise from L2/3 (Figure 1.6). L5 neurons project to subcortical structure. L6 cells connect back to the thalamus. Top-down input projects heavily to L1 (Harris and Mrsic-Flogel, 2013).

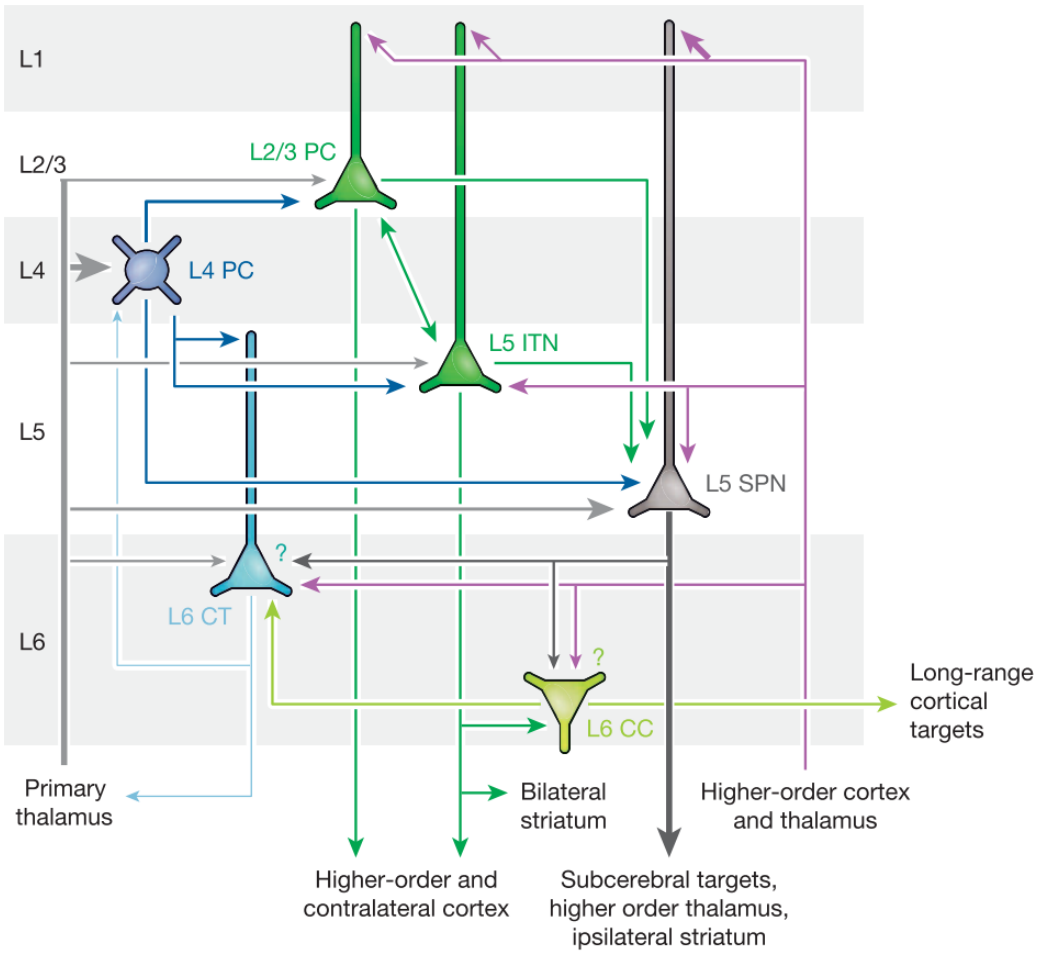

Figure 1.6: Connectivity between cortical principle cells.

Schematic depiction of cortical microcircuit of excitatory pyramidal cells. Colour coded cells represent major cell types of sensory cortices. Line thickness represents strength of pathway. Abbreviations: CC, Corticocortical cells, CT, Corticothalamic cells, ITN, Intratelencephalic neurons; SPN, Subcerebral projection neurons; PC, Principle cells. Figure reproduced from Harris and Mrsic-Flogel, 2013.

The function of excitatory neurons is to integrate and transmit information. Cortical interneurons, on the other hand, modulate the response profile and thereby shape the way pyramidal cells integrate information (Blackwell and Geffen, 2017; Wood et al., 2017b). Three major classes of cortical interneurons have been identified: Parvalbumin-positive interneurons (PV), somatostatin-positive interneurons (SOM) and vasopressin-positive interneurons (VIP). Inhibitory interneurons do not only link up with excitatory cells but also form reciprocal connections with each other. PV cells mainly target

the soma of principal cells, whereas SOM cells mostly connect to the dendritic tree (Figure 1.7). These diverse connections control how the network processes information. Variance in the composition of these cell types can cause different integration of thalamocortical or cortico-cortical input and hence lead to non-linear response patterns (Blackwell and Geffen, 2017). Recent evidence suggests that cortical inhibition sharpens frequency selectivity, increase signal-to-noise ratio of tone-evoked responses and facilitates cortical adaptation (Liu et al., 2007; Natan et al., 2017; Schinkel-Bielefeld et al., 2012). It has also been suggested that the stimulus history is important for the differential function of interneurons (Natan et al., 2017; Phillips et al., 2017a). Furthermore, inhibitory interneurons control complex behaviour by mediating learning (Letzkus et al., 2011). Thus, cortical responses to complex stimuli depend on a multitude of variables and can be highly variable based on the state of the cortical network.

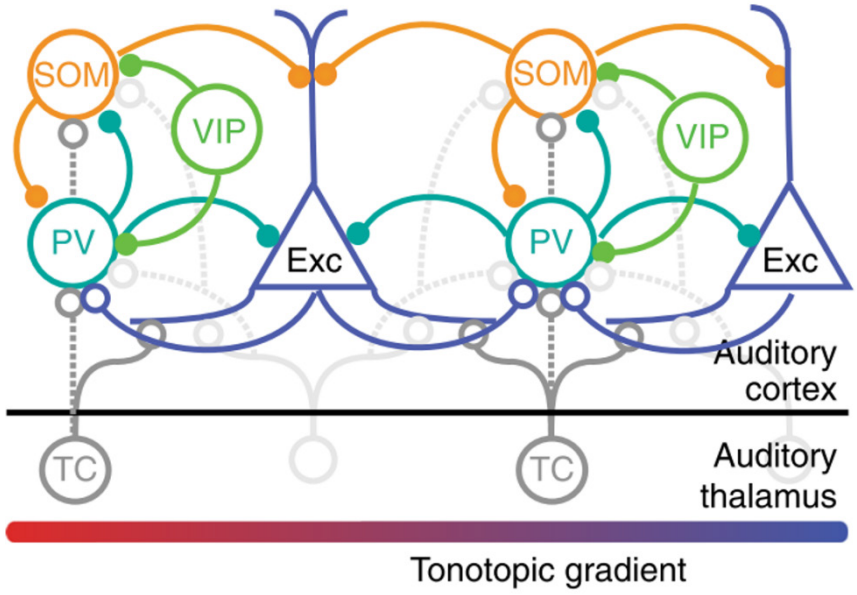

Figure 1.7: Schematic of connectivity between excitatory neurons and different types of inhibitory interneurons in the auditory cortex.

Colour code indicated different classes of neurons. Abbreviations: Exc: Excitatory neurons; PV: parvalbumin-positive interneurons; SOM: somatostatin-positive interneurons; VIP: vasopressin- positive interneurons; TC: Thalamocortical projection neurons. Figure reproduced from Blackwell and Geffen, 2017.

1.4.3.2 Corticocortical communication

To better understand how auditory perceptual grouping might be implemented into the cortical circuit, communication between cortical areas is an important factor to consider. Even though much of this work has been done in the visual domain, similar mechanisms might be important for information transfer between auditory cortical areas.

Neural networks process the temporal dynamics of spike trains with millisecond precision (Kayser et al., 2010), however, how cortical areas interact with each is still under debate. Several theories have previously been proposed:

Inter-areal signal transmission can be detected through spike correlations (Nowak et al., 1995, 1999). Thus, early theories have proposed a mechanism based on a temporal coding mechanisms (spiking synchrony), with local and distributed impact of synchronous input (Engel et al., 1991; Gray et al., 1989; Singer, 1999). However, experimental data suggest no increased synchrony during perceptual binding of moving elements in area MT, indicating no functional significance of neural synchrony (Thiele and Stoner, 2003).

Another framework argues that effective communication between cortical areas is achieved by “communication through coherence” (Fries, 2005, 2009). Since cortical excitability oscillates, it is hypothesized that only coherently activated neuronal populations interact effectively. Oscillatory synchrony allows spikes from cortical cells in area A to arrive in area B during a time window of excitability. This increases the likelihood of eliciting action potentials in area B neurons. Oscillatory interactions between neuronal clusters can be measured with local field potentials (LFP), which reflect the integrated synaptic current at the recording site (Kajikawa and Schroeder, 2011). The inter-areal flow of information in the cortex has been associated with different oscillatory frequencies. Specifically, feedforward interactions convey sensory input through synchronisation in the theta (~4Hz) and gamma (40-90 Hz) range, whereas feedback connections are associated with oscillation synchrony in the alpha (5-15) and beta (15-30Hz) range (Bastos et al., 2015; van Kerkoerle et al., 2014). This distinction in oscillatory information transfer has also been linked to the predictive coding framework (Friston, 2005; Heilbron and Chait, 2018), where each cortical area generates predictions about the world based on

sensory input that get passed on to the next step in the cortical hierarchy. The prediction error, the difference between a generated model and the actual sensory input is sent to downstream, higher-order brain areas. There it is used to refine and update predictions that then get passed back to upstream cortical areas, lower in the processing hierarchy. That way, expectation and prediction errors play a fundamental role in sensory processing and ultimately perception. Gamma oscillatory synchrony (>40Hz) between cortical areas has been linked to predictions error transmission (feedforward) whereas alpha/beta oscillations (<30Hz) seem to correspond to updated prediction signals (feedback) from higher cortical centres (Chao et al., 2018). This framework is compatible with the canonical cortical microcircuit, suggesting layer-specific generators for the different oscillatory signals. The microcircuitry within a cortical column allows for interactions between feedforward and feedback transmissions, needed for the generation of the prediction error (Bastos et al., 2012).

Another plausible mechanism of information transfer is based on selective and flexible information transfer through a “communication subspace”. In this framework, instead of sending the combined information present within a cortical area, only a selected subset of lower dimensional information is passed on to the next stage of cortical processing (Semedo et al., 2019).

Finally, cortical activity fluctuates between up- and down-states across layers, caused by changes in neuromodulatory and glutamatergic input and the allocation of attention (Engel et al., 2016; Harris and Thiele, 2011). The current state of the cortex is controlled locally and has profound impact on neuronal responsiveness and latency of responses (Engel et al., 2016; Hasenstaub et al., 2007). Periods of vigorous spiking activity (Up-state) can also be thought as information packets. These packets are usually 50-200ms of sustained firing, depend heavily on the cortical state and are thought to be the basic building blocks of cortical processing (Luczak et al., 2015).

1.5 Auditory cortical information processing

Perceptual organisation will ultimately rely on the information processing strategies of the neocortex. This section reviews what we know about cortical pathways, cortical responses to simple sounds and finally neuronal correlates of stream segregation.

1.5.1 Dual pathway model of information processing

Cortical connectivity and neuronal response properties of macaques have suggested that auditory cortical areas process sensory information in two separate streams (Kaas and Hackett, 1999; Rauschecker and Tian, 2000).

A ventral stream that runs through the lateral belt (ML, AL), parabelt, superior temporal gyrus, temporal pole towards the frontal cortex (areas 10, 12, 45, 46) which is mainly engaged in object processing. Cells in those areas are highly selective for stimulus category.

In contrast, a dorsal stream from caudal belt areas connects through the posterior parietal areas with frontal areas 8a and 46. Neurons in the posterior auditory cortex show high selectivity for spatial location (Miller and Recanzone, 2009; Rauschecker and Tian, 2000; Romanski et al., 1999a).

Functional imaging has revealed analogous parallel processing of sensory input in humans (Ahveninen et al., 2006; Alain et al., 2001; Arnott et al., 2004).

Similarly, the visual system uses several parallel pathways to process sensory information. A magnocellular and parvocellular pathway conveys visual information through the lateral geniculate nucleus to the primary visual cortex. From there, information gets relayed into a ventral and dorsal cortical processing stream that each process different dimensions of the sensory input. The dorsal pathway extends from V1 through areas MT and MST to the posterior parietal cortex and is involved in the processing of motion and spatial location. On the other hand, the ventral pathway runs through V4 towards the inferior temporal cortex and is contributes to object recognition. There is, however, significant cross talk between both pathways that facilitates scene analysis (Kandel et al., 2000).

Thus, parallel, hierarchical processing of sensory input might be a universal property of sensory cortices. Similar to the visual processing chain, the auditory system splits spatial and object processing. However, subcortical computations are much more complex compared to the visual information flow, suggesting important differences between sensory modalities. In addition, there is mounting evidence for a more distributed nature of auditory feature representation (Bizley et al., 2009; Ortiz-Rios et al., 2017; Stecker and

Middlebrooks, 2003), which questions the strict distinction between ventral ‘what’ and dorsal ‘where’ processing and rather suggests more dynamic information encoding with widespread crosstalk between auditory cortical areas.

1.5.2 How neurons encode stimulus features

Neuronal responses to simple sounds (e.g. sinusoidal tone, short noise burst), that do not require cognitive resources to be perceived, can vary distinctively between neurons of the same auditory cortical field. Therefore, the cortical representation of a percept, as opposed this sensory representation, is likely to be more complex. Thus, how might the cortical representation of a percept to look like?

The majority of neurons respond with a transient sound onset response before settling in on a sustained firing level throughout the presentation of the sound. However, some neurons might only respond with an onset transient, build up firing rate over time or show a sustained response throughout stimulus presentation (Bizley et al., 2005; Recanzone, 2000). Figure 1.8 illustrates these differences for three example multi-units. Firing rate dynamics seem to be similar for tone and noise stimuli (Recanzone, 2000). However, across primary and non-primary fields, bandpass noise bursts generally evoke greater response magnitudes with shorter latencies compared to tone stimuli. This might be due to the integration of more excitatory input to one cell. Best frequencies seem to be unaffected regardless of tone or noise stimulation (Kajikawa et al., 2011).

Response latencies vary widely across cortical areas and stimuli, with shorter latencies in the posterior-medial fields of the auditory cortex (Camalier et al., 2012; Kikuchi et al., 2010). Even core areas, A1 and R, show significantly different onset latencies and response profiles with longer latencies and less synchronised response dynamics in rostral fields (Bendor and Wang, 2008; Camalier et al., 2012; Kuśmierek and Rauschecker, 2009a; Scott et al., 2011) suggesting differences in the temporal integration window between those fields. Neurons also respond in a highly variable manner to the offset of an sound, with some cells strongly responding to sudden silence whereas other neurons return to baseline firing (Recanzone, 2000; Sołyga and Barkat, 2019). The auditory cortex represents these temporal response properties in a spatial, tonotopic manner, with parallel processing networks for onset and offset responses (Liu et al., 2019).

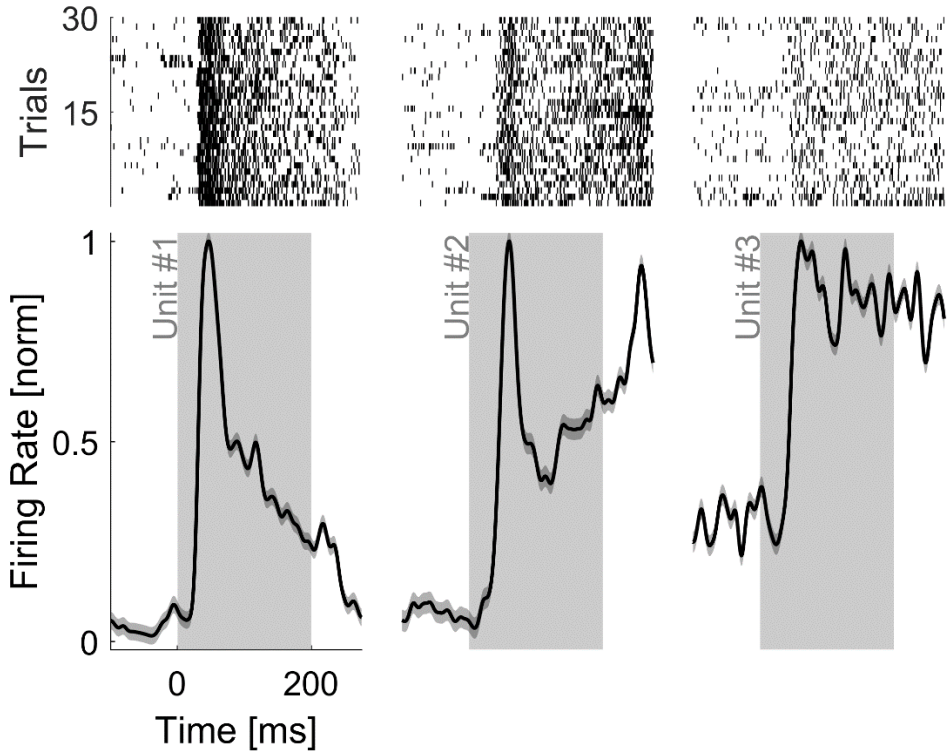

Figure 1.8: Spiking response of three example multi-units to white noise bursts.

Raster plots are shown on top. Each row corresponds to a trial, each point within a trial to a single spike. Spike density functions for each condition shown below. Shaded areas indicate standard error of mean. Firing rates are normalised to the maximum response, averaged over trials. Grey background corresponds to the stimulus presentation period. Left: Marked onset response, sustained activity slowly fading away. Middle: Marked response to onset and offset with strong sustained firing in between. Right: No onset or offset transient but sustained firing in response to stimulus.

More sophisticated tone signals, like harmonic complex tones, can induce a complex response profile that varies from pure tone stimulation. Complex tuning profiles suggest that the selectivity of neurons partly depend on the stimulus choice (Feng and Wang, 2017). When comparing responses to sinusoidal tones and natural sounds across primary and non-primary auditory cortex, firing rates to complex, harmonic sounds show shorter latencies in the lateral belt. This suggests clear differences in feature selectivity between core and belt areas with facilitated processing of harmonic sounds in lateral, non-primary areas (Kikuchi et al., 2014). However, other evidence shows that similarities in sound processing between primary and non-primary auditory cortex outweigh the differences, which suggests strong parallel processing of

auditory input (Eggermont, 1998). In addition, behavioural relevance (Brosch et al., 2005, 2011; Fritz et al., 2003; Recanzone et al., 1993; Yin et al., 2014), stimulus history (Phillips et al., 2017b; Ulanovsky, 2004; Ulanovsky et al., 2003), passive exposure (Chang and Merzenich, 2003; Noreña et al., 2006) and sensory context (Barczak et al., 2018; Williamson et al., 2016) dramatically influence the response properties of neurons across the auditory cortex.

Taken together, neuronal responses to simple sounds can vary distinctly in their firing rate modulation and onset latency based on the presented stimulus and task demands. Since these responses seem to be the building block of percepts, the same properties would presumably apply for more complex cortical object representations.

1.5.3 Neuronal basis of stream formation

In the last decades, the neuronal basis of stream segregation has been thoroughly investigated using simple alternating or synchronous ABAB tone sequences, where A and B correspond to pure tones of different frequencies (Figure 1.1). Electrophysiological recordings in the auditory cortex of different animals have provided a physiological foundation for the psychophysical findings described above.

When passively listening to an ABAB sequence (A set to best frequency), where both tones drive the cells sufficiently, multi-unit responses show marked suppression to B-tones with increasing presentation rates, frequency separation and tone duration (Fishman et al., 2001, 2004). These results suggest that the percept of a single, coherent auditory stream is facilitated when a neuronal population responds strongly to both tones. In contrast, if the two tones activate different populations of neurons, two distinct streams are usually perceived. This idea became known as the population-separation model (Fishman et al., 2017). The idea behind this model is that downstream secondary auditory, frontal and/or parietal areas read out the number of peaks occurring on the topographic representation of cortical activity to determine if there is enough evidence for stream segregation. However, the role of tonotopic separation of the neuronal responses during stream formation remains debated as the tonotopic gradient is a one-dimensional cortical map that might be fully activated by complex sounds. Moreover, according to this model, the relative timing of A and B tones should not impact the behavioural percept. However, recent work has demonstrated that temporally coherent tone elements cause a one-stream percept

(Elhilali et al., 2009a). Thus, other grouping cues (e.g. temporal coherence (Elhilali et al., 2009a), inharmonicity (Fishman and Steinschneider, 2010), spatial location (Wood et al., 2019)) might be better suited to explain perceptual grouping in real-life scenarios. In addition, studies by Fishman and colleagues (Fishman et al., 2001, 2004) presented tone sequences passively without any behavioural feedback. Without behavioural measures, it is speculative to evaluate neuronal responses to passively presented stimuli as neural correlates of a percept. Despite those limitations, the evidence suggests that spatially segregated responses contribute to stream formation as the ability of neurons to signal temporal incoherence across frequencies is mostly determined by the frequency selectivity of the cells (Fishman et al., 2001, 2004). In addition, alternating sequences also cause greater effective tonotopic separation compared to synchronously presented tones (Fishman et al., 2017). Despite all this evidence, information processing based on the population separation model is limited in complex acoustic scenes with multiple overlapping sources, masked sounds and low signal to noise ratios. Hence, tonotopic separation of neuronal responses might only be a facilitating factor instead of the primary cause for stream segregation.

Another model of stream formation is known as the temporal coherence model (Shamma et al., 2011). It states that temporally coherent elements, e.g. correlated tones that vary or remain constant together, will be bound into one perceptual stream (Shamma and Micheyl, 2010; Shamma et al., 2011). Hence, the relative timing of sound elements is critical for perceptual grouping. Psychophysical evidence strongly supports the role of temporal coherence as a fundamental grouping cue (Elhilali et al., 2009a; Teki et al., 2013). Neurophysiological studies exploring the underlying neuronal mechanisms show that neurons in the primary auditory cortex respond, on average, in a comparable way to alternating and synchronous sequences regardless of the spatial separation along the tonotopic axis (Elhilali et al., 2009a). However, given that the behavioural percept might vary considerably between these conditions, it appears that the average rate does not represent a correlate of auditory streaming. However, synchronous sequences lead to rapid changes in spectrotemporal receptive fields, the response amplitudes and the spiking

correlations of neurons when the sound has behavioural relevance, i.e. the animal is attending the sequence (Lu et al., 2017). These results suggest that cells tuned to the features of the sound sequence form rapid and mutually excitatory (cooperative) connections. In contrast, suppressive (competitive) connections arise when neurons are incoherently driven by alternating tones (Lu et al., 2017; Shamma et al., 2011). Complex frequency tuning properties of neurons might facilitate information processing in this context. Cortical neurons with multipeak tuning have been shown to be widely distributed across the auditory cortex (Feng and Wang, 2017; Kadia and Wang, 2003; Kikuchi et al., 2014; Sutter and Schreiner, 1991). In addition, a mechanism based on temporal coherence would explain the enhanced spatial separation of responses to alternating tone sequences (Fishman et al., 2017) and is also in line with context-dependent adaptation of cortical tuning (David et al., 2012; Fritz et al., 2003, 2005). Thus, temporal coherence might be a more flexible and powerful grouping cue for stream segregation which is applicable to every sound dimension and therefore allows parsing of complex auditory scenes.

However, up to this point, neuronal grouping mechanisms still need to be demonstrated for complex stimuli. As indicated before, stream segregation/figure-ground segregation in real-life scenarios will most likely integrate information from many grouping cues. Even though the mentioned studies above are important for the understanding of basic encoding strategies of the auditory cortex, they are barely translatable to natural acoustic scenes. In contrast to ABAB tone sequences, figure-ground stimuli can better model natural acoustic scenes. The next section reviews the basics of figure-ground processing and introduces the motivation for the current electrophysiological study.

1.5.4 Neuronal basis of figure-ground segregation

Across sensory modalities, figure-ground segregation refers to the perception of an object, the figure, against a nondescript background. In vision, perceptual grouping relies predominantly on physical cues like luminance, colour, shape and motion (Roelfsema, 2006). Visual figure-ground segregation has been associated with feedback interactions between higher cortical centres and primary visual cortex (Lamme and Roelfsema, 2000; Poort et al., 2016; Self et al., 2019), with layer-specific differences in cortical processing (Self et al., 2013). Evidence points towards recurrent

figure processing, starting with an early, bottom-up, pre-attentive boundary detection and later top-down, attention-dependent region filling and ground suppression (Lamme, 1995; Poort et al., 2012; Roelfsema et al., 2002; Self et al., 2019). The higher-order percept of the figure seems to be introduced into the lowest levels of visual processing, causing an evolving representation of the figure that is further driving an iterative process (Jones et al., 2015; Poltoratski et al., 2019; Self and Roelfsema, 2015).

In the auditory domain, our understanding of stream segregation grew during the last decades, however, the actual neuronal processing behind figure-ground segregation remains mostly unknown. Predictions about the nature of figure-ground processing can be made based on previous fMRI, MEG and EEG studies, which found a clear effect of figure saliency, localised to non-primary auditory cortex (O'Sullivan et al., 2015a; Teki et al., 2011, 2016; Tóth et al., 2016). It appears that, in contrast to the visual system, auditory figure-ground processing does not involve primary cortical areas. Neurons in non-primary auditory cortex, however, seem to detect figures and respond with changes in their firing rate. Based on the aforementioned studies, the magnitude of this change should be based on the saliency of the figure.

Detection mechanisms of objects/figures have to be different across modalities. In contrast to the visual system, spatial location is not mapped within the auditory cortical organisation but has to be computed based on interaural time- and level differences (van der Heijden et al., 2019; King et al., 2007; Ortiz-Rios et al., 2017; Recanzone, 2001). Since there is no clear-cut spatial boundary between different auditory objects, perceptual grouping based on spatial location has to work in a very different way. In addition, boundary detection of the figure in the frequency space is not meaningful as natural stimuli can cover distinct frequency bands across the spectrum, all of which need to be grouped into one percept. As indicated before, it appears that temporal coherence is the main grouping cue across different sound dimensions. A visual correlate of temporal coherence may be the tracking of a moving object over time which is already beyond simple figure segregation and requires higher order computations (Born et al., 2000; Kourtzi et al., 2002; Recanzone et al., 1997). Hence, there seem to be major differences in the segregation and

grouping aspects between visual and auditory system. However, recent studies imply that the underlying object-processing strategy might be comparable. Similar to visual figure-ground processing in the LGN (Jones et al., 2015), one study found correlates of simple stream segregation as early as the cochlear nucleus (Pressnitzer et al., 2008). Here, the data closely resembled activity patterns of the auditory cortex (Fishman et al., 2004, 2012), with multi-second adaptation of neuronal responses. This indicates that some aspects of auditory scene analysis might be processed at a very early stage in the auditory processing hierarchy. Strong corticofugal feedback connections between auditory cortex and subcortical structures exist (Bajo and King, 2013; Bajo et al., 2006, 2010; Homma et al., 2017), providing the auditory system with similar conditions for recurrent figure processing across the processing hierarchy. In addition, cortical information flow in the auditory cortex resembles visual processing with two distinct processing streams (Kaas and Hackett, 1999; Rauschecker and Tian, 2000; Romanski et al., 1999a) that mostly rely on stepwise transfer of auditory information (Scott et al., 2015). This implies some sort of hierarchy, even though cortical areas are heavily interconnected and receive parallel input (Hackett et al., 2014; Kikuchi et al., 2010; Scott et al., 2015, 2017).

1.5.5 Attentional modulation of perceptual organisation

Directing attention describes the process of prioritising and preferentially processing sensory input (Desimone and Duncan, 1995; Harris and Thiele, 2011). Attention can be shifted towards a spatial location (Posner, 1980), an object (Duncan, 1984) or feature (Rossi and Paradiso, 1995). Selective attention refers to the filtering of behavioural irrelevant information that is an essential step between perception and behavioural action (Desimone and Duncan, 1995; Driver, 2001; Johnston and Dark, 1986). Auditory attention can be voluntarily directed towards a sound object in a scene (endogenous, central, top-down). On the other hand, sound-based saliency (exogenous, reflexive, bottom-up) can draw the attentional focus towards highly salient sound objects, which is important for analysing the most important elements of a complex, natural scenes (Fritz et al., 2007b; Posner, 1980; Shinn-Cunningham, 2008).

The default mode of analysing auditory scenes is to listen to one object at a time, while other objects remain in the perceptual background (Shinn-Cunningham, 2008). This requires foreground-background decomposition or figure-ground analysis.

It is debated whether attention is involved in the process of sound segregation and perceptual grouping or if it is only needed for object/stream selection. Some evidence suggests that stream segregation can happen without attentional involvement (Sussman et al., 2007). Hence, primitive segregation is thought to be a bottom-up, pre-attentive, stimulus-driven mechanism, based on automated change detection (Bregman, 1990; Moore and Gockel, 2012). However, there is strong evidence for attentional involvement during perceptual grouping (Shamma et al., 2011). Cognitive load and sensory distractors have been shown to reduce stream segregation (Carlyon et al., 2001, 2003) and figure-ground segregation (Molloy et al., 2018). Top-down, schema-based segregation also requires prior knowledge or expectations to process and organise incoming information (Bregman, 1990; Moore and Gockel, 2012). Familiarity of sound sequences has also been shown to improve performance (Bey and McAdams, 2002).

As indicated above, cortical activity fluctuates which has been associated with changes in cognitive functions such as arousal and attention (Harris and Thiele, 2011). Thus, attentional engagement might enhance sound segregation compared to passive experimental paradigms by involving more relevant neuronal networks. On the neuronal level, evidence points towards a close interaction between stimulus-driven sensory processing and top-down attention. The neural representation of the percept is likely a feature-driven mechanism whereas attentional processes shape the arising responses. Temporally coherent/locally synchronous neural activity optimises information processing of task-salient features of sound through flexible changes of the receptive field or shorter response latencies (Elhilali et al., 2009b; Fritz et al., 2005; Lu et al., 2017). The flow of information between sensory and higher order cortical areas is behaviourally gated. Rapid, selective, persistent and task-related changes might facilitate the transfer of information to the next stage (Fritz et al., 2010). Shamma, Elhilali and Micheyl (2011) summarise these findings in a model of stream formation that covers the following steps:

1) Initial feature analysis of auditory information that creates a multi-dimensional cortical representation that results in a rich set of neuronal responses that encode all acoustic properties of the sound.

2) Coherence analysis by correlating temporal output of feature-selective neurons. Sorting neuronal responses based on the degree of coherence gives rises to distinct perceptual streams.
3) Top-down selective attention is complementing the feed-forward processing by inducing rapid receptive field plasticity, prioritising processing for coherent sound features and modulating neuronal ensemble activity (e.g. phase coherence).

In addition to explaining stream formation for complex stimuli like speech or music, this model also reconciles feature and object-based auditory attention, by using selectively attended features as an anchor point to bind other temporally coherent object features with it (Shamma et al., 2011).

In summary, segregation of sound objects is probably possible in a pure bottom-up manner but in real-life scenarios it is likely that the state of arousal and attentional involvement play an important role in perceptual organisation. In complex environments, interactions between object formation and object selection is key for perception.

1.6 Key problems addressed

The behavioural detection performance to auditory figures has been previously addressed in humans (O'Sullivan et al., 2015a; Teki et al., 2011, 2013; Tóth et al., 2016). Furthermore, areal involvement (Teki et al., 2011), temporal dynamics (O'Sullivan et al., 2015a; Teki et al., 2016; Tóth et al., 2016) and susceptibility to cognitive load (Molloy et al., 2018) of figure-ground processing has been studied in human subjects before. However, systematic, invasive investigation of spiking activity is currently not feasible in humans. Thus, to further understand the basic coding mechanisms of the auditory system, an animal model needs to be established that allows examination of neuronal mechanisms in the auditory cortex.

This thesis contains three different studies: In study 1, I tested the figure detection performance of macaque monkeys. Based on the evidence reviewed above, I hypothesized that macaques can detect auditory figures and that figure coherence will have an impact on the perceptual detection performance. Behavioural results were comparable to the studies mentioned above, showing increasing performance to

higher salient figures. In addition, the results suggested marked effects of figure coherence on reaction times.

Study 2 demonstrated the areal organisation of figure-ground processing using fMRI in awake subjects. Based on the current knowledge of the cortical organisation of macaques, I hypothesized that the areal organisation between humans and non-human primates is comparable. Indeed, I showed that the involvement of the secondary, non-primary auditory cortex, localised to the anterolateral belt and parabelt, was similar to human cortical activation in non-primary auditory regions. Taken together, behavioural and imaging findings suggest the rhesus macaques as an excellent animal model for auditory figure-ground analysis.

Finally, in the third study, I investigated the impact of auditory figures on cortical multi-unit activity across the auditory hierarchy. According to the literature, neuronal responses should encode figure coherence. Moreover, figure-ground modulation of neuronal responses should be sensitive to attentional involvement. In addition, I hypothesized that neuronal responses will point towards distributed and recurrent figure processing, similar to what has been found in the visual system. The results suggested that a subset of cells across core and belt areas respond to auditory figures. Furthermore, I have identified differences in the encoding of perceptual saliency across cortical fields.

Taken together, the work presented in this thesis has established a new animal model for auditory figure-ground analysis and has shed some light on the underlying neuronal mechanisms in response to complex, stochastic figure-ground signals.

Despite the organisation of the chapters in this thesis, fMRI data were acquired first in naïve animals, before behavioural training and assessments took place. Subsequently, electrophysiological experiments were conducted (Figure 1.9). Only one subject participated in all three studies.

Figure 1.9: Timeline of experimental data acquisition.

Functional imaging data were acquired first using the original SFG stimuli (Teki et al., 2011) and naïve animals. Subsequently, subjects were trained in the figure detection task with redesigned stimuli (see methods section, chapter 2). After behavioural assessment, electrophysiological recordings were performed in two trained monkeys using the same redesigned stimuli.

2 Behavioural correlates of figure-ground segregation in macaque monkeys

2.1 Summary

Segregating the key features of the natural world within crowded visual or sound scenes is a critical aspect of everyday perception. In order to abstract relevant information from sound, the auditory system has to detect, extract, segregate and group information from the same sound source. In this chapter, I show that the detection performance of macaques to auditory figures increases with figure coherence, similar to what has been reported for humans. I also show marked coherence effects on reaction times. These results, in combination with other evidence, qualifies macaques as an animal model for auditory figure-ground segregation that allows investigation of the neuronal mechanisms in a way that is not possible in humans.

2.2 Introduction

The parsing of sensory scenes into objects is crucial for survival. Natural acoustic scenes are highly complex, where different sound sources can emit overlapping spectrotemporal cues. Before a percept of a sound object can arise, the auditory system has to detect related sound elements, segregate them from the background and group them together (Bizley and Cohen, 2013).

Figure-ground segregation describes the process of extracting auditory objects from a complex scene. This process is utterly important for survival (e.g. hearing a predator amid other natural sounds) as well as everyday human scenarios (e.g. having a conversation at a cocktail party or hearing a car approaching). Since it is not yet possible to get insight into the neuronal coding of auditory scenes in humans, an animal model might reveal the underlying central mechanisms the brain uses to organise complex acoustic scenes. Macaque monkeys are a good choice since they share similar audiograms (Jackson et al., 1999), detection of tones in quiet (Heffner and Heffner, 1986), detection of tones in noise (Dylla et al., 2013) and similar pitch perception (Joly et al., 2014a) with humans. Macaques also show homologous organisation of

the auditory cortex that allows comparison with that of humans (Baumann et al., 2013; Leaver and Rauschecker, 2016). Thus, rhesus monkeys might be an ideal candidate to investigate the cortical processing if auditory scene analysis.

I have tested the (auditory) figure detection performance of two rhesus macaques. In this study, I have used a redesigned version of the Stochastic Figure-Ground stimulus (Teki et al., 2011). The SFG stimuli consisted of multiple randomly generated frequency elements, where a foreground object, arising from the grouping of different frequency elements over time, can only occur if coherently repeated elements are present in a number of frequency channels (Figure 2.1). The correlated tones stand out against the background of uncorrelated tones which causes a perceptual pop-out effect. This stimulus captures a high-level acoustic process that requires the auditory system to integrate over frequency and time in order to extract the target (Shamma et al., 2011). Since this stimulus is devoid of species-specific meaning, such as speech, it can be used to compare detection performance across species.

The ability to detect auditory figures has been sufficiently tested in humans. All previous studies report that that detection performance increase with figure coherence (O'Sullivan et al., 2015a; Teki et al., 2011, 2013; Tóth et al., 2016). Thus, more salient auditory figures can be easier detected. Furthermore, figure duration is another important factor that has an impact on the detection probability. The more figure chords are presented to the subjects, the higher the hit rate (Teki et al., 2011, 2013; Tóth et al., 2016). Psychometric curves demonstrate that the detection performance plateaus at a high level for longer figures with high coherence level (O'Sullivan et al., 2015a; Teki et al., 2011). The representation of the figures is very robust. Performance always increases with duration and saliency, even when the chord duration is reduced, the chords are interrupted by noise or when the transition between background and figure is remove and only the figure part of the signal is presented. Modelling of the behavioural data is consistent with a grouping mechanism based on temporal coherence between the frequencies comprising the figure (Teki et al., 2013).

None of the previous studies that acquired behavioural data of auditory figure-ground segregation report the response times of their subjects. Hence, no conclusions can be made about the relation between the temporal dynamics of central figure processing and object perception.

In this chapter, I report behavioural data from two macaque monkeys that perform a figure detection task. The detection performance to auditory figures will give an indication if macaques are a suitable animal model for figure-ground segregation. Results shown in this chapter are published (Schneider et al., 2018b).

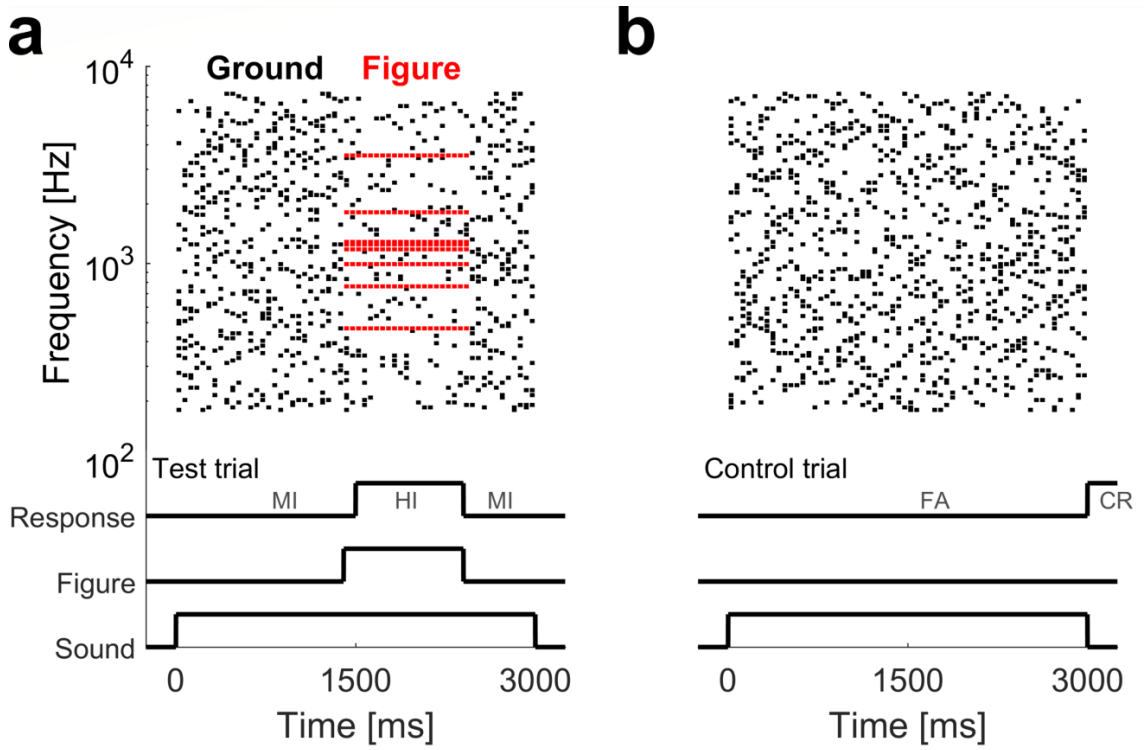

Figure 2.1: Schematic depiction of two example SFG stimuli used for behavioural experiments.

(**a**) Schematic spectrogram of an example SFG stimulus. Stimulus consist of 60 randomly generated chords. Each chord contains 15 frequency elements. Stimulus duration was 3000ms. Onset times were pseudorandom in a range from 300ms to 2000ms. Line plot below indicates the 900ms long behavioural response window for the displayed stimulus as well as the behavioural outcome for touch bar release inside (HI – hit) and outside (MI – miss) of this time window. (**b**) Example control stimulus without figure. No touch bar release until sound presentation finished (CR – correct rejection), otherwise trial is counted as false alarm (FA).

2.3 Methods

All procedures performed in this study were approved by the UK Home Office (project license: 70/7976) and by the Animal Welfare and Ethical Review Body at Newcastle University. All experiments comply with the UK Animals Scientific Procedures Act (1986) on the care and use of animals in research,

with the European Communities Council Directive on the protection of animals used in research (2010/63/EC) and with the US National Institute of Health Guidelines. We support the principles of the consortium on Animal Research Reporting of In Vivo Experiments (ARRIVE).

2.3.1 Animals

Two adult macaques (*Macaca mulatta*), M1 (Male, 11yrs, 11kg) and M2 (Female, 5yrs, 6kg), were used in this study. Animals were kept under fluid-controlled conditions. The controlled access to water was within ranges which do not negatively affect animal's physiological or psychological welfare (Gray et al., 2016).

2.3.2 Stimuli

Stochastic Figure-Ground stimuli were created at a sampling rate of 44.1 kHz with MATLAB (The Mathworks Inc., Natick, MA). Signals consisted of a sequence of 50ms long chords, defined as a sum of randomly selected multiple pure tone elements, selected from a pool of 129 evenly spaced frequencies (1/24 octave between successive frequencies) on a logarithmic scale between 179 Hz and 7246 Hz. The onset and offset of each tone were shaped by a 10ms raised-cosine ramp. Some stimuli included a sequence of repeated elements in a specified number of frequency channels ('figure'). The remaining signals comprised randomly shuffled elements only ('control').

Stimuli contained 60 chords (3s in duration) and had a fixed number of elements per chord (n = 15). In contrast to the earlier studies (Teki et al., 2011, 2013, 2016) and to the stimuli used in the imaging experiments (see chapter 3), we did not change the number of frequency elements per chord to ensure consistent broadband power across chords. Since coherent elements were not added on top but incorporated into the existing stream of chords, we eliminated any sound level cues at the onset of the figure. The coherence level of the figure was defined as the number of frequency channels that constitute the figure and was varied between 4, 6, 8, 10 and 12 elements. Figure onset times were randomised between 0.3 and 2 seconds. Figure and control stimuli were presented in a randomised order.

2.3.3 Behavioural training

All subjects were naïve to the behavioural detection task. By means of positive reinforcement, we established a bar release – reward relationship. A fixed target stimulus was then paired via operant conditioning. This target was a plain figure (duration: 1000ms, coherence: 10) without any distractor elements. After monkeys responded proficiently to the sound, we introduced the SFG background tones. The signal to noise ratio was gradually decreased by increasing the sound level of the ground signal. Subsequent to this introductory phase, the ground sound intensity was set to a fixed level (65dB) whereas the figure sound level was incremented to give subjects an extra cue to the target. These sound level increments were then gradually decreased until subjects could detect the figures without any intensity cues. As a last step, figure coherence was manipulated in order to assess the animal's performance. The entire training period took about 8 months of daily training.

2.3.4 Experimental design

To make inferences about the figure-ground perception of macaques in crowded acoustic scenes, we designed a figure detection task as a Go/No-Go paradigm. For behavioural testing, macaques sat in a primate chair (Christ Instruments) and initiated trials by touching a touch bar, placed in front of them. Two free-field speakers (Yamaha Monitor Speaker MS101 II), located at approximately 45 degree to the left and right of the animal (distance: ~65cm from ear), delivered the stimuli at ˜65dB SPL via an Edirol UA-4FX external USB-Soundcard. The experiment was controlled with a custom-made MATLAB (2015b) script, including PsychToolbox 3.0 functions through a LabJack U3-HV interface.

Before each session, a new set of stimuli was created (n = 1000). For each trial, a stimulus file was randomly drawn from this pool of stimuli. If the monkey responded correctly during the figure presentation period ('Hit'), a fluid reward was administered through a gravity-based reward system. The amount of reward was dependent on the reaction time of the respective trial. Faster responses led to higher reward volumes. Inter-trial intervals (ITI) were set to 1s. In case the monkeys missed to respond to a target, no reward was administered

but a 3s penalty time-out was imposed in addition to the ITI. Stimuli were terminated as soon as the subjects responded or after the target sound ended. Trials with stimuli containing a figure comprised 60% of all trials. The remaining 40% were catch trials (control condition) in which only the ground stimulus was presented. In these catch trials, subjects needed to hold the bar for the entire length of the stimulus (3s). In case of a correct rejection of the trial (bar not released), a fixed reward was given. The amount of juice earned on those trials was greater than during detection trials, since monkeys had to hold the bar up to two seconds longer. Similar to the miss of a figure, false alarms resulted in no reward but a 3s penalty time-out in addition to the ITI. Each behavioural session lasted around two hours (average number of trials per session: M1 = 1000, M2 = 873). Data were acquired, saved and analysed using MATLAB.

2.3.5 Statistical analysis

For data analysis, signal detection theory was applied. In total, data from 52 behavioural sessions were included in this analysis (M1: 23, M2: 29). Performance was evaluated based on hit and false alarm rates, which are the basis for d' calculation, a measure of discriminability between responses to different stimuli. Computation of d' values was done by using the formula below:

$$d' = Z(Hit\ rate) - Z(False\ alarm\ rate)$$

where Z is the z-transform of hit/false alarm rate respectively, which is defined as the inverse of the cumulative Gaussian distribution (MATLAB: norminv). Due to the changed stimulus design, control stimuli no longer have a varying number of added elements. Hence, false alarm rates were pooled across conditions. Since d' values take hit rates as well as false alarm rates into account, they provide a measure of all possible responses to both detection- and catch trials. Mean d' values across all sessions for each coherence condition were the basis for the assessment of the behavioural performance. Trials with responses below 0.4s after stimulus onset were excluded from the analysis (M1: 1.67%, M2: 1.38%). Reaction times (relative to figure onset) were corrected for sound output latency of the operating system. 95% confidence intervals were calculated via bootstrapping (MATLAB: bootci, 5000 repetitions). For statistical testing, data of both subjects were pooled as we were

interested in the overall trend of the responses. Effects of coherence were tested across sessions with a repeated measures ANOVA for d-prime values, mean reaction times and responses variability, respectively. Response variability was assessed with the Coefficient of Variation (CV), a ratio of standard deviation divided by the mean. Normal distribution was evaluated with a one-sample Kolmogorov-Smirnov test. A Mauchly sphericity test assessed if the assumption of sphericity was violated. If that was the case, a conservative lower bound correction was applied to the degrees of freedoms and p-values of the repeated measures ANOVA. To assess the effect of figure coherence on reaction times on a trial-by-trial basis, a general linear mixed effects model was constructed with figure coherence defined as a fixed effect. Random intercepts were included for each subject and session to account for repeated measurements: Reaction time ~ Coherence + (1 | Session) + (1 | Monkey). This model was tested against an intercept-only model without coherence as a factor by means of maximum likelihood ratio tests: Reaction time ~ 1 + (1 | Session) + (1 | Monkey).

2.4 Results

Behavioural experiments tested if macaques can segregate complex auditory figures. Two monkeys were trained to perform an active figure detection task. The stimulus design was altered compared to earlier studies (see methods, Teki *et al.*, 2011, 2013, 2016) as the false alarm rate of subjects increased with the number of added elements, suggesting responses to sound level changes instead of temporal coherence (Figure 2.2). Each chord of the redesigned stimuli contained 15 frequency elements, effectively removing sound level cues between chords. Because control stimuli did no longer differ between conditions, the cumulative false alarm rate was used for analysis (Figure 2.3).

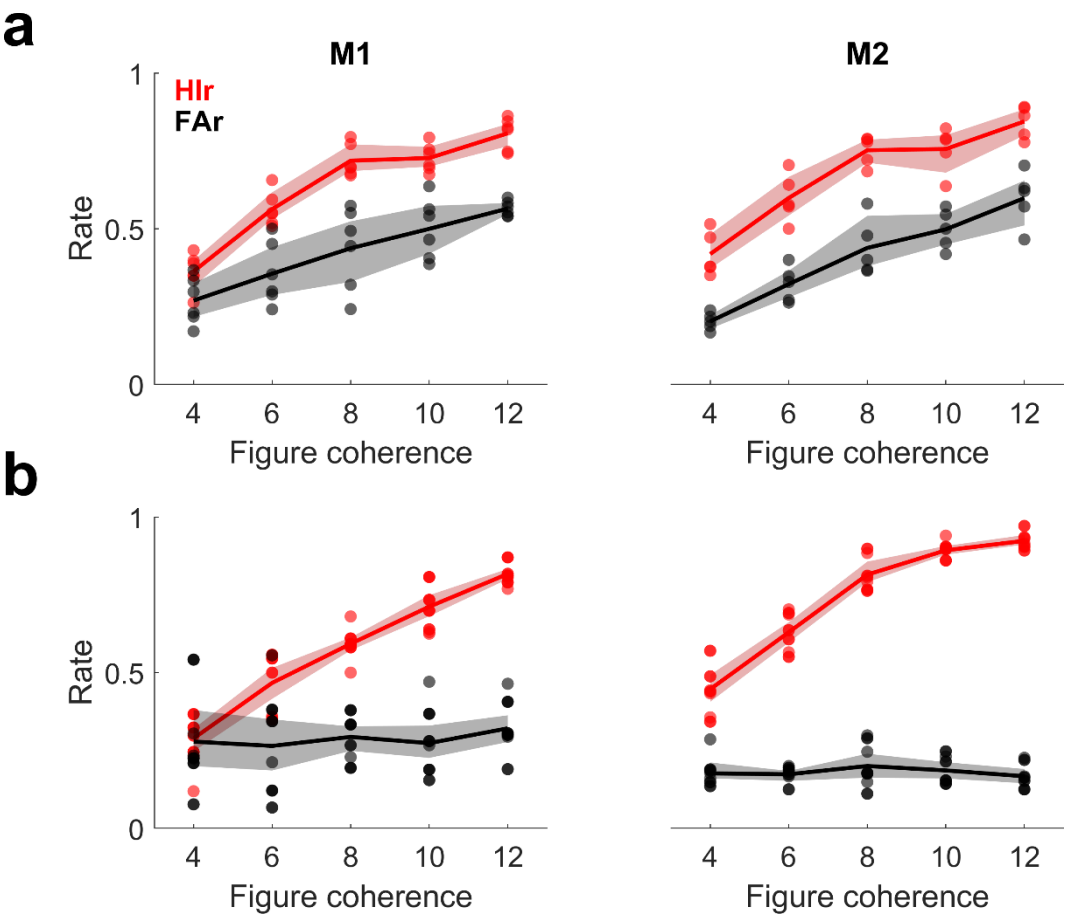

Figure 2.2 Behavioural performance in example training sessions with original (a) and redesigned (b) SFG stimuli.

Colour-coded data points show hit rate (HIr, red) and false alarm rate (Far, black) for each included training session. Average rate is shown as solid line. Shaded areas correspond to the bootstrapped 95% confidence interval of the mean.

Proficiency on the task was indicated by the mean hit rates to the most salient condition with figures comprising 12 coherent frequency elements (M1: 0.86, M2: 0.92). The reaction time (RT) distributions showed a clear peak for both subjects (Figure 2.3 a, M1: Median RT: 0.5344 s; M2: Median RT: 0.4603s), indicating competent detection of auditory figures. Average performance measurements are summarised in Table 2.1 and Table 2.2 for M1 and M2, respectively. Hit rates (Figure 2.3 b) increased as a function of figure coherence (Repeated measure (RM) ANOVA, Sphericity violation ($\chi^2(9) = 32.34$, $p = 1.72e^{-4}$), Lower bound correction applied: $F_{1, 50} = 933.03$, $p = 5.26e^{-34}$), suggesting that the number of coherent elements has an impact

on figure detection. False alarm rates were low across sessions in both subjects (M1: 0.22±0.044, M2: 0.12±0.042), indicating that monkeys could competently withhold responses to stimuli without a figure. D-prime values mirrored the trend of hit rates with increasing values for more salient figures (Figure 2.3 c). The main effect of figure coherence was significant (RM-ANOVA: $F_{4, 200} = 743.66$, $p = 8.13e^{-119}$), confirming that figure coherence is an important factor during perceptual organisation. Sphericity was not violated in this test ($\chi^2(9) = 8.28$, $p = 0.51$). For hit rates and d-prime values, post hoc-tests (Table 2.3) revealed that all conditions are significantly different in M1 whereas M2's results only differed up to a coherence level of 10. This suggests that M2's detection performance plateaus from there. Furthermore, we found decreasing reaction times and response variability with increasing saliency of the figures (Figure 2.3 d and e, Mean RT: RM-ANOVA, Sphericity violation ($\chi^2(9) = 78.51$, $p = 3.19e^{-13}$), Lower bound correction applied: $F_{1, 50} = 253.89$, $p = 3.12e^{-21}$; Response variability (Coefficient of variation): RM-ANOVA, Sphericity violation ($\chi^2(9) = 43.2$, $p = 1.98e^{-06}$), Lower bound correction applied: $F(1, 50) = 39.22$, $p = 8.53e^{-08}$). Post hoc tests (Table 2.3) showed clear differences for higher coherence levels in both monkeys that seem to break down in M1 but not M2. The coefficient of variation, used to measure response variability, showed no effects in M1 but a clear significant downwards trend in M2 for coherence level higher than 4 elements. A general linear mixed effects model performs significantly better when figure coherence is included as predictor ($\chi^2(1) = 2018$, $p = 0$) and confirms the impact of figure coherence on reaction times on a trial-by-trial basis ($t_{(21015)} = -46.026$, $p = 0$).

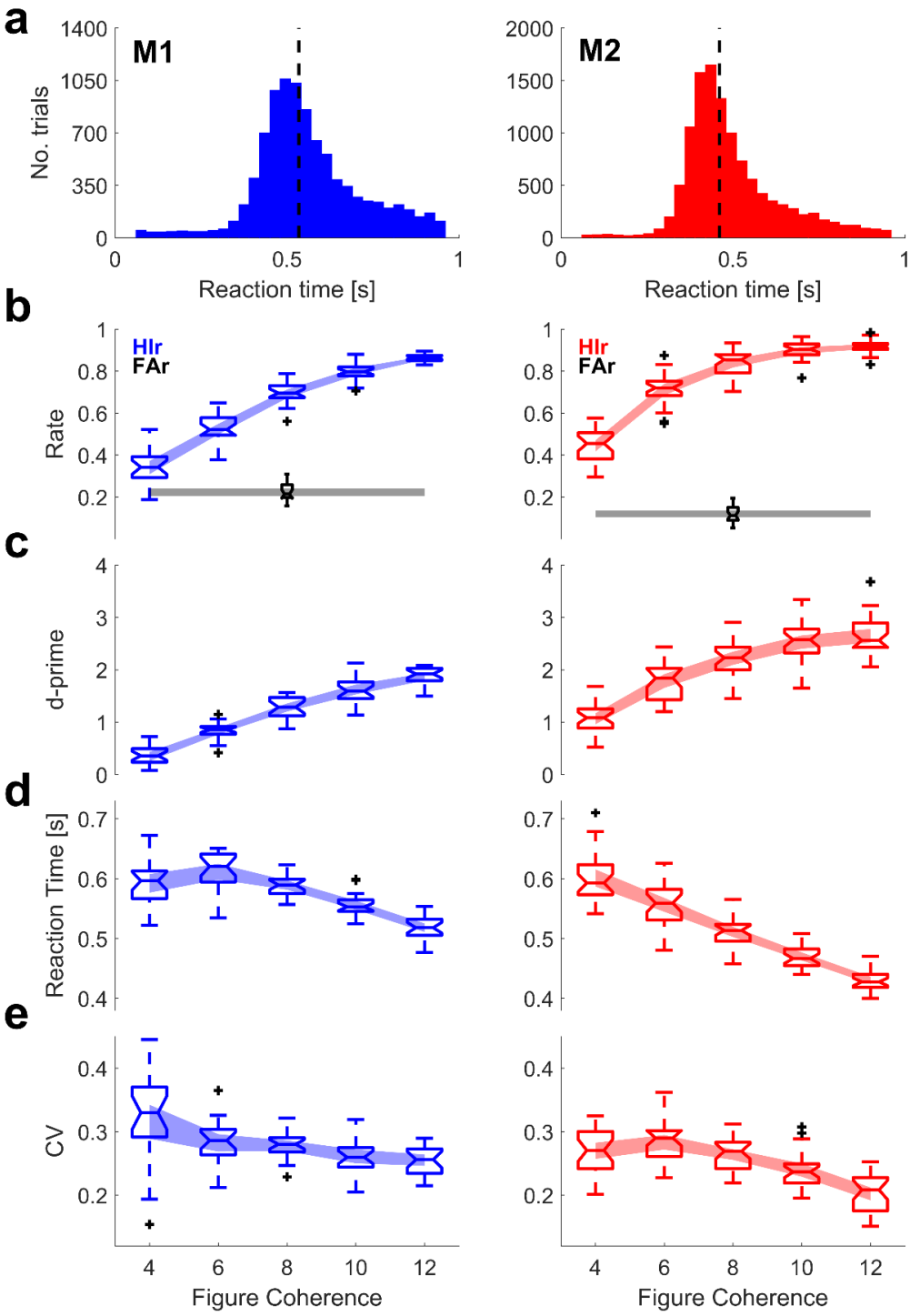

Figure 2.3: Behavioural performance of M1 (left) and M2 (right)

(**a**) Reaction time distribution of all trials pooled across coherence level and sessions. Black dashed line corresponds to median. (**b**) Distribution of hit (HIr, coloured) and false alarm rate (FAr, black) across sessions shown for each coherence level. Shaded area corresponds to bootstrapped 95% confidence interval of the mean. (**c-e**) Same conventions as in (**b**). (**c**) Distribution of d-prime. (**d**) Distribution of mean reaction times. (**e**) Distribution of coefficient of variation (CV) for reaction times.

Table 2.1: Summary of M1's behavioural results in the figure-detection task. The mean is shown for each tested coherence level.

Coherence level	4	6	8	10	12
Mean hit rates	0.3405	0.5248	0.6935	0.7979	0.8633
Mean false alarm rates			0.2228		
Mean d-prime	0.3496	0.8334	1.2815	1.6164	1.8703
Mean reaction time [s]	0.5914	0.6129	0.5888	0.5548	0.5179
Mean Coefficient of Variation	0.3188	0.2822	0.2778	0.2605	0.2550

Table 2.2: Summary of M2's behavioural results in the figure-detection task. The mean is shown for each tested coherence level.

Coherence level	4	6	8	10	12
Mean hit rates	0.4444	0.7118	0.8381	0.9012	0.9173
Mean false alarm rates			0.1186		
Mean d-prime	1.0654	1.7820	2.2206	2.5317	2.6289
Mean reaction time [s]	0.5993	0.5563	0.5110	0.4696	0.4294
Mean Coefficient of Variation	0.2701	0.2828	0.2637	0.2386	0.2022

Table 2.3: Summary of conducted post-hoc tests between adjacent coherence levels for both monkeys. Significantly different conditions are highlighted in red for a Bonferroni-corrected alpha of 0.0125.

Tested coherence level		4 vs 6	6 vs 8	8 vs 10	10 vs 12
Hit rates	**M1**	$2.70e^{-05}$	$2.70e^{-05}$	$2.70e^{-05}$	$4.03e^{-05}$
	M2	$2.56e^{-06}$	$4.33e^{-06}$	$1.95e^{-05}$	0.0314
D-prime	**M1**	$2.70e^{-05}$	$2.70e^{-05}$	$2.70e^{-05}$	$4.59e^{-05}$
	M2	$2.56e^{-06}$	$4.33e^{-06}$	$2.37e^{-05}$	0.0369
Reaction times	**M1**	0.0126	0.0126	$2.70e^{-05}$	$2.70e^{-05}$
	M2	$2.60e^{-05}$	$1.08e^{-05}$	$2.85e^{-06}$	$2.56e^{-06}$
Coefficient of Variation	**M1**	0.0074	0.7151	0.0208	0.2871
	M2	0.1444	0.0058	$3.75e^{-04}$	$3.80e^{-05}$

Reaction time distributions differed distinctly between tested conditions. Not only did the number of hit trials vary (see hit rate, Figure 2.3 b) but the shape of the distribution seemed to change distinctly with coherence (Figure 2.4). A ROC analysis was used to compare reaction time distributions of all coherence conditions. This revealed marked differences between the reaction time distributions across coherence conditions for both monkeys (Figure 2.4). Here the distribution of response times to figures with twelve coherent elements was tested against all other coherence level. This was done in order to have the most proficient response characteristic as reference for all other distributions. The higher the difference in coherence level, the more the reaction time distributions differed (Figure 2.4 inset), indicating less overlap and a different shape between the RT distributions. This, in combination with the increasing mean RT and response variability for lower coherence levels (Figure 2.3 d), suggests that the confidence with which monkeys can detect less salient figures decreases rapidly. Humans can detect figures with four elements given an adequate figure duration (Teki et al., 2011, 2013). Thus, macaques might be less sensitive to auditory figures as they seem to require more coherent elements to reach similar detection performance to humans. Despite this, the behavioural performance indicates that macaques can perceive auditory figures in noisy acoustic scenes and that behavioural performance increases with target to masker ratio, as is the case for human listeners.

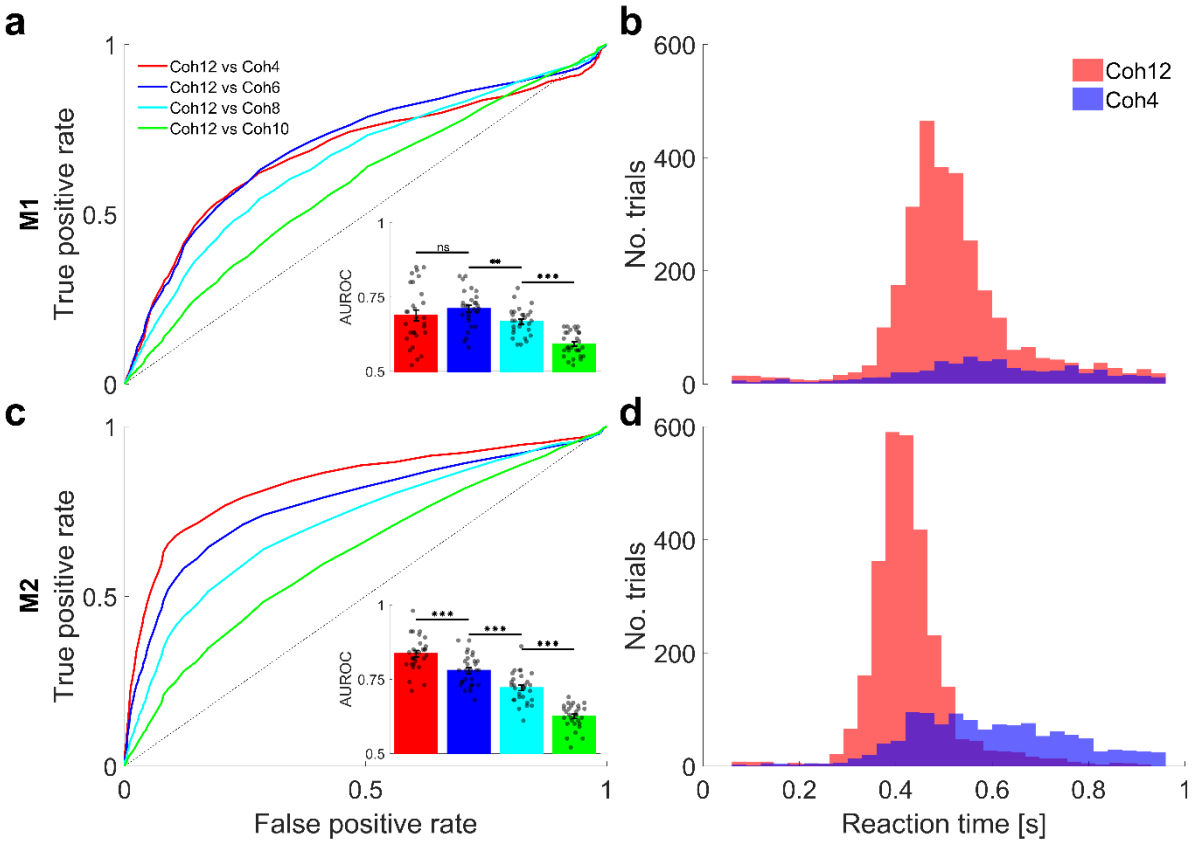

Figure 2.4: Receiver operating characteristic curve

(**a**) M1. Average ROC curve (colour-coded) contrasts pairs of reaction time distributions of different coherence level. Inset shows area under ROC curve for each condition. Same colour code applies. Test results between conditions indicated above bars: * p< 0.05, ** p < 0.01, *** p< 0.001. (**b**) Reaction time distribution for trials of different coherence level. Trials of highest (12, red) and lowest (4, blue) tested coherence level are shown. (**c+d**) M2. Same conventions as (**a+b**).

2.5 Discussion

Data in this chapter demonstrate that macaques do perceive auditory figures in a similar way to humans. In line with results from previous human studies (O'Sullivan et al., 2015a; Teki et al., 2011, 2013, 2016; Tóth et al., 2016), I show that figure detection performance depends on the amount of temporally coherent frequency elements. Figure coherence not only impacts the rate of detection but also influences the timescale and confidence of the arising percept. Here, I have shown that lower figure coherence corresponds, on average, to longer reaction times and higher response variability. Reaction time differences between coherence levels likely represent the varying timescale over which evidence is accumulated and percepts develop, however, it could also be the case that motor preparation takes longer if subjects are uncertain.

During the perceptual learning phase, I initially presented the original stimuli used by Teki and colleagues (Teki et al., 2011), where each chord comprised a pseudorandom number of frequency elements plus extra (coherent or shuffled) elements on top of the existing ground stream. Using this stimulus design, animals strongly responded to changes in sound level (Figure 2.2). This caused the false alarm rates to rise as a function of added elements. Because of this, stimuli in this chapter (and chapter 4) have been redesigned with an equal number of frequency elements for each chord to match the overall broadband power across time. This approach proved to be very effective as false alarm rates remained flat across coherence conditions.

By incorporating the figure into a fixed number of elements per chord, I effectively change the target to masker intensity ratio. Thus, I cannot make any claims whether the segregation effect shown above are caused by temporal coherence of changed statistical properties of the stimulus. Changes in the regularity of auditory patterns have been shown to affect behaviour and neural responses strongly (Barascud et al., 2016; Barczak et al., 2018; Sohoglu and Chait, 2016a, 2016b; Southwell et al., 2017). However, the cause of segregation is not the main focus of this thesis, where I try to define the underlying changes in the neural network during figure-ground segregation.

Recognition of regular auditory patterns has been shown to occur after 1.5 cycles (Barascud et al., 2016), which indicates that the auditory system should be able to detect auditory patterns after 2 chords. However, the chords of the stochastic figure-ground stimulus are shorter than the auditory patterns used in the studies investigating regularity processing. MEG figure-ground effect latencies have been reported to occur after 150-200ms for 25ms chords (Molloy et al., 2018; Teki et al., 2016). For all experiments in this thesis, I have used chords that were 50ms long. Hence, effect latencies might be higher than previously found. In line with reported MEG effect latencies (Teki et al., 2016), reaction time distributions shown here indicate that the neural response latency depends on the figure coherence level.

There is a performance difference between monkeys. Generally, M2 detection performance is better with higher hit rates and faster reaction times. This could be due to the age difference between monkeys. Alternatively, subjects could be at different stages in their learning curve. Even though monkeys seem to be fully trained, M2‘s results indicate more competent behaviour.

3 Areal organisation of figure-ground processing

3.1 Summary

Previous functional imaging work in humans suggests the involvement of non-primary auditory cortex during stimulus-driven figure-ground segregation, but systematic investigations of these mechanisms on the cellular level are not feasible in humans. There is, however, substantial evidence that suggests macaque monkeys are a good animal model for auditory scene analysis. In this chapter, I report functional MRI data that was acquired during passive presentation of Stochastic Figure-Ground stimuli to naïve macaques. A Figure vs Control contrast shows significantly changed blood oxygenation in anterolateral, non-primary auditory cortex in response to auditory figures. Similar to the behavioural results, these findings are in line with reports of human brain activity in response to the same type of stimulus and therefore enable us to investigate figure-ground processing at the neuronal level.

3.2 Introduction

Natural scenes are filled with a multitude of objects whose sensory representation might overlap in different stimulus dimensions (e.g. spatial location, frequency, time). Figure-ground analysis is critical to making sense of the natural world. This is a particularly challenging problem in the auditory system, where different sound objects emanating from the same spatial location, have to be dynamically decoded using spectrotemporal features that are difficult to segregate from noisy backgrounds (Bregman, 1990; Shamma et al., 2011).

I have assessed the perception (chapter 2) and neural representation of auditory figure-ground stimuli in the macaque monkey. As established in chapter 2, macaques show strong physiological and perceptual similarities to humans (Joly et al., 2014a). Macaques also show homologous organisation of the auditory cortex that allows comparison with that of humans (Dylla et al., 2013; Heffner and Heffner, 1986; Jackson et al., 1999; Joly et al., 2014a). Furthermore, the organisation of the auditory cortex seems to be homologous to humans (Baumann et al., 2013; Leaver and Rauschecker, 2016). The aim of this study was to define the areal organisation of figure-ground analysis in auditory cortex.

We used a stimulus in which a figure emerges from a noisy background, similar in design to earlier studies in humans (Teki et al., 2011, 2013). The paradigm captures a high-level acoustic process that requires grouping over frequency and time in complex sounds devoid of species-specific meaning, such as speech. The SFG stimuli consisted of multiple randomly generated frequency elements, where a foreground object, arising from the grouping of different frequency elements over time, can only occur if coherently repeated elements are present in a number of frequency channels. A series of human behavioural and modelling experiments is consistent with a grouping mechanism based on temporal coherence between the frequencies comprising the figure (Teki et al., 2013). Human imaging experiments using fMRI (Teki et al., 2011) and MEG (Teki et al., 2016) demonstrated activity in non-primary auditory cortex and intraparietal sulcus (IPS) that accompanied perceived figures. Whether the same would hold neurobiologically in an animal model is unknown.

Previous research suggests that the emergence of auditory figures causes changes in brain responses that scale with figure coherence (O'Sullivan et al., 2015a; Teki et al., 2016). Thus, higher figure coherence causes higher response amplitudes. Furthermore, it has been shown that cognitive load greatly affects the ability to segregate figures, which manifests itself in a lower evoked MEG field strength (Molloy et al., 2018).

Data in this part of the thesis was acquired and analysed in collaboration with Pradeep Dheerendra[1]. In this chapter we report fMRI data that were acquired during the passive presentation of SFG stimuli in awake macaque monkeys. The figure coherence was set to a highly salient level. No task requirements were forced onto subjects, hence no cognitive load was involved during this paradigm. The resulting brain response is used to assess the areal organisation of stimulus-driven figure-ground segregation. Results shown in this chapter are published (Schneider et al., 2018b).

[1]Work done by Pradeep Dheerendra (PD) and Felix Schneider (FS). Data acquisition M1: PD; Data acquisition M2: FS; Data pre-processing: PD; Data analysis: PD + FS

3.3 Methods

All procedures performed in this study were approved by the UK Home Office (project license: 70/7976) and by the Animal Welfare and Ethical Review Body at Newcastle University. All experiments comply with the UK Animals Scientific Procedures Act (1986) on the care and use of animals in research, with the European Communities Council Directive on the protection of animals used in research (2010/63/EC) and with the US National Institute of Health Guidelines. We support the principles of the consortium on Animal Research Reporting of In Vivo Experiments (ARRIVE).

3.3.1 Animals

Two adult macaques (*Macaca Mulatta*), Monkey 1 (Male, 11yrs, 9kg) and Monkey 2 (Male, 11yrs, 11kg), participated in these experiments. Animals were kept under fluid-controlled conditions. Fluid control was within ranges which do not negatively affect animal's physiological or psychological welfare (Gray et al., 2016).

3.3.2 Stimuli

Stochastic Figure-Ground stimuli with a sampling rate of 44.1 kHz were created with MATLAB (The Mathworks Inc., Natick, MA). Signals consisted of a sequence of 50ms long chords, defined as a sum of multiple pure tone elements that were not harmonically related. The onset and offset of each tone was shaped by a 10ms raised-cosine ramp. Some stimuli included a sequence of repeated elements within several frequency channels ('figure'). The remaining signals comprised randomly shuffled elements only ('control').

Stimuli consisted of 120 chords (6s in duration) in total. For each of these chords a random number of 5 to 15 tonal ground elements was drawn from a pool containing 129 evenly spaced frequencies (1/24 octave between successive frequencies) on a logarithmic scale between 179 Hz and 7246 Hz. The number of bands that contribute to the figure ('coherence') was set to a constant value (n = 10). SFG stimuli used for imaging had extra coherent or shuffled elements added on top of the ground signal after two seconds for the following 40 chords (2s in duration). Stimulus design is consistent with previous studies (Teki et al., 2011, 2013, 2016). Figure and control stimuli were presented in a randomised order.

3.3.3 Experimental design

For functional imaging scans, macaques were transferred into a custom-made, MRI-compatible scanner chair. During the session, awake animals were head restrained by means of an implanted head post. The details of the surgical procedures are described in Thiele et al., (2006). Single-shot echo-planar images were acquired with an actively shielded, vertical 4.7T MRI scanner (BrukerBiospec 47/60 VAS) equipped with a Bruker BGA-38S gradient system with an inner-bore diameter of 38 cm (BrukerBioSpin GmbH, Ettlingen, Germany). One volume transmit coil and two 4 channel receiver coils were used. A sparse imaging paradigm was applied to avoid the interfering effect of the high intensity noise generated by the MRI scanner. Shimming was performed with the MAPSHIM algorithm (Kanayamay et al., 1996) which measures B0 field inhomogeneity to apply first and second order corrections to it. The applied sequence was a GE-EPI with 2x GRAPPA acceleration with the following parameters: TR = 10s, TA=2011ms, TE= 21ms, flip angle (FA) of 90º, receiver spectral bandwidth of 200 kHz, field of view (FOV) of 9.6 x 9.6 cm^2, with an acquisition matrix of 96 x 96, an in plane resolution and slice thickness of 1.2 mm and 32 slices. The TR duration was sufficient to avoid recording the BOLD response to the gradient noise of the previous scan. Per scan 360 volumes were acquired (of which 90 volumes baseline/silence).

In total, 135 stimuli per condition (control i.e. ground only or figure) were created and presented in pseudo-randomized manner (see Figure 3.1). The same stimuli were used for all scans and all subjects. Sounds were presented using Cortex software (Salk institute) at an RMS sound pressure level (SPL) of 75 dB via custom adapted electrostatic headphones based on a Nordic NeuroLab system (NordicNeuroLab, Bergen, Norway). These headphones feature a flat frequency response up to 16 kHz and are free from harmonic-distortion at the applied SPL. SPL was verified using an MR-compatible condenser microphone B&K Type 4189 (Bruel&Kjaer, Naerum, Denmark) connected by an extension cable to the sound level meter Type 2260 (same company). A structural scan was acquired at the end of each functional scanning session. Anatomical MR images are T1-weighted (T1w) images, consisting of a 2D magnetization-prepared rapid gradient-echo (MPRAGE)

sequence with a 180° preparation pulse, TR = 2000ms, TE = 3.74ms, TI = 750ms, 30° flip angle, receiver bandwidth = 50KHz, an in-plane resolution of 0.67 x 0.67 mm^2 with a slice thickness of 0.6mm (Voxel size: 0.67x0.67x0.6mm). Structural scans covered the same field of view as the functional scans.

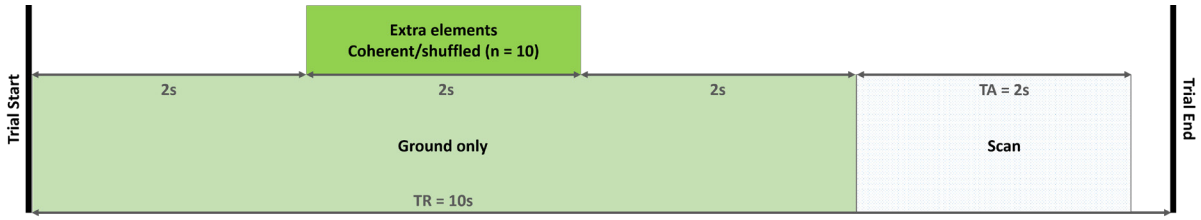

Figure 3.1: Schematic of imaging paradigm.

Stimulus presentation (green, 6s) comprised ground signal (light green) and extra elements (dark green), which were either coherent or shuffled. Each chord of the ground signal contained a randomly drawn number of tonal elements (n = 5 to 15). Time of repetition was 10s, image acquisition time was 2s long. For 90 scan volumes, no stimulus was presented (silence condition).

3.3.4 Statistical analysis

MR images were first converted from the scanner's native file format into a common MINC file format using the Perl script pvconv.pl (http://pvconv.sourceforge.net/). From MINC format, it was converted to NIfTI file format using MINC tools. Imaging data were then analysed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/ - Wellcome Trust Centre for Neuroimaging).

In the pre-processing steps, the volumes within a session are realigned and resliced to incorporate the rigid body motion compensation. Next, image volumes from multiple sessions were combined by realigning all volumes to the first volume of the first session. This data was then spatially smoothed using a Gaussian kernel with full-width-at-half-maximum (FWHM) of 3 mm. We chose such a low FWHM as the analysis in this chapter is based on single-subject data that do not require correction for variability in the cortical macrostructure. A standard SPM regression model was used to partition components of the BOLD response at each voxel. The two conditions, figure and control, were modelled as effects of interest and convolved with a hemodynamic boxcar response function. Next, the time series was high pass filtered with a cut-off of

1/120 Hz to remove low-frequency variations in the BOLD signal. Finally, this data was adjusted for global signal fluctuations, also known as global scaling to account for differences in system responses across multiple sessions. A general linear model analysis (Friston et al., 1994) of the combined sessions included the motion parameters, the voxel-wise response estimates and the regression coefficients. The t-values for two contrasts (Figure vs Control, Sound vs Silence) were calculated. We performed single subject inference in these two subjects. Data were thresholded at $p<0.001$ (uncorrected for multiple comparisons across the brain). Results from monkey M2 survived $p<0.05$ (family wise error corrected across the brain) and it showed a pattern similar to that presented here. Data were co-registered and displayed in standard space (D99, Saleem and Logothetis, 2012).

The total number of included scans for the two monkeys was as follows (M1: 12, M2: 10). Sessions with obvious large imaging artefacts, high signal differences between hemispheres and/or insufficient baseline activity in the sound vs silence contrast were not included in the analyses (M1: 6, M2: 4 sessions).

3.3.5 Probabilistic maps

The applied probabilistic maps are an estimate of functional areas of the auditory field in standard space (D99) (Saleem and Logothetis, 2012) based on the tonotopic gradients of six macaques (not included in this study), with the probabilistic map threshold set at 0.5, equivalent to at least 3 animals overlapping in the location of the auditory cortical fields. Isofrequency lines from mirror reversals between core (A1/R) and belt areas (ML/AL) were extended laterally to approximate the border between rostral and caudal parabelt. Core-belt boundaries were estimated by tone vs. bandpass noise responses. Belt-parabelt boundaries were assigned based on the breakdown of the tonotopic gradient. For each functional area, all voxels have an assigned value, representing the probability that a given voxel fell within this field. By thresholding these maps to 0.5, we made sure that each voxel is in at least 50% of the scanned population within the boundaries of the respective functional field.

3.4 Results

We acquired fMRI data from two naïve monkeys during passive exposure to the original SFG stimuli (Teki et al., 2011, 2013). Using the same stimulation enabled us to compare the BOLD modulation between humans and macaques. Functional imaging data were recorded before the same animals were trained in the active figure detection task (chapter 2) with an adapted stimulus design (to avoid responses to sound intensity changes, see Figure 2.2). In this section, we contrast cortical responses in voxel-space to different stimulus categories: first, we assess the engagement of auditory cortex for Sound vs Silence. Subsequently, we compare Figure vs Control conditions.

As expected, sound evoked activation ($p < 0.001$, uncorrected for multiple comparisons) engages nearly the entire auditory cortex (Figure 3.2, Table 3.2 and Table 3.4). We find the strongest activation in the primary core regions and the adjacent lateral belt. The contrast strength appears to be weaker in medial cortical regions (CM, MM, RM, RTM). Hence, this confirms 1) stimulus presentation that both subjects perceive to a sufficient degree, 2) that Stochastic Figure-Ground stimuli strongly drive the auditory cortex and 3) that there is no functional auditory cortical abnormality that could perceptually impair the animal. Subcortical regions like the IC and MGN show strong sound evoked BOLD modulation (Figure 3.2). The percentage of significantly sound-driven voxels in those structures cannot be determined as the only way to define the respective region of interest is by taking the Sound vs Silence contrast into account.

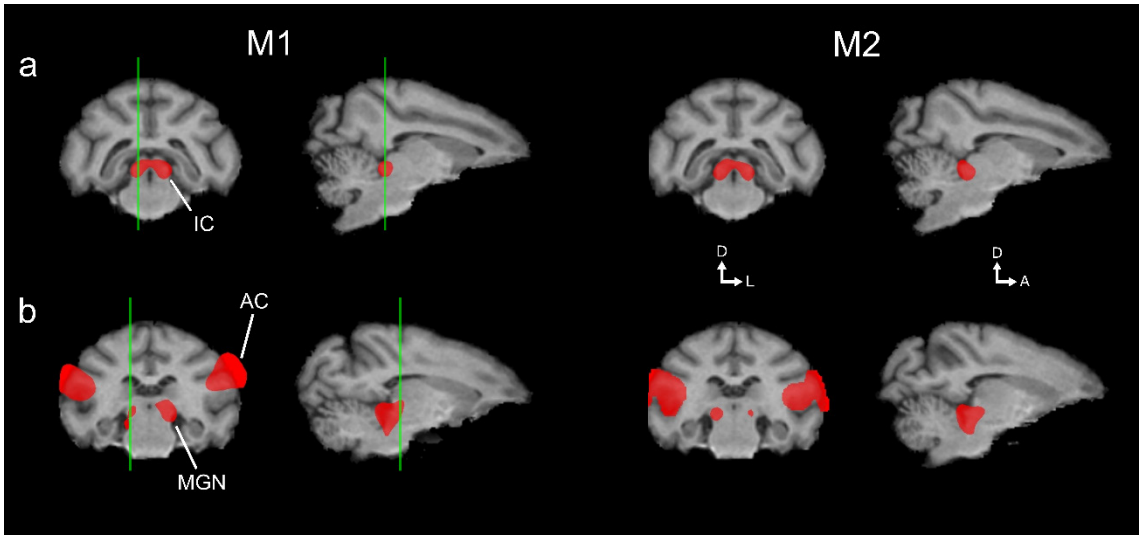

Figure 3.2: Sound vs Silence contrast overlaid on standard brain.

Sound vs Silence contrast for M1 (left) and M2 (right) illustrated in red. (**a**) Sound evoked BOLD modulation of inferior colliculus (IC) for threshold of T > 15. (**b**) Sound evoked BOLD modulation of medial geniculate nucleus (MGN) and auditory cortex (AC) for a threshold of T > 4. Green lines denote the location of sagittal and coronal slices, respectively. The same slices are shown for both monkeys.

A contrast for Figure vs Control ($p < 0.001$, uncorrected, Figure 3.3) revealed significant BOLD changes along the convexity of the superior temporal gyrus and in the rostral parts of the superior temporal plane, demonstrating bilateral involvement of higher-level auditory regions rostro-laterally to the auditory core. The observed pattern of significant BOLD signal changes is consistent between subjects. In order to assign a functional area to the peak BOLD response, we illustrate the Figure vs. Control contrast alongside the probabilistic functional maps of auditory cortical fields, derived from tonotopic gradients of six macaques. This comparison reveals that the main activation during a perceived figure is located in the rostral parabelt (RPB) and the rostro-lateral belt (RTL) for both monkeys (Figure 3.4, Table 3.1 and Table 3.3). In the rostral parabelt of M2, more than 40% of voxels show a significantly different brain response in both hemispheres (Table 3.5). Area RTL has an even higher proportion of significantly activated voxels, with more than 80% of voxels in each hemisphere responding to auditory figures. Similar to M2, M1 shows significantly different responses in RTL and RPB but only unilaterally and to a

much lower degree. We find about 16% significantly activated voxels in the right RPB and about 6% of voxels in left RTL. Significant clusters also extend to the rostral superior temporal gyrus (STGr, no probabilistic map available) in both monkeys. Furthermore, in M2, significantly different BOLD modulation was found in the rostro-temporal core (RT), the anterolateral belt (AL) and the caudal parabelt (CPB). Very weak activation can be seen in the middle lateral belt (ML) and the rostral core (R) unilaterally. Generally, the result demonstrate that T-values ramp up towards the rostro-lateral parts of the auditory field (Figure 3.4 and Table 3.3). No figure-ground modulation was found in subcortical nuclei, suggesting a purely cortical mechanism of sound segregation. Based on those results, I conclude that figure-ground processing happens in rostral parts of the auditory ventral stream in untrained animals.

Table 3.1: Coordinates of maximum Figure vs Control contrast in M1 and M2 for each hemisphere. Data are displayed relative to interaural line.

Subject	Hemisphere	X [mm]	Y [mm]	Z [mm]
M1	L	29	13	14
	R	-29	18.5	12
M2	L	29	21.5	10.5
	R	-27	17	12

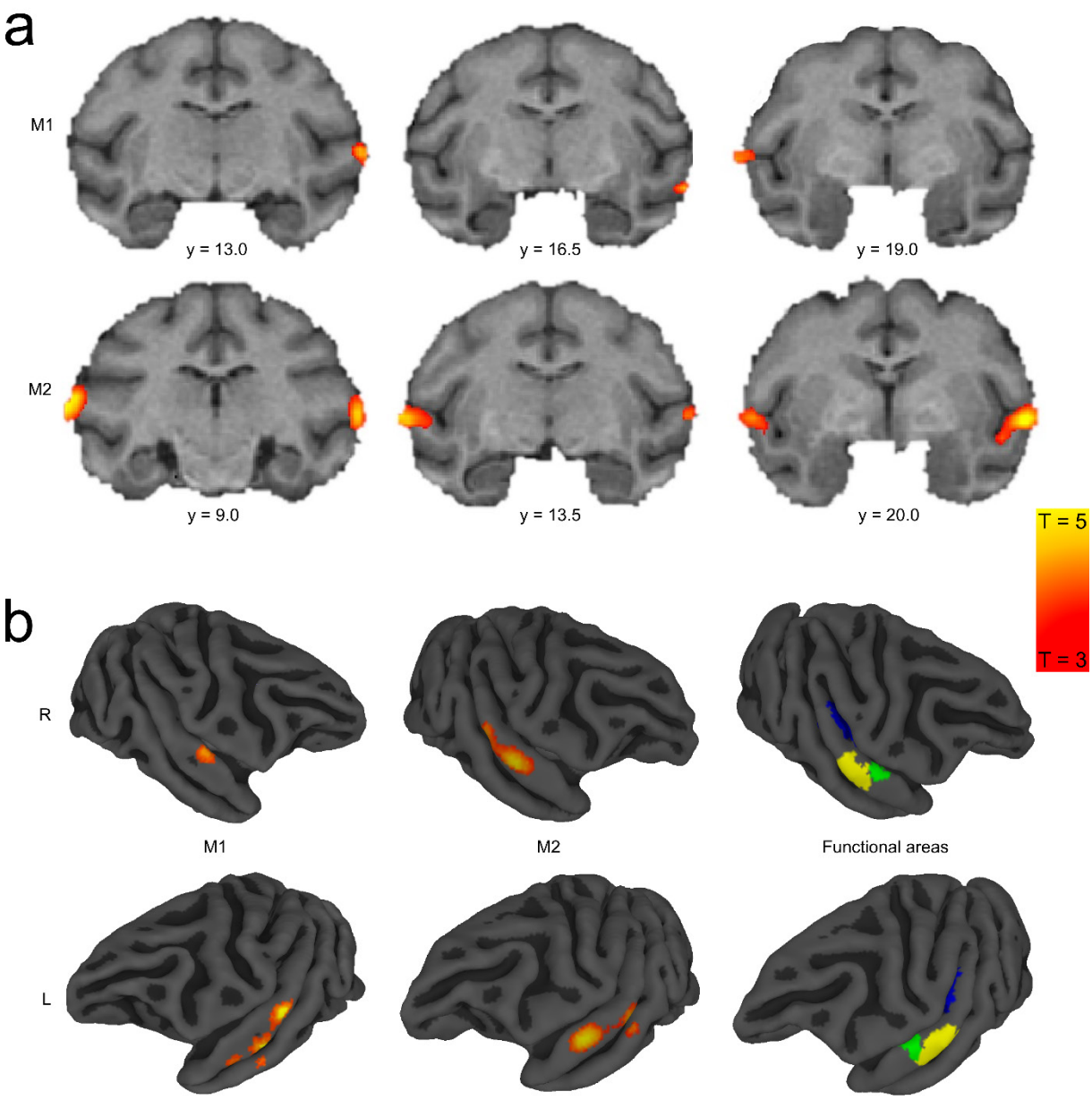

Figure 3.3: Figure vs Control contrast overlaid on standard brain.

(a) Series of coronal MR images from posterior (left) to anterior (right) with Figure vs Control contrast overlay (3 < T < 5) for subject M1 (above) and M2 (below). Position of slices relative to interaural line in [mm] is indicated below slices. (b) Figure vs Control contrast overlaid on right (above) and left (below) brain surface of M1 (left) and M2 (middle). Colour-coded probabilistic maps of functional areas overlaid on standard brain (right). Functional areas: A1 - Primary auditory cortex (blue), RPB – Rostral parabelt (yellow), RTL - Lateral rostrotemporal area (green). Brain extraction (**a**) done by Fabien Balezeau. 3D rendering of results (b) done by Michael Ortiz-Rios.

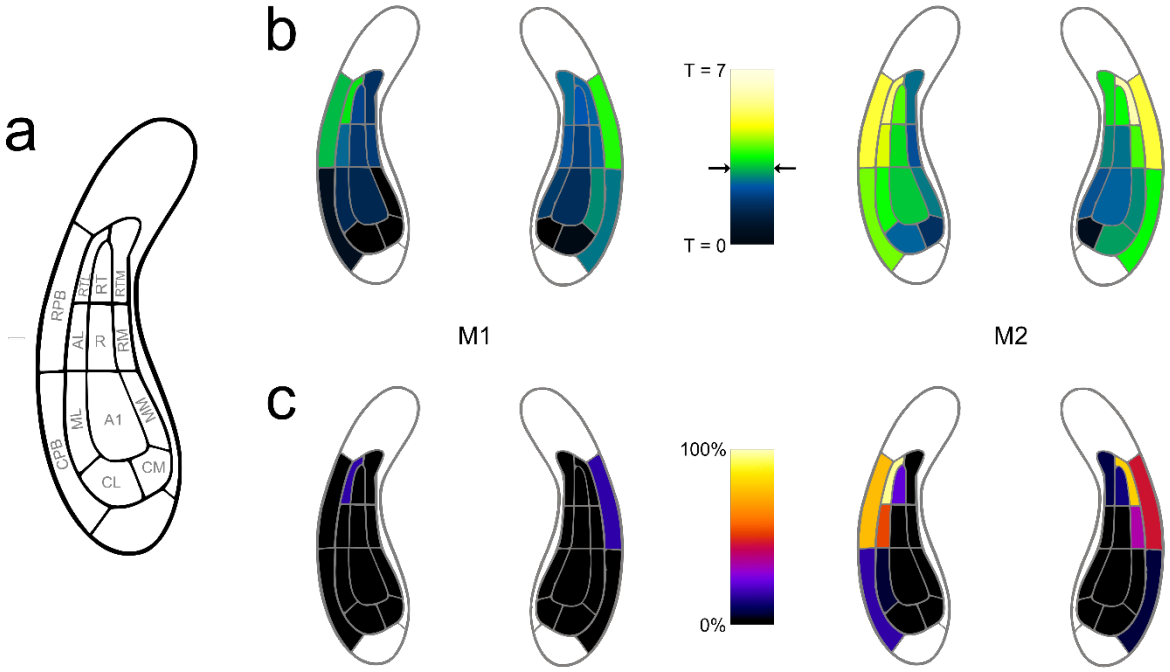

Figure 3.4: Involvement of auditory areas in figure-ground processing.

(a) Map of macaque auditory cortex. (b) Maximum T-values for Figure vs Control contrast overlaid on auditory fields for M1 (left) and M2 (right). Data based on probabilistic maps. Significance level of T = 3.09 ($p < 0.001$) is indicated by black arrows. (c) Fraction of significant voxels per auditory field. Cortical maps were colour-coded by Fabien Balezeau.

Table 3.2: Maximum T-value of the Sound vs Silence contrast shown for each cortical area in left (LH) and right hemisphere (RH).

Area \ Subject	M1		M2	
	LH	RH	LH	RH
A1	42.99	29.04	45.33	46.66
AL	31.81	22.05	30.5	33.08
CL	20.41	16.56	9.49	21.16
CM	15.56	4.88	4.69	15.3
CPB	31.49	31.17	13.8	16.27
ML	44.3	23.34	25.54	21.32
MM	6.7	7.56	38.62	26.35
R	25.62	25.4	40.99	38.7
RM	1.01	8.63	12.27	9.94
RPB	27.15	36.19	20.09	22.47
RT	10.34	8.54	28.94	25.42
RTL	21.53	6.67	27.73	21.94
RTM	4.56	2.83	14.5	10.63

Table 3.3: Maximum T-value of the Figure vs Ground contrast shown for each cortical area in left (LH) and right hemisphere (RH).

Area \ Subject	M1		M2	
	LH	RH	LH	RH
A1	0.81	0.94	2.93	1.89
AL	1.98	1.88	5	4.01
CL	-0.65	0.23	1.91	2.53
CM	-0.55	-0.22	1	0.28
CPB	0.29	2.06	4.31	3.43
ML	0.82	0.23	3.55	2.29
MM	-0.99	1.14	2.16	1.46
R	1.16	1.29	3.15	2.07
RM	1.23	1.59	1.6	2.24
RPB	2.78	3.66	5.68	5.71
RT	1.43	1.73	4.07	3.49
RTL	3.29	1.9	5.99	6.44
RTM	0.91	2.04	1.99	3.21

Table 3.4: Fraction of significantly activated voxel for Sound vs Silence contrast shown for each cortical area in left (LH) and right hemisphere (RH). Threshold was set to T > 3.09 (p < 0.001, uncorrected for multiple comparisons across the brain).

Subject / Area	M1		M2	
	LH	RH	LH	RH
A1	96.85	94.4	97.11	100
AL	100	100	100	87.05
CL	94.52	76.83	22.58	74.49
CM	12.3	2.38	1.37	29.76
CPB	100	100	100	96.72
ML	100	100	100	98.02
MM	5.98	12.23	88.04	54.26
R	62.58	100	97.35	96.58
RM	0	23.12	15.93	6.13
RPB	100	89.91	100	66.47
RT	60.11	56.32	95.74	73.16
RTL	100	23.11	100	99.56
RTM	15.31	0	31.63	7

Table 3.5: Fraction of significantly activated voxel for Figure vs Ground contrast shown for each cortical area in left (LH) and right hemisphere (RH). Threshold was set to T > 3.09 (p < 0.001, uncorrected for multiple comparisons across the brain).

Subject / Area	M1		M2	
	LH	RH	LH	RH
A1	0	0	0	0
AL	0	0	45.51	30.57
CL	0	0	0	0
CM	0	0	0	0
CPB	0	0	13.08	1.97
ML	0	0	2.14	0
MM	0	0	0	0
R	0	0	0.66	0
RM	0	0	0	0
RPB	0	15.58	75.32	41.99
RT	0	0	19.15	6.32
RTL	7.53	0	99.32	80.89
RTM	0	0	0	1

3.5 Discussion

This chapter shows correlates of stimulus-driven figure-ground segregation in the auditory cortex of macaque monkeys. Pradeep Dheerendra and I demonstrate that cortical responses to auditory figures engage non-primary auditory cortex, mostly along the anterolateral part of the auditory field. Contrast strength appears as low in the posteromedial part of the auditory field and then gradually ramps up towards the anterior belt and parabelt regions. This finding indicates homologous processing of figure segregation to humans (Teki et al., 2011).

Electrophysiological experiments have established that visual object detection, segmentation and recognition is caused by recurrent cortical processing from early subcortical stages (Jones et al., 2015) throughout the cortical visual hierarchy (Poort et al., 2016; Roelfsema, 2006; Self et al., 2019) to the inferior temporal cortex (DiCarlo et al., 2012). Here, we show a pattern of cortical involvement in the non-primary anterolateral belt and parabelt that is in line with earlier imaging studies (Molloy et al., 2018; Teki et al., 2011, 2016) and reflects neural correlates of the perceptual organisation of the sound scene. Neuronal correlates of auditory scene analysis have previously been found in primary auditory cortex for two-tone paradigms (Fishman et al., 2001, 2004, 2017; Lu et al., 2017), however, we demonstrate a system involving circumscribed parts of the rostro-lateral belt and parabelt cortex (Figure 3.4), at a high level in the cortical hierarchy in macaques (Hackett et al., 2014; Kaas and Hackett, 2000; Scott et al., 2015) for complex figure-ground segregation. Subcortical structures as well as the primary auditory cortex do not show a modulated BOLD response, which suggests fundamentally different figure-ground processing compared to the visual system. The functional organisation reported here corresponds to activation along the ventral auditory processing stream (Rauschecker and Tian, 2000; Romanski et al., 1999a). In line with these results, previous evidence suggests that the most anterior regions of the ventral processing stream represent a complete acoustic signature of auditory objects (Leaver and Rauschecker, 2010).

In our paradigm, a segregated figure is an auditory object that consist of repeating frequency elements. We have argued that the detection of the SFG stimulus requires a mechanism that can integrate across different frequency bands, in order to detect any temporal coherence between them (Teki et al., 2013). A possible mechanism of figure-ground analysis is based on single neurons in higher level cortical areas, with inputs arising from a combination of units in primary auditory cortex that exhibit either narrowband or multi-peaked tuning. In support of this, neuronal responses to sounds with harmonically related components have been described in primate core (Feng and Wang, 2017) and belt areas (Kikuchi et al., 2014). However, a neuronal mechanism that supports the present results requires single neurons to respond to multiple frequencies that do not have a simple mathematical relationship to each other.

One imaging study suggests harmonic and non-harmonic multipeak tuning occurs in large parts of the ventral auditory stream (Moerel et al., 2013). However, these responses are averaged over thousands of cells. Thus, fMRI does not allow for disambiguation of neuronal mechanisms from a population code.

The necessary broadband tuning for figure-responsive units is well described in the belt cortex (Kikuchi et al., 2014; Rauschecker, 2004). Broadband responses in the parabelt are likely, given that they occur at a higher level in the auditory hierarchy (Hackett et al., 2014; Kaas and Hackett, 2000; Scott et al., 2015), but receptive fields of parabelt neurons have not been extensively characterised (Kajikawa et al., 2015).

From first principles, such neurons might be expected at a higher level in the auditory hierarchy: we therefore predict the existence of such units in the rostro-lateral belt and parabelt. Teki and colleagues found that non-primary cortical activity varied parametrically with figure coherence (Teki et al., 2011). Hence, the auditory cortex responded stronger to higher figure coherence. This indicates that firing rates of individual neurons are sensitive to figure coherence. Thus, it could be the case that the presence of auditory figures is signalled with a rate code, but that the cells in the anterolateral belt and parabelt respond in a relative invariant manner to the individual frequencies of that figure. However, following segregation, the grouping of repeated elements and detection of the figure could cause some form of top-down modulation in upstream brain areas like A1 in the form of neural entrainment (Barczak et al., 2018).

Attention should also be paid to the previously identified involvement of the intraparietal sulcus in stream segregation (Cusack, 2005) and figure-ground processing (Teki et al., 2011, 2016). However, since participants were not asked to make perceptual reports, it remains unclear whether the IPS activity reflects perceptual object processing (Teki et al., 2011). Contrary to these studies, we were not able to show a BOLD response modulation in the IPS, which could partly be due to the cranial implants of the animals that can lead to signal dropouts. On the other hand, the absence of IPS activity in macaques could point towards a higher-level cognitive (e.g. attentional) involvement in human subjects that might not have been present in the tested, naïve monkeys. Trained animals that assign the target stimulus behavioural meaning might display a more diverse response pattern including frontal (Elgueda et al., 2019), parietal (Zhong et al., 2019) or hippocampal (Itskov et al., 2012) brain areas. Alternatively, a species differences in figure-ground processing cannot be ruled out.

One reason why fMRI might not show clear Figure vs Control BOLD changes in primary cortical fields could be due to misalignment of the probabilistic maps. The Sound vs Silence contrast is suspiciously low in medial cortical field (Table 3.2 and Table 3.4) which suggests that the co-registration of the contrast data to the standard brain might have introduced some error. Since this trend can be found bilaterally, the problem seems to be a scale instead of a shift issue. If this is the case, then the areal assignment is biased towards lateral cortical fields. On the other hand, primary auditory cortex involvement was not found in human fMRI studies either (Teki et al., 2011), suggesting that perceptual organisation relies predominantly on non-primary auditory cortex.

In summary, our data suggest that a fundamental form of figure-ground analysis relies on non-primary auditory cortex in the macaque monkey. Our approach has allowed us to investigate grouping over frequency-time space using stimuli that are not species-specific, but that require grouping mechanisms relevant to the extraction of species-relevant sounds from noise. This work predicts specific neuronal responses to figure-ground analysis in rostro-lateral auditory areas and forms the basis for invasive electrophysiological investigation in the macaque that is not possible in humans.

4 Neuronal correlates of figure-ground processing

4.1 Summary

Previous imaging work has shown the involvement of non-primary auditory sites during complex, pre-attentive figure-ground segregation in humans and macaques, even during passive stimulus presentation. In this chapter, I investigate the neuronal basis for figure-ground segregation across the auditory cortex of macaque monkeys. I report figure-ground modulation in single- and multi-units. Specifically, I show that auditory cortical figure-ground processing is not limited to higher cortical areas, but also takes place in the primary core. Thus, A1 neurons can detect temporally coherent elements that do not have a simple mathematical relationship to each other. In anterolateral fields, neuronal responses scale with perceptual saliency of the figure. Modulation latencies in posterior and anterior auditory cortex seem to be similar which indicate simultaneous processing of auditory input across the cortical hierarchy. Figure-ground modulation is present even without behavioural detection, however, differences in neuronal responses also seem to affect object perception.

4.2 Introduction

Figure-ground segregation of natural scenes is essential for directing behaviour, independent of the sensory modality. The coding mechanisms of sensory brain regions during perceptual segregation can be examined with extracellular recordings. Previous investigations into the neuronal correlates of auditory scene analysis have focussed mainly on narrowband 2 tone (ABAB) stimuli (Elhilali et al., 2009a; Fishman et al., 2001, 2004, 2017; Lu et al., 2017). Although these studies have led to new insights regarding the neuronal basis of stream segregation, these predictable, narrowband pure tone sequences do not reflect natural stimulus characteristics with overlapping spectral and temporal components between sound objects. Stochastic figure-ground stimuli model the natural scene because spectrotemporal integration is required to extract auditory figures. In this chapter, I report, to the best of my knowledge, the first extracellularly recorded unit responses to complex figure-ground stimuli.

The literature regarding the psychophysics and functional imaging of auditory scene analysis and figure-ground segregation has been reviewed in previous chapters.

In a nutshell, emerging auditory sources within a complex sound or changes in the statistical structure of the stimulus can be rapidly detected (Sohoglu and Chait, 2016a, 2016b). Human listeners are highly sensitive to auditory figures with detection performance increasing as a function of figure coherence (O'Sullivan et al., 2015a; Teki et al., 2011, 2013; Tóth et al., 2016). Regular and predictable acoustic patterns cause larger brain responses compared to complex, random scenes. In addition to broad activation of the auditory cortical network, frontal areas seem to be involved in regularity processing (Barascud et al., 2016; Sohoglu and Chait, 2016a). Other evidence points towards sustained representation of auditory objects and streams in non-primary auditory cortex (Gutschalk, 2005; Leaver and Rauschecker, 2010). Brain responses localised to non-primary auditory cortex and the intraparietal sulcus have been associated with figure-ground segregation, notably without any involvement of primary core areas (Molloy et al., 2018; Teki et al., 2011, 2016). MEG responses reveal a figure-related negativity that is impaired under high cognitive load suggesting that figure segregation depends on computational resources that are shared across sensory modalities (Molloy et al., 2018).

Whether attention is required for sound segregation is currently still unclear. Cortical responses do not require attention in order to detect deviants in sound streams (mismatch negativity, object-related negativity). Hence, it has been argued that attention might not be a necessary prerequisite of stream segregation (Dyson and Alain, 2004; Shamma and Micheyl, 2010; Sussman et al., 2005). In a series of fMRI and MEG experiments, in which attention was directed towards a visual distractor task, Teki and colleagues found evidence for segregation of temporally coherent figure elements without top-down attention (Teki et al., 2011, 2016), which further backs the claim that sound segregation can happen before attentional selection. In contrast, primary and non-primary cortical responses are strongly affected by selective attention. In situations of competing sounds, attentional modulation enhances the cortical representation of attended streams and suppresses responses to 'irrelevant', not attended information (Bidet-Caulet et al., 2007; Woldorff et al., 1993). Furthermore, noise correlation between primary auditory cortical neurons are highly susceptible to effects of selective attention which decorrelates neurons

with similar feature tuning. This then enhances the resolution for feature processing along the relevant dimension (Downer et al., 2015, 2017). Decorrelation of neuronal tuning curves causes more efficient coding of sensory information that can change spike rates, signal-to-noise ratios and interstimulus variance along relevant sensory dimensions (Lu et al., 2019). Decorrelation has been shown to influence population sensitivity in the visual system as well (Cohen and Maunsell, 2009). In addition, focussed attention to any sound dimension can act as a gateway to perceptual binding of other temporally coherent sound dimensions (Shamma et al., 2011). This implies that selective attention can alter cortical feature processing to facilitate attended features and all associated sound dimensions. Other work suggests a complex top-down/bottom-up interaction between attention and stimulus parameters that could clock neuronal responses and induce receptive field plasticity (Elhilali et al., 2009b). Rapid plasticity of spectrotemporal receptive fields in response to attended auditory stimuli has been demonstrated (Lu et al., 2017). Moreover, even A1 responses shift from pure sensory encoding to behaviourally-driven sound representation when stimuli are attended in a Go/No-Go task design (Bagur et al., 2018). Attentional effects during task engagement can not only be found in A1 but are more and more enhance further up the cortical hierarchy (Elgueda et al., 2019). Further imaging work suggests that the overall cognitive load is an important factor for figure segregation in the auditory cortex. High load in a visual task leads to impaired detection performance of auditory figures and reduced MEG responses (Molloy et al., 2018), indicating that the availability of shared attentional resources aid figure-ground segregation. During an informational masking paradigm, which included the presentation of bistable stimuli, changes in brain responses were only observed when target stimuli were detected (Gutschalk et al., 2008b). Taken together, most evidence points towards attentional involvement or at least attentional facilitation of sound segregation.

Little is known about the electrophysiological basis of figure-ground processing. Using EEG, previous studies have identified clear figure-evoked brain activity that scales with figure coherence. Higher coherence levels cause larger EEG response amplitudes (O'Sullivan et al., 2015a; Tóth et al., 2016). The response dynamics to an emerging figure are characterised by an object-related negativity (ORN) and a P400 component, both of which are impacted by changes to coherence level and duration of the figure (Tóth et al., 2016). Behavioural engagement has a substantial impact on

the EEG activity, with higher global field power during active listening (O'Sullivan et al., 2015b). Responses in hit and miss trials indicate that ORN and P400 are a signature of figure emergence and the perceptual decision, respectively. Across coherence levels, peak latencies for the ORN were found around 250ms from figure onset. P400 peaks were detected around 300ms thereafter (Tóth et al., 2016).

In previous chapters of this thesis, I have established that macaques are a good animal model to investigate auditory scene analysis that exhibit homologous cortical organisation (Baumann et al., 2013), comparable audiograms (Dylla et al., 2013; Jackson et al., 1999), equivalent pitch perception (Joly et al., 2014a) and, crucially, similar figure-detection performance as well as homologous cortical involvement during figure-ground segregation (see chapter 2 & 3). Because this evidence implies highly similar underlying brain mechanisms, I investigated cortical responses during figure-ground segregation in primary and non-primary auditory cortex of macaque monkeys. Multi-unit responses to SFG signals allow inferences about the population code the brain uses to segregate auditory objects from complex, natural scenes. Any resulting insights will most likely be transferable to auditory processing in humans.

4.3 Methods

4.3.1 Animals

Two adult macaques, Monkey 1 (Male, 11yrs, 11kg) and Monkey 2 (Female, 6yrs, 7kg), participated in this study. Detailed description of the animal training can be found in chapter 2. A circular PEEK chamber (17mm ID) was implanted over the left hemisphere with a 10 degree (Monkey 1) or 15 degree (Monkey 2) medial tilt to allow access to the left auditory cortical areas (Figure 4.1). Structural and functional MRI scans were used to position the chamber. The chamber implantation procedure is described elsewhere (Thiele et al., 2006). During testing periods, animals were kept under fluid-controlled conditions. Fluid control was within ranges which do not negatively affect animal's physiological or psychological welfare (Gray et al., 2016).

All procedures performed in this study were approved by the UK Home Office (Project License: 70/7976) and by the Animal Welfare and Ethical Review Body at Newcastle University. All experiments comply with the UK Animals Scientific Procedures Act (1986) on the care and use of animals in research, with the European Communities Council Directive on the protection of animals used in research (2010/63/EC).

4.3.2 Figure detection task

A detailed description of this stimulus and the task paradigm can be found in chapter 2. Monkeys were seated in a primate chair (Christ Instruments) in a sound-attenuated chamber with a touch bar and a grey screen in front of them. Trials were initiated by bar touch. After a 500ms baseline period, a stochastic figure-ground stimulus was presented. The monkeys responded to the presence of a figure by touch bar release. Independent of the behavioural outcome of the trial, sounds were kept on for the entire stimulus duration (3 secs). Visual feedback was given immediately after response. The colour of the screen changed either to green for correctly performed trials or to red for error trials. Reward was administered after the stimulus presentation period for correct trials. In each recording session, 20 randomly selected stimuli, of which 60% contained a figure, were presented in a pseudo-random order to ensure an equal number of presentations. We presented signals with two coherence levels (figures composed of 8 & 12 frequencies, equal probability). For monkey 2 we recorded 17 additional sessions with a lower figure saliency (figures composed of 4 & 8 frequencies, equal probability).

4.3.3 Acoustic stimuli

For most recording sessions, we presented a battery of stimuli:

Stochastic figure-ground stimuli

For a detailed description of the stochastic figure ground stimulus, see chapter 2. Only figures with two coherence levels (either 4&8 or 8&12) were presented. The number of stimulus repetitions varied based on the number of performed trials. A new set of 20 randomly selected stimuli was presented in each session.

Pure Tones (PT)

A total of 14 pure tones (200ms long, half-octave step-width [180Hz – 16292Hz]) was presented during every recording session. A 10ms cosine on- and off-ramp was applied to all signals. Tones were presented at three different intensities (50dB, 60dB, and 70dB SPL). A minimum of 10 repetitions per stimulus condition was obtained in each session.

Band-pass noise (BPN)

13 frozen band-pass noise bursts (200ms) were presented in each recording session. Passbands were half an octave wide, covering the range between the pure tones described above (centre frequency is mean of adjacent PT frequencies). A 10ms cosine on- and off-ramp was applied to all signals. Similar to the pure frequencies, noise bursts were presented at 3 different intensities (50dB, 60dB, and 70dB SPL) with a minimum of 10 repetitions per stimulus condition.

Click trains (CLK)

Monophasic, 200ms long click trains with varying frequencies (25 Hz, 50 Hz, 75 Hz and 100Hz) were presented at 80dB SPL. Each pulse had a duration of 2ms. A minimum of 10 repetitions per conditions was obtained for a number of recordings with Monkey 2 (36/101 recordings, 89% of recorded channels).

White noise (WN)

For some recordings of M1 (39/153) and all recordings of M2, we presented 200ms long white noise bursts at 80dB SPL. Stimuli were not frozen but created online during each recording session. A minimum of 30 white noise bursts was recorded. Responses to white noise bursts were not further analysed for this thesis.

PT, BPN, CLK stimuli were presented in an alternating block design. Each block started with the presentation of pure tones, followed by band-pass noise and ended with click trains. Per block, each stimulus was presented once. Within each block, the presentation order was randomised.

4.3.4 Neurophysiological recordings

Single- and multi-units as well as local field potentials were recorded by advancing one to four microelectrodes (0.2-5MΩ) into the auditory cortex by means of a remotely (CMS Drive, NAN Instruments) or manually controlled Microdrive (MO97 Oil Hydraulic Micromanipulator, Narishige). Epoxylite-coated tungsten electrodes (FHC, Bowdoin, ME), custom-built glass-coated tungsten electrodes or 16-channel electrode arrays (V-probe, Plexon, Dallas, TX) were used for recordings. Stainless steel guide tubes (23ga or 26ga) were used to penetrate the dura mater. Custom-made PEEK grids (1x1mm or 0.8x0.8mm) were oriented approximately parallel to the anteroposterior axis and served as spatial reference for the electrode position. The signal was referenced to the guide tube or electrode shaft (V-probe), amplified, filtered (LFP: 1-300Hz, Units: 600-9000Hz), digitised (LFP: 1kHz, Units: 32kHz) and recorded via a 32-channel Digital Lynx SX acquisition system (Neuralynx, Cheetah 5.6 software). Anatomical landmarks (lateral sulcus), noise bursts and natural sounds were used to identify that the auditory cortex was reached.

Stimulus presentation, behavioural control and reward administration was controlled with an in-house program written in Python 2.7, which is partly based on Psychopy (Peirce, 2007) on Ubuntu 16.04 LTS via a DAQ-LabJack U6-Pro Interface. Recording sessions started with the figure detection task. The battery of sounds used to assess the units' tuning was presented after the subject stopped working. A microphone (Audio Technica U841R with AT8531 power module) placed in front of one of the speakers (Creative GigaWorks T20 Series II) recorded the sound environment within the sound-attenuated booth. This signal was used to correct the sound onset timestamps offline for every trial by adding the delay period between timestamps and physical sound onset.

4.3.5 Data analysis

Animal behaviour

Behavioural performance was evaluated via d-prime, a sensitivity index that provides a measure of separation between signal and noise distribution and takes all possible behavioural responses into account (Stanislaw and Todorov, 1999). D-prime was calculated in the following way: d' = Z(Hit rate) - Z(False alarm rate), where Z is the inverse of the standard normal cumulative distribution function of hit rate and false

alarm rate, respectively. Additionally, reaction times were analysed by comparing the mean as well as the coefficient of variation between coherence levels. The coefficient of variation, a measure of data dispersion, is the ratio of standard deviation divided by the mean. Effects of reaction time and response variability were tested across all included recording sessions (n = 155) with a repeated-measures ANOVA. To assess the effect of figure coherence on reaction times on a trial-by-trial basis, a linear mixed effects model was constructed with figure coherence defined as a fixed effect. Random intercepts were included for each subject and session to account for repeated measurements: Reaction time ~ Coherence + (1 | Session) + (1 | Monkey). This model was tested against an intercept-only model without coherence as a factor by means of maximum likelihood ratio tests: Reaction time ~ 1 + (1 | Session) + (1 | Monkey).

Neuronal response estimation

The envelope of the multi-unit activity (MUAe) was calculated based on the analogue, band-pass filtered signal (600-9000Hz). After rectification of the time course, the signal was low-pass filtered using a third-order Butterworth filter with a 200Hz cut-off frequency. Subsequently, it was down sampled to 1kHz and saved as a new file. After import, data were baseline normalised, using the 400ms window prior to sound onset.

Example units that were displayed with spiking activity were manually spike-sorted using SpikeSort3D (version 2.5.3). Spike density functions were computed by fitting a Gaussian curve with a width of 5ms to each detected spike. Subsequently, spike-wise Gauss curves were summed, and data were averaged over stimulus repetitions.

Spatial maps

To create spatial maps of the recording field, neural responses to pure tone stimuli were evaluated using a 2-factorial ANOVA [frequency x intensity] and inspection of the signal-to-noise ratio. Signal-to-noise ratio contrasted the neural response after sound onset (10ms – 150ms) with the average baseline activity 200ms before stimulus presentation. A 50ms sliding window was used

to estimate SNR along time. Signal-to-noise onset ratio of the cell was then defined as the average difference between activity across all sliding window positions and baseline measurements. Units were included if the signal-to-noise ratio across all trials exceeded 3 or if the ANOVA yielded a frequency specific effect ($p < 0.05$). Trials with movement artefacts, identified by saturated LFP signal, were excluded from further analysis.

The best frequency of a unit was determined by taking the maximum of the average response to pure tones across trials for each condition [frequency x intensity]. The data was then averaged across sound intensities, smoothed with a smoothing spline (smoothing parameter 0.98) and the best frequency was assigned to the peak of the resulting curve. Spatial maps were created by rounding recording coordinates to integers and averaging best-frequencies and peak latencies of all included multi-units for each coordinate. The tonotopic low frequency gradient reversal was used to subdivide the recording field into the anterior and posterior area. Based on structural MRI, tonotopy, latencies to pure tones and either histology (M1) or responses to click trains (M2) we estimated which cortical areas contribute to anterior and posterior recording field.

Inclusion criteria

For the figure-detection task, a minimum of 10 repetitions per stimulus was required for inclusion. MUA was compared between figure and control trials by averaging the multi-unit (envelope) activity in the 300-100ms window before the behavioural response (figure trials) or during a pseudorandom 200ms time interval (control trials, based on figure onset distribution). Units that were sound responsive and showed a significant difference between average firing rates in hit and correct rejection trials (2-sample t-test, $p < 0.01$) were classified as figure-responsive and included into the analysis. Sound responsiveness was assessed by comparing the spectral power during sound presentation and baseline period. Only if the neuronal response showed a significant enhancement at 20Hz (50ms duration of SFG stimulus chord), units were classified as sound responsive.

Modulation latency

To determine the onset of the figure-ground modulation (FGM), the mean firing rate for each SFG stimulus time bin was extracted. Subsequently, a difference curve for a given figure stimulus was calculated with all control stimuli that were presented (Figx – Ctr1:n). All difference curves were then pooled and the mean firing rate for each time bin was bootstrapped (5000 repetitions). We defined the onset of the figure-ground effect as the first significant sample ($p < 0.01$) after figure onset that was followed by at least four consecutively significant time bins (5ms in total).

Figure-ground modulation

To quantify how reliable multi-units can discriminate between figure and control trials neuronal d-prime was calculated: dAB = (mA–mB)/s, where mA and mB are the mean responses in stimulus conditions A and B, and s is the pooled standard deviation. In addition to this parametric measure, I also illustrate the non-parametric area under the receiver operating characteristics (AUROC) that demonstrate FGM based on a binary classifier.

Time frequency analysis

Time-frequency decomposition of the LFP signal was computed using a Morlet-wavelet analysis across 100 linearly spaced frequencies between 7Hz and 100Hz, with a logarithmically spaced number of cycles that ranged from 4 (lowest frequency) to 10 cycles (highest frequency). For each frequency, we transformed both signal and wavelet into the frequency dimension using the fast Fourier transform (FFT). Subsequently, the amplitude-normalised wavelet and the signal were convoluted by applying the inverse Fourier transform over the point-wise multiplication between both vectors. The difference matrix between figure and control condition was then z-scored by subtracting the mean and dividing by the standard deviation of shuffled trials. After visual inspection of the pooled decision aligned maps, a 300ms long window (-350 to -50ms before behavioural response) was drawn in the alpha/beta (7-30Hz) and gamma range (35-100Hz). For statistical testing, power in this window was averaged across frequency and time.

Inter-trial phase coherence

Inter-trial phase coherence was computed by extracting the frequency-specific phase angle and measuring the unit-length vector of each trial. Inter-trial phase coherence was defined as the average length of the vectors across trials.

Phase locked responses

LFP phase locking in response to click trains was assessed by calculating a Fast Fourier Transform over the baseline and sound presentation period for each trial. FFT's were then averaged across trials for each condition. If the spectral power in the frequency bands of the click trains [25, 50, 75, 100 Hz] exceeded 3 standard deviations of the baseline power, the response was classified as phase locked.

Post-hoc tests

Post-hoc tests of MUA between different stimulus conditions were done using the non-parametric paired Wilcoxon signed rank test. P-values were false-discovery-rate (FDR) corrected.

4.4 Results

4.4.1 Figure coherence is decisive factor for perception

Two macaque monkeys were trained to detect auditory figures in a noisy background (see chapter 2, Figure 4.1 a+b). For the electrophysiological experiments, only figures that were highly salient to both humans and macaques were presented (Schneider et al., 2018b; Teki et al., 2013). Both subjects performed the figure detection task with high performance that was similar to human behaviour (Figure 4.1 c-e). A main effect of figure coherence was found on detection performance (Repeated measures ANOVA, $F_{1, 152} = 565.68$, $p < 0.01$) and reaction times (Trial-based: Linear mixed effects model, $t_{(52885)} = -83.242$, $p < 0.01$, Average-based: Repeated measures ANOVA, $F_{1, 152} = 978.49$, $p < 0.01$), indicating that the number of coherent elements is a critical factor for perception. Detection performance and average reaction times were significantly different for each monkey (Figure 4.1 c + d, Wilcoxon signed rank test; d-

prime: M1, $p < 0.01$; M2, $p < 0.01$; Reaction time: M1, $p < 0.01$; M2, $p < 0.01$). A linear mixed effects model of reactions times performs significantly better with coherence as predictor ($\chi^2_{(1)} = 6508.9$, $p < 0.001$). We also found a main effect of coherence on response variability, however, the coefficient of variation showed a coherence effect for M2 only (Figure 4.1 e, Repeated measures ANOVA, $F_{1, 152} = 12.217$, $p < 0.01$; Wilcoxon signed rank test M1, $p = 0.1982$; M2, $p < 0.01$).

D-prime values reported in this chapter are higher than in the previously reported behavioural study (Table 4.1), indicating that the monkeys had not reached the end of the learning curve when the behavioural data in chapter 2 was acquired.

Table 4.1: Average behavioural performance for each condition. Shown are d-prime (d') values, rounded median reaction times and the coefficient of variation (CV) for each coherence condition.

Monkey		Coh8	Coh12
M1	d'	1.7282	2.5651
	RT [ms]	683	568
	CV	0.2222	0.2160
M2	d'	2.5686	3.0140
	RT [ms]	650	548
	CV	0.1991	0.1854

4.4.2 Recorded units mostly located in core fields

Multi-unit activity was recorded in primary and surrounding non-primary auditory cortex of two adult macaque monkeys. For each recorded unit, frequency selectivity, peak latencies and phase locking capabilities were extracted. This information was then used to create a spatial map of the recording field and formed the basis for the subdivision of the recording field into anterior and posterior auditory cortex (Figure 4.1 g-i). Moreover, these maps indicate whether an area has mostly primary sensory cortex-like response

properties or whether unit responses in this field resemble those of secondary belt areas.

The basis for these spatial maps was cortical spiking activity in response to pure tones. Figure 4.2 and Figure 4.3 illustrate spiking responses of an example multi-unit to pure tones and band-pass noise. For this unit, tone response amplitudes clearly increase with sound intensity. Maximum responses are easily identifiable for both pure tones (4073Hz) and noise bursts (3476Hz), which are the best (centre) frequency for each of the respective categories. Although band-pass noise stimulation leads to greater neuronal activation across frequency bands, we focussed on responses to sinusoidal tone stimulation as they show higher selectivity.

Tonotopic maps show the average best frequency at each recording location. In both subjects, a high-low-high gradient of best-frequencies can be identified (Figure 4.1 g). The low frequency gradient reversal determines the boundary between area A1 and R (Baumann et al., 2013; Joly et al., 2014b; Kuśmierek and Rauschecker, 2009b). This boundary was used subdivide the recording field into an anterior and a posterior extent for further analysis. The posterior high-frequency extent is very small in M1, suggesting that the chamber location did not allow full coverage of A1. Even though on average a tonotopic gradient can be identified, there is high variability in best frequencies across the recording field (Figure 4.4). Maps are smoothed but due to the limited number of recordings at each site (especially in M1), the tonotopic gradient is not as clear as seen in other studies (Kikuchi et al., 2014, 2019; Kuśmierek and Rauschecker, 2009b, 2014).

Latency maps in both subjects show regions with short peak latencies in the posteromedial parts of the recording field (Figure 4.1 h), suggesting that these recording locations correspond to core fields with predominant input from the ventral division of the medial geniculate nucleus (Camalier et al., 2012; Scott et al., 2017). Lateral and anterior regions show on average longer response latencies with higher variability across recording channels, indicating higher cortical centres (Camalier et al., 2012). Low latency regions in M1 stretch across the division boundary, suggesting primary regions on both sides. In M2, low latency areas are confined to the posterior recording field. The average peak latency for M1 is 30.65ms and 29.38ms for anterior and posterior auditory cortex, respectively. M2 shows a clearer difference between areas with 34.47ms (anterior) and 23.64ms (posterior).

In Figure 4.2 and Figure 4.3, I have shown spike density functions that are based on fitting Gaussian curves to single spikes. This results in subtle differences in the response dynamics. For example, the resulting curve might look smoother than the actual extracellular current fluctuation has been. Moreover, spiking responses are highly variable across stimuli. The average response strength also depends on the spike threshold and overall recording quality. To avoid these issues, I decided to report the envelope of the multi-unit activity (MUA) instead of a binary, suprathreshold signal for the population activity for the remaining chapter. The basis of the multi-unit envelope is neuronal spiking. Hence, this signal is a direct reflection of the cortical activity surrounding the electrode tip. However, since it is based on the raw analogue signal, it incorporates the entire recorded population response and preserves the actual response dynamics. On a population level, it leads (in this case) to more reliable effects compared to thresholded spike signals.

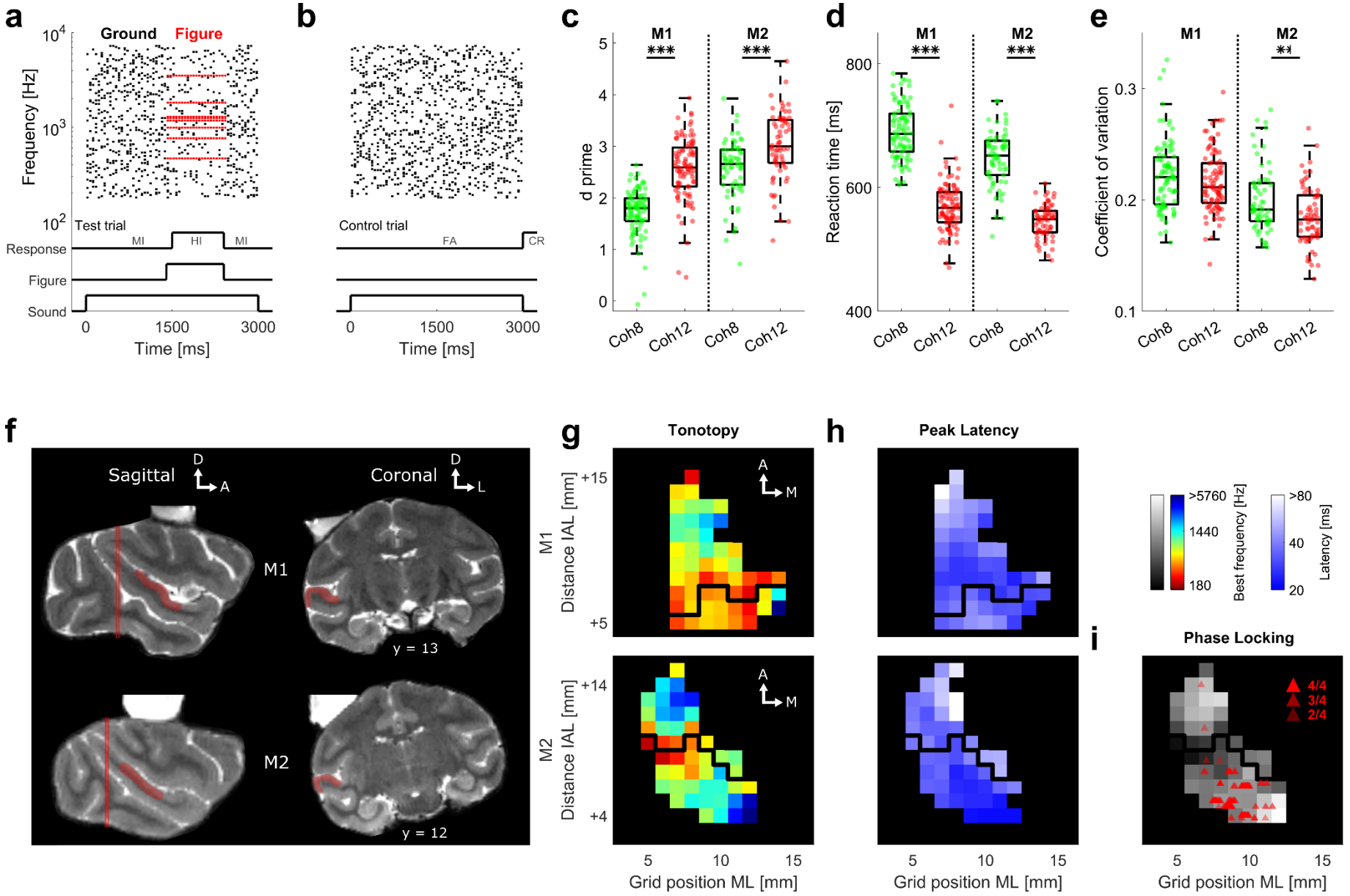
a
Ground
Figure
Frequency [Hz]
Test trial
MI
HI
MI
Response
Figure
Sound
Time [ms]
b
Control trial
FA
CR
Time [ms]
c
M1
M2
d prime
Coh8
Coh12
d
Reaction time [ms]
e
Coefficient of variation
f
Sagittal
Coronal
M1
M2
y = 13
y = 12
g
Tonotopy
Distance IAL [mm]
Grid position ML [mm]
h
Peak Latency
Best frequency [Hz]
Latency [ms]
i
Phase Locking
4/4
3/4
2/4

Figure 4.1: Summary of experimental paradigm, behavioural performance and recording field.

(**a**) Schematic spectrogram of an example SFG stimulus. Line plot below indicates the 900ms long behavioural response window for the displayed stimulus as well as the behavioural outcome for touch bar release inside (HI – hit) and outside (MI – miss) of this time window. (**b**) Example control stimulus without figure. No touch bar release until sound presentation finished (CR – correct rejection), otherwise trial is counted as false alarm (FA). (**c-e**) Behavioural detection performance of Monkey 1 (left) and Monkey 2 (right) for a coherence level of eight (green) and twelve (red) elements, respectively. Only sessions with more than 200 trials were included. (**c**) Average d-prime values. (**d**) Mean reaction time. (**e**) Response variability (coefficient of variation). (**f**) Structural T2 MRI of both subjects. Red vertical lines indicate location of interaural line. Distance of coronal sections from interaural line in [mm] is indicated below. Recording chamber is filled with saline for visibility. Both recording chambers have a medial tilt (10deg for M1, 15deg for M2) to allow easier access to the lateral auditory cortex. Auditory cortex is highlighted in in red. (**g**) Best frequency maps for M1 (top) and M2 (bottom). Colour code indicates average best frequency across the surface of the superior temporal gyrus. Y-coordinates show distance to the interaural line (IAL). X-coordinates show the grid position. Maps are smoothed with a 2x2mm Gaussian kernel. The black line indicates the division boundary between anterior and posterior recording field based on low frequency gradient reversal. (**h**) Latency map for M1 (top) and M2 (bottom). Colour code illustrates average peak latency for each grid position. (**i**) Location of channels that exhibit significant LFP phase locking (red triangles) overlaid on best M2's frequency map. Strength of phase locking (No. of click train frequencies that elicit phase locking) is indicated by transparency of triangle.

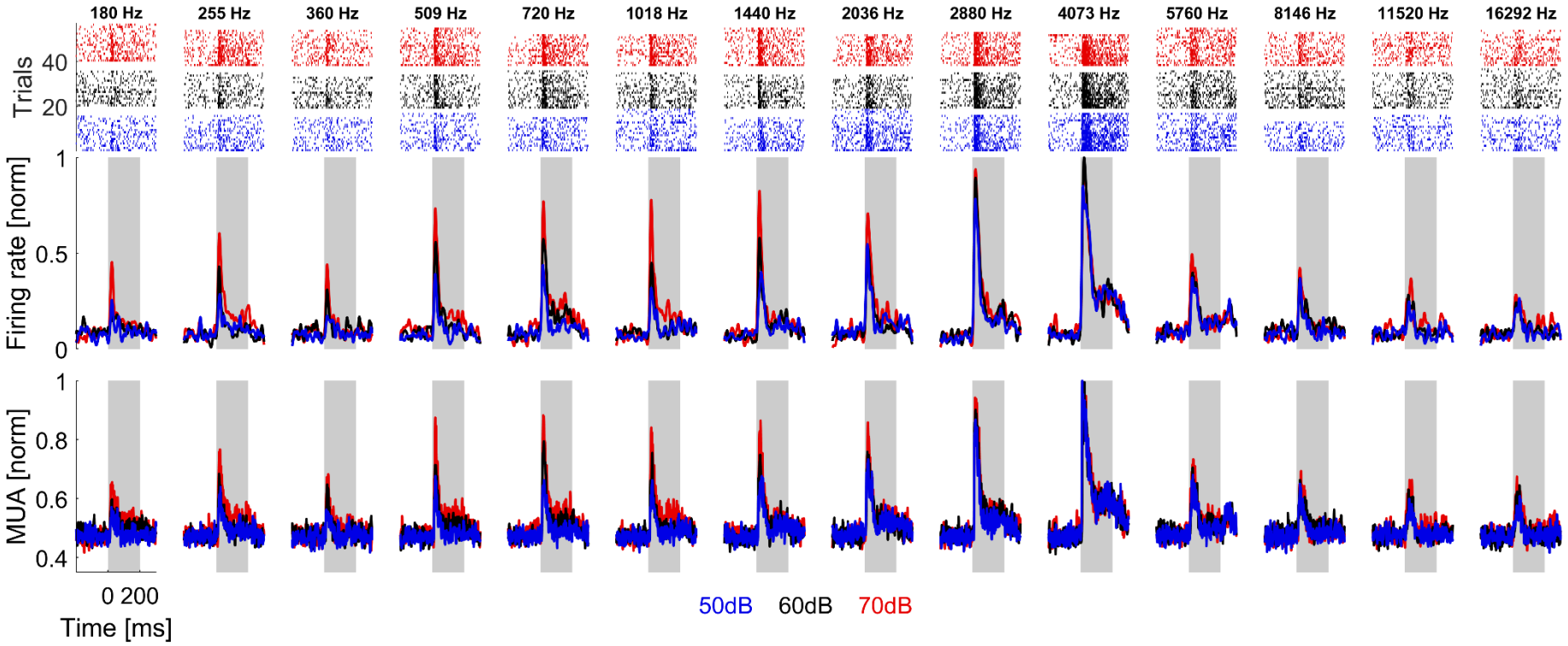

Figure 4.2: Neuronal responses of an example multi-unit to 14 pure tones.

Suprathreshold spiking (top & middle row) and multi-unit envelope (bottom row) shown. Columns represent neuronal responses to different stimulus frequencies. The colour code denotes stimulus intensity [50, 60, 70 dB SPL]. 20 repetitions of each intensity were presented but trials with saturated LFP were excluded. Raster plots are shown on top. Each row corresponds to a trial, each point within a trial to a single spike. Spike density functions for each condition shown below. Firing rates are normalised to the maximum response, averaged over trials (i.e. best frequency). Shaded areas correspond to the stimulus presentation period. The corresponding multi-unit envelope response is illustrated in the bottom row with similar conventions.

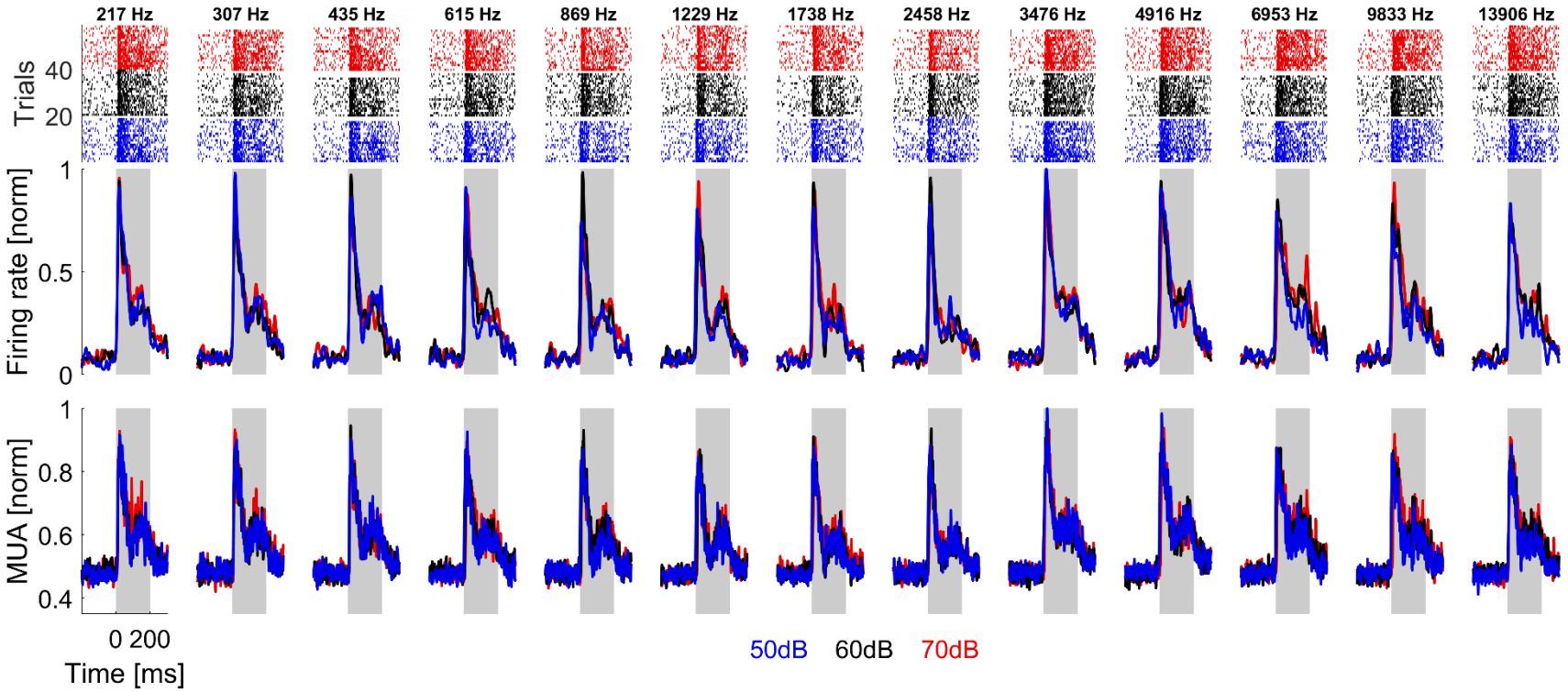

Figure 4.3: Neuronal responses of an example multi-unit to 13 bandpass noise bursts.

Same unit as in Figure 4.2. Frequencies indicate centre frequency of passband. Otherwise, similar conventions as seen in Figure 4.2.

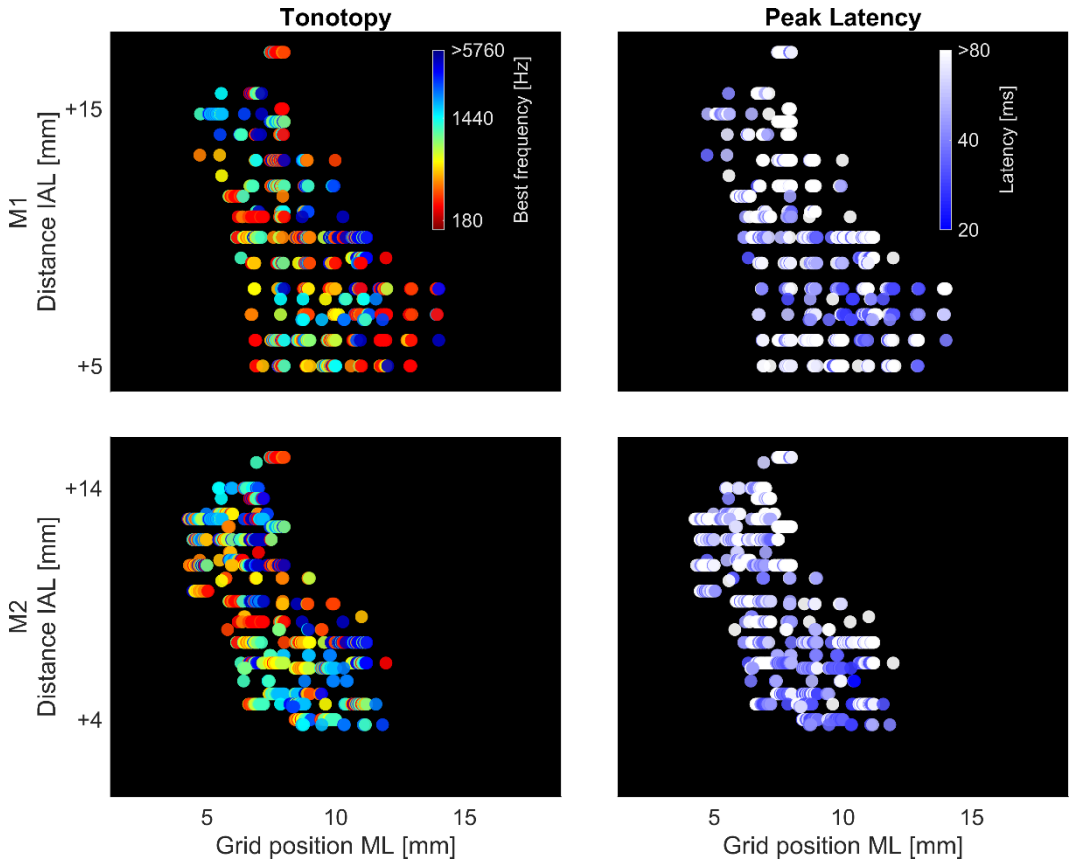

Figure 4.4: Individual channel responses.

Best frequency (left) and peak latency (right) of each recorded channel with significantly different tone vs baseline responses for M1 (top) and M2 (below). Recording locations are corrected for electrode depth and angle of the chamber.

An interesting property of primary auditory areas is the ability to strongly phase-lock to periodically repeating stimuli, like click trains (Billig et al., 2019; Lu et al., 2001; Oshurkova et al., 2008; Steinschneider et al., 1998; Wang et al., 2008). This characteristic can be used to distinguish primary from non-primary auditory cortex, as has previously been done in humans (Billig et al., 2019; Brugge et al., 2009). Spiking activity in the caudomedial belt area of the macaque has also been shown to phase-lock to click trains, however, to the tested frequency range in this thesis only a very small percentage of neurons responded with excitatory rate increases to the stimuli (Oshurkova et al., 2008). Hence, recording sites that exhibit phase-locking to click trains are likely located in primary cortical areas.

Responses to click trains with varying frequency were recorded for M2 only. Phase locking can be found in both spiking and LFP. However, oscillatory responses

at click train frequency seem to be more robust in the LFP and were only visible in the PSTH at lower stimulus frequencies. This is in line with previous research (Lu et al., 2001). This section describes the spatial organisation of phase locked LFP responses along the cortical surface of the superior temporal gyrus. The strength of phase locking was taken into account by quantifying how many click frequencies the unit responded to with phase-locked responses.

Figure 4.5 illustrates responses to click trains of an example unit that exhibits phase-locking to all click train frequencies. It is therefore likely that this unit was located in primary auditory cortex. Figure 4.1 i shows the location of all channels that exhibit phase locking to at least two frequencies. A spatial cluster in the posterior recording region can be identified, suggesting primary core areas at the posteromedial recording locations. Overall, 274 of 621 channels (44.12%) exhibit phase locked LFP responses. In anterior auditory regions, no reliable phase locked LFP responses were found. Only 31/292 recording channel (10.62%) show phase locked responses to at least one click train. Just 0.68% of channels are responsive to more than one stimulus frequency. In contrast, 174/329 channel (52.89%) in the posterior field respond with phase locked LFP responses to click trains, with 31.31% being responsive to more than one frequency. Combined with the tonotopic gradient and latency information, this provides strong evidence that the posterior recording field in M2 comprises large parts of area A1. In contrast, the absence of phase locking in combination with longer peak latencies indicated non-primary areas in the anterior extent of the recording field, as the rostral core areas R and RT show phase locking abilities to some degree.

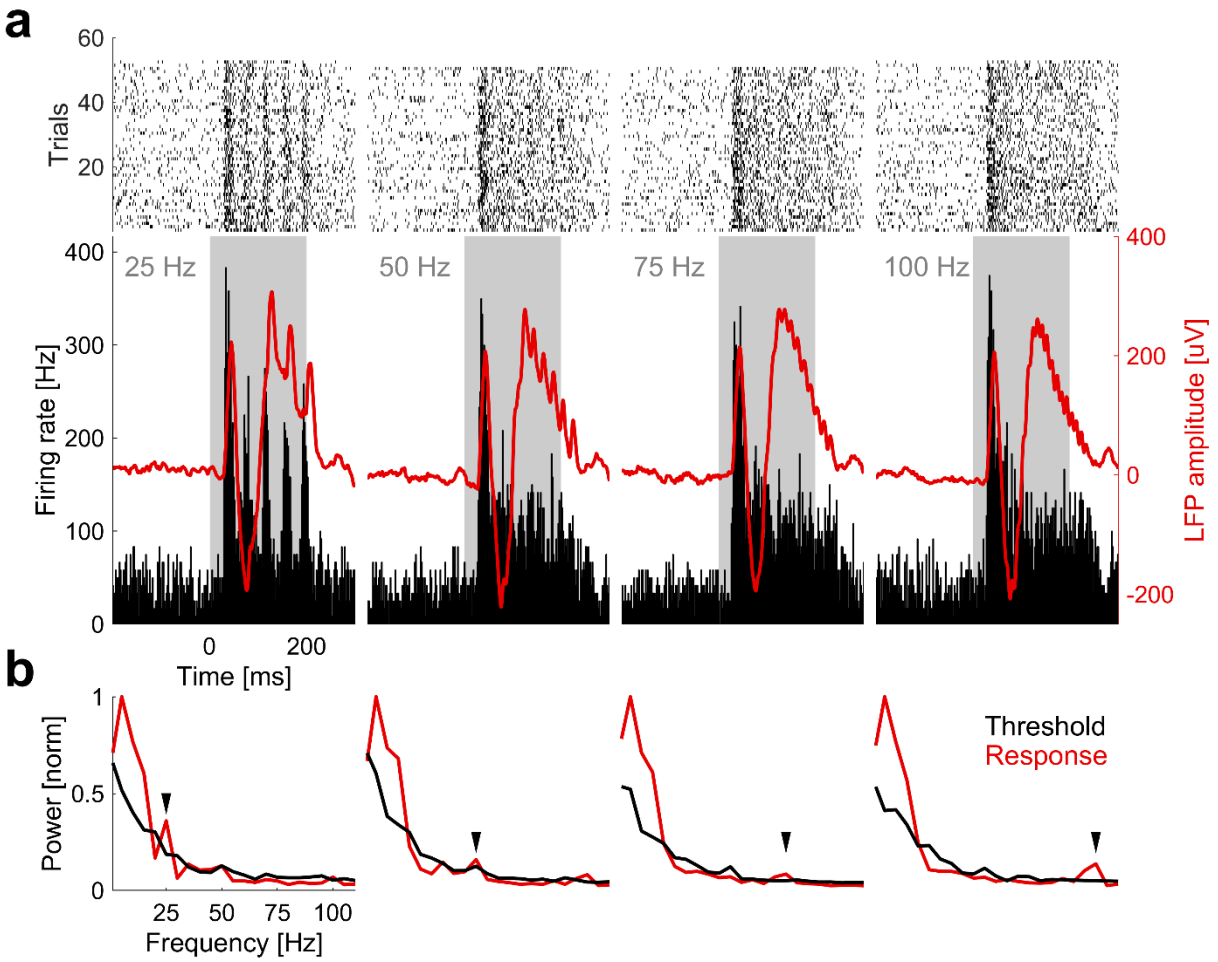

Figure 4.5: Spiking and LFP response of an example multi-unit to click trains.

Same unit as shown in Figure 4.2. (**a**) Raster plots are shown on top. Each row corresponds to a trial, each point within a trial to a single spike. Different plots correspond to different stimulus frequencies. 60 repetitions were presented per click train frequency. Trials with saturated LFP were excluded. PSTH with a 2ms bin width plotted below. Overlaid red line shows raw LFP trace. Shaded areas indicate the stimulus presentation period. Click train frequency is indicated in grey. (**b**) Average FFT of raw LFP response to click trains during stimulus presentation period. Click train frequency is indicated by the black arrow. Significance threshold (black line) was determined by calculating the spectral power of the baseline period and adding two standard deviations. All click rates (4/4,100%) induce phase-locked responses.

Although no responses to click trains were recorded in M1, a histological investigation yielded valuable information regarding the location of recording sites (Figure 4.6). Parvalbumin staining was used to identify primary koniocortex as it nicely stains the prominent granular layer in the auditory core (Hackett et al., 1998a; Jones et al., 1995). Electrode tracks can be seen best in slices with Gallyas and Nissl staining. There is no obvious damage to the superior temporal plane where the auditory cortex is located, however, the majority of slices show tracks in the parietal lobe, suggesting that most recordings were done in the target region (Figure 4.6 c). There are some sparse white matter tracks ventral to the auditory cortex, indicating that the electrode went too deep in a few sessions. Most of the white matter tracks aim at the core areas with some electrode traces that point towards the lateral belt. Thus, the histology indicates that the recording field stretched across primary and lateral, non-primary areas in M1. However, the track pattern suggests that data from core areas was acquired in most sessions. The realignment of recording sites with structural MRI suggested that most of A1 was accessible in M1 (Figure 4.7). Despite this, the following points suggest that most of the recorded units originate from a more anterior cortical region:

(1) Lack of a high-frequency region in the posterior part of the recording field (Figure 4.1 g)
(2) Location of the recording chamber with respect to the brain (Figure 4.1 f)
(3) Sustained response dynamics to SFG stimuli in posterior recording field (Figure 4.8 d)

Thus, the available data suggest that M1's posterior field comprises low frequency regions of A1 and R. Hence, the division boundary may be placed within area R. The anterior field includes a large part of core regions (R, RT) with some contributions of lateral belt regions (AL).

M2 shows clearer tonotopic gradient and latency organisation. In combination with structural MRI scans and click train responses, I concluded that the posterior field likely represents much of A1. In contrast, the absence of phase locking and longer response latencies suggest that the anterior extent covers partly lateral, non-primary belt areas (AL), partly core regions (R).

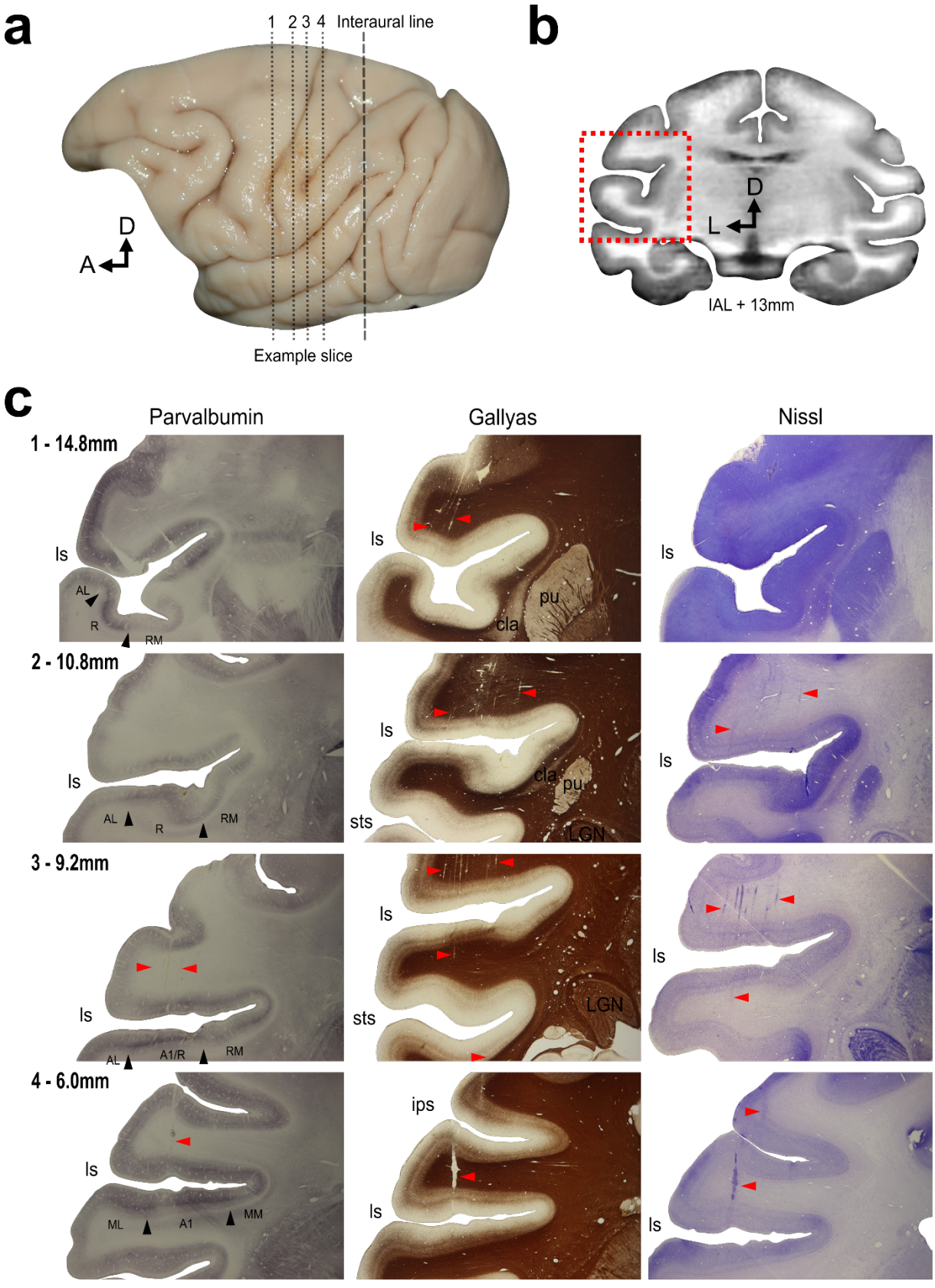
a
1 2 3 4 Interaural line
D
A
Example slice
b
D
L
IAL + 13mm
c
Parvalbumin
Gallyas
Nissl
1 - 14.8mm
ls
AL
R
RM
pu
cla
2 - 10.8mm
sts
LGN
3 - 9.2mm
A1/R
4 - 6.0mm
ips
ML
A1
MM

Figure 4.6: Histology of M1.

(**a**) Extracted brain with marked interaural line (dashed) and approximated slice location (dotted). A macaque brain atlas (Saleem and Logothetis, 2012) was used to align the example slice locations with the brain. Anatomical markers such as the shape of the superior temporal plane, IPS and the Claustrum were as a guide. (**b**) Coronal MR image (T1), approximately 13mm anterior to the interaural line. Red box indicates field of view for brain slices below. (**c**) Stained, coronal brain slices (50um each) showing parts of the left parietal and temporal lobe. Approximated distance from interaural line indicated in black. Note that these distances were measured after the brain has shrunk due to the preservation procedure. Electrode traces marked with red arrows. Approximated core boundaries are indicated by black arrows. Slicing and staining of the brain was done by Claudia Distler (University of Bochum) who also photographed the slices. Abbreviation: cla, Claustrum; IAL: Interaural line; ips, Intraparietal Sulcus; LGN, Lateralal geniculate body; ls, Lateral Sulcus; pu, Putamen; Auditory areas: A1, Primary auditory cortex; AL, anterolateral belt; ML, Middle lateral belt; MM, Middle medial belt; R, Rostral core; RM, Rostromedial belt

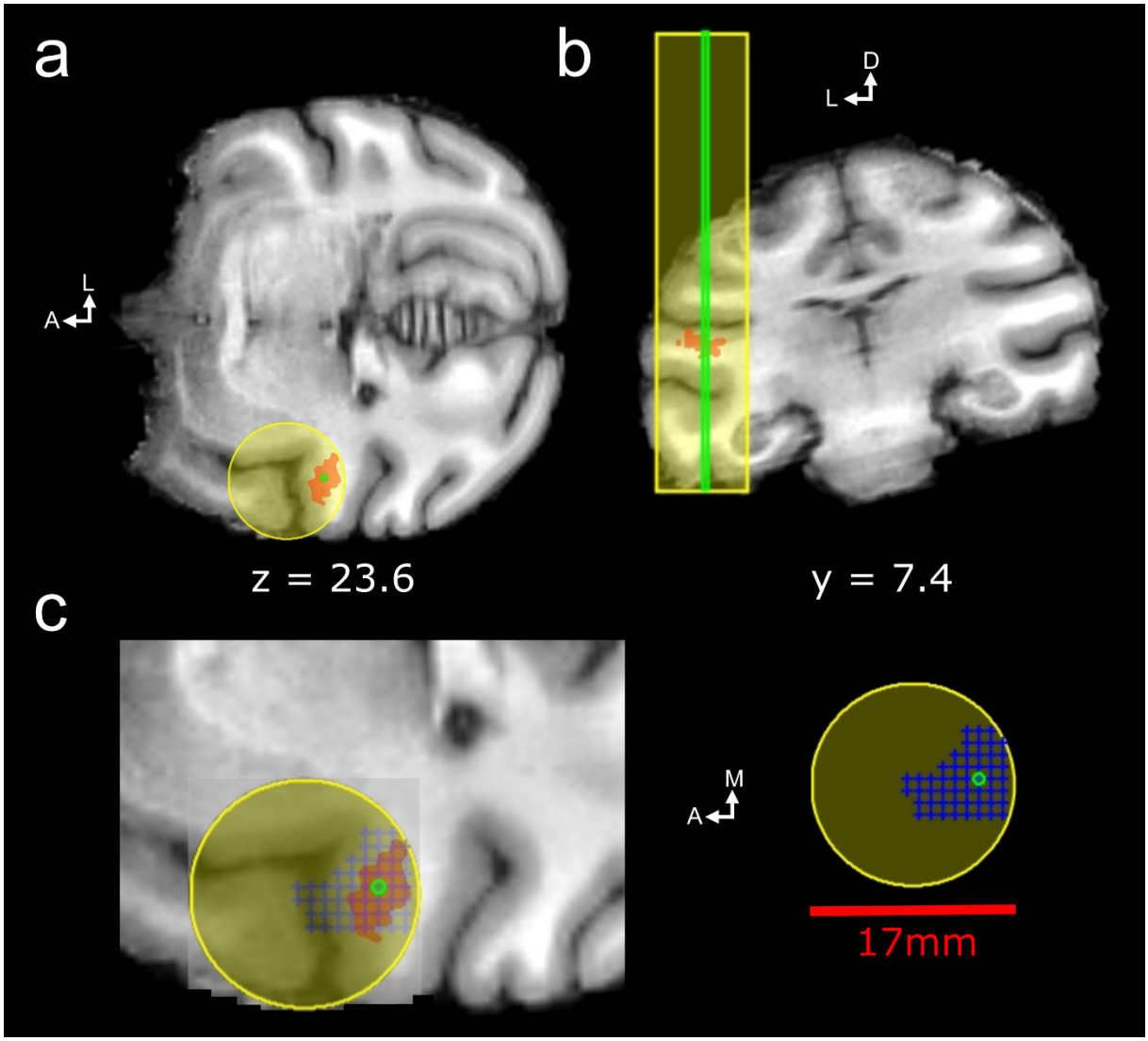

Figure 4.7: Coregistration of probabilistic field maps and recording grid in M1.

(**a**) Horizontal MRI section (T1 contrast) of M1's brain. The inner chamber diameter (17mm) is reconstructed with a yellow circle. Probabilistic map of A1 shown in red. An arbitrary reference location in A1, close to the tonotopic gradient reversal, is indicated in green. Distance from interaural line is shown below in [mm]. (**b**) Coronal section. Same conventions as in (**a**). (**c**) Zoomed version of (**a**) focussing on the left temporal lobe with recording grid locations overlaid on the chamber. Only grid coordinates from which data were acquired are shown.

4.4.3 Figure-ground modulation in early auditory cortex

Auditory cortical activity was recorded in response to SFG stimuli. Multi-unit responses strongly follow the temporal envelope of the stimulus, which is evident by a strong 20Hz oscillation of the neuronal signal. The average population activity is illustrated in Figure 4.8 a. Since there are obvious differences in the response dynamics and the location of the recording sites, population activity is displayed for each monkey individually. The Fast Fourier Transform across control stimuli shows enhanced spectral power at the frequency of chord presentation (Figure 4.8 a inset, Chord duration: 50ms, hence 20Hz), indicating that cortical responses track individual chords across the auditory cortex. Thus, multi units show a strong phasic response to each chord.

A subset of recorded units showed significantly modulated activity in response to auditory figures (Figure 4.8 b, Figure vs Control, $p < 0.01$, M1: n = 99, 29.64%; M2: n = 228, 36.71%). Responses to different stimuli were highly variable (Figure 4.9), however, on average, the figure-onset aligned population signals of responsive units reveal a slowly evolving ramp-up in MUA in both subjects (Figure 4.8). Averaged over the first four chords after figure onset, MUA does not significantly change between stimulus conditions (Two-factorial ANOVA [monkey, condition], Condition: $F_{1,651} = 0.73$, $p = 0.3938$, Monkey: $F_{1,651} = 1.73$, $p = 0.1895$). In the time window 201-400ms after figure onset (chord 5-8), however, firing rate averages in response to figures do differ compared to the control condition (2-factorial ANOVA [monkey, condition]: Condition: $F_{1,651} = 11.34$, $p < 0.01$, Monkey: $F_{1,651} = 3.05$, $p = 0.081$).

Critically, figure-ground modulation can be found in single neurons too (Figure 4.9), however, the sample size of well isolated single units was too small for an extended analysis. In case of the example unit shown, the average firing rate shows a significant difference between figure and control trials in the first 400ms after figure onset (Wilcoxon signed rank test, Coh8 vs Control: $p < 0.001$; Coh12 vs Control: $p < 0.001$). Moreover, this unit shows a significant difference between responses to different coherence levels (Wilcoxon signed rank test, Coh8 vs Coh12: $p < 0.001$). This effect cannot be explained by the number of frequency elements in the receptive field of the unit as the slope of a straight

line fit (least squares) to the number of frequency elements across chords is not significantly different from zero (Wilcoxon signed rank test, Coh8: p = 0.2588; Coh12 p = 0. 5656). Thus, temporal coherence of frequency elements across chords is likely driving this tonic increase in firing rate.

Critically, modulated multi-units are widely distributed across the recording field with no sign of spatial clustering towards anterolateral regions (Figure 4.8 c). This indicates an involvement of the earliest cortical stages, namely, the primary auditory cortex (A1) during figure-ground segregation.

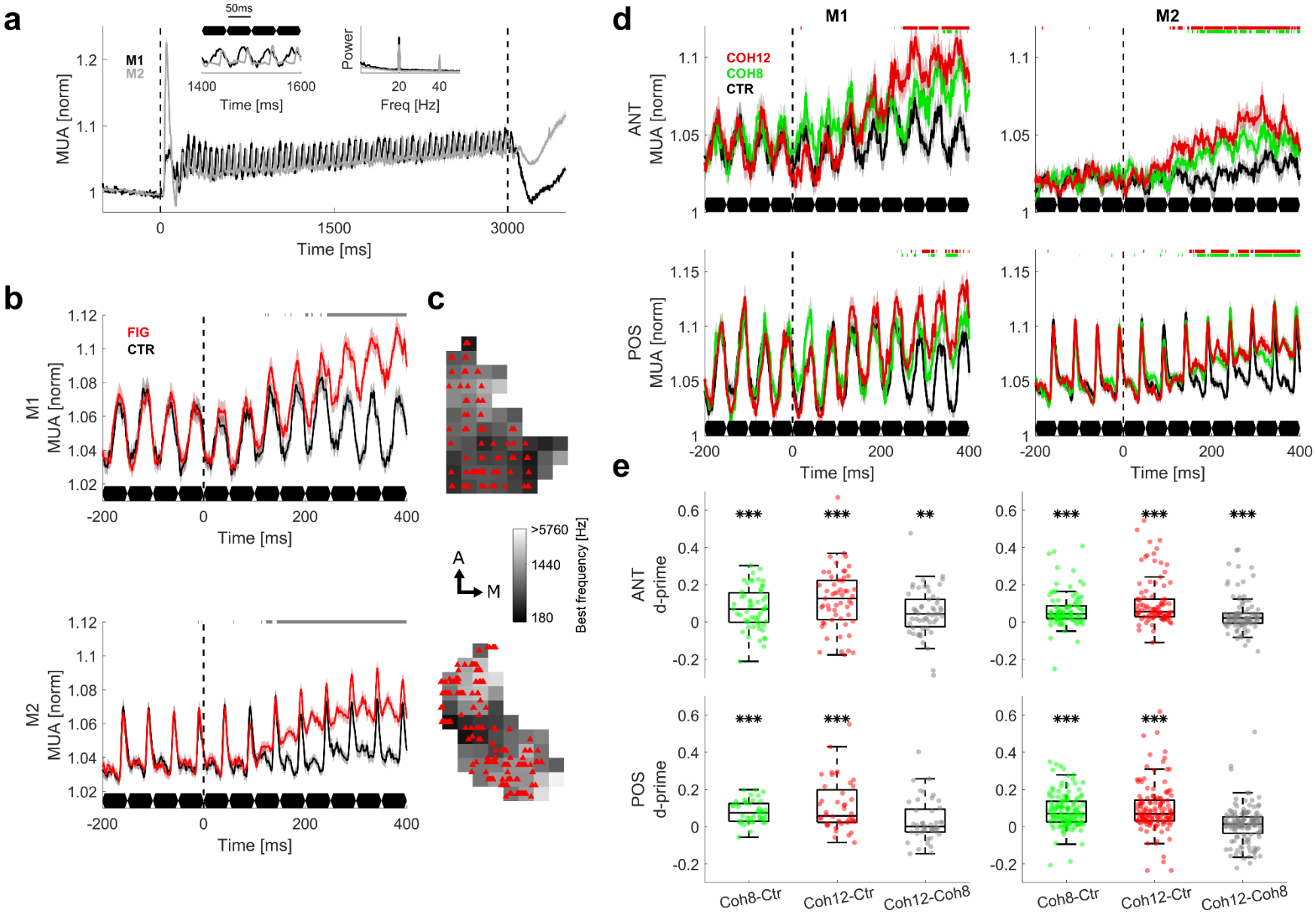
a
50ms
M1
M2
MUA [norm]
Time [ms]
Power
Freq [Hz]
b
FIG
CTR
M1
M2
MUA [norm]
Time [ms]
c
A
M
>5760
1440
180
Best frequency [Hz]
d
M1
M2
COH12
COH8
CTR
ANT
POS
MUA [norm]
Time [ms]
e
ANT
POS
d-prime
Coh8-Ctr
Coh12-Ctr
Coh12-Coh8

Figure 4.8: Average population responses to SFG stimuli.

(**a**) Average MUA to control stimuli for both M1 (black) and M2 (green). Inset shows zoomed response to four chords and the FFT of the average response. Normalised to the maximum power. (**b**) Figure-onset aligned population time course of responsive units for M1 (top) and M2. Average MUA to auditory figures (red) and control condition with no coherent elements (black). Shaded regions represent the standard error of mean. Figure onset is indicated with dashed line. Chords are outlined in black below. Significantly different responses for Figure vs Control conditions are depicted in grey above (Wilcoxon-signed rank test, $p < 0.05$, FDR corrected). (**c**) Spatial maps of the recording field indicate the location in individual responsive units. (**d+e**) Figure-onset aligned population time course of responsive units for anterior (above) and posterior (below) recording field. (**d**) Average MUA to twelve (red), eight (green) and no coherent elements (black). Similar conventions as (**b**). (**e**) Color-coded boxplots show neuronal d-prime averaged across chords 5 to 8 (201 to 400ms) after figure onset. Stars indicate significance (Wilcoxon signed rank test): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

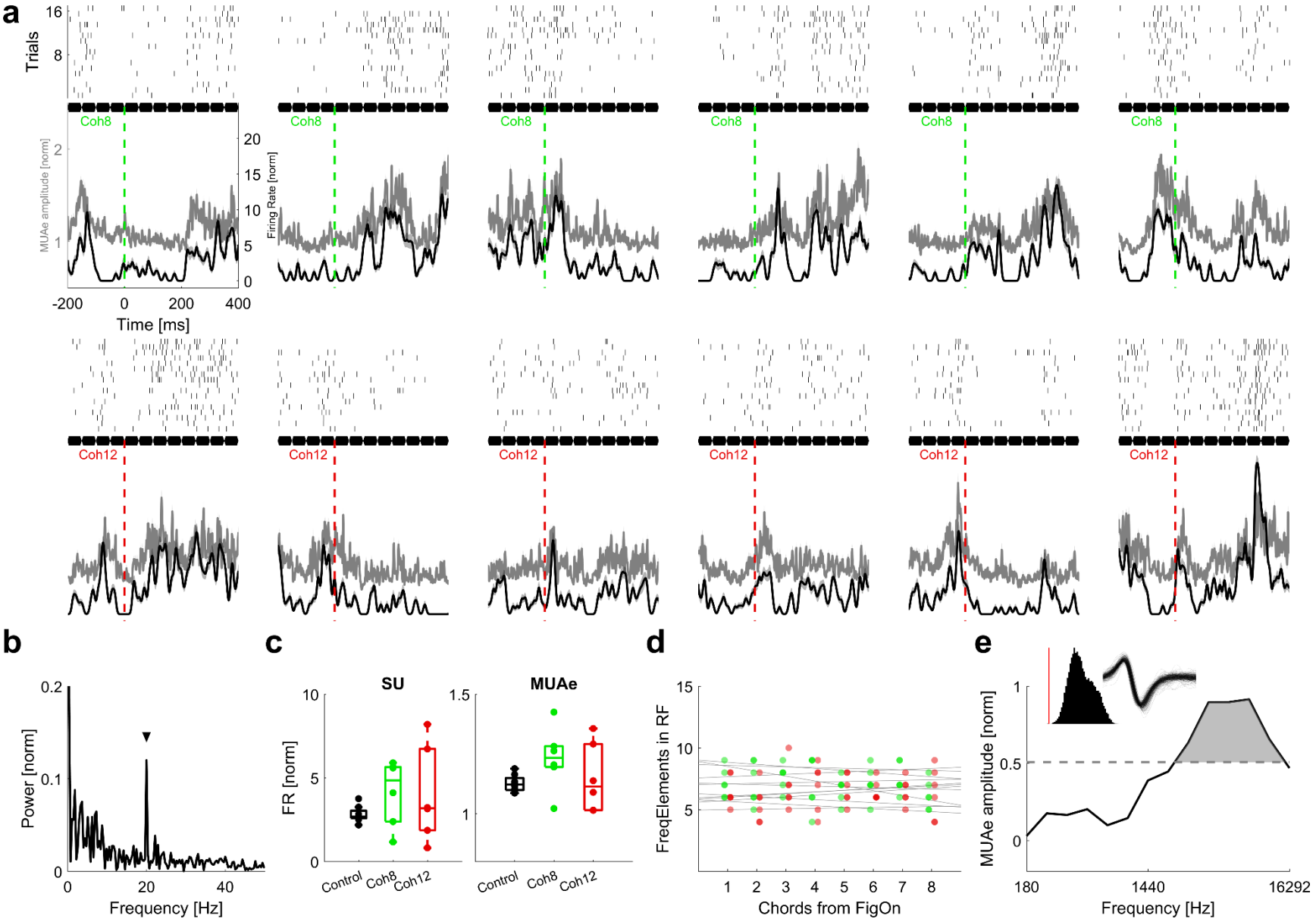
a
Trials
16
8
Coh8
Coh12
MUAe amplitude [norm]
Firing Rate [norm]
-200
0
200
400
Time [ms]
b
Power [norm]
Frequency [Hz]
c
SU
MUAe
FR [norm]
Control
Coh8
Coh12
d
FreqElements in RF
Chords from FigOn
e
MUAe amplitude [norm]
180
1440
16292
Frequency [Hz]

Figure 4.9: Spiking response of an example single unit to individual SFG stimuli.

Average response of an example single unit (SU, black) and the multi-unit envelope (MUA, grey) to individual SFG stimuli. (**a**) Twelve different test stimuli were presented, 50% of which contained figures with 8 coherent elements (top row). The remaining stimuli contained figures with a coherence level of 12 elements (bottom row). Different plots correspond to individual figure-onset aligned responses to different SFG stimuli. Figure onset is indicated by a dashed line at time zero. Raster plots are shown on top. Each row corresponds to a trial, each point within a trial to a single spike. Spike density functions and the envelope of MUA are shown below. Responses are baseline normalised, averaged over all trials. Plots are colour-coded according to figure coherence (coh8 = green; coh12 = red). Single chords of the SFG stimulus are indicated in black. (**b**) FFT of averaged SFG stimuli. Normalised to the maximum power. 20Hz peak (black arrow) indicates rhythmic responses to each presented chord. (**c**) Mean SU and MUA amplitude for stimuli of each condition averaged in time windows 200 to 400ms after figure onset. Due to the small sample size (n = 6 for figures stimuli, n = 8 for control stimuli) no statistical test is shown. (**d**) Quantification of frequency elements that fall into the frequency-response area of the unit for the first eight chords after figure onset shown for each stimulus. Coloured points show stimulus wise distribution for each chord. Grey lines demonstrate stimulus-wise linear regression of elements in RF. (**e**) Pure tone tuning curve of the single unit. Normalised to maximum response, then averaged across sound intensities. Dashed line indicates half maximum. “Responsive” area indicated in grey. Insets show the waveform and log-transformed inter-spike interval (ISI) histogram. The red line indicated an ISI of 1ms.

4.4.4 Response differences between anterior and posterior recording field

To investigate whether figure processing differs along the cortical hierarchy, I compared responses between cortical areas. Population activity differs between anterior and posterior recording field. The average response properties of the anterior region show strong similarities across monkeys. However, the temporal response dynamics in the posterior recording field varies drastically across subjects (Figure 4.8 d). In M1, the average MUA strongly resembles the response profile of the anterior recording field, suggesting that responses might indeed come from similar areas. Average response amplitudes of units located in the posterior recording field are higher but not significantly different (Normalised MUA: antAC = 1.0447 ± 0.1003; posAC = 1.061 ± 0.0817, 2-sample t-test: $p = 0.3882$). In contrast, the average MUA in the posterior field of M2 exhibits sharp onset transients to each presented chord after an average latency of 42ms. Response amplitudes are significantly higher compared to the anterior field (Normalised MUA: antAC = 1.0232 ± 0.0844; posAC = 1.0559 ± 0.0851, 2-sample t-test: $p < 0.01$). Taken together, this suggests marked response differences that might be caused by a more primary-like unit pool in the posterior part in contrast to a mixture of primary and non-primary units on the anterior recording field.

To further assess the observed modulation in population responses, d-prime values were calculated for each significantly modulated unit. When contrasting the average figure and control responses, MUA is significantly higher in stimuli that contain a figure across the recording field cluster (Figure 4.8 e, Wilcoxon signed rank test, $p < 0.001$ for all conditions). However, MUA modulation based on figure coherence is only found in anterior regions (Wilcoxon signed rank test, antAC: M1: $p < 0.01$, M2: $p < 0.001$; posAC: M1: $p = 1$, M2: $p = 1$), suggesting the encoding of perceptual saliency in the anterior auditory cortex. Furthermore, these results are confirmed by the area under the receiver operating characteristics (AUROC), a non-parametric measure of neuronal discriminability. Coherence-dependent modulation is mostly found in the anterior recording region (Figure 4.11). When pooled across fields, only onset-aligned data are significantly different from 0.5 (Figure 4.10, Wilcoxon signed rank test, Onset: M1: $p < 0.01$, M2: $p< 0.001$; Decision: M1: $p = 0.0628$,

M2: p = 0.0559), suggesting a transient encoding of information about perceptual saliency.

The number of frequency elements in the receptive field of the tested units does not significantly vary from chord to chord (Figure 4.12 a), demonstrated by the fact that on the population level the slope of a straight line fit (least squares) to the number of frequency elements across chords is not significantly different from zero (Wilcoxon signed rank test, p = 1 for all conditions). This further suggests that the observed effect on neuronal firing is driven by the temporal coherence of the figure elements.

There is no relationship between the figure-ground effect magnitude and the width of the tuning curve of the units (Figure 4.12 c, Wilcoxon signed rank test, p = 1 for all conditions), suggesting that broadband tuning does not facilitate figure-processing. However, multi-unit tuning might not reveal such effects.

The modulation latency is similar across anterior and posterior auditory cortex with no main effect of figure coherence (Figure 4.12 b, 2-factorial ANOVA [field x coherence], Field: $F_{1,\ 629} = 2.48$, p = 0.116, Coherence: $F_{1,\ 629} = 1$, p = 0.3179), suggesting similar processing timescales for highly salient auditory figures.

Table 4.2: Rounded median figure-ground modulation latency [ms] for both subjects. Shown for each condition and recording field with standard error of mean.

	ANT		POS	
	Coh8	**Coh12**	**Coh8**	**Coh12**
M1	95 ± 11	98 ± 10	96 ± 12	109 ± 15
M2	143 ± 10	121 ± 10	117 ± 8	99 ± 7

4.4.5 Figure-ground modulation without behavioural detection

Neuronal responses in hit and miss trials were compared to investigate if elevated spiking responses do occur even without behavioural detection. MUA in hit trials is higher compared to miss trials in one animal (Figure 4.10, Wilcoxon signed rank test, Onset-aligned: M1: p < 0.001, M2: p = 1; Decision-aligned: M1: p < 0.001, M2: p = 0.1), suggesting that higher cortical activity can cause enhanced object perception. A median split of reaction times revealed that, for some units, MUA differs between slow and fast reaction time trials (2-sample t-test, p < 0.05, ANT: M1 = 31.25%, M2 = 21.89%; POS: M1 = 23.53%, M2 = 25.34%). Most of the concerned

units (ANT: M1 = 100%, M2 = 98.51%; POS: M1 = 91.76%, M2 = 100%) had higher population activity in fast reaction time trials compared to trials with slow reaction times. This further suggests a link between auditory cortical population responses and perception.

MUA between correct rejection (CR) trials and miss (MI) trials is significantly different in both subjects (Figure 4.10, Wilcoxon signed rank test, $p < 0.001$ for all conditions). Cortical population responses in miss trials are higher, suggesting that even without behavioural detection, neurons respond to coherent elements. Thus, primary auditory cortex can detect auditory figures even when subjects are distracted or during low arousal periods (Figure 4.11 d, see M2, posterior field).

Recordings with single contact electrodes are oftentimes biased towards the superficial cortical layers. This problem can be circumvented with (linear) electrode arrays. Data in this chapter has been recorded with both electrode types. Depth information, aligned to the first sound-responsive site, show that figure-responsive units seems to be relatively evenly distributed across the cortical depth (Figure 4.13). For multi-contact recordings (Linear 16-channel V-Probe, Plexon), structural MRI scans were used to align the recording coordinates with the brain in order to estimate the penetration angle of the probe. Due to the angle of the recording chamber (Medial tilt, M1: 10 degrees, M2: 15 degrees) and the curvature of the superior temporal gyrus (see Figure 4.1 f and Figure 4.6), most of the electrodes did not penetrate the auditory cortex perpendicularly (Table 4.3). Thus, no inferences can be made about layer-specific effects.

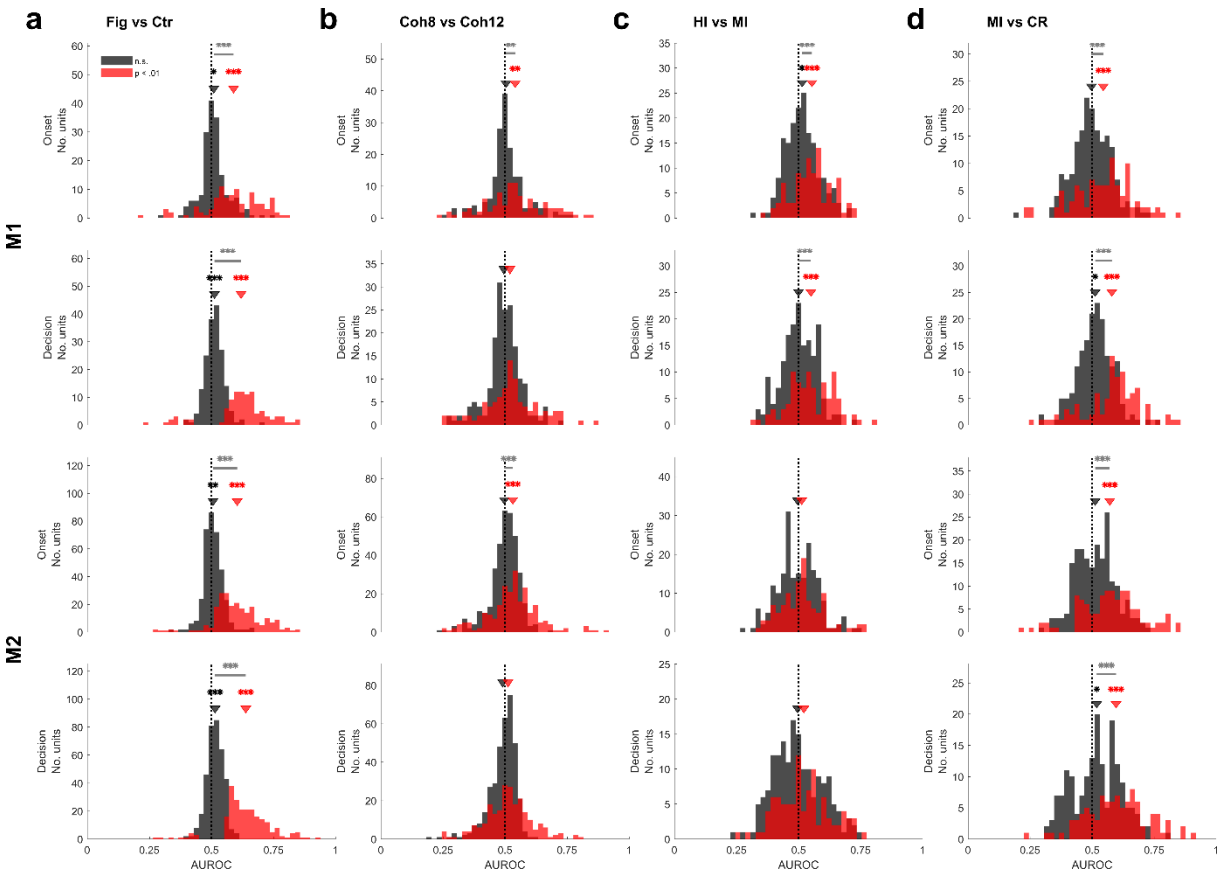

Figure 4.10: Auditory cortical response modulation for M1 (above) and M2 (below).

Histograms show distributions of area under receiver operating characteristic (AUROC) for significantly modulated ($p < 0.01$, red) and unmodulated units (black). Averages of onset-aligned (top row, 201 to 400ms) and decision-aligned data (bottom row, -300 to – 100ms) were used for AUROC calculation. (**a**) Figure-ground modulation. (**b**) Modulation based on figure coherence. (**c**) Modulation based on behavioural detection of temporally coherent elements. Only recordings with at least 20 MI trials were included. (**d**) Modulation based on temporal coherence without detection. Arrows indicate mean of distribution. Data were tested against 0.5 with a Wilcoxon signed rank test. Black and red distribution were tested against each other with two-sided Wilcoxon rank sum test. Colour-coded stars indicate significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

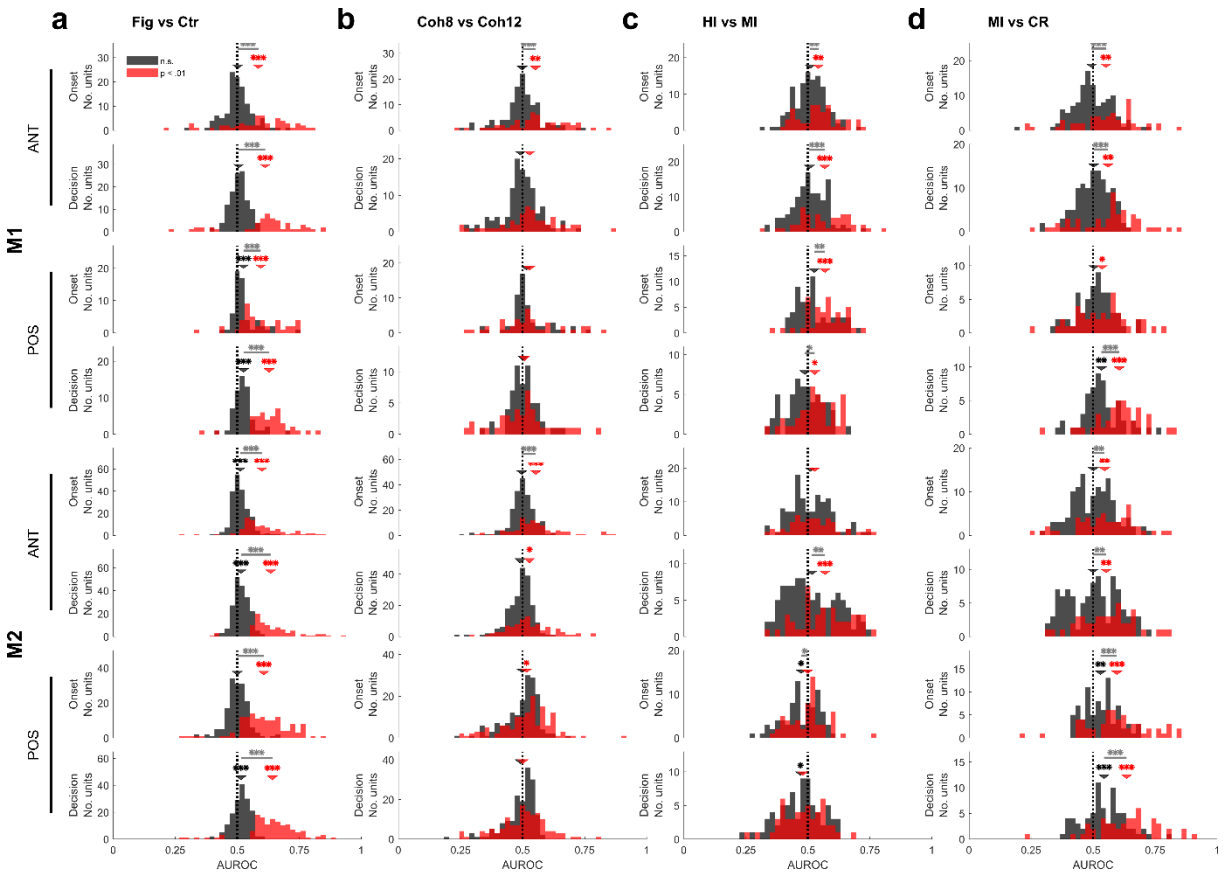

Figure 4.11: Response modulation illustrated for anterior (top) and posterior (bottom) recording field of each monkey.

Similar conventions as in Figure 4.10

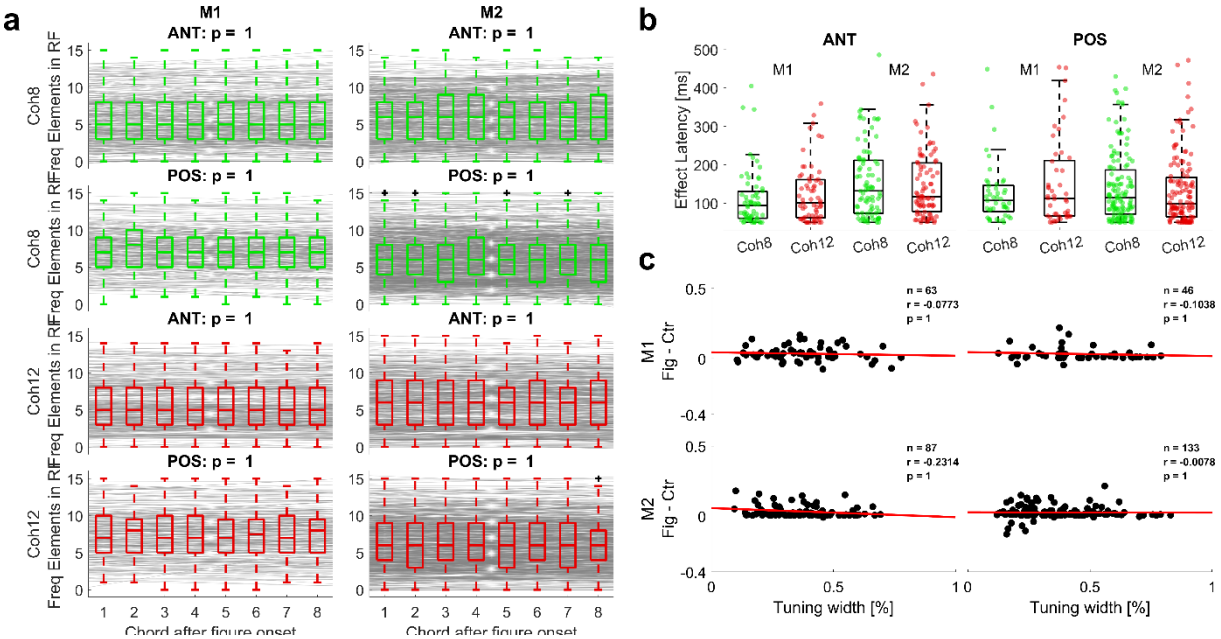

Figure 4.12: Summary of control analyses.

(**a**) Quantification of frequency elements that fall into the frequency-selective area (receptive field, RF) of each unit for the first eight chords after figure onset. Data are shown for each coherence level (coh8: green, coh12: red), recording field (anterior vs posterior) and subject. The half-maximum was used to classify pure tone responses into either the frequency-selective or unresponsive category (see Figure 4.9 e). For each individual stimulus, the number of frequency elements that fall into the RF of the unit was counted for the first eight chords after figure onset, which was the basis for a linear regression (shown in grey here, see Figure 4.9 d). The slopes of this regression were tested against zero (Wilcoxon signed rank test). The resulting FDR-corrected p-values are displayed above. (**b**) Figure-ground effect latency for anterior (left) and posterior auditory cortex (right) for both subjects. Effect latencies for figures with 8 (green) and 12 coherent elements (red) shown for units with significant figure vs control response ($p < 0.01$). Latencies were determined by bootstrapping the mean firing rate of each time bin (5000 repetitions). The effect onset was defined as the first significantly different time bin that was followed by at least four additional significantly different time bins (5ms, $p < 0.01$). The smallest possible latency value was set to 50ms (Duration of one chord). Raw data points are shown for sessions where this procedure was able to extract a latency value. (**c**) Correlation between tuning width and figure-ground modulation for M1 (top) and M2 (below). Anterior (left) and posterior field (right) shown. MUA differences between figure and control condition plotted as a function of tuning width, expressed as percentage of the tested frequency space. Each point represents one unit. Data are pooled across coherence conditions. The red line shows the least-squares line. Correlation coefficient and p-value are shown within each plot.

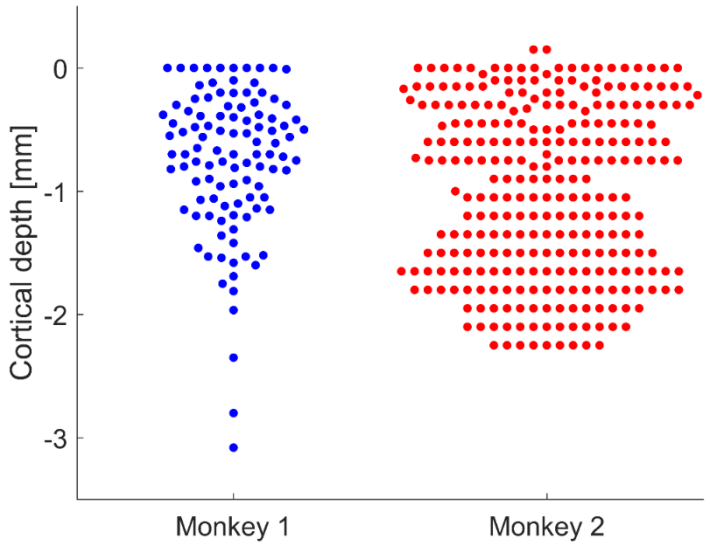

Figure 4.13: Cortical depth of figure responsive units for both subjects.

Depth coordinates are aligned to the first sound-responsive site. Anatomical landmarks (laterals sulcus), natural sounds and noise bursts were used to determine if the auditory cortex was reached. Superficial locations correspond to upper layers of the auditory cortex.

Table 4.3: Penetration angle of electrode array for each recording session. In most session two 16-channel probes were used. Inter-electrode spacing was 150um.

Recording ID	Penetration angle [deg]		
	Sagittal [AP]	Coronal [ML]	Probe
2019-06-17_15-44-25	77	53	16-chan Plexon
	95	60	16-chan Plexon
2019-06-18_09-49-48	50	65	16-chan Plexon
	90	35	16-chan Plexon
2019-06-19_11-21-28	53	59	16-chan Plexon
	96	50	16-chan Plexon
2019-07-02_11-59-08	60	55	16-chan Plexon
	90	74	16-chan Plexon
2019-07-03_11-01-30	75	72	16-chan Plexon
	95	49	16-chan Plexon
2019-07-04_14-20-16	84	52	16-chan Plexon
	95	77	16-chan Plexon
2019-07-05_11-56-10	77	63	16-chan Plexon
	96	61	16-chan Plexon
2019-07-08_11-50-04	72	59	16-chan Plexon
	96	45	16-chan Plexon
2019-07-09_10-12-54	61	63	16-chan Plexon
	95	60	16-chan Plexon
2019-07-10_11-22-42	78	66	16-chan Plexon
	90	45	16-chan Plexon
2019-07-11_13-33-44	73	54	16-chan Plexon
	96	45	16-chan Plexon
2019-07-12_11-16-08	58	61	16-chan Plexon
	83	48	16-chan Plexon
2019-07-15_12-02-58	50	65	16-chan Plexon
	101	52	16-chan Plexon
2019-07-16_11-30-46	56	70	16-chan Plexon
	107	42	16-chan Plexon
2019-07-17_11-43-09	45	75	16-chan Plexon
2019-07-18_11-28-26	76	55	16-chan Plexon
	100	42	16-chan Plexon
2019-07-19_10-44-01	63	58	16-chan Plexon
	90	35	16-chan Plexon
2019-07-24_11-12-09	74	56	16-chan Plexon
	90	45	16-chan Plexon
2019-07-25_11-27-24	75	72	16-chan Plexon
	77	47	16-chan Plexon

4.4.6 Coherence-dependent LFP power differences

LFP activity was analysed to investigate if figure-ground segregation impacts oscillatory responses. Time-frequency decomposition of control stimuli shows enhanced power at 20Hz and 40Hz (Figure 4.14 a). In addition, inter-trial phase coherence is higher at 20Hz and harmonics (Figure 4.14 b). In line with the spectral decomposition of the MUA signal (Figure 4.8 a), this confirms population responses to individual chords of the SFG stimulus.

The decision aligned time-frequency difference between hit and correct rejection trials across monkey's shows changes in alpha (7-12Hz), beta (13-30Hz) and gamma band (>30Hz). Across the auditory cortex, alpha/beta power is significantly suppressed shortly before the decision, whereas gamma power is enhanced (Wilcoxon signed rank test, Alpha/beta: Mean power difference = -1.3804 ± 0.04, $p < 0.001$; Gamma: Mean power difference = 0.6375 ± 0.02, $p < 0.001$). Critically, these power differences cannot be explained by the touch bar release of the monkey (Figure 4.14 d). A permutation-based significance analysis for each pixel revealed similar changes across monkeys and recording subfields and suggests coherence dependent modulation of oscillatory power across frequency bands (Figure 4.15 and Figure 4.16). Based on the observed activity pattern, two 200ms long windows were drawn in the alpha/beta- and broadband gamma range (Figure 4.14 c). Main effects of figure coherence were found for both alpha/beta and gamma range (2-factorial ANOVA [Coherence x Field], Alpha/Beta: Coherence: $F_{1,1291} = 33.57$, $p < 0.001$, Field: $F_{1,1291} = 14.61$, $p < 0.001$; Gamma: Coherence: $F_{1,1291} = 10.55$, $p < 0.01$, Field: $F_{1,1291} = 9.78$, $p < 0.01$). Post-hoc tests suggests that coherence-dependent alpha/beta modulation seems to be present across auditory cortex whereas gamma power differences seem to be present in the anterior recording field only (Figure 4.17).

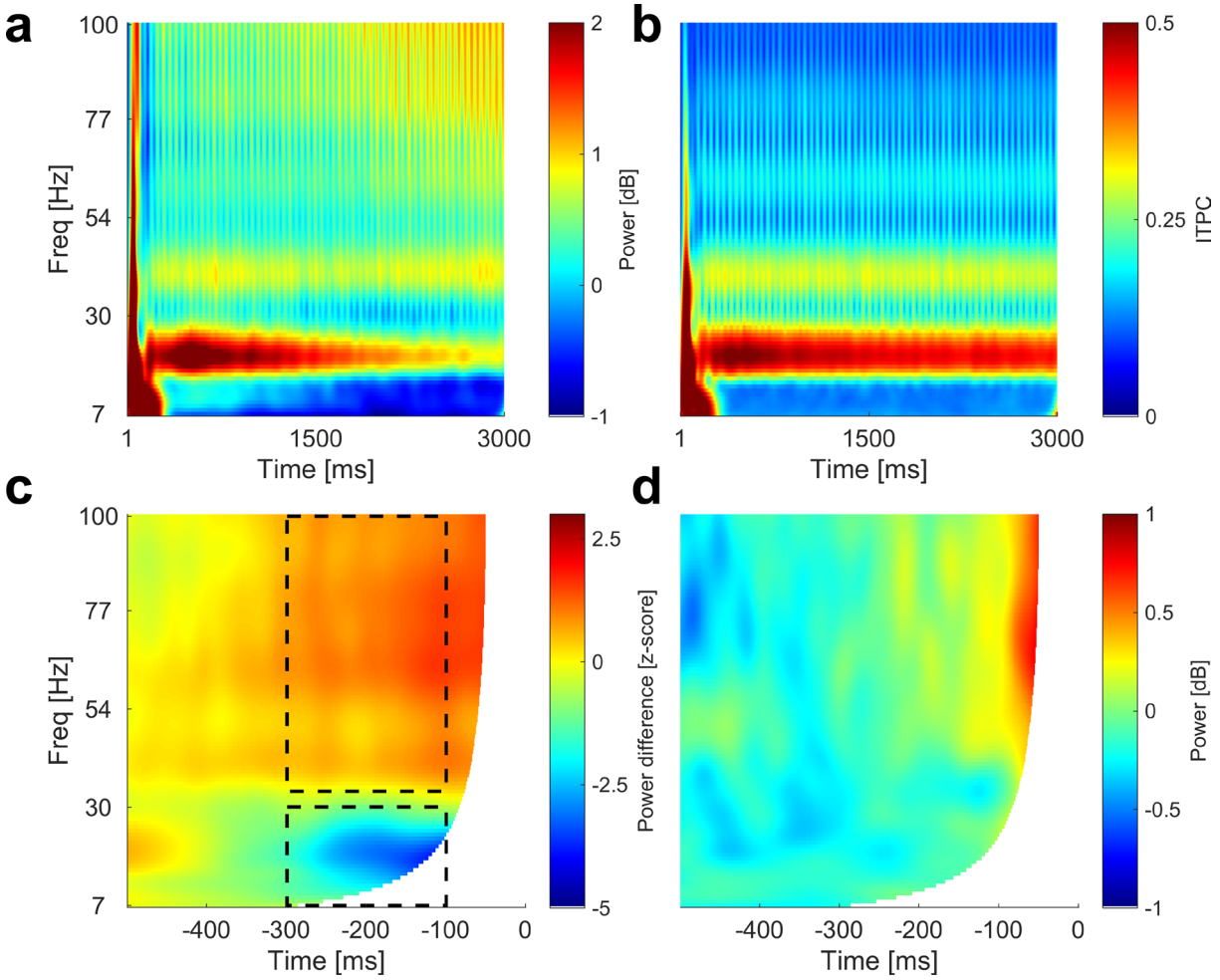

Figure 4.14: Summary of time-frequency decomposition averaged across both monkeys.

(**a**) Baseline normalised, decibel-converted time-frequency power across all control stimuli. (**b**) Inter-trial phase coherence across all control stimuli. (**c**) Decision aligned time-frequency analysis across coherence conditions for difference between hit and correct rejection trials. ½ wavelet was excluded to avoid motion artefact contribution. Dashed boxes indicate position of window of interest for alpha/beta (7-30Hz) and gamma range (35-100Hz). (**d**) Self-paced bar release control experiment with M2. The monkey was rewarded for touch bar release without any auditory stimulation. The plot shows the average movement aligned LFP response of well-driven auditory multi-units (n = 30) from across the recording field. Each frequency band was individually normalised by the mean power across trials in the time window 1000ms prior to response. ½ wavelet was excluded to avoid motion artefact contribution.

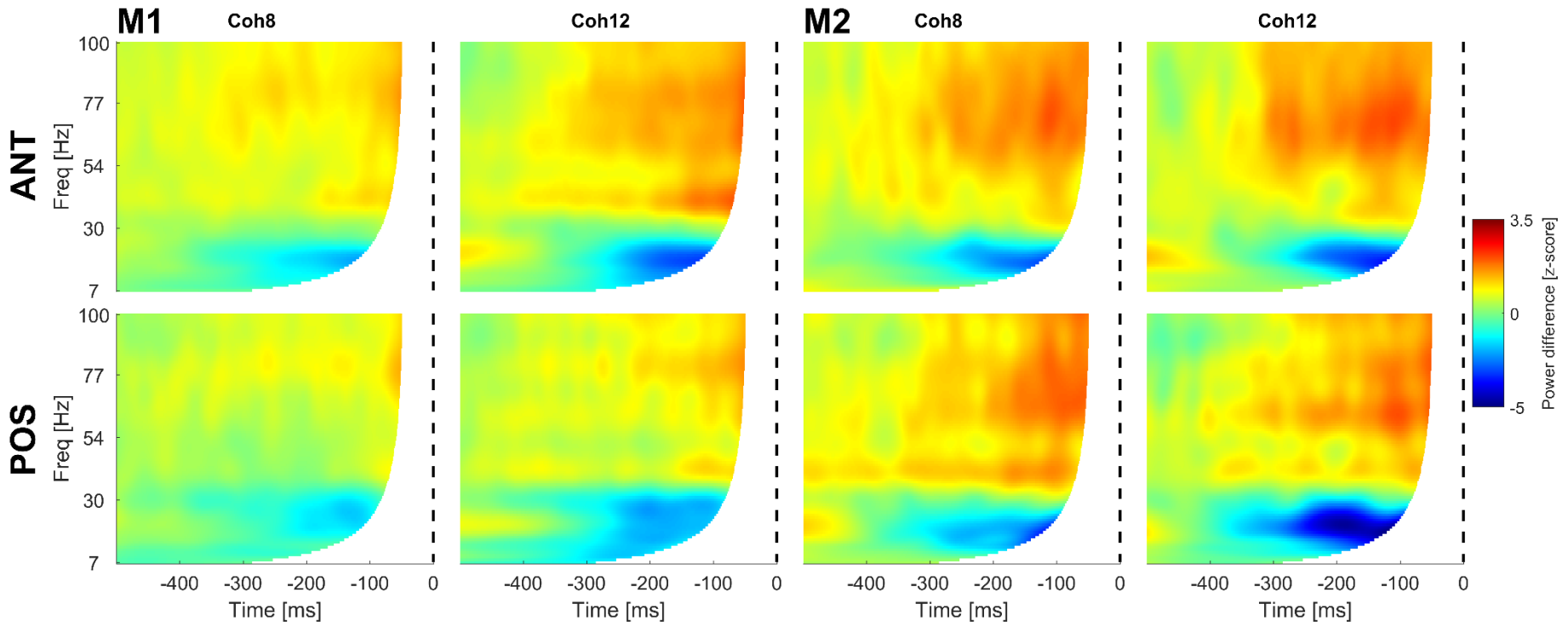

Figure 4.15: Individual time-frequency analysis for both subjects and coherence conditions.

Decision aligned time-frequency decomposition of hit minus correct rejection trials [z-score] for M1 (left) and M2 (right) shown for recording channels in anterior (top row) and posterior (bottom row) recording field.

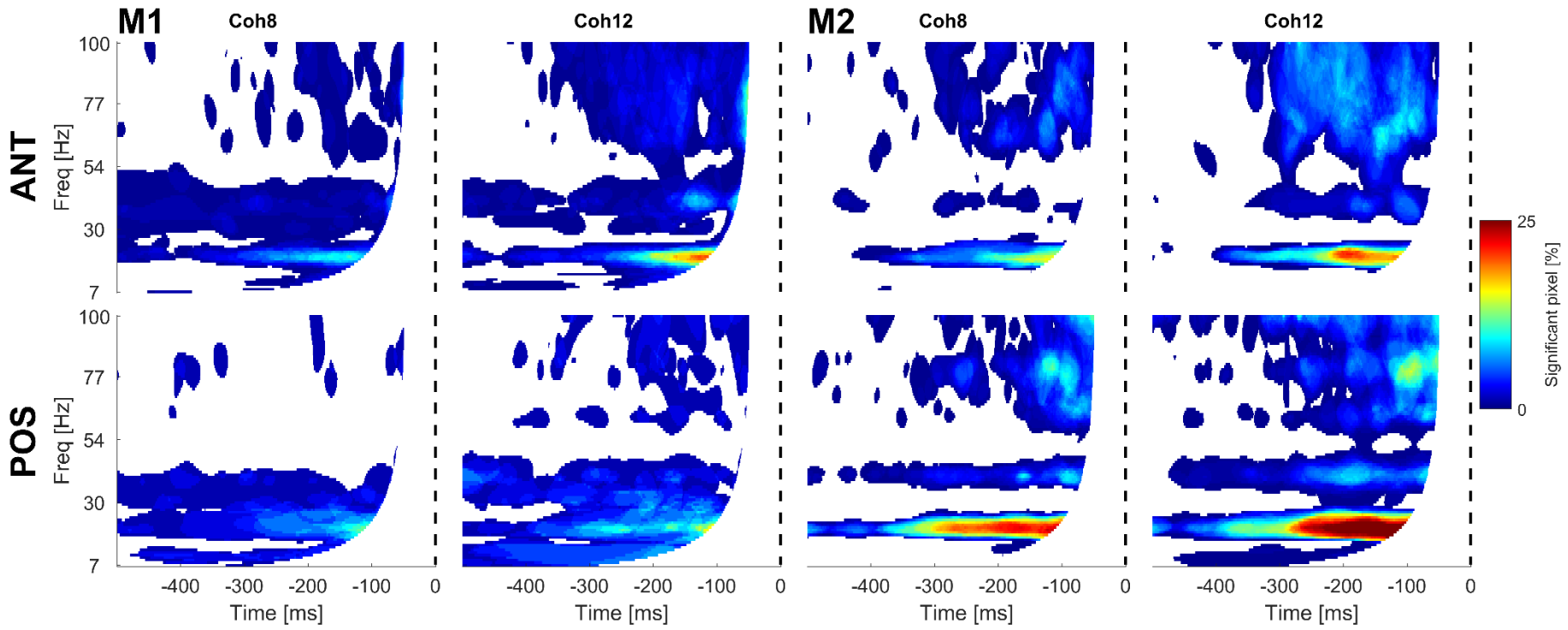

Figure 4.16: Fraction of significantly activated pixels for decision aligned time-frequency analysis.

Permutation-based significance testing based on extreme values was conducted for each channel. Same conventions as Figure 4.15.

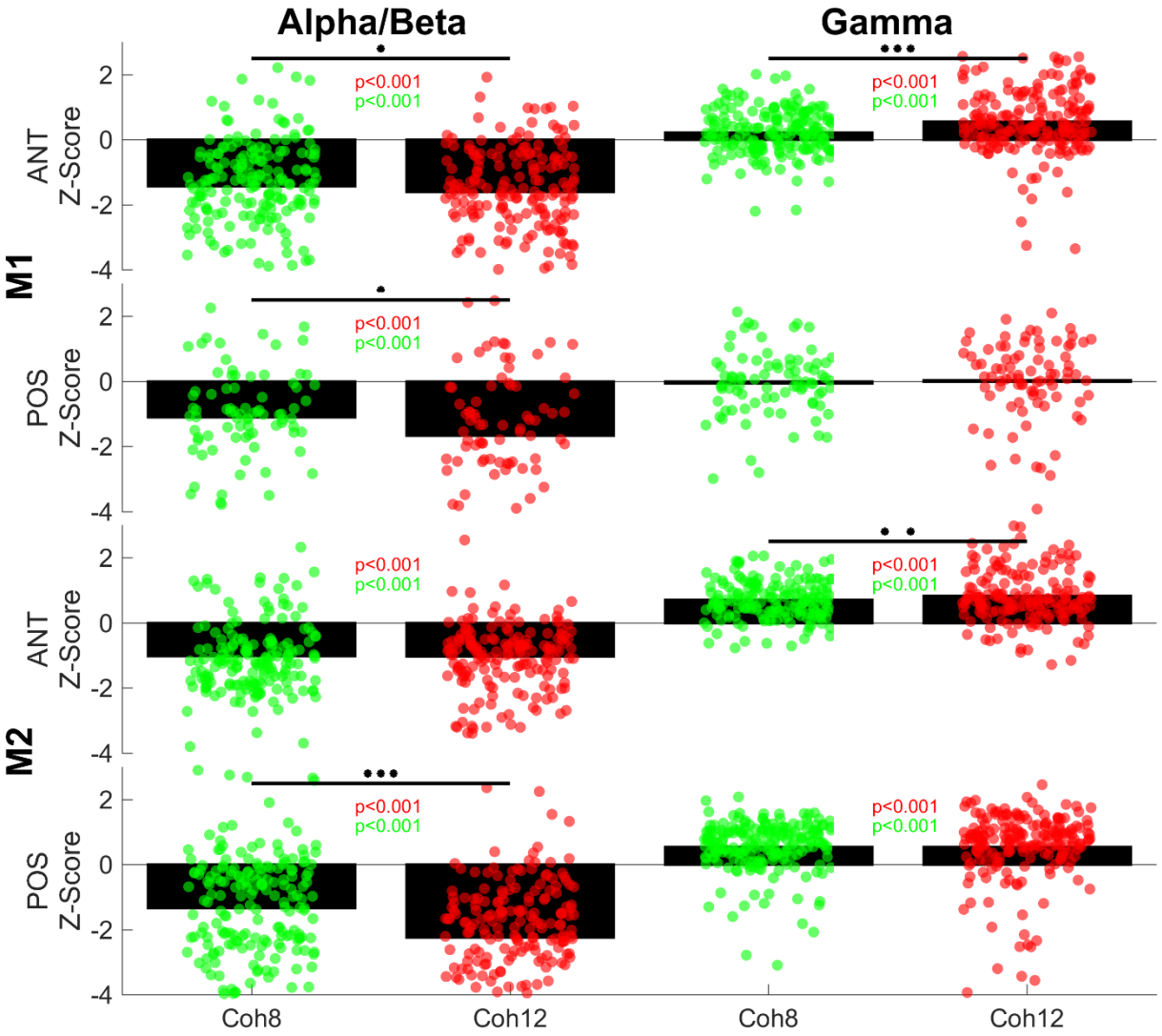

Figure 4.17: Average power difference (hit minus correct rejection) in pre-defined time windows in alpha/beta (left) and gamma range (right) shown for M1 (top) and M2 (below).

Bar plot shows average power in predefined time windows across recording channels. Coherence level is colour coded (Coh8, green; Coh12, red). Average power of each recording channel overlaid as grey circle. P-value for each condition is shown colour-coded next to bar (Wilcoxon-signed ranked test against zero). Test result of Wilcoxon signed rank test between coherence levels within one area shown in black above. Stars indicate significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.5 Discussion

This chapter reports the first neurophysiological data of auditory cortical activity in response to complex Stochastic Figure-Ground stimuli. Here, I have demonstrated that multi units across the auditory hierarchy, including the primary auditory cortex, can detect auditory figures. Furthermore, I established that figure coherence is an important factor that impacts neuronal responses differently across cortical subfields. Lastly, I showed that auditory figures cause figure-ground modulation even without behavioural detection.

Stochastic Figure-Ground stimuli are rapidly changing, broadband signals that strongly drive cortical activity. Temporal response dynamics reveal that multi-units can robustly track individual complex chords. This phasic response shows high chord-to-chord amplitude variability that is likely due to the frequency selectivity of the unit and subtle changes in sound intensity due to the nonlinearity of the speakers. Strong variability to broadband stimuli (random and natural) have been shown to induce high variability over time that occurs with high temporal precision (Kayser et al., 2010). Response variability to different auditory figures also suggests object specific feature selectivity that contributes to sound segregation. Generally, MUA seems to slightly ramp up over time (Figure 4.8 a), even in the absence of figures, which might be an indication of reward expectancy or movement planning. Reward feedback has been shown to be represented in auditory cortex (Brosch et al., 2011). Likewise, strong motor-related (but inhibitory) modulation of auditory responses has been shown (Schneider and Mooney, 2018; Schneider et al., 2014, 2018a). Since a touch bar release is part of the trial structure, motor preparation might partially explain the ramping population response.

Auditory figures produce robust changes in neuronal firing for a subpopulation of multi-units (Figure 4.8). The figure-evoked increase in MUA is not caused by changes in the frequency content of the receptive field (Figure 4.12). Hence, grouping cues like temporal coherence or changes in stimulus regularity are likely to be the driving force behind the changes in cortical responses. Responsive units are widely distributed across the auditory cortex with no indication of spatial clustering (Figure 4.8). This indicates that figure-ground segregation is not just the product of a higher order subfields in the auditory cortex. In fact, the data demonstrate that even A1

neurons can detect temporally coherent frequency elements that have no simple mathematical relationship to each other. Earlier studies have argued that stream segregation depends on tonotopically organised responses in primary auditory cortex (Fishman et al., 2001, 2004). Data reported in this chapter suggests that the A1 population can indeed segregate target sounds from irrelevant information. However, a population separation model would not explain these results since temporally coherent, randomly chosen frequency elements could theoretically cause responses along the entire tonotopic gradient at the same time. Using streams of repeated noise embedded in random noise, Saderi et al. (2020) suggested that neurons in the primary auditory cortex exploit gain differences between foreground and background representations to achieve figure-ground segregation (Saderi et al., 2020). According to this study, repeated noise elements generally cause stimulus-specific adaptation, however, foreground representations were enhanced compared to the background. They argue that perceptual grouping emerges in A1 and is refined in downstream cortical areas. Other studies have also demonstrated an involvement of primary auditory fields in target representation (Christison-Lagay et al., 2017; Mesgarani et al., 2014) suggesting A1 as critical component in sound segregation. These findings are in line with results presented in this thesis. However, our stimuli cause, on average, a tonic increase in population activity without stimulus-specific adaptation. Whether the observed A1 MUA modulation presented here indicates regularity detection, coherence analysis or already a full object representation cannot be determined based on this experimental design. However, the distributed nature of responsive units and the qualitative difference in coherence processing suggests that figure-ground segregation will likely depend on a complex interplay between bottom-up and top-down processing.

Auditory responses to synchronous (Lu et al., 2017) and regular tone sequences (Barascud et al., 2016) require a build-up period. Similar mechanism will likely be involved here where integration over the spectrotemporal space has to happen before a percept can arise. Figure-ground effect latencies in the auditory cortex vary vastly between units. On average, it takes two chords for most units to detect the changes in stimulus statistics, which is in the same temporal range to previously reported MEG and EEG data (O'Sullivan et al.,

2015a; Teki et al., 2016). However, modulation latencies can be higher than 400ms (8 chords). This indicates that some units respond on different timescales, which could be due to a cortical circuit, feedback connections or simply a small overall figure-ground effect that makes latency estimation imprecise. Similar modulation latencies (Figure 4.12) across the auditory cortex suggest no coherence-driven differences in figure processing.

It has been argued before that figure-ground segregation relies on a pure bottom-up mechanism (Teki et al., 2011). This might be true for highly salient sound objects (like an explosion) or in complete silence. However, in real-world acoustic scenes, where sensory systems are constantly confronted with input, it is likely that most of the time top-down attention is involved in perceptual decision making. Impaired figure-ground perception has been demonstrated during high visual load (Molloy et al., 2018), linking perceptual organisation to available cognitive resources. This work demonstrates firing rate differences between behavioural categories (hit vs. miss trials), which suggests that early cortical population activity influences perceptual detection. SFG stimuli have been passively presented in previous experiments and no significant BOLD change was found in primary core regions (Schneider et al., 2018b; Teki et al., 2011, 2016). This suggests that primary cortical population activity is a crucial component for behavioural detection. Furthermore, data presented here demonstrate different population responses between miss and correct rejection trials. Higher MUA in miss trials indicates that cortical neurons can detect temporally coherent elements independent of behavioural detection. These findings support previous functional imaging data shown in chapter 3 (Schneider et al., 2018b), where non-primary cortical figure-ground modulation was demonstrated without any task engagement. Task engagement has been shown to have dramatic effects on the neuronal representation of sound. Selective attention generally causes firing rate suppression compared to the passive presentation of identical stimuli (Bagur et al., 2018; Otazu et al., 2009). This has also been shown for MEG responses during the presentation of regular and random patterns (Sohoglu and Chait, 2016a). Since no attentional suppression effect has been found here, it remains uncertain if the observed increase in MUA corresponds to enhanced target representation in the same way. However, auditory regularity causes enhanced MEG responses even when attention is drawn toward visual distractor stimuli (Barascud et al., 2016), suggesting different response profiles for the processing of complex sounds like the SFG stimulus. Taken

together, the comparison between functional imaging data and electrophysiology has to be interpreted with caution. Based on the data in this thesis it remains uncertain whether passive presentation of SFG signals in naïve animals causes an actual percept of an auditory object.

A crucial difference of this experiment compared to previous imaging work is the redesigned SFG stimulus. Earlier imaging studies have used randomly varying number of frequency elements between chords. Temporally coherent (figure) or random (control) elements were added on top of this varying signal. Here, in order to avoid intensity differences between individual chords, the number of frequency elements was capped to 15 frequencies per chord. Coherent figure elements were incorporated into and not added to the existing stimulus. This led to an improved stimulus presentation with relatively reliable sound intensity from chord to chord. The downside of this change in stimulus design is that the observed figure-ground effect can be attributed to either temporal coherence or changed target-to-masker ratio (i.e. less randomness), as changes in stimulus regularity have been shown to affect cortical responses (Barascud et al., 2016; Sohoglu and Chait, 2016a). Even though I cannot fully disentangle the underlying cause of the neuronal response, the interpretation regarding figure-ground segregation remains the same: The auditory input changes in some dimension which causes segregation and grouping and eventually a perceptual pop-out effect.

This chapter also confirms that the number of coherent figure elements has an impact on the cortical response which has been shown in earlier studies (O'Sullivan et al., 2015a; Teki et al., 2016; Tóth et al., 2016). Since Teki and colleagues found a parametric effect of figure coherence (Teki et al., 2011), I have argued in the previous chapter that auditory figures might be encoded with a rate code and that higher figure coherence causes higher increments in neuronal responses. That is indeed what the data show. Although multi-units in the posterior part of the recording field show on average no difference between figure coherence levels, anterior AC shows a link between the number of coherent figure elements and firing rate. Thus, information about the perceptual saliency of the figure is already represented in the anterior parts of the auditory cortex. Again, whether this 'saliency' is evoked by more regularity or temporal coherence with respect to the random background cannot be determined based

on this experimental design. However, an earlier study with different stimulus design shows similar effects that have to be based on temporal coherence (Teki et al., 2016).

I have also argued that neurons with broad tuning curves might respond in a non-linear way to temporally coherent frequencies but might not show a similar selectivity to the individual frequencies that comprise the figure. Data reported in this chapter show that there is no link between the width of the pure tone tuning curve and the magnitude of the figure-ground effect, suggesting different integration mechanisms. However, multi-unit tuning might not reveal such effects.

The analysis of LFP's suggest power suppression in the alpha/beta range and enhanced power for gamma oscillations prior to behavioural response. This is in line with visual figure-ground experiments (van Kerkoerle et al., 2014), where a figure in the receptive field causes power suppression in the alpha range for the figure minus ground difference. Furthermore, this paper also demonstrates increased power in the gamma range. Only weak beta modulation was found in this study. In addition, higher LFP power in the gamma band was associated with stronger multi-unit activity (van Kerkoerle et al., 2014). This indicates that oscillatory cortical activity responds in a similar fashion across sensory modalities.

The areal assignment of M1's recording field is not entirely clear. Histological investigation partially confirms probabilistic maps of cortical subfields suggesting that most recordings were done in the primary core. However, there is a discrepancy regarding the accessible area of A1. Probabilistic maps (Figure 4.7) clearly suggest that nearly all of A1 should have been accessible. Functional indicators like tonotopy (Figure 4.1) and histological brain slices (Figure 4.6), however, suggest that only a small region in the low frequency area of A1 was accessible. The reason for this discrepancy between MRI and histology/electrophysiology could be due to the way the probabilistic maps are created and used to approximate the location of the auditory subfield. Maps are estimated (in AP direction) based on the tonotopic gradients of a small number of other subjects (not included here) that are co-registered to the standard space. Tonotopic gradients can show high inter-subject variability. Moreover, M1's brains had to be morphed into the same space to allow areal assignment (see chapter 3). Thus, there might be considerable room for error. In this specific case, histology and functional response properties of the recorded cells are the most reliable indicator of the actual recording sites. Based on that information, most recordings in M1 were conducted in the primary core. However, only a small fraction of the low

frequency part of A1 was accessible and the bulk of the data was recorded in area R, RT and the adjacent lateral belt. That could also mean that M1's division boundary might not separate A1 and R but could be more anteriorly located within the rostral field. Overall, the available information indicates that the recording field in M1 was more anterior compared to that in M2. Since figure responses in M1's posterior field show a trend towards scaled responses depending on the figure coherence, I hypothesise that this saliency information gradually arises from posterior auditory field towards the anterior regions.

The same monkeys that were behaviourally tested in chapter 2, were used for electrophysiological recordings. Behavioural performance reported here was improved compared to data shown in chapter 2, which is an indication that subjects were still on a learning curve while behavioural data was acquired. I have argued earlier that there might be a difference in the sensitivity to auditory figures between macaques and humans (Schneider et al., 2018b). Behavioural results in this chapter suggest that monkeys have similar perceptual detection performance to humans. Thus, there is no species difference regarding figure-ground segregation thresholds.

5 General discussion

In this thesis, I have investigated figure-ground processing on a behavioural, systemic and neuronal level in an animal model, the rhesus macaque. In chapter 2, I have described the behavioural detection performance of macaque monkeys, where I have demonstrated similar perceptual performance to humans. Moreover, I presented a reaction time analysis in response to SFG stimuli, showing marked effects of figure coherence. In chapter 3, I have evaluated the systemic organisation of BOLD signal changes during passive, stimulus-driven figure-segregation. In line with human imaging results, I have shown involvement of non-primary auditory sites during the presentation of stimuli with an auditory figure. Because the macaque cortex is much better understood than the human counterpart, I could precisely locate the focal point of this cortical modulation to the anterolateral belt and parabelt. Chapter 4 showed neuronal responses to SFG stimuli. I have reported increased population activity in response to auditory figures for a subset of neurons across the cortical hierarchy. This specifically includes core areas A1 and R. Furthermore, I have revealed differences in object processing, with anterior cortical areas exhibiting a modulation of MUA based on figure coherence. Together, these results indicate similar figure-ground processing across species and suggest specific cortical processing strategies. In this section, I will interpret and relate these findings to our current knowledge of auditory scene analysis.

5.1 Temporal scale of figure ground segregation

Behavioural results (chapter 2 + 4) have revealed similar figure detection performance between macaques and humans. Across primate species, higher coherence causes enhanced detection (O'Sullivan et al., 2015a; Teki et al., 2011, 2013; Tóth et al., 2016). Thus, one can assume that the perception of these complex signal is similar. Whether the underlying segregation cue is the same cannot be determined due to the changed stimulus design, however, this should not affect the central processing of the figure once segregation has occurred. For complex acoustic scenes, the default setting of the auditory system is to assume a single sound stream until enough evidence has been gathered for the segregation of sound sources (Moore and Gockel, 2012). For two tone stream segregation, the probability of fission into two streams seem to increase with time and depends on a variety of parameters like

presentation rate and frequency separation between tones (Moore and Gockel, 2012). There is no reason to assume that the segregation of sound objects from a complex scene would not require a build-up period. Complex, regular sounds induce a significantly different neural responses after about 1.5 cycles (Barascud et al., 2016). For figure-ground segregation, EEG and MEG experiments suggest an evoked figure response after 150-200ms for highly salient figure, with longer latencies for lower coherence level (O'Sullivan et al., 2015a; Teki et al., 2016). In line with those studies, electrophysiological experiments (chapter 4) revealed significantly modulated firing rates after about 2-3 chords after figure onset. This confirmed that the auditory system can reliably detect coherent or regular elements after only a few chords/cycles.

I have demonstrated clear reaction time differences between coherence levels, with longer response times for lower coherence levels (Figure 2.3 and Figure 4.1). Median reaction times for the most salient figure condition was around 550ms for both subjects. However, significantly different population responses can be found much earlier (Figure 4.12). Given an average motor response to pure tones after approximately 300ms (at 70dB SPL, Pfingst *et al.*, 1975), additional processing steps downstream of the auditory cortex seems to delay the behavioural responses. This could imply a recurrent processing loop where evidence needs to be integrated over time. Neurons in the auditory cortex might then signal the presence of an object and attentional selection is done in higher cortical centres like the IPS before a decision can be made. In addition, the coherence-dependent scaling of MUA suggests that other cortical centres downstream of the superior temporal plane are able to detect less salient auditory figures that are not identified by core or belt regions. The parabelt region is likely a good candidate for this task that might provide reliable object representations on which attentional selection can be done.

No latency difference between coherence levels was found, suggesting equal time scales of cortical processing independent of the magnitude of change in stimulus statistics. Given the increasing reaction times to less salient figures, it seems that more evidence needs to be acquired by the auditory system before a stable object percept occurs.

Taken together, figure-ground segregation is a demanding cognitive process which requires a build-up period that seems to be modulated by

attention (O'Sullivan et al., 2015a). Reaction time distributions and figure-ground effect latencies suggest multiple, distributed steps of processing with auditory cortical responses at the beginning of the processing chain.

5.2 Cortical coding of auditory figures

Previous human imaging studies (Teki et al., 2011, 2016) as well as fMRI experiments presented here (chapter 3) strongly suggest the involvement of non-primary auditory cortex during passive, stimulus-driven figure-ground segregation across primate species. In both monkeys, the maximum BOLD change was observed in the anterolateral belt and parabelt regions (Figure 3.4). No significant change in BOLD signal was found in primary auditory cortex or subcortical structures. Thus, the data seemed to suggest that a combination of neurons in primary core cortex with high frequency selectivity for one or multiple frequencies project to single cells in anterolateral belt and parabelt areas. Those cells would then integrate responses across the frequency spectrum and detect temporally coherent elements. Electrophysiological results contradicted this idea. Extracellular recordings across the auditory cortical hierarchy not only revealed figure-ground modulation of neuronal responses in the primary auditory cortex, they also confirmed that the figure-ground effect magnitude seems to be similar between anterior and posterior recording field. Thus, A1 neurons can already detect temporally coherent elements that have no simple mathematical relationship to each other.

Previous electrophysiological investigations have suggested an involvement of the primary auditory cortex in target representation for artificial and natural noise stimuli (Christison-Lagay et al., 2017; Mesgarani et al., 2014; Saderi et al., 2020; Schneider and Woolley, 2013; Town et al., 2019). These findings point towards a critical role of A1 in complex sound processing. Similar to our results, one primate study demonstrated that a subset of neurons in the auditory core can encode target-to-noise ratio, with higher firing rates for more salient targets (Christison-Lagay et al., 2017). Another study nicely demonstrated that illusory continuity of tones is encoded in the primate primary auditory cortex (Petkov et al., 2007). Thus, the early auditory cortex seems to be heavily involved in perceptual encoding. Whether A1 is the first station that exhibit target representation remains unclear. Neuronal correlates of streaming have been found as early as the cochlear nucleus (Pressnitzer et al., 2008)

suggesting that perceptual grouping relies on subcortical centres or on a corticofugal loop. Another study suggests increasingly noise-invariant representations of target sounds along the processing hierarchy due to forward suppression (Schneider and Woolley, 2013) with higher proportion of noise representation in subcortical brain areas.

Object-specific feature selectivity seems to be a contributing factor for segregation on the cortical level as data confirm high variability between responses to different auditory figures (see example unit in Figure 4.9). Given the randomness of the stimulus, the observed pattern of figure-responsive neurons (Figure 4.8) indicates that, on the level of the auditory cortex, distributed ensemble activity represents the identity of auditory objects. It might be the case that both core and belt areas do not yet encode abstract, invariant object information but rather signal changes in the auditory scene.

Responsive cluster seem to signal auditory figures with a population rate code that builds up over time, suggesting recurrent cortical processing that can be explained by the temporal coherence model (Shamma et al., 2011). Neuronal responses in hit and miss trials are different in one subject (Figure 4.10), with hit trials eliciting higher firing rate modulations. A difference between hit and miss responses was also found with EEG (Tóth et al., 2016). This confirms that higher population responses affect object perception and points towards a critical role of attention/arousal in figure-ground segregation. These findings also suggest that previous imaging experiments (Teki et al., 2011, 2016) reflect bottom-up object detection. In addition, the difference between correct rejection and miss trials suggests that the auditory cortex can detect figures independent of attentional selection. This further confirms fMRI imaging data (chapter 3) and backs up claims about the stimulus-driven nature of figure segregation.

Comparing figure-onset aligned multi-unit responses revealed differences between anterior and posterior auditory cortex. No difference in neuronal responses to varying coherence levels was found in the posterior recording field. In contrast, neurons in the anterior recording field encoded figure coherence with the modulation magnitude. Higher MUA increments are caused by higher coherence levels. Thus, this suggests a qualitative difference in cortical processing between anterior and posterior auditory cortex. Anterior

auditory regions seem to code for figure characteristics like the target-to-masker intensity ratio ('perceptual salience'), whereas posterior fields only signal a change in the statistical properties of the sensory input. A hierarchical relationship between anterior and posterior auditory cortex has been proposed before (Kikuchi et al., 2010), suggesting that results point towards increased complexity of sensory representations. However, latencies of the figure-ground effect are similar across the auditory cortex suggesting parallel processing instead of a serial/hierarchical flow of information that gets more complex over time. This points towards a distributed analysis strategy of the auditory cortex. Neuronal figure-ground processing in the parabelt region remains a mystery but is likely the central hub between frontal and parietal regions and cortical fields in the superior temporal plane (Rauschecker and Tian, 2000; Romanski et al., 1999b).

This thesis confirms that primary auditory core areas are involved in figure-ground processing with comparable effects to visual perceptual grouping. Similar to the reported neuronal activity in the auditory cortex, multi-unit activity in the primary visual cortex responds with an increase in firing rate to figures in the sustained period after the onset transient (Lamme, 1995; Poort et al., 2012, 2016; Roelfsema et al., 2002; Self et al., 2019). The latency and build-up of auditory figure-ground modulation point towards recurrent, incremental cortical processing, as it seems to be the case in the visual system (Roelfsema et al., 2002; Self et al., 2019). Decision-aligned LFP signals have revealed suppression in alpha/beta bands and enhanced power in the gamma range for both monkeys. This pattern of oscillatory activity has also been shown before for visual figure-ground segregation (van Kerkoerle et al., 2014). Using EEG recordings during the rapid presentation of either (visual) noise or high-level object stimuli, enhanced gamma oscillation have been linked to object percepts (Castelhano et al., 2013). Gamma oscillations during the presentation of SFG stimuli have also been found in human patients with implanted ECoG arrays (Gander et al., unpublished data). Taken together, these findings suggest a shared neural mechanism across sensory modalities, even though sensory input has to be processed in a very different way (e.g. spatial location). How features that are represented in different parts of the cortex are combined and how object information is integrated into an auditory salience map (Kaya and Elhilali, 2017; Kayser et al., 2005) cannot be answered based on this data set. Visual research has identified the posterior parietal cortex as integration hub for sensory, cognitive and motor-related

signals. Specifically, the lateral intraparietal area (LIP) has been suggested to act as a priority map, with enhanced neuronal responses selectively to task-relevant stimuli, locations or features (Bisley and Goldberg, 2010; Colby et al., 1996; Gottlieb, 2007; Gottlieb et al., 1998). Auditory perceptual organisation has been linked to the IPS (Cusack, 2005; Teki et al., 2011, 2016), where area LIP is located. The theoretical concept of visual saliency representations proposed for LIP (Gottlieb, 2007) is very similar to the model of auditory stream formation (Shamma et al., 2011). Given the reported similarities in figure-ground processing (chapter 4), it is not unlikely that similar overarching coding principles are shared between sensory modalities. I speculate that the processing hierarchy in a recurrent loop might look like this:

1) Parallel feature processing in subcortical nuclei
2) Coherence analysis and object/stream representations across auditory cortical areas
3) Attentional selection of available auditory objects/streams in parietal cortex

In this thesis, I have investigated one aspect of auditory scene analysis: figure-ground segregation. As outlined above, this process is crucial to real world listening, where the brain needs to detect, segregate and group spectrotemporally overlapping sound elements in order to create stable percepts and ultimately guide behaviour. Perceptual grouping in the visual domain is a well-established field with understood neuronal mechanisms. Auditory research regarding scene decomposition lacks at least two decades behind. The work presented here has cleared the way for further systematic investigations of this exciting topic, using stimuli that model natural scenes and allow cross-species comparisons. The following paragraph suggests extensions to the current paradigms that would make this line of research even more relevant to solving the complex sound processing and the cocktail party problem.

5.3 Future directions

Previous imaging work in combination with the behavioural and fMRI study presented in this thesis have allowed investigations of figure-ground

segregation on the neuronal level. The electrophysiological work presented here has illuminated neuronal responses across a vast extend of the auditory cortex using a very simple experimental setup without any sophisticated manipulation. Hence there is a huge potential for future research to further investigate the neuronal behaviour during figure-ground segregation in complex, nature-like scenes.

All studies described in this thesis presented stimuli from a stationary position to both ears. Posterior auditory cortical areas are selective to the spatial location of stimuli (Rauschecker and Tian, 2000; Romanski et al., 1999a). While the anterior auditory fields seem to code for perceptual saliency of the auditory object, I speculate that the posterior cortical areas are selective to spatial figure location as well as temporal features (like ramping or modulated figures). This could be easily tested with a speaker ring that presents background noise in 360 degrees around subjects, while temporally coherent elements are only presented at a certain location. This would further clarify the functional distinction in auditory processing across the superior temporal plane and shed light on the distributed nature of object processing in complex scenes.

Furthermore, functional imaging experiments strongly suggests the anterolateral belt and parabelt regions as central hubs in figure-ground processing. It would be interesting to know whether deactivating those areas, e.g. via optogenetic constructs, pharmacological intervention or cooling loops, causes a breakdown in behavioural performance as well as impaired firing rate modulation in the primary core. This would point towards a central role of the parabelt regions in object segregation and would also imply that A1 figure-ground responses are caused by a feedback signal from higher cortical areas. Conversely, inactivation the primary auditory cortex would help to understand whether downstream neurons require A1 responses in order to exhibit response modulation. Recordings from the parabelt would also confirm whether perceptual invariance is already encoded at the level of the auditory cortex or whether higher cortical areas, presumably frontal and are involved in this process.

In addition, the data presented here suggest that figure-ground modulation is dependent on arousal or attentional deployment towards the stimuli. Attentional effects have also been demonstrated in previous EEG experiments (O'Sullivan et al., 2015a; Tóth et al., 2016). Moreover, cognitive load has been shown to affect figure-ground segregation (Molloy et al., 2018). To reveal the actual role of top-down cognitive

control in the neuronal coding properties during figure ground segregation, future research should include cues that point the attentional focus either towards or away from the auditory stimulation. Alternatively, cues could label either ground or figure components as behavioural targets. Such paradigms would further clarify if and how attention, decision making and working memory contribute to object segregation and selection.

The stimulus history has an effect on current stimulus processing (Ho et al., 2019; Phillips et al., 2017b). Further work could be done in this direction to assess how changes in complex auditory scenes influence the segregation of single objects and how this alters the cortical processing. This would also link nicely to the predictive coding framework and would allow investigations into the processing of predicted or surprising auditory objects.

Another interesting direction of research would be the investigation of local cortical circuitry. The depth information in chapter 4 suggest no obvious layer-specific clustering of figure-responsive units, however, penetrations angles were highly variable. Laminar recordings could illuminate differences between different layers. Furthermore, interneurons have been shown to play a major role in cortical processing (Blackwell and Geffen, 2017; Natan et al., 2017; Wood et al., 2017b) and local microcircuits have been proposed to contribute to auditory object categorisation (Tsunada and Cohen, 2014; Tsunada et al., 2012). Since results shown here suggest no relationship between the figure-ground effect magnitude and the pure tone tuning curve, the question remains how the tonic increase in firing rate comes about. It might be worth investigating the role of different cells classes during the processing of complex auditory input.

Oscillatory changes of population activity have been demonstrated in the LFP shortly before subjects make a perceptual decision. Changes in the alpha/beta and gamma range seem to be reliable across the cortical hierarchy, however, it would be useful to assess how populations in anterior and posterior auditory cortex interact. Granger causality or spike-field coherence between recording contacts may be suitable means.

An involvement of the IPS has been strongly suggested in human fMRI (Cusack, 2005; Teki et al., 2011, 2016) experiments. I could not find a significant BOLD change in the IPS that could be due to a multitude of reasons,

like the cranial implants or even a species difference. It might be the case that active task engagement is necessary to drive neurons in that regions. Future experiments could investigate if and how the IPS modulates auditory scene analysis. Furthermore, work on the contribution of prefrontal cortex and hippocampus in auditory object processing are necessary to further reveal how the neuronal object representation changes across the cortical hierarchy.

It also remains an open question if and how subcortical structures like the MGN, IC or even cochlear nucleus contribute to auditory figure-ground segregation. Recordings from different levels in the ascending auditory hierarchy, e.g. dorsal CN, central IC or ventral MGN would be invaluable to determine where the first point of convergence for temporally coherent elements is.

Lastly, the current experimental paradigm uses a Go/No-go task. This required a touch bar release in figure conditions but no behavioural response during control trials. Other behavioural task designs, like a two-alternative forced choice task, might be a more elegant way to better control for motor components in the neural activity. Furthermore, natural acoustic scenes do not just contain one target and noise, but a variety of potential targets mixed together. Incorporating other grouping cues like harmonicity into the stimulus design could help to further understand how neurons value different competing grouping cues and how those responses contribute to resulting percept.

6 References

Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B.G., et al. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. Proc. Natl. Acad. Sci. *103*, 14608–14613.

Alain, C., Arnott, S.R., Hevenor, S., Graham, S., and Grady, C.L. (2001). “What” and “where” in the human auditory system. Proc. Natl. Acad. Sci. *98*, 12301–12306.

Arnott, S.R., Binns, M.A., Grady, C.L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. Neuroimage *22*, 401–408.

Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., and Bizley, J.K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. Neuron *97*, 640-655.e4.

Bagur, S., Averseng, M., Elgueda, D., David, S., Fritz, J., Yin, P., Shamma, S., Boubenec, Y., and Ostojic, S. (2018). Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex. Nat. Commun. *9*, 2529.

Bajo, V.M., and King, A.J. (2013). Cortical modulation of auditory processing in the midbrain. Front. Neural Circuits *6*, 1–12.

Bajo, V.M., Nodal, F.R., Bizley, J.K., Moore, D.R., and King, A.J. (2006). The Ferret Auditory Cortex: Descending Projections to the Inferior Colliculus. Cereb. Cortex *17*, 475–491.

Bajo, V.M., Nodal, F.R., Moore, D.R., and King, A.J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat. Neurosci. *13*, 253–260.

Bandyopadhyay, S., Shamma, S.A., and Kanold, P.O. (2010). Dichotomy of functional organization in the mouse auditory cortex. Nat. Neurosci. *13*, 361–368.

Barascud, N., Pearce, M.T., Griffiths, T.D., Friston, K.J., and Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc. Natl. Acad. Sci. *113*, E616–E625.

Barczak, A., O’Connell, M.N., McGinnis, T., Ross, D., Mowery, T., Falchier, A., and Lakatos, P. (2018). Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. Proc. Natl. Acad. Sci. *115*, E7605–E7614.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical Microcircuits for Predictive Coding. Neuron *76*, 695–711.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron *85*, 390–401.

Baumann, S., Petkov, C.I., and Griffiths, T.D. (2013). A unified framework for the organization of the primate auditory cortex. Front. Syst. Neurosci. *7*, 11.

Baumann, S., Joly, O., Rees, A., Petkov, C.I., Sun, L., Thiele, A., and Griffiths, T.D.

(2015). The topography of frequency and time representation in primate auditory cortices. Elife *4*.

Beauvois, M.W. (1997). Time decay of auditory stream biasing.

Bee, M.A., and Klump, G.M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J. Neurophysiol. *92*, 1088–1104.

Bendor, D., and Wang, X. (2008). Neural Response Properties of Primary, Rostral, and Rostrotemporal Core Fields in the Auditory Cortex of Marmoset Monkeys. J. Neurophysiol. *100*, 888–906.

Besle, J., Mougin, O., Sánchez-Panchuelo, R.-M., Lanting, C., Gowland, P., Bowtell, R., Francis, S., and Krumbholz, K. (2019). Is Human Auditory Cortex Organization Compatible With the Monkey Model? Contrary Evidence From Ultra-High-Field Functional and Structural MRI. Cereb. Cortex *29*, 410–428.

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. Percept. Psychophys. *64*, 844–854.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., and Bertrand, O. (2007). Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex. J. Neurosci. *27*, 9252–9261.

Billig, A.J., Herrmann, B., Rhone, A.E., Gander, P.E., Nourski, K. V., Snoad, B.F., Kovach, C.K., Kawasaki, H., Howard, M.A., and Johnsrude, I.S. (2019). A Sound-Sensitive Source of Alpha Oscillations in Human Non-Primary Auditory Cortex. J. Neurosci. *39*, 8679–8689.

Bisley, J.W., and Goldberg, M.E. (2010). Attention, Intention, and Priority in the Parietal Lobe. Annu. Rev. Neurosci. *33*, 1–21.

Bizley, J.K., and Cohen, Y.E. (2013). The what, where and how of auditory-object perception. Nat. Rev. Neurosci. *14*, 693–707.

Bizley, J.K., Nodal, F.R., Nelken, I., and King, A.J. (2005). Functional Organization of Ferret Auditory Cortex. Cereb. Cortex *15*, 1637–1653.

Bizley, J.K., Walker, K.M.M., Silverman, B.W., King, A.J., and Schnupp, J.W.H. (2009). Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. J. Neurosci. *29*, 2064–2075.

Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., and Schnupp, J.W.H. (2013). Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. Curr. Biol. *23*, 620–625.

Blackwell, J.M., and Geffen, M.N. (2017). Progress and challenges for understanding the function of cortical microcircuits in auditory processing. Nat. Commun. *8*, 2165.

Blasdel, G., Lund, J., and Fitzpatrick, D. (1985). Intrinsic connections of macaque striate cortex: axonal projections of cells outside lamina 4C. J. Neurosci. *5*, 3350–3369.

Born, R.T., Groh, J.M., Zhao, R., and Lukasewycz, S.J. (2000). Segregation of Object and Background Motion in Visual Area MT. Neuron *26*, 725–734.

Bregman, A.S. (1978). Auditory streaming is cumulative. J. Exp. Psychol. Hum.

Percept. Perform. *4*, 380–387.

Bregman, A.S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound (Cambridge: MIT Press).

Brosch, M., Selezneva, E., and Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J. Neurosci. *25*, 6797–6806.

Brosch, M., Selezneva, E., and Scheich, H. (2011). Representation of Reward Feedback in Primate Auditory Cortex. Front. Syst. Neurosci. *5*, 5.

Brugge, J.F., Nourski, K. V, Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., and Howard, M.A. (2009). Coding of Repetitive Transients by Auditory Cortex on Heschl's Gyrus. J. Neurophysiol. *102*, 2358–2374.

Callaway, E.M. (1998). LOCAL CIRCUITS IN PRIMARY VISUAL CORTEX OF THE MACAQUE MONKEY. Annu. Rev. Neurosci. *21*, 47–74.

Camalier, C.R., D'Angelo, W.R., Sterbing-D'Angelo, S.J., de la Mothe, L.A., and Hackett, T.A. (2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. Proc. Natl. Acad. Sci. *109*, 18168–18173.

Carlyon, R.P., Cusack, R., Foxton, J.M., and Robertson, I.H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. *27*, 115–127.

Carlyon, R.P., Plack, C.J., Fantini, D.A., and Cusack, R. (2003). Cross-Modal and Non-Sensory Influences on Auditory Streaming. Perception *32*, 1393–1402.

Castelhano, J., Rebola, J., Leitão, B., Rodriguez, E., and Castelo-Branco, M. (2013). To Perceive or Not Perceive: The Role of Gamma-band Activity in Signaling Object Percepts. PLoS One *8*, e66363.

Chang, E.F., and Merzenich, M.M. (2003). Environmental Noise Retards Auditory Cortical Development. Science (80-.). *300*, 498–502.

Chao, Z.C., Takaura, K., Wang, L., Fujii, N., and Dehaene, S. (2018). Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. Neuron *100*, 1252-1266.e3.

Cherry, E.C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. J. Acoust. Soc. Am. *25*, 975–979.

Christison-Lagay, K.L., and Cohen, Y.E. (2014). Behavioral correlates of auditory streaming in rhesus macaques. Hear. Res. *309*, 17–25.

Christison-Lagay, K.L., Bennur, S., and Cohen, Y.E. (2017). Contribution of spiking activity in the primary auditory cortex to detection in noise. J. Neurophysiol. *118*, 3118–3131.

Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. *12*, 1594–1600.

Colby, C.L., Duhamel, J.R., and Goldberg, M.E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. J.

Neurophysiol. *76*, 2841–2852.

Constantinople, C.M., and Bruno, R.M. (2013). Deep Cortical Layers Are Activated Directly by Thalamus. Science (80-.). *340*, 1591–1594.

Cusack, R. (2005). The Intraparietal Sulcus and Perceptual Organization. J. Cogn. Neurosci. *17*, 641–651.

Cusack, R., Decks, J., Aikman, G., and Carlyon, R.P. (2004). Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. J. Exp. Psychol. Hum. Percept. Perform. *30*, 643–656.

Darwin, C.J. (1997). Auditory grouping. Trends Cogn. Sci. *1*, 327–333.

David, S. V, Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. *109*, 2144–2149.

Deike, S., Heil, P., Böckmann-Barthel, M., and Brechmann, A. (2012). The Build-up of Auditory Stream Segregation: A Different Perspective. Front. Psychol. *3*.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.

Diamond, M.E., von Heimendahl, M., Knutsen, P.M., Kleinfeld, D., and Ahissar, E. (2008). “Where” and “what” in the whisker sensorimotor system. Nat. Rev. Neurosci. *9*, 601–612.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How Does the Brain Solve Visual Object Recognition? Neuron *73*, 415–434.

Douglas, R.J., and Martin, K.A.C. (2004). NEURONAL CIRCUITS OF THE NEOCORTEX. Annu. Rev. Neurosci. *27*, 419–451.

Downer, J.D., Niwa, M., and Sutter, M.L. (2015). Task Engagement Selectively Modulates Neural Correlations in Primary Auditory Cortex. J. Neurosci. *35*, 7565–7574.

Downer, J.D., Rapone, B., Verhein, J., O’Connor, K.N., and Sutter, M.L. (2017). Feature-Selective Attention Adaptively Shifts Noise Correlations in Primary Auditory Cortex. J. Neurosci. *37*, 5378–5392.

Driver, J. (2001). A selective review of selective attention research from the past century. Br. J. Psychol. *92*, 53–78.

Duncan, J. (1984). Selective attention and the organization of visual information. J. Exp. Psychol. Gen. *113*, 501–517.

Dykstra, A.R., Cariani, P.A., and Gutschalk, A. (2017). A roadmap for the study of conscious audition and its neural basis. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160103.

Dylla, M., Hrnicek, A., Rice, C., and Ramachandran, R. (2013). Detection of Tones and Their Modification by Noise in Nonhuman Primates. J. Assoc. Res. Otolaryngol. *14*, 547–560.

Dyson, B.J., and Alain, C. (2004). Representation of concurrent acoustic objects in

primary auditory cortex. J. Acoust. Soc. Am. *115*, 280–288.

Eggermont, J.J. (1998). Representation of Spectral and Temporal Sound Features in Three Cortical Fields of the Cat. Similarities Outweigh Differences. J. Neurophysiol. *80*, 2743–2764.

Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S. V, Shamma, S.A., and Fritz, J.B. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. Nat. Neurosci.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., and Shamma, S.A. (2009a). Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. Neuron *61*, 317–329.

Elhilali, M., Xiang, J., Shamma, S.A., and Simon, J.Z. (2009b). Interaction between Attention and Bottom-Up Saliency Mediates the Representation of Foreground and Background in an Auditory Scene. PLoS Biol. *7*, e1000129.

Engel, A.K., Konig, P., and Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding (cat/visual cortex/correlation analysis/synchronization/cell assembly).

Engel, T.A., Steinmetz, N.A., Gieselmann, M.A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. Science (80-.). *354*, 1140–1144.

Feldman, J. (2003). What is a visual object? Trends Cogn. Sci. *7*, 252–256.

Feng, L., and Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. Proc. Natl. Acad. Sci. *114*, E840–E848.

Fishman, Y.I., and Steinschneider, M. (2010). Neural Correlates of Auditory Scene Analysis Based on Inharmonicity in Monkey Primary Auditory Cortex. J. Neurosci. *30*, 12480–12494.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. *151*, 167–187.

Fishman, Y.I., Arezzo, J.C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. *116*, 1656–1670.

Fishman, Y.I., Micheyl, C., and Steinschneider, M. (2012). Neural mechanisms of rhythmic masking release in monkey primary auditory cortex: implications for models of auditory scene analysis. J. Neurophysiol. *107*, 2366–2382.

Fishman, Y.I., Kim, M., and Steinschneider, M. (2017). A Crucial Test of the Population Separation Model of Auditory Stream Segregation in Macaque Primary Auditory Cortex. J. Neurosci. *37*, 0792–17.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. *9*, 474–480.

Fries, P. (2009). Neuronal Gamma-Band Synchronization as a Fundamental Process

in Cortical Computation. Annu. Rev. Neurosci. *32*, 209–224.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. *360*, 815–836.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., and Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. *2*, 189–210.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.

Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hear. Res. *206*, 159–176.

Fritz, J.B., Elhilali, M., and Shamma, S.A. (2007a). Adaptive Changes in Cortical Receptive Fields Induced by Attention to Complex Sounds. J. Neurophysiol. *98*, 2337–2346.

Fritz, J.B., Elhilali, M., David, S. V., and Shamma, S.A. (2007b). Auditory attention—focusing the searchlight on sound. Curr. Opin. Neurobiol. *17*, 437–455.

Fritz, J.B., David, S. V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. *13*, 1011–1019.

Gaucher, Q., Panniello, M., Ivanov, A.Z., Dahmen, J.C., King, A.J., and M Walker, K.M. (2019). Complexity of frequency receptive fields predicts tonotopic variability across species. BioRxiv.

Gottlieb, J. (2007). From Thought to Action: The Parietal Cortex as a Bridge between Perception, Action, and Cognition. Neuron *53*, 9–16.

Gottlieb, J.P., Kusunoki, M., and Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. Nature *391*, 481–484.

Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature *338*, 334–337.

Gray, H., Bertrand, H., Mindus, C., Flecknell, P., Rowe, C., and Thiele, A. (2016). Physiological, Behavioral, and Scientific Impact of Different Fluid Control Protocols in the Rhesus Macaque (Macaca mulatta). ENeuro *3*.

Griffiths, T.D., and Warren, J.D. (2004). What is an auditory object? Nat. Rev. Neurosci. *5*, 887–892.

Guest, H., Munro, K.J., Prendergast, G., Millman, R.E., and Plack, C.J. (2018). Impaired speech perception in noise with a normal audiogram: No evidence for cochlear synaptopathy and no relation to lifetime noise exposure. Hear. Res. *364*, 142–151.

Guo, W., Chambers, A.R., Darrow, K.N., Hancock, K.E., Shinn-Cunningham, B.G., and Polley, D.B. (2012). Robustness of Cortical Topography across Fields, Laminae,

Anesthetic States, and Neurophysiological Signal Types. J. Neurosci. *32*, 9159–9172.

Gutschalk, A. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. J. Neurosci. *25*, 5382–5388.

Gutschalk, A., Micheyl, C., and Oxenham, A.J. (2008a). Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol. *6*, e138.

Gutschalk, A., Micheyl, C., and Oxenham, A.J. (2008b). Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol. *6*, e138.

Hackett, T.A. (2011). Information flow in the auditory cortical network. Hear. Res. *271*, 133–146.

Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1998a). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. J. Comp. Neurol. *394*, 475–495.

Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1998b). Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. J. Comp. Neurol. *400*, 271–286.

Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Res. *817*, 45–58.

Hackett, T.A., Preuss, T.M., and Kaas, J.H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. *441*, 197–222.

Hackett, T.A., De La Mothe, L.A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C.E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J. Comp. Neurol. *502*, 924–952.

Hackett, T.A., Rinaldi Barkat, T., O'Brien, B.M.J., Hensch, T.K., and Polley, D.B. (2011). Linking Topography to Tonotopy in the Mouse Auditory Thalamocortical Circuit. J. Neurosci. *31*, 2983–2995.

Hackett, T.A., de la Mothe, L.A., Camalier, C.R., Falchier, A., Lakatos, P., Kajikawa, Y., and Schroeder, C.E. (2014). Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. Front. Neurosci. *8*, 1–21.

Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. Nature *503*, 51–58.

Harris, K.D., and Thiele, A. (2011). Cortical state and attention. Nat. Rev. Neurosci. *12*, 509–523.

Hasenstaub, A., Sachdev, R.N.S., and McCormick, D.A. (2007). State Changes Rapidly Modulate Cortical Neuronal Responsiveness. J. Neurosci. *27*, 9607–9622.

Haywood, N.R., and Roberts, B. (2010). Build-up of the tendency to segregate auditory streams: Resetting effects evoked by a single deviant tone. J. Acoust. Soc. Am. *128*, 3019–3031.

Heffner, H.E., and Heffner, R.S. (1986). Hearing loss in Japanese macaques

following bilateral auditory cortex lesions. J. Neurophysiol. *55*, 256–271.

van der Heijden, K., Rauschecker, J.P., de Gelder, B., and Formisano, E. (2019). Cortical mechanisms of spatial hearing. Nat. Rev. Neurosci.

Heilbron, M., and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? Neuroscience *389*, 54–73.

Higgins, N.C., Little, D.F., Yerkes, B.D., Nave, K.M., Kuruvilla-Mathew, A., Elhilali, M., and Snyder, J.S. (2020). Neural correlates of perceptual switching while listening to bistable auditory streaming stimuli. Neuroimage *204*, 116220.

Hind, S.E., Haines-Bazrafshan, R., Benton, C.L., Brassington, W., Towle, B., and Moore, D.R. (2011). Prevalence of clinical referrals having hearing thresholds within normal limits. Int. J. Audiol. *50*, 708–716.

Ho, H.T., Burr, D.C., Alais, D., and Morrone, M.C. (2019). Auditory Perceptual History Is Propagated through Alpha Oscillations. Curr. Biol. *29*, 4208-4217.e3.

Holmes, E., and Griffiths, T.D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. Sci. Rep. *9*, 16771.

Homma, N.Y., Happel, M.F.K., Nodal, F.R., Ohl, F.W., King, A.J., and Bajo, V.M. (2017). A Role for Auditory Corticothalamic Feedback in the Perception of Complex Sounds. J. Neurosci. *37*, 6149–6161.

Huang, C.L., and Winer, J.A. (2000). Auditory thalamocortical projections in the cat: Laminar and areal patterns of input. J. Comp. Neurol. *427*, 302–331.

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. J. Physiol. *195*, 215–243.

Itatani, N., and Klump, G.M. (2014). Neural correlates of auditory streaming in an objective behavioral task. Proc. Natl. Acad. Sci. *111*, 10738–10743.

Itskov, P.M., Vinnik, E., Honey, C., Schnupp, J., and Diamond, M.E. (2012). Sound sensitivity of neurons in rat hippocampus during performance of a sound-guided task. J. Neurophysiol. *107*, 1822–1834.

Jackson, L.L., Heffner, R.S., and Heffner, H.E. (1999). Free-field audiogram of the Japanese macaque (Macaca fuscata). J. Acoust. Soc. Am. *106*, 3017–3023.

Johnston, W.A., and Dark, V.J. (1986). Selective Attention. Annu. Rev. Psychol. *37*, 43–75.

Joly, O., Baumann, S., Poirier, C., Patterson, R.D., Thiele, A., and Griffiths, T.D. (2014a). A perceptual pitch boundary in a non-human primate. Front. Psychol. *5*, 1–5.

Joly, O., Baumann, S., Balezeau, F., Thiele, A., and Griffiths, T.D. (2014b). Merging functional and structural properties of the monkey auditory cortex. Front. Neurosci. *8*.

Jones, E.G., Dell'Anna, M.E., Molinari, M., Rausell, E., and Hashikawa, T. (1995). Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. J. Comp. Neurol. *362*, 153–170.

Jones, H.E., Andolina, I.M., Shipp, S.D., Adams, D.L., Cudeiro, J., Salt, T.E., and Sillito, A.M. (2015). Figure-ground modulation in awake primate thalamus. Proc. Natl. Acad. Sci. *112*, 7085–7090.

Kaas, J.H., and Hackett, T.A. (1999). “What” and “where” processing in auditory cortex. Nat. Neurosci. *2*, 1045–1047.

Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. *97*, 11793–11799.

Kaas, J.H., Hackett, T.A., and Tramo, M.J. (1999). Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. *9*, 164–170.

Kadia, S.C., and Wang, X. (2003). Spectral Integration in A1 of Awake Primates: Neurons With Single- and Multipeaked Tuning Characteristics. J. Neurophysiol. *89*, 1603–1622.

Kajikawa, Y., and Schroeder, C.E. (2011). How Local Is the Local Field Potential? Neuron *72*, 847–858.

Kajikawa, Y., Camalier, C.R., de la Mothe, L.A., D’Angelo, W.R., Sterbing-D’Angelo, S.J., and Hackett, T.A. (2011). Auditory cortical tuning to band-pass noise in primate A1 and CM: A comparison to pure tones. Neurosci. Res. *70*, 401–407.

Kajikawa, Y., Frey, S., Ross, D., Falchier, A., Hackett, T. a, and Schroeder, C.E. (2015). Auditory Properties in the Parabelt Regions of the Superior Temporal Gyrus in the Awake Macaque Monkey: An Initial Survey. J. Neurosci. *35*, 4140–4150.

Kanayamay, S., Kuhara, S., and Satoh, K. (1996). In vivo rapid magnetic field measurement and shimming using single scan differential phase mapping. Magn. Reson. Med. *36*, 637–642.

Kandel, E.R., Schwartz, J.H., and Jessell, T.M. (2000). Principles of Neural Science (Mc-Graw-Hill).

Kavanagh, G.L., and Kelly, J.B. (1988). Hearing in the ferret (Mustela putorius): effects of primary auditory cortical lesions on thresholds for pure tone detection. J. Neurophysiol. *60*, 879–888.

Kaya, E.M., and Elhilali, M. (2017). Modelling auditory attention. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160101.

Kayser, C., Petkov, C.I., Lippert, M., and Logothetis, N.K. (2005). Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. Curr. Biol. *15*, 1943–1947.

Kayser, C., Logothetis, N.K., and Panzeri, S. (2010). Millisecond encoding precision of auditory cortex neurons. Proc. Natl. Acad. Sci. *107*, 16976–16981.

van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc. Natl. Acad. Sci. *111*, 14332–14341.

Kikuchi, Y., Horwitz, B., and Mishkin, M. (2010). Hierarchical Auditory Processing Directed Rostrally along the Monkey’s Supratemporal Plane. J. Neurosci. *30*, 13021–13030.

Kikuchi, Y., Horwitz, B., Mishkin, M., and Rauschecker, J.P. (2014). Processing of harmonics in the lateral belt of macaque auditory cortex. Front. Neurosci. *8*, 1–13.

Kikuchi, Y., Kumar, S., Baumann, S., Overath, T., Gander, P.E., Sedley, W., Patterson, R.D., Petkov, C.I., and Griffiths, T.D. (2019). The distribution and nature of responses to broadband sounds associated with pitch in the macaque auditory cortex. Cortex *120*, 340–352.

King, A.J., Bajo, V.M., Bizley, J.K., Campbell, R.A., Nodal, F.R., Schulz, A.L., and Schnupp, J.W. (2007). Physiological and behavioral studies of spatial coding in the auditory cortex. Hear. Res. *229*, 106–115.

Knyazeva, S., Selezneva, E., Gorkin, A., Aggelopoulos, N.C., and Brosch, M. (2018). Neuronal Correlates of Auditory Streaming in Monkey Auditory Cortex for Tone Sequences without Spectral Differences. Front. Integr. Neurosci. *12*.

Kourtzi, Z., Bülthoff, H.H., Erb, M., and Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. Nat. Neurosci. *5*, 17–18.

Kujawa, S.G., and Liberman, M.C. (2009). Adding Insult to Injury: Cochlear Nerve Degeneration after “Temporary” Noise-Induced Hearing Loss. J. Neurosci. *29*, 14077–14085.

Kumar, G., Amen, F., and Roy, D. (2007). Normal hearing tests: is a further appointment really necessary? J. R. Soc. Med. *100*, 66–66.

Kuśmierek, P., and Rauschecker, J.P. (2009a). Functional Specialization of Medial Auditory Belt Cortex in the Alert Rhesus Monkey. J. Neurophysiol. *102*, 1606–1622.

Kuśmierek, P., and Rauschecker, J.P. (2009b). Functional Specialization of Medial Auditory Belt Cortex in the Alert Rhesus Monkey. J. Neurophysiol. *102*, 1606–1622.

Kuśmierek, P., and Rauschecker, J.P. (2014). Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. J. Neurophysiol. *111*, 1671–1685.

Lamme, V. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. *15*, 1605–1615.

Lamme, V.A.F., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. *23*, 571–579.

Leaver, A.M., and Rauschecker, J.P. (2010). Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. J. Neurosci. *30*, 7604–7612.

Leaver, A.M., and Rauschecker, J.P. (2016). Functional Topography of Human Auditory Cortex. J. Neurosci. *36*, 1416–1428.

Letzkus, J.J., Wolff, S.B.E., Meyer, E.M.M., Tovote, P., Courtin, J., Herry, C., and Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. Nature *480*, 331–335.

Linden, J.F. (2003). Columnar Transformations in Auditory Cortex? A Comparison to Visual and Somatosensory Cortices. Cereb. Cortex *13*, 83–89.

Liu, B.-H., Wu, G.K., Arbuckle, R., Tao, H.W., and Zhang, L.I. (2007). Defining

cortical frequency tuning with recurrent excitatory circuitry. Nat. Neurosci. *10*, 1594–1600.

Liu, J., Whiteway, M.R., Sheikhattar, A., Butts, D.A., Babadi, B., and Kanold, P.O. (2019). Parallel Processing of Sound Dynamics across Mouse Auditory Cortex via Spatially Patterned Thalamic Inputs and Distinct Areal Intracortical Circuits. Cell Rep. *27*, 872-885.e7.

Lu, K., Xu, Y., Yin, P., Oxenham, A.J., Fritz, J.B., and Shamma, S.A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. Nat. Commun. *8*, 13900.

Lu, K., Liu, W., Dutta, K., Zan, P., Fritz, J.B., and Shamma, S.A. (2019). Adaptive Efficient Coding of Correlated Acoustic Properties. J. Neurosci. *39*, 8664–8678.

Lu, T., Liang, L., and Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat. Neurosci. *4*, 1131–1138.

Luczak, A., McNaughton, B.L., and Harris, K.D. (2015). Packet-based communication in the cortex. Nat. Rev. Neurosci. *16*, 745–755.

Malmierca, M.S. (2013). Anatomy and Physiology of the Mammalian Auditory System. In Encyclopedia of Computational Neuroscience, (New York, NY: Springer New York), pp. 1–36.

Marie, D., Jobard, G., Crivello, F., Perchey, G., Petit, L., Mellet, E., Joliot, M., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2015). Descriptive anatomy of Heschl's gyri in 430 healthy volunteers, including 198 left-handers. Brain Struct. Funct. *220*, 729–743.

Merzenich, M.M., and Brugge, J.F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. Brain Res. *50*, 275–296.

Mesgarani, N., David, S. V., Fritz, J.B., and Shamma, S.A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. Proc. Natl. Acad. Sci. *111*, 6792–6797.

Micheyl, C., Tian, B., Carlyon, R.P., and Rauschecker, J.P. (2005). Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. Neuron *48*, 139–148.

Miller, G.A., and Heise, G.A. (1950). The Trill Threshold. J. Acoust. Soc. Am. *22*, 637–638.

Miller, L.M., and Recanzone, G.H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. Proc. Natl. Acad. Sci. *106*, 5931–5935.

Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., and Formisano, E. (2013). Processing of Natural Sounds: Characterization of Multipeak Spectral Tuning in Human Auditory Cortex. J. Neurosci. *33*, 11888–11898.

Molloy, K., Lavie, N., and Chait, M. (2018). Auditory figure-ground segregation is impaired by high visual load. J. Neurosci. 2518–18.

Moore, B.C.J., and Gockel, H.E. (2012). Properties of auditory stream formation. Philos. Trans. R. Soc. B Biol. Sci. *367*, 919–931.

Morel, A., Garraghty, P.E., and Kaas, J.H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J. Comp. Neurol. *335*, 437–459.

Natan, R.G., Rao, W., and Geffen, M.N. (2017). Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. Cell Rep. *21*, 878–890.

Nelken, I., Bizley, J.K., Nodal, F.R., Ahmed, B., King, A.J., and Schnupp, J.W.H. (2008). Responses of Auditory Cortex to Complex Stimuli: Functional Organization Revealed Using Intrinsic Optical Signals. J. Neurophysiol. *99*, 1928–1941.

Niwa, M., Johnson, J.S., O'Connor, K.N., and Sutter, M.L. (2012). Activity Related to Perceptual Judgment and Action in Primary Auditory Cortex. J. Neurosci. *32*, 3193–3210.

van Noorden, L.P.A.S. (1975). Temporal Coherence in the Perception of Tone Sequences, Institute for Perception Research, (Eindhoven).

Noreña, A.J., Gourévitch, B., Aizawa, N., and Eggermont, J.J. (2006). Spectrally enhanced acoustic environment disrupts frequency representation in cat auditory cortex. Nat. Neurosci. *9*, 932–939.

Nowak, L.G., Munk, M.H., Nelson, J.I., James, A.C., and Bullier, J. (1995). Structural basis of cortical synchronization. I. Three types of interhemispheric coupling. J. Neurophysiol. *74*, 2379–2400.

Nowak, L.G., Munk, M.H.J., James, A.C., Girard, P., and Bullier, J. (1999). Cross-Correlation Study of the Temporal Interactions Between Areas V1 and V2 of the Macaque Monkey. J. Neurophysiol. *81*, 1057–1074.

O'Sullivan, J.A., Shamma, S.A., and Lalor, E.C. (2015a). Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. J. Neurosci. *35*, 7256–7263.

O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., and Lalor, E.C. (2015b). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb. Cortex *25*, 1697–1706.

Ortiz-Rios, M., Azevedo, F.A.C., Kuśmierek, P., Balla, D.Z., Munk, M.H., Keliris, G.A., Logothetis, N.K., and Rauschecker, J.P. (2017). Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. Neuron *93*, 971-983.e4.

Oshurkova, E., Scheich, H., and Brosch, M. (2008). Click train encoding in primary and non-primary auditory cortex of anesthetized macaque monkeys. Neuroscience *153*, 1289–1299.

Otazu, G.H., Tai, L.-H., Yang, Y., and Zador, A.M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. Nat. Neurosci. *12*, 646–654.

Oxenham, A.J. (2016). Predicting the Perceptual Consequences of Hidden Hearing Loss. Trends Hear. *20*, 233121651668676.

Peirce, J.W. (2007). PsychoPy—Psychophysics software in Python. J. Neurosci. Methods *162*, 8–13.

Petkov, C.I., O'Connor, K.N., and Sutter, M.L. (2007). Encoding of Illusory Continuity in Primary Auditory Cortex. Neuron *54*, 153–165.

Pfingst, B.E., Hienz, R., and Miller, J. (1975a). Reaction time procedure for measurement of hearing. II. Threshold functions. J. Acoust. Soc. Am. *57*, 431–436.

Pfingst, B.E., Hienz, R., Kimm, J., and Miller, J. (1975b). Reaction−time procedure for measurement of hearing. I. Suprathreshold functions. J. Acoust. Soc. Am. *57*, 421–430.

Phillips, E.A.K., Schreiner, C.E., and Hasenstaub, A.R. (2017a). Cortical Interneurons Differentially Regulate the Effects of Acoustic Context. Cell Rep. *20*, 771–778.

Phillips, E.A.K., Schreiner, C.E., and Hasenstaub, A.R. (2017b). Diverse effects of stimulus history in waking mouse auditory cortex. J. Neurophysiol. *118*, 1376–1393.

Pickles, J.O. (2012). An Introduction to the Physiology of Hearing (Emerald).

Pienkowski, M. (2017). On the Etiology of Listening Difficulties in Noise Despite Clinically Normal Audiograms. Ear Hear. *38*, 135–148.

Poltoratski, S., Maier, A., Newton, A.T., and Tong, F. (2019). Figure-Ground Modulation in the Human Lateral Geniculate Nucleus Is Distinguishable from Top-Down Attention. Curr. Biol. *29*, 2051-2057.e3.

Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., Neumann, H., and Roelfsema, P.R. (2012). The Role of Attention in Figure-Ground Segregation in Areas V1 and V4 of the Visual Cortex. Neuron *75*, 143–156.

Poort, J., Self, M.W., Van Vugt, B., Malkki, H., and Roelfsema, P.R. (2016). Texture Segregation Causes Early Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. Cereb. Cortex *26*, 3964–3976.

Posner, M.I. (1980). Orienting of attention. Q. J. Exp. Psychol. *32*, 3–25.

Pressnitzer, D., and Hupé, J.-M. (2006). Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. Curr. Biol. *16*, 1351–1357.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I.M. (2008). Perceptual Organization of Sound Begins in the Auditory Periphery. Curr. Biol. *18*, 1124–1128.

Rauschecker, J.P. (2004). Processing of Band-Passed Noise in the Lateral Auditory Belt Cortex of the Rhesus Monkey. J. Neurophysiol. *91*, 2578–2589.

Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. *97*, 11800–11806.

Recanzone, G.H. (2000). Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. Hear. Res. *150*, 104–118.

Recanzone, G.H. (2001). Spatial Processing in the Primate Auditory Cortex. Audiol. Neuro-Otology *6*, 178–181.

Recanzone, G., Schreiner, C., and Merzenich, M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J. Neurosci. *13*, 87–103.

Recanzone, G.H., Wurtz, R.H., and Schwarz, U. (1997). Responses of MT and MST Neurons to One and Two Moving Objects in the Receptive Field. J. Neurophysiol. *78*, 2904–2915.

Recanzone, G.H., Guard, D.C., and Phan, M.L. (2000). Frequency and Intensity Response Properties of Single Neurons in the Auditory Cortex of the Behaving Macaque Monkey. J. Neurophysiol. *83*, 2315–2331.

Roelfsema, P.R. (2006). CORTICAL ALGORITHMS FOR PERCEPTUAL GROUPING. Annu. Rev. Neurosci. *29*, 203–227.

Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., and Bosch, H. (2002). Figure—Ground Segregation in a Recurrent Network Architecture. J. Cogn. Neurosci. *14*, 525–537.

Rogers, W.L., and Bregman, A.S. (1993). An experimental evaluation of three theories of auditory stream segregation. Percept. Psychophys. *53*, 179–189.

Rogers, W.L., and Bregman, A.S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. Percept. Psychophys. *60*, 1216–1227.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999a). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. *2*, 1131–1136.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999b). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. *2*, 1131–1136.

Rossi, A.F., and Paradiso, M.A. (1995). Feature-specific effects of selective visual attention. Vision Res. *35*, 621–634.

Saderi, D., Buran, B.N., and David, S. V (2020). Streaming of repeated noise in primary and secondary fields of auditory cortex. J. Neurosci. JN-RM-2105-19.

Saenz, M., and Langers, D.R.M. (2014). Tonotopic mapping of human auditory cortex. Hear. Res. *307*, 42–52.

Saleem, K.S., and Logothetis, N.K. (2012). Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., and Ohl, F.W. (2007). The cognitive auditory cortex: Task-specificity of stimulus representations. Hear. Res. *229*, 213–224.

Schinkel-Bielefeld, N., David, S. V., Shamma, S.A., and Butts, D.A. (2012). Inferring the role of inhibition in auditory processing of complex natural stimuli. J. Neurophysiol. *107*, 3296–3307.

Schneider, D.M., and Mooney, R. (2018). How Movement Modulates Hearing. Annu. Rev. Neurosci. *41*, 553–572.

Schneider, D.M., and Woolley, S.M.N. (2013). Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. Neuron *79*, 141–152.

Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature *513*, 189–194.

Schneider, D.M., Sundararajan, J., and Mooney, R. (2018a). A cortical filter that learns to suppress the acoustic consequences of movement. Nature *561*, 391–395.

Schneider, F., Dheerendra, P., Balezeau, F., Ortiz-Rios, M., Kikuchi, Y., Petkov, C.I., Thiele, A., and Griffiths, T.D. (2018b). Auditory figure-ground analysis in rostral belt and parabelt of the macaque monkey. Sci. Rep. *8*, 17948.

Schnupp, J., Nelken, I., and King, A. (2011). Auditory Neuroscience (Cambridge: MIT Press).

Scott, B.H., Malone, B.J., and Semple, M.N. (2011). Transformation of Temporal Processing Across Auditory Cortex of Awake Macaques. J. Neurophysiol. *105*, 712–730.

Scott, B.H., Leccese, P.A., Saleem, K.S., Kikuchi, Y., Mullarkey, M.P., Fukushima, M., Mishkin, M., and Saunders, R.C. (2015). Intrinsic Connections of the Core Auditory Cortical Regions and Rostral Supratemporal Plane in the Macaque Monkey. Cereb. Cortex *27*, bhv277.

Scott, B.H., Saleem, K.S., Kikuchi, Y., Fukushima, M., Mishkin, M., and Saunders, R.C. (2017). Thalamic connections of the core auditory cortex and rostral supratemporal plane in the macaque monkey. J. Comp. Neurol. *525*, 3488–3513.

Selezneva, E., Scheich, H., and Brosch, M. (2006). Dual Time Scales for Categorical Decision Making in Auditory Cortex. Curr. Biol. *16*, 2428–2433.

Self, M.W., and Roelfsema, P.R. (2015). Scene perception in early vision: Figure-ground organization in the lateral geniculate nucleus. Proc. Natl. Acad. Sci. U. S. A. *112*, 6784–6785.

Self, M.W., van Kerkoerle, T., Supèr, H., and Roelfsema, P.R. (2013). Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. Curr. Biol. *23*, 2121–2129.

Self, M.W., Jeurissen, D., van Ham, A.F., van Vugt, B., Poort, J., and Roelfsema, P.R. (2019). The Segmentation of Proto-Objects in the Monkey Primary Visual Cortex. Curr. Biol. *29*, 1–11.

Semedo, J.D., Zandvakili, A., Machens, C.K., Yu, B.M., and Kohn, A. (2019). Cortical Areas Interact through a Communication Subspace. Neuron *102*, 249-259.e4.

Shamma, S.A., and Micheyl, C. (2010). Behind the scenes of auditory perception. Curr. Opin. Neurobiol. *20*, 361–366.

Shamma, S.A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. Trends Neurosci. *34*, 114–123.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. Trends Cogn. Sci. *12*, 182–186.

Singer, W. (1999). Neuronal Synchrony: A Versatile Code for the Definition of

Relations? Neuron *24*, 49–65.

Smiley, J.F., Hackett, T.A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D.C., and Schroeder, C.E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J. Comp. Neurol. *502*, 894–923.

Smith, P.H., Uhlrich, D.J., Manning, K.A., and Banks, M.I. (2012). Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. J. Comp. Neurol. *520*, 34–51.

Sohoglu, E., and Chait, M. (2016a). Detecting and representing predictable structure during auditory scene analysis. Elife *5*, 1–17.

Sohoglu, E., and Chait, M. (2016b). Neural dynamics of change detection in crowded acoustic scenes. Neuroimage *126*, 164–172.

Sołyga, M., and Barkat, T.R. (2019). Distinct processing of tone offset in two primary auditory cortices. Sci. Rep. *9*, 9581.

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., and Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory patterns. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160105.

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. Behav. Res. Methods, Instruments, {&} Comput. *31*, 137–149.

Stecker, G.C., and Middlebrooks, J.C. (2003). Distributed coding of sound locations in the auditory cortex. Biol. Cybern. *89*, 341–349.

Steinschneider, M., Reser, D.H., Fishman, Y.I., Schroeder, C.E., and Arezzo, J.C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. J. Acoust. Soc. Am. *104*, 2935–2955.

Sussman, E.S., Bregman, A.S., Wang, W.J., and Khan, F.J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cogn. Affect. Behav. Neurosci. *5*, 93–110.

Sussman, E.S., Horváth, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. Percept. Psychophys. *69*, 136–152.

Sutter, M.L., and Schreiner, C.E. (1991). Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. J. Neurophysiol. *65*, 1207–1226.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T.D. (2011). Brain Bases for Auditory Stimulus-Driven Figure-Ground Segregation. J. Neurosci. *31*, 164–171.

Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T.D. (2013). Segregation of complex acoustic scenes based on temporal coherence. Elife *2*, 1–16.

Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T.D., and Chait, M. (2016). Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. Cereb. Cortex *26*, 3669–3680.

Thiele, A., and Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. Nature *421*, 366–370.

Thiele, A., Delicato, L.S., Roberts, M.J., and Gieselmann, M.A. (2006). A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. J. Neurosci. Methods *158*, 207–211.

Tóth, B., Kocsis, Z., Háden, G.P., Szerafin, Á., Shinn-Cunningham, B.G., and Winkler, I. (2016). EEG signatures accompanying auditory figure-ground segregation. Neuroimage *141*, 108–119.

Town, S.M., Brimijoin, W.O., and Bizley, J.K. (2017). Egocentric and allocentric representations in auditory cortex. PLOS Biol. *15*, e2001878.

Town, S.M., Wood, K.C., and Bizley, J.K. (2019). Signal processing in auditory cortex underlies degraded speech sound discrimination in noise. BioRxiv 833558.

Tsunada, J., and Cohen, Y.E. (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. Front. Neurosci. *8*, 1–10.

Tsunada, J., Lee, J.H., and Cohen, Y.E. (2011). Representation of speech categories in the primate auditory cortex. J. Neurophysiol. *105*, 2634–2646.

Tsunada, J., Lee, J.H., and Cohen, Y.E. (2012). Differential representation of auditory categories between cell classes in primate auditory cortex. J. Physiol. *590*, 3129–3139.

Tsunada, J., Liu, A.S.K., Gold, J.I., and Cohen, Y.E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. Nat. Neurosci. *19*, 135–142.

Ulanovsky, N. (2004). Multiple Time Scales of Adaptation in Auditory Cortex Neurons. J. Neurosci. *24*, 10440–10453.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. Nat. Neurosci. *6*, 391–398.

Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., and von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. Psychol. Bull. *138*, 1172–1217.

Wang, X., Lu, T., Bendor, D., and Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. Neuroscience *154*, 294–303.

Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. Nat. Rev. Neurosci. *5*, 279–290.

Weinberger, N.M. (2012). Plasticity in the Primary Auditory Cortex, Not What You Think it is: Implications for Basic and Clinical Auditory Neuroscience. Otolaryngology *S3*, 997–1003.

Williamson, R.S., Ahrens, M.B., Linden, J.F., and Sahani, M. (2016). Input-Specific Gain Modulation by Local Sensory Context Shapes Cortical and Thalamic Responses to Complex Sounds. Neuron *91*, 467–481.

Winer, J.A., and Lee, C.C. (2007). The distributed auditory cortex. Hear. Res. *229*,

3–13.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., and Bloom, F.E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc. Natl. Acad. Sci. *90*, 8722–8726.

Wood, K.C., Town, S.M., Atilgan, H., Jones, G.P., and Bizley, J.K. (2017a). Acute Inactivation of Primary Auditory Cortex Causes a Sound Localisation Deficit in Ferrets. PLoS One *12*, e0170264.

Wood, K.C., Blackwell, J.M., and Geffen, M.N. (2017b). Cortical inhibitory interneurons control sensory processing. Curr. Opin. Neurobiol. *46*, 200–207.

Wood, K.C., Town, S.M., and Bizley, J.K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. Nat. Commun. *10*, 3019.

Yin, P., Fritz, J.B., and Shamma, S.A. (2014). Rapid Spectrotemporal Plasticity in Primary Auditory Cortex during Behavior. J. Neurosci. *34*, 4396–4408.

Zhong, L., Zhang, Y., Duan, C.A., Deng, J., Pan, J., and Xu, N.-L. (2019). Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. Nat. Neurosci.