A primate model of human cortical analysis of auditory objects

Pradeep Dheerendra

Institute of Neuroscience, Newcastle University

Thesis submitted for the degree of:

Doctor of Philosophy

September 2018

Abstract

The anatomical organization of the auditory cortex in old world monkeys is similar to that in humans. But how good are monkeys as a model of human cortical analysis of auditory objects? To address this question I explore two aspects of auditory object processing: segregation and timbre. Auditory segregation concerns the ability of animals to extract an auditory object of relevance from a background of competing sounds. Timbre is an aspect of object identity distinct from pitch. In this work, I study these phenomena in rhesus macaques using behaviour and functional magnetic resonance imaging (fMRI). I specifically manipulate one dimension of timbre, spectral flux: the rate of change of spectral energy.

I present this thesis in five chapters. Chapter 1 presents background on auditory processing, macaque auditory cortex, models of auditory segregation, and dimensions of timbre. Chapter 2 presents an introduction to fMRI, the design of the fMRI experiments and analysis of fMRI data, and macaque behavioural training techniques employed. Chapter 3 presents results from the fMRI and behavioural experiments on macaques using a stochastic figure-ground stimulus. Chapter 4 presents the results from the fMRI experiment in macaques using spectral flux stimulus. Chapter 5 concludes with a general discussion of the results from both the studies and some future directions for research.

In summary, I show that there is a functional homology between macaques and humans in the cortical processing of auditory figure-ground segregation. However, there is no clear functional homology in the processing of spectral flux between these species. So I conclude that, despite clear similarities in the organization of the auditory cortex and processing of auditory object segregation, there are important differences in how complex cues associated with auditory object identity are processed in the macaque and human auditory brains.

Abbreviations

A1	Primary auditory cortex
AC	Auditory Cortex
AAC	Auditory Association Cortex
AFM	Auditory Field Map
AL	Anterior Lateral area
BOLD	Blood Oxygenation Level Dependent signal
CiS	Circular Sulcus
CL	Caudal Lateral area
CM	Caudal Medial area
CPB	Caudal Parabelt area
ECoG	Electrocorticography
EEG	Electroencephalography
EPI	Echo-Planar Imaging
fMRI	functional Magnetic Resonance Imaging
FWHM	Full width at half maximum
GLM	General Linear Model
HG	Heschl’s Gyrus
HRF	Haemodynamic Response Function
Hz	Hertz
IC	Inferior Colliculus
IPS	Intra Parietal Sulcus

KA	Koniocortical area
MDS	Multi-Dimensional Scaling
MEG	Magnetoencephalography
MGB	Medial Geniculate Body
ML	Medial Lateral area
MM	Medial Middle area
MRI	Magnetic Resonance Imaging
MUA	Multi Unit Activity
NHP	Non-Human Primates
PAC	Primary Auditory Cortex
PaA	Parakoniocortical area
PaI	Para-insular area
ProA	Prokoniocortical area
PET	Positron Emission Tomography
PRT	Positive Reinforcement Training
PT	Planum Temporale
R	Rostral area
ROI	Region of Interest
RM	Rostral Medial area
RPB	Rostral Parabelt area
RT	Rostral Temporal area
RTL	Rostral Temporal Lateral area

RTM	Rostral Temporal Medial area
SFG	Stochastic Figure Ground Stimulus
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
SPM	Statistical Parametric Maps
STG	Superior Temporal Gyrus
STP	Superior Temporal Plane
STS	Superior Temporal Sulcus
SUA	Single Unit Activity
T1w	T1 (Longitudinal relaxation time)-weighted
T2w	T2 (Transverse relaxation time)-weighted
TA	Time to Acquire or Acquisition time
TE	Time to Echo
TI	Inversion time
TR	Time to Repeat
Tpt	Temporo-parietal area
Ts3	Temporalis superior 3

Acknowledgements

I would like to thank my supervisor Professor Tim Griffiths for his constant support, encouragement and guidance.

I am supported by Wellcome Trust, UK, four-year PhD studentship program (Grant No: WT102561/Z/13/Z). I am also supported by the Overseas Research Scholarship from the Faculty of Medical Sciences, Newcastle University. I thank them both for giving me this opportunity to pursue my Masters and PhD in Neuroscience.

I would like to thank the staff of the Comparative Biology Centre for the excellent animal care they provide, the training they gave me in handling the NHPs, and the teaching (ASPA module 1-3 coursework, module 21-22 on animal surgery) and support they provided to facilitate my procurement of a UK Home Office personal license to work with macaques.

I had help from Dr Fabien Balezeau and Dr Simon Baumann in the data collection during fMRI macaque experiments. I would like to thank my colleague Felix Schneider for taking forward the macaque behaviour work. I also thank Dr Olivier Joly for helping me parcellate the monkey auditory cortex.

I would like to thank Professor Alexander Thiele, Dr Alwin Gieselmann, and Dr Christian Brandt, for their guidance on macaque training. I also thank Dr Sukhbinder Kumar for his help in stimulus construction and discussions on fMRI data analysis. I would like to thank Dr David Hunter for sharing his expertise by providing technical support during fMRI data acquisition. I also thank Professor Adrian Rees and Professor Chris Petkov for their guidance.

I indebted to my friends Dr Thomas Hall, Dr Krishnan Srinivasarengan, and Dr Damar Susilaradeya for their help and support at times when a helping hand mattered the most. I am also indebted to the wonderful organization which is the National Health Service (NHS). I thank my parents for their support and encouragement.

Declaration

The work presented in this thesis is original and carried out by Pradeep Dheerendra except where explicitly acknowledged in the text.

Pradeep Dheerendra Date 21-09-2018

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Auditory Object

A visual object may be intuitive to define and understand, however objects perceived via other senses, including auditory, olfactory, and tactile, are important. So it becomes relevant to define an object in a way that is agnostic of the sensory modality. However, an object cannot be considered independently of the sense data responsible for its perception (Kant, 1929). Objects in any sensory modality can be defined as physical entities responsible for a coherent perceptual whole that is distinct from other aspects of the environment (Griffiths and Warren, 2004). Further, 'Object analysis' refers to the process of perceiving the world that encompasses the different tasks of extraction, representation, abstraction, maintenance and integration of the various objects across all sensory modalities (Griffiths et al., 2012).

Consider an acoustically rich habitat where sounds from various sources overlap in time and frequency. The spectral and temporal characteristics of these naturally occurring sounds are complex. Despite this complexity, humans and animals are able to handle the composite waveform to extract objects of relevance. The process through which the auditory system accomplishes the transformation of an acoustic signal to an object-based representation is termed as 'auditory scene analysis' (Bregman, 1990) and 'auditory segregation' is one of its fundamental aspects.

The analysis of auditory-object is defined as the "computational result of the auditory system's ability to detect, extract, segregate, and group the spectrotemporal regularities in the acoustic environment into stable perceptual units". In other words, it is the "perceptual consequence of the auditory system's interpretation of acoustic events" (Bizley and Cohen, 2013). Auditory objects can be represented as complex shapes within the spectrotemporal representation of a sound (Griffiths and Warren, 2004). Such shapes are determinant of the quality or 'timbre' of auditory objects. While pitch corresponds to the repetition rate that allows us to compare and put sounds in order on a scale, timbral properties distinct from pitch allow humans and animals to distinguish different voices and sounds (Krumhansl, 1989).

Primates encounter a range of species-specific calls apart from calls from other species and environmental sounds. These calls and sounds may refer to objects or events in the environment and may convey information about food, predators, social

relationships, caller identity apart from the emotional state of the caller (Ghazanfar and Hauser, 1999). Thus the auditory system must segregate and analyse the spectrotemporal structure of the auditory objects to extract the invariant acoustic cues that convey meaning (Wang, 2000, Zoloth and Green, 1979, Beecher et al., 1979, May et al., 1989). In this work, I will focus on two aspects of auditory objects processing in monkeys namely auditory segregation and timbral analysis.

1.2 Auditory scene analysis

In a natural habitat, sounds from multiple sources may occur simultaneously, overlapping in frequency as well as time. Despite this acoustic complexity, humans and other animals have evolved to be able to extract effortlessly from the acoustic mixture sounds of relevance to guide their behaviour (Bregman, 1990). Colloquially this is also known as the ‘cocktail party problem’ (Cherry, 1953). This extraordinary ability of the auditory system to break apart and organize a composite acoustic wave into its constituent sources is explored in this thesis.

The term ‘Auditory scene analysis’ is defined as the process by which the auditory system accomplishes the transformation of an acoustic signal from low-level sensory information into high level object-based perceptual representation (Bregman, 1990). Auditory segregation or perceptual organization is a fundamental aspect of scene analysis that refers to the task of parsing an acoustic scene to perceptually extract a specific auditory object from a background of competing sounds into a specific stream. This process either activates old representations or leads to the formation of new objects using a listener’s experience and knowledge of the auditory environment. Thus, an auditory stream is a sequence of sounds that are grouped together by perceptual properties (Moore and Gockel, 2002). After streams are formed, one can selectively attend and track just one at a time (Sussman et al., 2007), but one can choose to switch between them at will.

1.2.1 Auditory segregation

The process of auditory segregation exploits the cues based on commonalities across sounds in the environment. These commonalities could be based on cues that occur at the same time (in Gestalt terms 'common fate') or remain consistent across time (in Gestalt terms 'proximity'). For instance, sounds that start or change at the same time or modulate in frequency or amplitude, in the same manner, are likely to be produced by the same source. Likewise, sequential sounds with similar pitch or frequency are likely to come from the same source than sounds from dissimilar pitch or frequency. Thus, there are two kinds of processes under auditory scene analysis, those that deal with the perceptual organization of either simultaneously occurring acoustic elements or sequential occurring acoustic elements.

In the first case, the simultaneous organization acts as a grouping cue as it presents a vertical boundary in the spectro-temporal domain. It has been investigated by

manipulating one specific property of multiple concurrent components. In humans, earlier studies have shown that segregation to perceive concomitant sources could be achieved by manipulating properties such as spatial location (McDonald and Alain, 2005), onset asynchrony (Lipp et al., 2010), harmonicity (McDonald and Alain, 2005, Lipp et al., 2010) and common amplitude modulation (Hall et al., 1984).

Auditory stream segregation refers to this process of forming a segregated percept of auditory sources by a sequential grouping of acoustic elements over time (Shamma et al., 2011). For instance, speech and music are perceived as coherent streams that can be selectively attended to and followed over time. In humans it has been shown using a sequence of tones that segregation can be achieved based on the grouping of regularities in frequency (van Noorden, 1975), temporal synchrony (Elhilali et al., 2009a), timbre (Singh, 1987, Iverson, 1995), harmonicity (Moore et al., 1986), inter-aural time difference (Stainsby et al., 2011, Darwin and Hukin, 1999), temporal envelope (Grimault et al., 2002), fundamental frequency (Vliegen and Oxenham, 1999), asynchrony (Darwin and Carlyon, 1995), phase spectrum (Roberts et al., 2002), and spatial position (McDonald and Alain, 2005, Hill et al., 2011), and ear of entry (Darwin and Carlyon, 1995).

1.2.2 Previous investigations

Table 1-1 presents a brief summary of some of the important studies of auditory segregation and the models proposed.

1.2.2.1 Van Noorden Paradigm

A sequence (see Figure 1-1 a) of alternating tones (ABAB…) or tone-triplets (ABA…) with a difference in frequency of tones A and B were employed by van Noorden (1975) to demonstrate the effect of frequency separation as well as presentation rate on auditory stream segregation. At slow presentation rate, the sequence is perceived as a single stream but at faster presentation rates, the sequence is perceived as a split stream of separate higher and lower frequencies. Similarly, if the frequency separation between A and B tones is increased then the probability of perceiving the sequence as separate streams increases. Here, the attended stream becomes foreground while the other stream is background. So this is a type of scene analysis.

Despite the fact that these two streams are presented as a single sequence, they are perceived as either single integrated stream i.e. a single source (this is known as

‘fusion’) or two separated streams i.e. disparate sources (this is known as ‘fission’) depending upon the tone presentation rate and tone frequency separation. The percept can be ‘bistable’ i.e. it switches between integrated or segregated percepts, for values of presentation rate and frequency separation that are intermediate between fission and fusion boundaries. The integrated percept is facilitated by bottom-up grouping mechanisms while the segregated percept is furthered by top-down attention processes to separate the two streams. There is a gradual increase in the probability of perceiving separate streams upon listening for a long duration though at the beginning the probability of integrated percept is much higher. This phenomenon is known as ‘build-up of streaming’. However, this process is susceptible to listener’s attention shifts and sudden changes in the sequence properties which resets the percept.

Bregman postulated two types of brain mechanisms underlying auditory segregation. First, a primitive bottom-up mechanism that encodes sensory characteristics of stimuli after which these attributes are utilized for grouping based on Gestalt principles described earlier. Second, higher level processes that have to be learned through experience and these schemas enable recognition of patterns in the incoming auditory stimuli. Both these processes are influenced by top-down specific attention devoted to them.

Auditory scene analysis was first explicitly studied in a non-human animal by Hulse et al. (1997) using European starling because starlings are competent at responding to their conspecifics vocalizations amongst others. Using operant training methods Starlings were trained to correctly identify stimuli containing the starling song compared to stimuli that did not. Next, MacDougall-Shackleton et al. (1998) trained starlings to discriminate between galloping rhythm of ABA- sequence from either of the individual isochronous patterns of bursts (A-A- or B---B---). After the birds achieved sufficient accuracy, novel probe stimuli (ABA- sequences with varying frequency difference between A and B tones) were introduced on a small proportion of trials. If birds reported percept of these probe stimuli similar to individual isochronous stimuli, then it was hypothesized that they perceived a segregated percept. The proportion of reports indicating isochronous patterns increased with increasing frequency difference between A and B tones of the probe stimuli. Thus stream segregation was the best explanation for the starlings’ behaviour in this experiment. Similar behavioural results were reported in finches as well. So one can

conclude that scene analysis is a general biological phenomenon, and that it doesn't require a mammalian brain.

Fay (1992) classically conditioned goldfish (*Carassius auratus*) to respond to a simple mixture of two tones and then tested for generalization to stimuli that consisted of a single pure tone. Generalization phenomenon permits an estimate of the extent to which a novel probe sound generalizes from the training sound after conditioning to it. It can be interpreted as the degree of equivalence or perceptual similarity of the probe and training stimuli. Segregation of tones from the mixtures would result in a generalization gradient with peaks at the component frequencies of the mixture. Goldfish behaviour showed a two-peaked generalization function implying that they can segregated mixture into streams. Fay (1998) carried out an auditory stream segregation experiment on goldfish by classically conditioning to a simultaneous mixture of two pulse trains, the high-frequency pulse repeated at a high rate, and the low-frequency pulse repeated at a low rate. They demonstrated that goldfish correctly associated a particular spectral envelope with its repetition rate i.e., information about the two mixed pulse trains was obtained independently. Thus auditory stream segregation provided the best description of these results. So fishes, in general, are capable of stream segregation.

Streaming studies in frogs make use of characteristics of their natural calls which elicit a phonotactic approach behaviour towards the acoustic source that matches their mating calls. The advertisement call in Cope's Gray treefrog (*Hyla chrysoscelis*) has a frequency range from 1 to 2.8 kHz with a pulse rate of 35-50 pulses per second. Nityananda and Bee (2011) employed this species to investigate stream segregation. They used stimuli consisting of target pulses presented at the natural rate interleaved with distractor pulses of similar frequency. The approach response towards the sound source increased with increase in frequency difference between target and distractor. This demonstrated stream segregation in frogs.

Stream segregation in rats (*Rattus*) was studied using ABA triplet. Wistar rats were trained to discriminate between segregated slow isochronous rhythms corresponding to 'B' tones, fast isochronous rhythms corresponding to 'A' tones and galloping rhythms corresponding to the integrated percept of ABA tones. Noda et al. (2013) reported that rats perceive segregated stream if the frequency difference between the A and B tones is more than 12 tones but perceive an integrated percept if the

frequency difference is 6 semitones or less. This study was the first demonstration that rats are able to perform auditory stream segregation.

Behavioural results showing that ferrets (*Mustela putorius*) were able to perform stream segregation based was demonstrated by Ma et al. (2010). Ferrets were trained to report the frequency shift in B tones in an ABAB tone paradigm that differed in frequency. They found in ferrets that as the frequency difference between 'A' and 'B' tones increased, the threshold for detecting the frequency shift decreased. From human psychophysics, we know that the sensitivity for detecting frequency shifts of B tones in ABAB paradigm is better if 'A' and 'B' tones are processed in separate streams. Thus this study demonstrates that ferrets are able to segregate sounds into separate auditory streams.

Thus it seems likely that all vertebrate animals have the capacity for stream segregation and thus some form of auditory scene analysis. So scene analysis can be studied in a comparative context.

Using this streaming paradigm, macaque neurophysiology (Fishman et al., 2001) was conducted to infer the neuronal basis of stream segregation. This study employed ABAB sequence and manipulated the presentation rate to record the multiunit activity in primary auditory cortex (PAC or A1 core region) where the A tones are presented at the best frequency of the single unit, thus, B tones are the non-best frequency in nature but (up to 12 semitones away). At low presentation rates, the evoked response to 'B' tones was similar to that of 'A' tones and, thus, the evoked responses were generated at the same rate as stimulus presentation. However, at fast presentation rates, the evoked responses to 'B' tones were relatively suppressed in comparison with that of 'A' tones and the evoked responses occurred at half the rate of the stimulus presentation. This suggests a basis for the representation of a segregated percept, which was interpreted in terms of a differential suppression of responses to tones via forward masking. The proposed model suggests that segregated percept due to frequency separation primarily arises out of the spatial separation of the responses to the individual tones due to tonotopy: adaptation and forward masking of responses lead to segregated responses at faster presentation rates. However, as these recordings were carried with no active behaviour the association between perceptual state and neuronal responses is not directly demonstrated. But Izumi (2002) showed that macaques could perform

auditory stream segregation based on frequency separation. Further, Christison-Lagay and Cohen (2014) used the ABAB sequence to study streaming percept in rhesus macaques and report behavioural results that suggest that macaques are able to segregate similar to humans. Since responses to these tones in non-primary auditory regions were not recorded in this experiment, the observations were based on recordings from PAC only. Further Fishman et al. (2004) showed an effect of tone duration where increasing it increases the differential suppression of B tones. These experiments have been interpreted in terms of adaptation-based models, in which physical separation between neuronal populations in A1 coding for different elements of the stream facilitated segregation.

Micheyl et al. (2005) recorded responses from single units in PAC of awake rhesus macaques for ABA tone triplet sequences. They showed a temporal build-up of streaming as a function of frequency separation and presentation rate. They proposed a model to predict the probability of integrated versus segregated percepts using the statistical variability of neuronal responses. If the spike count for a tone triplet exceeds a threshold determined by maximizing the fit between the data and model, then integrated percept is predicted while if the spike count for a single tone of the triplet exceeds the threshold then a segregated percept is predicted.

Multiunit activity was recorded from forebrain of awake European starlings (*Sturnus vulgaris*) while ABA- tone paradigm with a difference in frequency between A and B tones was presented. Bee and Klump (2004) reported that larger frequency difference and shorter tone repetition time between A and B tones resulted in neuronal response pattern that was similar to macaque neurophysiology results.

Neural responses from the auditory cortex of mustached bats (*Pteronotus parnellii*) were recorded while they were presented with ABAB tone paradigm with a difference in frequency between ‘A’ and ‘B’ tones. Kanwal et al. (2003) reported increased suppression in the neuronal responses with a decrease in tone repetition time between tones. These results are line with the results from macaque neurophysiology.

Single unit spike responses were recorded from the auditory cortex of anaesthetized guinea pigs (*Cavia porcellus*) while they were presented with ABAB tone paradigm with a difference in frequency between A and B tones. Scholes et al. (2015) reported

suppression of responses to off-best frequency B tones which was consistent with findings from macaque recordings discussed earlier.

Local field potentials were recorded from the auditory cortex of anaesthetized rats (*Rattus*) while they were presented with van Noorden tone paradigm. Noda et al. (2013) reported that amplitude and the phase of the cortical oscillatory activity in the gamma band are important for auditory stream segregation.

Neurophysiological investigations into streaming have been carried out in other animal models, including cats (*Felis catus*) (Brosch and Schreiner, 1997), ferrets (*Mustela putorius*) (Elhilali et al., 2009a), and treefrogs (*Hyla*) (Bee, 2015). These investigations concurred with the literature discussed above.

However, the results from anaesthetized guinea pigs (Pressnitzer et al., 2008) obtained while presenting with ABA- tone paradigm using single unit recordings in the ventral part of the cochlear nucleus showed frequency selectivity and forward suppression in line with the results from primary auditory cortex of other animals. This implied that stream segregation occurred as early as in the cochlear nucleus.

1.2.2.2 Informational Masking paradigm

Another paradigm (see Figure 1-1 b) employed to understand auditory segregation is 'Informational Masking' (IM) paradigm (Gutschalk et al., 2008, Kidd et al., 1994) associated with central processes where the threshold for detection of target increases due to the perceptual similarity of masker to target. This is different from energetic masking associated with the auditory periphery where an increase in detection threshold occurs due to overlapping activation in the cochlea or auditory nerve. This paradigm requires subjects to identify target signal which can be a tone sequence located amidst multi-tone masker that is presented simultaneously (Kidd et al., 1994), but with a spectral protective band around the target which facilitates segregation of target from the background. Despite the spectral spacing between target and masker, which is designed to minimize energetic masking, listeners still are unable to detect the target at times due to distraction from the maskers. The detection of the target is dependent on the spectral protection width and the masker density (Elhilali et al., 2009b, Gutschalk et al., 2008, Micheyl et al., 2007b).

Adaptation based mechanisms have been proposed to explain target detection in the above paradigm. Using magnetoencephalography (MEG) in humans contrasting

evoked fields in trails where target tones were detected vs undetected, Gutschalk et al. (2008) reported a response in PAC termed ‘awareness related negativity’. In line with this observation, a similar manipulation in an fMRI study (Wiegand and Gutschalk, 2012) yielded activations in medial Heschl’s Gyrus (HG). When attention state was manipulated between foreground (frequency deviant in target) and background (duration deviant in masker), MEG (Elhilali et al., 2009b) in humans revealed that attention modulates neural representation of foreground signal. In real life situations, sounds do not have a protective spectral gap around them but overlap both in time and frequency. Thus, this paradigm does not fully model the complexities of natural auditory scenes (Dykstra and Gutschalk, 2013).

1.2.2.3 Temporal coherence

The grouping of temporally coherent elements appears as an important aspect of how the auditory system solves the ‘binding’ problem. However, temporal coherence aiding in feature binding is not specific to auditory modality and has been proposed in other sensory modalities as well. For instance, a principle similar to temporal coherence has been suggested in visual modality (Alais et al., 1998, Blake and Lee, 2005). Further, temporal coherence between cortical areas corresponding to different modalities has been suggested to play a role in cross-modal binding (Mirbagheri et al., 2012), for instance, lip reading is thought to be one such instance.

When Elhilali et al. (2009a) employed sequentially presented synchronous tones, the behavioural results from humans showed that despite the frequency difference (e.g. greater than 10 semitones) between the two tones, subjects tend to report hearing one stream when tones were presented synchronously as opposed to their report of two streams when the same tones were presented asynchronously. These results could not be explained by the adaptation based models. Ferrets show a similar behavioural phenomenon (Ma et al., 2010) and parallel neurophysiology recordings led to the conclusion that the tonotopic separation between active neuronal populations is not a sufficient condition for forming a two-stream percept. Temporal coherence between elements in a stream was suggested as an alternate basis for streaming (Shamma and Micheyl, 2010).

More than half the world's population above 75 years suffer age-related hearing loss (Lin et al., 2011), also known as 'Presbycusis' and this can have a severe impact on quality of life (Ciorba et al., 2012). Patients report difficulty understanding speech amidst background noise (Dubno et al., 1984), for instance, when hearing someone talk in a noisy café. Since speech is an auditory object which has multiple frequency components that start and stop together, this symptom can be interpreted using the temporal coherence model of segregation. The age-related hearing loss in some subjects is characterised by hearing loss at higher frequencies, as well as poor frequency and time resolution (Wingfield et al., 2005). Loss of frequency selectivity due to broadened auditory filters results in reduced resolvability of frequency components. This results in reduced ability to perceptually isolate simultaneous harmonic sounds and thus reduced ability to extract individual properties of sound. In the context of temporal coherence, this reduced frequency resolution together with inaudible higher frequencies causes the temporal coherence of speech to reduce and this may explain the observed difficulty in perceiving speech in a noisy environment.

In summary, an important aspect of stream formation is the temporal relationship between different components of an auditory scene. This temporal aspect determines the segregation of different components into respective groups.

1.2.3 Synthetic stimulus

Conspecific vocalizations like human speech can be employed as stimuli for scene analysis. However, there are some associated disadvantages of using such natural communication sounds. Speech sounds have semantic content that elicits other cognitive processes (top-down) while a parametric control over its spectrotemporal properties is not possible in the same way as feasible for synthetic stimuli. Also, a synthetic stimulus could be employed as an 'audiogram for auditory scene analysis' to characterise deficits in understanding speech-in-noise whilst avoiding the bias of semantic clues present in the speech stimuli. More importantly, a synthetic stimulus has added advantage that it could be equally applied to humans as well as animals without any confounds. So I choose to employ synthetic stimulus in my experiments to address questions regarding the development of a primate model.

1.2.4 Stochastic Figure-Ground

A synthetic stimulus (see Figure 1-1 c) introduced in Teki et al. (2011) approximates the challenges of scene analysis faced in real life. This paradigm is used to investigate the temporal coherence model of auditory segregation. This stimulus consists of a 'figure' made of temporally coherent spectral elements that repeat in time against a background of randomly varying spectral elements. So the figure is heard as a warble within the ongoing background where both components overlap in spectro-temporal space without any spectral gap between them. Figure coherence is defined as the number of temporally coherent spectral elements that repeat in time. The saliency of the figure increases with figure coherence. The spectral properties of the figure vary with each exemplar and are only distinguishable from the ground by their fluctuation statistics.

Based on haemodynamic data in humans, Teki et al. (2011) showed brain bases underlying automatic, stimulus-driven auditory figure-ground decomposition. Significant activations were reported bilaterally in superior temporal sulcus (STS) and intraparietal sulcus (IPS), a non-auditory region, to increasing coherence. However, no significant activity was reported in primary auditory cortex contrary to reports from previous studies using other types of streaming paradigms.

Teki et al. (2013) conducted behavioural experiments in humans with normal hearing using SFG stimulus. The figure detection performance increased with increasing figure coherence and increasing figure duration (longest figure duration was 350 ms). This suggested that SFG stimulus taps low-level finely tuned segregation mechanisms. The listeners could reliably identify identical figures that were different from other figures against random background components (given sufficient coherence and duration) indicating that figure detection is associated with a grouping of coherent components as a distinct perceptual object rather than through the detection of some low-level changes in the stimulus. Figures were made of components that were smaller in duration to test whether the performance is affected by temporal scaling. They reported that performance largely depended on the number of repeating chords irrespective of the timescale. Figures interrupted by broadband noise were employed to infer whether figure detection was accomplished by low-level mechanisms sensitive to power increase in certain frequency bands. Robust performance for figure detection indicated that segregation mechanisms were robust to temporal perturbations. Figures were ramped in frequency instead of

keeping them fixed to understand the effect of the spectral perturbation on figure detection. They reported that segregation mechanisms were more susceptible to spectral rather than temporal perturbations but listeners could still integrate over dynamically changing rather than fixed figure components. In order to test whether segregation is mediated by deviation from adaptation to ongoing background statistics, isolated figures (without a preceding ground only segment) were employed. Robust detection of figures indicated that segregation could not be explained by detection of deviation of adaptation to statistics of ground components. The ability of listeners to identify figures interrupted by extended noise (up to 500 ms) indicated that underlying high-level segregation mechanisms are robust over long windows.

In summary, they showed that the figure detection in SFG stimuli is not associated with detection of any low-level changes. They also showed that the ability to detect figure increases with figure coherence. They show that these results are consistent with the predictions of a model of perceptual organization based on coherence.

To understand the dynamics of the temporal coherence model of stream segregation, O'Sullivan et al. (2015) employ electroencephalography (EEG) in humans and manipulate the attention. They showed an early effect of temporal coherence in the passive listening condition (lasting from 115 ms to 185 ms) which was larger and lasted longer during active listening condition (lasting from 115 ms to 265 ms). This study provided evidence for early and pre-attentive neural computation of temporal coherence that is enhanced by active analysis of the auditory scene. Using MEG (Teki et al., 2016) in naïve and distracted humans revealed robust evoked responses that reflected the emergence of the figure from the ground. The neural sources underlying this process were localized to planum temporale and intraparietal sulcus.

Using electrocorticography (ECoG) in humans with normal hearing and employing this SFG stimuli revealed induced high gamma band (above 80 Hz) activity for the transition from ground only components to a figure (i.e. for an emergence of a figure) in Superior Temporal Gyrus (STG), near the end of HG, across both hemispheres. This strongly suggests a role for bilateral parabelt homologues on the convexity of the STG (Griffiths, 2017) in the perceptual organization.

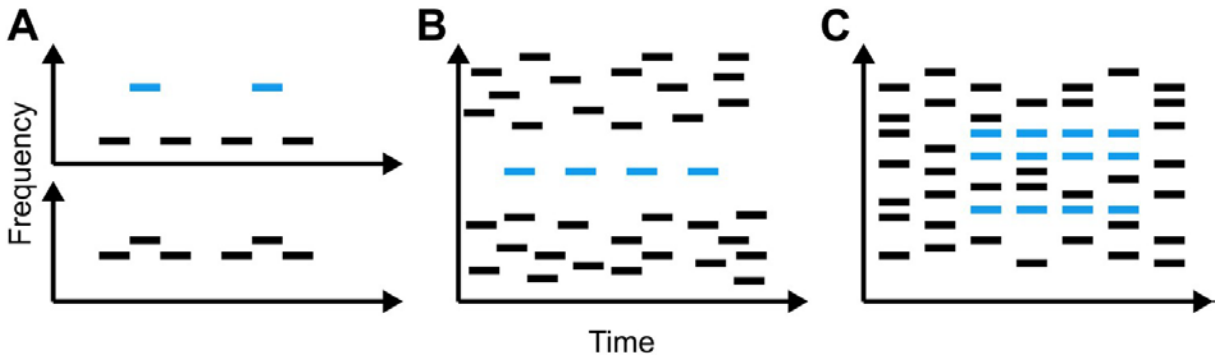


Figure 1-1 Different types of synthetic stimulus used in the study of auditory scene analysis (**A**) ABA stimulus from van Noorden paradigm (**B**) Informational masking stimulus paradigm (**C**) Stochastic Figure Ground stimulus. This image is reproduced from Dykstra and Gutschalk (2013)

Study	Species	Technique	Proposed Model	Comments
Van Noorden Paradigm				
Fishman et al. (2004)	Macaque	MUA	Spatial Separation	A1
Kondo and Kashino (2009)	Humans	fMRI	Spatial Separation	MGB
Cusack (2005)	Humans	fMRI	Multilevel model	IPS
Pressnitzer et al. (2008)	Guinea pigs	SUA	Spatial Separation	Periphery (VCN)
Elhilali et al. (2009a)	Ferrets	SUA	Temporal Coherence	A1
Informational Masking Paradigm				
Gutschalk et al. (2008)	Humans	MEG	Adaptation based model	PAC
Elhilali et al. (2009b)	Humans	MEG	Temporal Coherence	Effect of attention
Wiegand and Gutschalk (2012)	Humans	fMRI, MEG	Adaptation based model	Medial HG
Stochastic Figure Ground paradigm				
Teki et al. (2011)	Humans	fMRI	Temporal Coherence	STS, IPS
O'Sullivan et al. (2015)	Humans	EEG	Temporal Coherence	Effect of attention
Teki et al. (2016)	Humans	MEG	Temporal Coherence	PT, IPS
Griffiths (2017)	Humans	ECoG	Temporal Coherence	STG

Table 1-1 Summary of few important studies and the proposed models for perceptual organization across three main stimulus paradigms MGB – Medial Geniculate Body; VCN – Ventral Cochlear Nucleus;

1.3 Timbral analysis

1.3.1 Timbre

Timbre is a perceptual description of an auditory object. It is a key determinant of sound identity. For a lack of positive definition, it is defined as the property of sound that enables one to distinguish two sounds with an identical pitch, loudness, perceived duration and position within a soundscape (American-Standards-Association and Acoustical-Society-of-America, 1960). In other words, it allows one to distinguish between different musical instruments (Menon et al., 2002) even when they are playing the same pitch. It is timbre that allows us to distinguish two vowels sounds spoken at the same pitch, as well as different environmental sounds.

In music perception research, timbre is that aspect of music that is least understood, since there are detailed models of the pitch, loudness perception and coding of temporal information from psychoacoustics. The traditional definition of timbre would be found inadequate if the domain is broadened to include sounds other than those produced by musical instruments.

In human speech, timbral differences arise due to the filtering of periodic train produced by the vocal chords through the throat (velum), mouth (lips) and tongue i.e. the resonance imposed by them. Thus timbre acts as a principal determinant of phonetic identity in speech. Animal vocalizations have socioecological significance including for non-human primates like rhesus macaques, since conspecific vocalizations differ in timbre that enables identification of the caller and provides useful clues on age, gender, emotional and motivational state of the caller (Cheney and Seyfarth, 1990, Hauser, 1996) and can convey information about objects, like food, and events, like threats, in the environment.

Timbre is a multidimensional property of the acoustic structure (Licklider, 1951). It is determined by both spectral and temporal features of a sound. Thus extraction of timbre requires analysis of both spectral and temporal envelope of a sound source (Lyon and Shamma, 1996). This necessitates the need to understand the underlying dimensions of timbre.

1.3.2 Multi-dimensional scaling

Multi-dimensional scaling (MDS) studies allow inference on the underlying perceptual dimensions of a sound feature. MDS based methods work by mapping auditory

objects into Euclidean space, followed by a comparison of these mapped dimensions to acoustic properties of the stimulus. In these studies, a number of tones differing in timbre but identical in pitch, loudness, and perceived duration are presented in all possible pairs to subjects who have to rate how dissimilar the tones of each pair are on a scale. Then MDS algorithm is applied to this matrix of dissimilarity ratings where the algorithm tries to establish a monotonic relation between dissimilarity ratings and Euclidean distances among the sounds arranged in a geometric structure in 'n' dimensions, each sound is represented as a point (see Figure 1-2). Sounds with similar timbre are, thus, near one another in the space and those with dissimilar timbres are farther apart. A solution is selected after trying a different number of dimensions. This solution should be a compromise between having a small difference between ratings and not having more dimensions than can be readily interpreted in terms of their underlying perceptual and psychophysical relevance to the group of subjects tested. These methods were suggested to be more sensitive than discrimination paradigms to subtle perceptual differences in timbre perception (Samson et al., 2002).

1.3.3 Dimension of timbre

The dimensions of musical timbre have been studied using MDS methods (McAdams and Giordano, 2009, McAdams, 1999). The findings from MDS studies on timbre using computer-based synthesized sounds of musical instruments are summarized in Table 1-2. Using the MDS method, Wessel (1979) described two dimensions viz. brightness of the steady-state portion of the sound and rapidity of the attack and relative onsets of high and low spectral components. In a study on timbre by Grey (1977), the dimensions were found to be interpretable in terms of spectral energy distribution, the presence of synchronicity in the transients of higher harmonics along with a closely related amount of spectral fluctuation through time, and the presence of low-amplitude high-energy in the initial attack segment. Grey and Gordon (1978) suggested dimensions of timbre were spectral centroid, attack synchronicity or spectral flux and attack centroid. Iverson and Krumhansl (1993) reported two dimensions that correspond to spectral centroid and amplitude envelope. In the study by Krumhansl (1989), the proposed perceptual dimensions of timbre were spectral brightness, attack time, and spectral flux. The study by McAdams and Cunible (1992), the timbral dimensions were suggested to qualitatively correspond to the distribution of spectral energy, onset characteristics and degree of change in spectral

energy distribution (see Figure 1-2). Lakatos (2000) suggested three dimensions of timbre as amplitude envelope, spectral centroid, pitch strength or noisiness or spectral density.

Across studies, several different potential acoustic cues were proposed to correspond to timbre dimensions using various acoustic descriptors viz. spectral centroid, spectral deviation, spectral density, attack time, decay time, amplitude envelope, spectral flux, pitch strength, attack synchrony, attack centroid, noisiness. However, numerous MDS studies (Caclin et al., 2005, Burgoyne and McAdams, 2007, Lakatos, 2000, McAdams et al., 1995) concur in their findings that two essential dimensions of timbre are related to spectral envelope and temporal envelope. In summary, the timbral dimension of spectral envelope is broadly related to the centre of gravity of the long-term spectrum and corresponds to what musicians call ‘brightness’. The timbral dimension of the temporal envelope could be called ‘attack quality’ since it is broadly related to the rapidity of the attack and presence of inharmonic transients at the beginning of the tone. There is some disagreement on the third dimension of timbre. Krimphoff et al. (1994) suggested that it might correspond to ‘spectral fine structure’ which was defined as the standard deviation of time-averaged harmonic amplitudes from a spectral envelope. A recent study (Elliott et al., 2013) also suggested that most dimensions of timbre are not a result of purely spectral and temporal features but instead depend on spectrotemporal patterns. Thus, the third dimension of timbre could be related to a combined spectro-temporal property. In summary, two timbral dimensions that have consensus broadly correspond to the spectral and temporal envelope whilst the third dimension might correspond to an interaction between the spectral and temporal envelopes.

Study	Proposed dimensions
Grey (1977)	Spectral energy distribution; Synchronicity in the transients of higher harmonics; Rate of spectral fluctuation
Grey and Gordon (1978)	Spectral centroid; Attack synchronicity or spectral flux; Attack centroid
Wessel (1979)	Brightness; Attack; Relative onsets of high and low spectral components
Krumhansl (1989)	Brightness; Attack; Spectral flux
McAdams and Cunible (1992)	Spectral energy distribution; Onset characteristics; Degree of change in spectral energy distribution
Iverson and Krumhansl (1993)	Spectral centroid; Amplitude envelope
Lakatos (2000)	Spectral centroid; Amplitude envelope; Pitch strength or noisiness or spectral density

Table 1-2 Summary of studies and the proposed dimensions of timbre

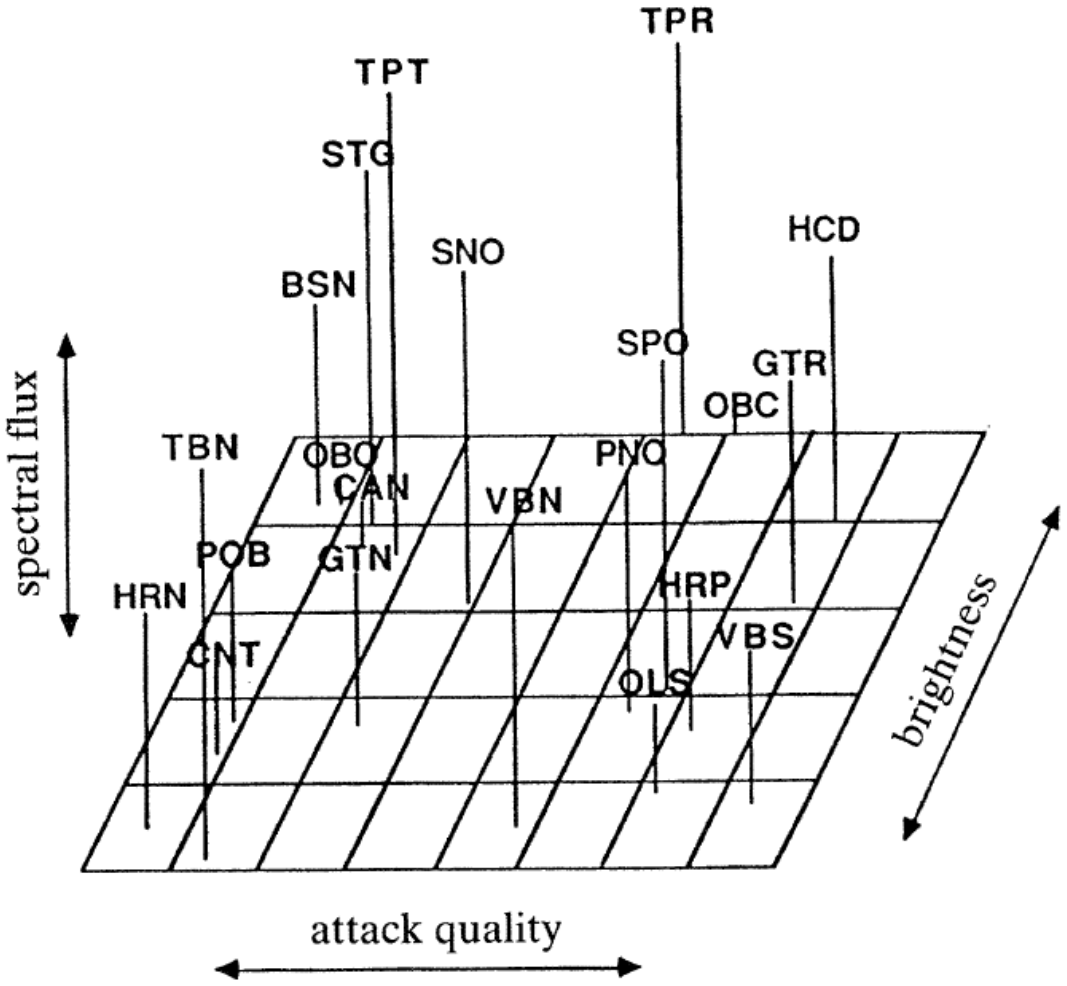


Figure 1-2 Dimensions of timbre derived from three-dimensional scaling solution for dissimilarity judgements on synthetic instrument tones represented as a point. The distances between these points on this Euclidean space correspond to the extent of perceptual similarity. Here the 3-letter acronyms correspond to the synthetic instruments. This image is reproduced from McAdams and Cunible (1992)

1.3.4 Spectral flux

Spectral flux is one of the key dimensions of timbre. It is defined as the rate of change of spectral energy over time. In other words, it is the degree of evolution of spectral shape over time. Thus, it corresponds to an interaction between the spectral and temporal envelopes. There is some disagreement on whether it is an important dimension of timbre. Caclin et al. (2005) found that spectral flux appeared as a less salient timbre parameter where its salience depended on the number of other dimensions that varied concurrently in the stimulus set.

Unlike brightness or attack time which correspond to static aspects of the spectrum, Spectral flux corresponds to dynamic changes in the spectrum of the sound. Thus spectral flux analysis would have relevance to the processing of animal vocalisation as well as human speech. For instance, speech has a high degree of spectral flux in phonemes compared to a lower spectral flux in syllables (Rosen, 1992). Next section discusses the importance of spectral flux analysis in the processing of macaque vocalisations. This motivates me to study the brain basis underlying spectral flux analysis in monkeys and compare it with humans.

However, comparison of brain basis underlying spectral flux processing requires a synthetic stimulus that affords a systematic manipulation. Spectral flux may be characterized as the spectro-temporal correlation between amplitude spectrums in adjacent time frames. Spectral flux can also be interpreted in terms of time window duration within which any two frames reach a minimum correlation. For instance, a rapid fluctuation in the spectrum corresponds to a low spectrotemporal correlation or short time window or high spectral flux while slow fluctuations in the spectrum correspond to a high spectrotemporal correlation or long time window or a low spectral flux.

1.3.5 Acoustic properties of macaque vocalisations

Being social, monkeys produce many species-specific calls or vocalisations used for communication that is context-dependent and essential for their survival. These calls may refer to objects or events in the environment (“what” information) like food, predators, social relationships, caller identity apart from the emotional state. Previously Rauschecker (1998) have classified macaque vocalisation into tonal, harmonic, and noisy categories based on the acoustic properties alone. Hauser (1998) classified rhesus macaque vocalisation into ten classes viz. coos, grunts,

harmonic arches, shrill bark, warble, scream, copulation scream, gecker, girney, and aggressive. These classes were defined based on the acoustical similarity of the vocalisations and their behavioural significance.

The spectrotemporal modulations that are present in these vocalisations can be characterised through a modulation power spectrum consisting of temporal modulations (Hz) and spectral modulations (cycles/Hz) which are aspects that influence timbre. Cohen et al. (2007) reported that most of the energy in the macaque vocalisations were at lower temporal and spectral modulation frequencies that rapidly decreased at higher frequencies as is usual of natural sounds. They also reported that most of the power for medium to higher spectral modulation frequencies were found at lowest temporal modulations as expected of animal vocalizations. This implied that the vocalisations that have a spectral structure are slower in nature. Further, they reported that the specific modulation power spectrum differed across classes of vocalisations which could be used to identify the specific kind of vocalisation. Thus it is evident that there are timbral changes associated with macaque vocalisation categories.

Animals could learn to attend mainly to spectrotemporal modulation regions of high variance across vocalisation categories as they convey information useful for categorisation of macaque vocalisations. Similarly, animals could also learn to ignore spectrotemporal modulation regions of low variance as they do not convey much information. So Cohen et al. (2007) systematically examined the spectrotemporal modulations regions that varied across the different macaque vocalisation categories. They reported that the variance was low at lower spectral and temporal modulations frequencies. However, they found that the variance was high both at medium spectral modulation frequencies between 2 to 5 cycles/kHz and high temporal modulation frequencies between 5 to 20 Hz. Further, they reported that the modulation frequencies with the highest within-category variance did not necessarily overlap with the modulation frequencies with the highest between-category variance enabling identification of details (age, gender etc.) of the caller in addition to the identification of the vocalisation type. So they concluded that intermediate spectrotemporal modulations conveyed the most information useful for classification of the macaque vocalisation into its specific category. Thus one can conclude that spectral flux analysis of the vocalisations conveys behaviourally relevant information to the monkeys.

Joly et al. (2012) analysed macaque vocalisation using a model that measured spectral and temporal modulations present in the sounds using spectrotemporal receptive fields that resembled the receptive fields of A1 neurons. These filters detected the presence of local modulations along the spectral axis (e.g. formants) or temporal axis (e.g. variations in amplitude) in the auditory spectrogram. In addition to time and frequency, the sounds were characterised using scale and rate parameters, where scale indexed the bandwidth of spectral modulations in cycles per octave while rate indexed the temporal envelope modulations in Hz. Macaque vocalisations had a bimodal distribution (see Fig 3 of this paper) in their rate-by-scale index which is sensitive to differences between acoustical nature of the sound stimuli. Some macaque calls (typically coos, most girneys, some screams) had a low rate-by-scale index (similar to human speech) with a distinct vocal quality while other macaque calls (typically shrill barks, few girneys, and some screams) had a high rate-by-scale index (unlike human speech) that had more noise like quality. Thus one can see that some macaque vocalisations that are behaviourally relevant are very dissimilar to human speech with high temporal envelope modulations requiring a much shorter time window for cortical analysis.

1.3.6 Mapping of the preferred window of temporal integration

Spectral flux can also be interpreted in terms of the duration of analysis window that is required to reach a minimum spectrotemporal correlation between any two frames within it. This warrants a literature survey on previous studies (in humans and macaques) that have examined the mechanisms for analysis of time windows (see Table 1-3). A preferred window of temporal integration of a given cortical area is defined as the minimum time period required to resolve two distinct acoustic events typically using the fluctuations in the neural discharge rates of that particular area.

Some of these studies manipulate the length of the segment within multiple segment sounds and others employ modulation of the temporal envelope of sounds while other studies used natural sounds to investigate anatomical organization for processing of different time windows.

Belin et al. (1998) have used pseudo-speech syllables to examine the processing of formant transitions of different durations, namely ‘rapid’ (40 ms) and ‘extended’ (200 ms). Using PET in humans, they show a lack of differential activations of the left auditory cortex to the different durations while the right auditory cortex demonstrates

a differential sensitivity to the slower duration. Unlike most of the other studies discussed here which employ synthetic stimuli, Belin et al. (1998) used signals derived from natural speech. This could be a reason for the difference in their results, compared to those seen in the following studies.

Giraud et al. (2000) have studied the cortical representation of temporal envelopes using continuous-acquisition fMRI in humans, and sinusoidally amplitude-modulated (4 to 256 Hz i.e. a window of 4 to 250 ms) white noise (0 to 10 kHz). Their results indicate that the auditory system is organized as a hierarchical filterbank where each processing level responds preferentially to a certain AM rate (8 Hz for PAC, 4-8 Hz for secondary auditory regions). Whilst they did not find a systematic spatial gradient for the AM frequencies, they do concur with other studies that PAC is sensitive to shorter window lengths than in secondary auditory regions.

Zatorre and Belin (2001) have used positron emission tomography (PET) in humans. They deliver synthetic stimuli, where increased temporal complexity is achieved via an increased rate of temporal switching between two tones (500Hz and 1kHz) from the slowest rate or longest window at 667 ms to the fastest rate or shortest window at 21 ms. Using analysis of covariation for the temporal parameter, they show a preference for shorter time windows in auditory core homologues as well as in anterior STG bilaterally while no areas seem to prefer longer time windows.

Jamison et al. (2006) employed the same stimulus as in Zatorre and Belin (2001) but only use extreme values of temporal rates, namely ‘temporal condition’ (21 ms) and ‘standard condition’ (667 ms). Using sparse fMRI in humans, they find similar results as the earlier study: namely that increased temporal variation is associated with HG and STG bilaterally. They also report that these responses to increased temporal variation are lateralized to the left hemisphere with the greatest effect seen in postero-medial HG.

Schönwiesner et al. (2005) manipulated temporal complexity using stimuli that differ in modulation rate, but not in bandwidth or energy. Their stimuli for temporal complexity consist of 3 spectral components that vary in temporal modulation rates from 5 Hz to 30 Hz (that is 200 ms to 33 ms windows). The increasing temporal modulation rate does not result in a monotonous two-tone staccato, since each stimulus is a broadband noise with changing spectral shape in contrast to the

melodic sequences of tones used in the above studies (Zatorre and Belin, 2001, Jamison et al., 2006). These synthetic stimuli are similar in acoustic complexity to those employed by me in this thesis. They record sparse fMRI in humans and report an increased preference for shorter time windows in the left STG (lateral from HG) which corresponds to AL auditory belt region of macaques. They find a flat relationship or a no-differential sensitivity to any specific time window in bilateral HG. They do not find a preference for longer time windows anywhere on the STP or STG.

Boemio et al. (2005) used synthetic stimuli that are generated by concatenating narrow band noise of different segmental durations, from 12 to 300 ms (spanning segmental transition to syllabic rate in human speech), where each segment is of 125 Hz bandwidth with centre frequency ranging from 1 to 1.5 kHz. They employ single-trial sparse fMRI in resting (but alert) human subjects. They report a preference for longer time windows in bilateral HG, STG, and STS. They do not find a preference for shorter time windows anywhere on the STP or STG.

Giraud et al. (2007) have performed simultaneous fMRI and EEG recording in the absence of acoustic stimuli (but in the presence of scanner noise) in humans. They report a posterior-anterior spatial distribution of spontaneous activity in the left auditory cortex with posterior regions of the auditory cortex (core and medial-belt homologues) showing evidence of ‘fast sampling mechanism’ and anterior regions of the auditory cortex (lateral-belt homologues) showing evidence of slower, integrative mechanisms.

Scott et al. (2011) recorded from single units from the auditory cortical fields in awake macaques. Synchronization of spike discharges to dynamic modulations to stimulus amplitude (rates: 0.7 to 200 Hz i.e. window of 5 ms to 1.5s) (similar to those present in macaque vocalisations and human speech), observed that the window of temporal integration in A1 was 20-30 ms while in R was 100 ms. They also reported that the onset latency to pure tones was longer in R (33 ms) than in A1 (20 ms) despite a parallel input to both regions from auditory thalamus. Further, the neurons in CM belt synchronize to amplitude modulation rates that were even higher than in A1 (88 Hz in CM vs 46 Hz in A1, though not statistically significant) while neurons in ML belt synchronize to much lower amplitude modulation rates than A1 but similar to R (12 Hz in ML and 10 Hz in R). Thus, this study suggests that postero-medial areas in

macaques have a shorter window while antero-lateral areas have a longer window of temporal integration.

Barton et al. (2012) used broadband noise (0 to 8000 Hz) that is amplitude modulated (rates: 2 to 256 Hz; i.e., window duration of 4 to 500 ms) to infer the periodotopic gradients in the auditory cortex of humans. Using a travelling wave method, modified to allow sparse-sampling on fMRI, and applying neither spatial nor temporal smoothing, nor motion correction, they observed a preference for faster modulation rates (or short windows) in medial HG and preference for slower modulation rates (or long windows) in lateral HG.

Herdener et al. (2013) used broadband noise (25 to 8000 Hz) that is amplitude modulated (rates: 2 to 32 Hz i.e. window duration of 31 to 500 ms) to examine the topography of temporal sound modulation rates in the auditory cortex of humans. Using sparse fMRI, they report a preference for faster modulation rates (or short windows) in medial HG and preference for slower modulation rates (or long windows) in lateral HG.

Santoro et al. (2014) have used natural sounds including human speech and non-speech sounds, musical instruments, environmental sounds and animal cries to infer how the natural sounds are encoded in the human auditory cortex. They suggest that the cortex derives multi-resolution representations of sounds through the combined analysis of spectral and temporal modulations in the spectrogram. The authors compute the modulation content of the auditory spectrogram through a bank of 2D modulation-selective filters, tuned to temporal modulation frequencies ranging from 1 to 27 Hz (i.e. a window of 37 ms to 1s). They propose that posterior auditory areas (core and medial belt homologues) preferentially encode coarse spectral information with high temporal precision (requiring a short analysis window), while anterior auditory areas (lateral belt homologues) favour fine spectral information that requires long analysis window, thus, low temporal precision.

Baumann et al. (2015) used broadband noise (25 Hz to 16 kHz) that was amplitude-modulated (rates: 0.5 to 512 Hz; i.e. window duration of 2 ms to 2 s) to examine ‘periodotopy’, the ordered spatial representation of temporal modulation rates in the auditory cortex. Using sparse fMRI in macaques, the authors reported results that are congruent to results in humans from Herdener et al. (2013) i.e. preference for short

time windows in postero-medial areas and long time windows in antero-lateral areas of the auditory cortex.

Erb et al. (2019) used identical methods to Santoro et al. (2014) to analyse the encoding of natural sounds in the macaque auditory cortex. Using fMRI in awake macaques, they report homologous large scale topographies for temporal modulations between macaques and humans. They report a preference for short time windows in posterior auditory areas and a preference for long time windows in anterior auditory areas.

Using a synthetic stimulus that characterized spectral flux, based on haemodynamic data in humans, Overath et al. (2008) reported bilateral sensitivity in planum temporale (PT) and anterior superior temporal gyrus (aSTG) to longer time windows while also reporting significantly right lateralized activity in superior temporal sulcus (STS). However, the brain basis underlying the processing of spectral flux has not been investigated in non-human primates as of yet.

Table 1-3 presents a summary of results from all studies in humans and macaques, with a variety of stimulus types, which infer a preference for the duration of the analysis window. In humans, five out of nine studies suggest a preference for short windows in the core homologues or postero-medial HG, while six out of nine studies suggest that the antero-lateral HG, or belt and parabelt homologues, prefer long windows. Two studies report a preference in STS to long windows while the rest do not make an observation on its preference. In macaques, there is a consensus in the literature that the preference for short windows in auditory core regions, and long windows in the belt and parabelt regions. Thus, there exist similarities in the anatomical organization of time window processing between macaques and humans.

Study	Stimuli	Window (ms)	Postero-medial AC		Antero-lateral AC	
			Left	Right	Left	Right
Studies in humans						
Belin et al. (1998)	Pseudo-speech	40 to 200	No Pref	Long	No Pref	Long
Zatorre and Belin (2001)	Narrowband 2-tones	21 to 667	Short	Short	Short	Short
Schönwiesner et al. (2005)	Narrowband 3-tones	33 to 200	No Pref	No Pref	Short	No Pref
Boemio et al. (2005)	Narrowband noise	12 to 300	Long	Long	Long	Long
Jamison et al. (2006)	Narrowband 2-tones	21 to 667	Short	Short	Short	Short
Overath et al. (2008)	Broadband tones	20 to 306	No Pref	No Pref	Long	Long
Barton et al. (2012)	AM broadband noise	4 to 500	Short	Short	Long	Long
Herdener et al. (2013)	AM broadband noise	31 to 500	Short	Short	Long	Long
Santoro et al. (2014)	Natural sounds	37 to 1000	Short	Short	Long	Long
Studies in macaques						
Scott et al. (2011)	Narrowband AM tone	5 to 1500	Short	Short	Long	Long
Baumann et al. (2015)	AM broadband noise	2 to 2000	Short	Short	Long	Long
Erb et al. (2019)	Natural sounds	37 to 1000	Short	Short	Long	Long

Table 1-3 Summary of results from studies in humans and macaques that inferred the preference for the duration of a temporal window of analysis. Short – Short window; Long – Long window; No Pref – No Preference.

1.4 Interaction between segregation and timbral analysis

Auditory system has to analyse the acoustic input to determine which set of simultaneous components/features came from the same acoustic source and should be segregated into a separate stream. The subset of acoustic features from a single source assigned to a separate stream need to be fused in perception and these features will determine the timbre of that stream. Similarly, the timbre of the simultaneous incident acoustic sources aid in the segregation of these sources into their specific auditory streams. Thus, these two perceptual processes of auditory segregation and timbral analysis are not isolated but interact with each other.

1.4.1 Influence of timbre on segregation

Studies (Bregman and Campbell, 1971, Warren et al., 1969, Broadbent and Ladefoged, 1959) have shown that listeners cannot correctly judge the temporal order of sounds having a different timbre that were played. Bregman and Campbell (1971) interpreted this result using perceptual organization and suggested that when two sound falls into different streams it is difficult to judge the temporal relationship between them. Further, Cusack and Roberts (2000) also showed that differences in timbre of sound can affect their perceptual organization. These studies showed that timbre can affect the perceptual organization of sounds.

1.4.2 Influence of spectral flux on segregation

Iverson (1995) conducted experiments to examine the influence of timbre (including spectral flux) on auditory segregation. Some of the observations from this study were that tones with a similar amount of spectral flux may stream less than tones with different amount of spectral flux, and tones with less spectral flux streamed more than tones with more spectral flux. Thus it concluded that tones segregated to the extent that they had dissimilar spectra, dissimilar attacks, and low spectral flux. So auditory stream segregation seem to be influenced by dynamic acoustic attributes including spectral flux.

1.4.3 Influence of segregation on timbre

van Noorden (1975) showed that in a rapidly repeating cycle a preceding pure tone will make audible an otherwise inaudible component of a complex tone. As a result of this stream segregation, the complex tone will now lose one of its components and hence it would have a different timbre than before.

Darwin et al. (1995) demonstrated that a sequence of precursor tones at the same frequency as ‘target’ component within a complex tone can lead to perceptual ‘capture’ of the target into the sequence. This reduces or eliminates the contribution of the target tone to the timbre and the pitch of the complex tone. This showed that auditory segregation influences timbral analysis.

1.4.4 Influence of temporal coherence on timbre

Bregman and Pinker (1978) showed that timbre (as measured using a richness rating) of a tone was judged ‘richest’ when it was exactly synchronous with another tone and this richness rating dropped off monotonously with increasing asynchrony irrespective of whether there was a lead or a lag between the two tones. This study showed the effect of temporal coherence on timbral perception.

1.4.5 Use of segregation to reveal timbral dimensions

Sounds from an acoustic source tend to retain their timbre and do not rapidly change their acoustic attributes over time. So successive acoustic events that have relatively similar timbre tend to segregate into one stream. Thus the degree of segregation can be used to judge the degree of dissimilarity between sounds of different timbres. This approach can be used instead of requiring participants to rate the dissimilarity of two sounds employed in the studies determining the dimensions of timbre.

Iverson (1995) used sound sequences alternating (ABA format) between two tones (at the same pitch and loudness) and asked the participants to judge the degree of segregation. MDS analysis on these segregation judgements, a measure of dissimilarity, can inform the underlying dimensions of timbre.

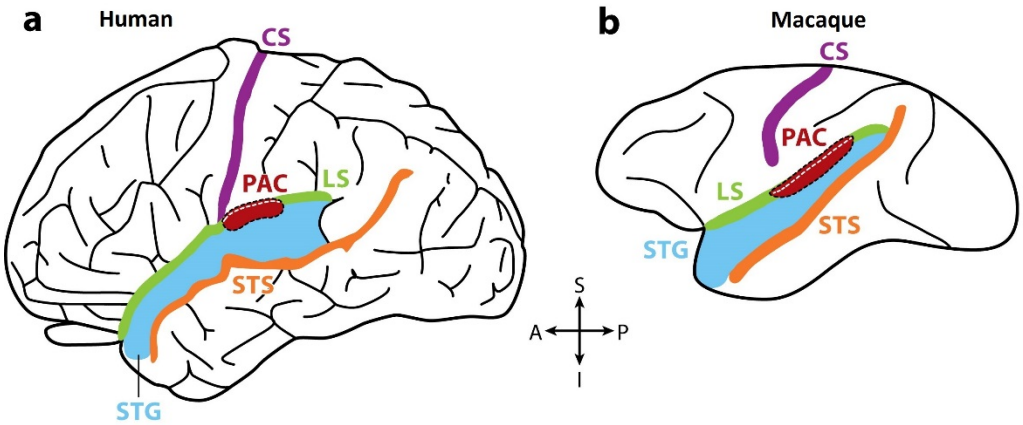
Singh and Bregman (1997) employed complex tones into galloping (ABA-) format where tones had different timbre but the same fundamental frequency (F0). The F0 difference between A and B was increased until the listener reported segregation. The F0 difference at the point of segregation was suggested as a potential tool that can indicate the magnitude of timbral differences (inversely proportional) due to different stimulus features.

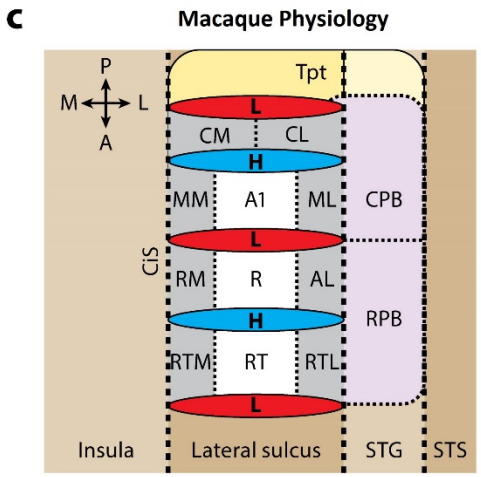
1.5 The auditory system

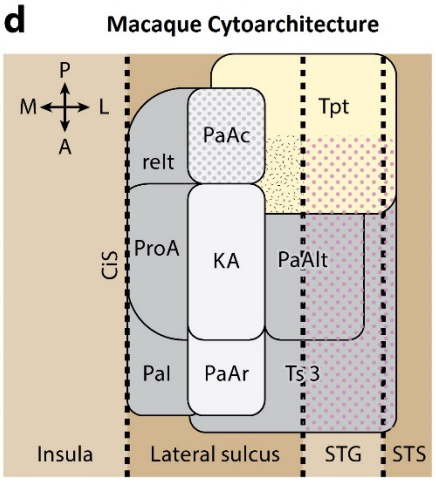
The range of auditory frequencies over which an individual can hear varies from one species to another. In rats the hearing ranges from 0.25 - 70 kHz, while it is 0.03 – 37 kHz in macaques (Jackson et al., 1999), and 0.02 – 20 kHz in humans. The primary function of the auditory system is to allow animals to perceive sounds in the environment. Sounds are characterized by frequencies that are tracked in time, and frequency representation serves as a major organizing principle of the auditory system. The processing of audible frequencies is a coordinated activity from the cochlea in the periphery to higher order areas in the association auditory cortex.

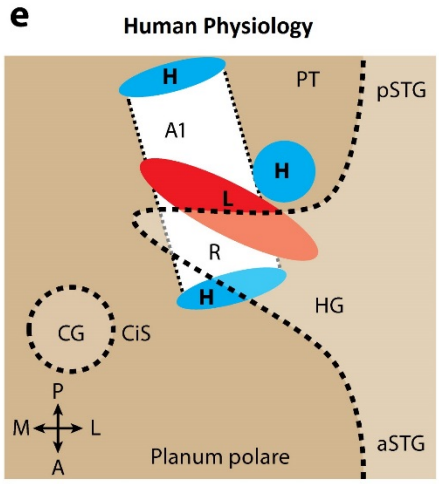
1.5.1 Information flow from the cochlea to the cortex

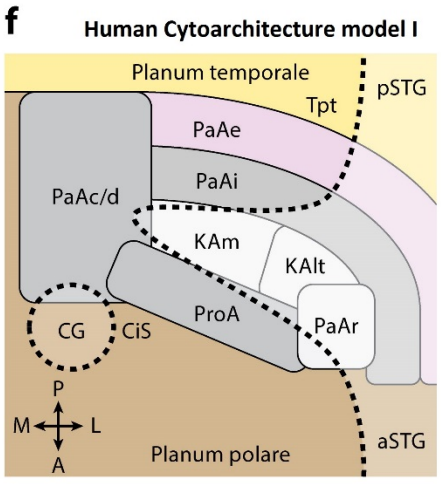
Sound waves are mechanically transmitted through the outer and middle ear to the cochlea in the inner ear where the hair cells of the organ of Corti is located. The organ of Corti spans the entire length of the basilar membrane. The mechanical properties of this basilar membrane change along its length which ensures that the tuning of the hair cells differ as a function of the distance from the oval window such that the base of the cochlea is tuned to higher frequencies while the apex is tuned to lower frequencies. Thus, the cochlea acts as a frequency analyser (von Bekesy, 1970). Further refinement of the frequency analysis mechanism is achieved by active mechanisms (Dallos, 1992). This information from the cochlea is transmitted from the inner hair cells by the auditory nerve fibres to the brainstem. In the cochlear nucleus, this frequency based information is transmitted in a number of parallel ascending pathways with different destinations (Schnupp et al., 2011). These auditory tracts converge onto the inferior colliculus (IC) in the auditory midbrain which acts as a relay station that sends information to the auditory cortex via the thalamus. The IC projects to the medial geniculate body (MGB), in the thalamus, via various nuclei in IC - ventral MGB receives inputs from central (IC_C) while dorsal MGB receives inputs from dorsal (IC_{DC}), and lateral (IC_L) nuclei of IC, and the magnocellular division of the MGB receives afferents from all nuclei of IC. The medial MGB mainly projects to the tonotopically organized core areas of the primary auditory cortex. The dorsal MGB projects to the belt areas that surround the core auditory cortex. The auditory cortex also receives sensory inputs related to non-auditory information (Hackett, 2011).

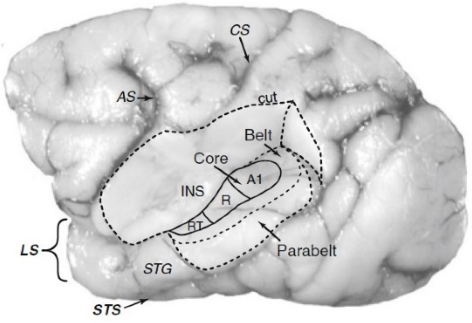

a
Human
CS
PAC
LS
STS
STG
b
Macaque
CS
PAC
LS
STS
STG
S
A
P
I


c
Macaque Physiology
P
M
L
A
Tpt
L
CM
CL
H
MM
A1
ML
CPB
CiS
L
RM
R
AL
H
RPB
RTM
RT
RTL
L
Insula
Lateral sulcus
STG
STS


d
Macaque Cytoarchitecture
P
M
L
A
Tpt
PaAc
relt
CiS
ProA
KA
PaAlt
Pal
PaAr
Ts3
Insula
Lateral sulcus
STG
STS


e
Human Physiology
H
PT
pSTG
A1
H
L
R
H
HG
CG
CiS
P
M
L
A
Planum polare
aSTG


f
Human Cytoarchitecture model I
Planum temporale
Tpt
pSTG
PaAe
PaAi
PaAc/d
KAm
KAlt
ProA
PaAr
CG
CiS
P
M
L
A
Planum polare
aSTG


g
Location of auditory areas in the cortex of the macaque monkey
CS
AS
cut
Belt
Core
A1
INS
R
RT
Parabelt
LS
STG
STS

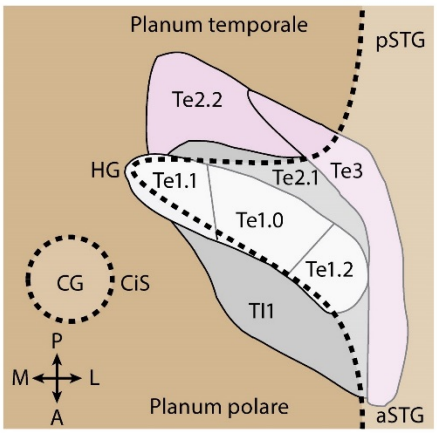

h
Human Cytoarchitecture model II
Planum temporale
pSTG
Te2.2
HG
Te1.1
Te2.1
Te3
Te1.0
Te1.2
CG
CiS
TI1
P
M
L
A
Planum polare
aSTG

Figure 1-3 Comparison of auditory cortex in humans and macaques – location, physiology and cytoarchitecture. (**a** & **b**) Location of the auditory cortex in humans and macaques. Lateral view of the left hemisphere in (**a**) humans and (**b**) macaque monkey. Primary auditory cortex (PAC) is marked in red; lateral sulcus (LS) is marked in green; superior temporal sulcus (STS) is marked in orange; superior temporal gyrus (STG) is marked in blue. (**c**) Model of macaque auditory cortex with auditory core (white), belt (grey), and parabelt (purple) regions showing high (H, blue) / low (L, red) tonotopic gradient reversals across auditory field maps from circular sulcus (CiS) to STS – A1: primary auditory cortex, AL: anterior lateral area, CL: caudal lateral, CM: caudal medial, CPB: caudal parabelt, ML: medial lateral, MM: medial middle, R: rostral, RM: rostral medial, RPB: rostral parabelt, RT: rostral temporal, RTL: rostral temporal lateral, RTM: rostral temporal medial, Tpt: Temporo-parietal area (**d**) Cytoarchitectonic model of monkey auditory cortex showing regions – KA: koniocortical area, PaAc: caudal parakoniocortical area, PaAlt: lateral parakoniocortical area, PaAr: rostral parakoniocortical area, PaI: parainsular area, ProA: prokoniocortical area, Tpt: Temporo-parietal area, Ts3: temporalis superior 3. (**e**) Model of human auditory cortex with tonotopy regions colour coded as blue for High (H) and red for Low (L). HG – Heschl's Gyrus; aSTG: anterior Superior Temporal Gyrus; pSTG: posterior Superior Temporal Gyrus; Results from Formisano et al. (2003) (**f & g**) Cytoarchitectonic models of the human auditory cortex. (**f**) Model I from Fullerton and Pandya (2007) showing the following regions viz. KAm: Koniocortical area medial; KAlt Koniocortical area lateral; PaAr: rostral parakoniocortical area; located on the HG. PaAc/d: caudal-dorsal parakoniocortical area; PaAe: lateral parakoniocortical area, external; PaAi: lateral parakoniocortical area, internal; located on the planum temporale (PT). ProA: prokoniocortical area; aSTG: anterior Superior Temporal Gyrus; CG: Circular Gyrus and CiS: Circular Sulcus located on the planum polare (PP); pSTG: posterior Superior Temporal Gyrus; Tpt: temporoparietal area within planum temporale (PT). (**g**) Location of auditory areas in the cortex of the macaque monkey. The cortex of the upper bank of the LS has been removed to reveal the auditory core, belt on the lower bank of LS. (**h**) Model II from Clarke and Morosan (2012) showing Te1.0, Te1.1, and Te1.2 (white) located on the HG, Tl1 (grey) located anteriorly to HG towards PP, Te2.1 (grey), Te2.2, Te3 (purple) located posteriorly to HG on the PT. This image is adapted from Brewer and Barton (2016). A – Anterior, L – Lateral, M – Medial, P – Posterior.

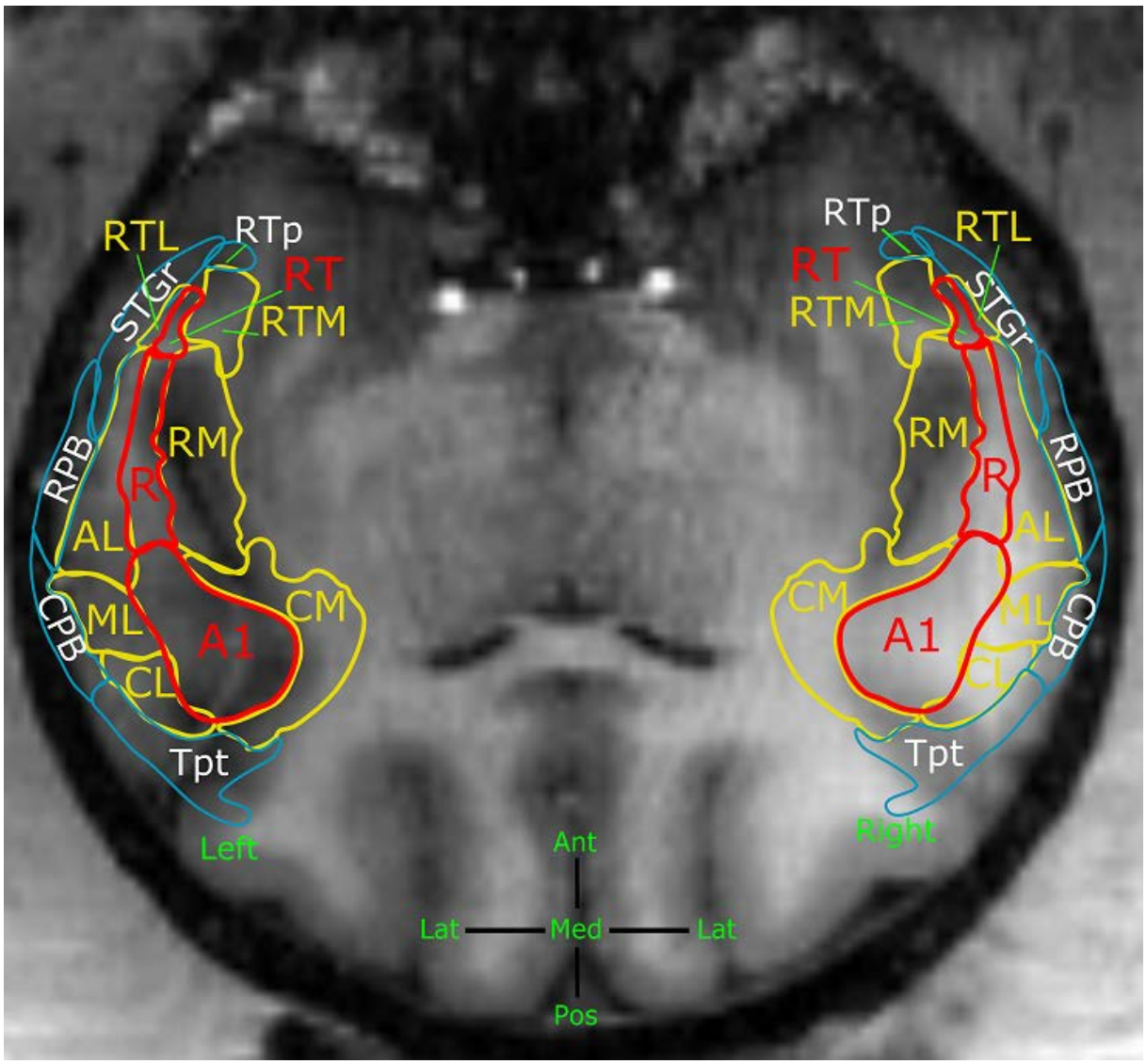


Figure 1-4 Schematic of the macaque auditory cortex showing different regions of interests overlaid on T1 structural axial image. T1 structural MRI of a macaque showing the axial section is tilted in pitch by 30 degrees such that the axial section passes through the auditory cortex on the Superior Temporal Plane (STP). Regions are outlined in different colours to indicate where it belongs in the cortical hierarchy: Red – core; Yellow – belt; Blue – parabelt. A1: primary auditory cortex, AL: anterior lateral area, CL: caudal lateral, CM: caudal medial, CPB: caudal parabelt, ML: medial-lateral, R: rostral, RM: rostral medial, RPB: rostral parabelt, RT: rostral temporal, RTL: rostral temporal lateral, RTM: rostral temporal medial, RTp: Rostral Temporal polare, STGr – Superior Temporal Gyrus rostral, Tpt: Temporo-parietal area. Ant – Anterior, Lat – Lateral, Med – Medial, Pos – Posterior.

1.5.2 Macaque as an animal model

Animal models are quite helpful as it would allow use of techniques that are not suitable for performing in humans like terminal procedures including transcardial perfusion (Hackett et al., 2001), retrograde and anterograde staining methods (Hackett et al., 2014), as well as destructive lesioning (Fritz et al., 2005) studies. Apart from these, single unit and intracellular recordings are not possible or at best opportunistic but not systematic in humans. This motivates us to search for a suitable animal model of human auditory perception.

Information about subcortical auditory processing comes from non-primate mammalian studies, mainly cats, bats and rodents since related primate studies are limited and usually complementary in nature. Thus, the principles of auditory subcortical organization are generalized across mammals. In contrast, functional specialization in the organization of the auditory cortex in primates and other mammals is better understood. This allows us to search for animal models of human auditory cortical organization. Though an apparently homologous A1 has been identified in all studied mammals, one needs to look at any further dissimilarity in the organization to arrive at the best possible animal model of human cortical processing.

In all studied primates, the tonotopic frequency representational sequence within A1 is high-to-low oriented from caudomedial-to-rostrolateral on STP. However in rats, gerbils and all carnivores, the A1 tonotopic low-to-high sequence is oriented from caudal-to-rostral (reversed to that in primates), while in ferrets this tonotopic orientation is shifted by roughly 90 degrees with respect to primates (Bizley et al., 2005). The differences in A1 orientation between primates and other species could be due to differences in sulcation, and formation of deep Sylvian fissure and insula in primates. Further, on the basis of anatomical and physiological criteria, there are 3-6 auditory cortical fields in rodents, 6-8 auditory cortical fields in cats, and more than 8 in monkeys. Thus mammals with highly developed auditory cortex seem to have a larger number of auditory cortical fields. The cortical field bordering low-frequency A1 and mirroring its topography is homologous to A1 in all primates. However, identification of such a homologous field in other mammals is uncertain and it is absent in marsupials and echo-locating bats. The cytoarchitectonic features of this rostral field in primates do not apply to low-frequency reversed field recorded in the same location in non-primate mammals, like the cat. Thus all studied primates appear to have homologous auditory fields, but these fields (especially rostral field R

and lateral field CL) do not have clear homologies in other mammals. Thus non-primate mammals are not the best possible animal models of human auditory cortical organization when a non-human primate is available for research purposes.

The complexity and refinement in the organization of the auditory cortex have increased gradually during primate evolution from New World and Old World monkeys to apes and humans. However, macaques are more suited as an animal model than any other mammal given their shared evolutionary lineage with humans with the exception of great apes in whom invasive experiments are not permitted due to ethical considerations. Thus, there have been anatomical, neurophysiological and behavioural studies to identify the structure-function relationships in many non-human primates including old world primates like macaques, and New World monkeys like marmosets (Hackett et al., 2001, Kaas and Hackett, 2000, Wang, 2000, Rauschecker, 1998). The anatomical homology of the human auditory cortex with macaques is more evident (Hackett et al., 2001) than other mammals however the exact functional homology is still under investigation (Baumann et al., 2013, Brewer and Barton, 2016) and disagreement on the extent of this homology continues (Moerel et al., 2014).

The anatomical and the neurophysiological organization of the auditory cortex in primates has been determined from studies in the monkeys amongst which macaques are generally considered as the best available model of human auditory processing (Morel et al., 1993, Heffner and Heffner, 1990). Chimpanzees are close to humans and show anatomical similarities in auditory cortex (Gannon et al., 1998), but are not available for invasive research. The anatomical organization of the macaque auditory cortex is similar to that in humans (Papez, 1929, Galaburda and Pandya, 1983, Galaburda and Sanides, 1980). In both humans and macaques, the auditory cortex is located in the superior temporal plane (STP) in the depth of a lateral fissure (see Figure 1-3 a & b). Species-specific calls play an important role in the behaviour of wild macaques. Species belonging to primate semiorder *Strepsirrhini* are usually adept at detecting higher frequencies while species belonging to primate semiorder *Haplorhini* are usually adept at detecting lower frequencies. However, macaques (0.03 - 37 kHz) and humans (0.02 – 20 kHz) are sensitive to a similar range of frequencies, despite the slight sensitivity of macaques to higher frequencies than humans (Jackson et al., 1999). Further, macaques show hearing loss which is similar

to that noted in humans after lesion of their auditory cortex (Heffner and Heffner, 1986a, Heffner and Heffner, 1986b).

Macaques are suited as an animal model of human auditory cortical organization for many reasons. The tonotopic rostro-caudal axis in humans is similar to macaques (Formisano et al., 2003). Further, the organization of the auditory cortex where a central tonotopic core region is surrounded by less clearly tonotopic belt regions is seen in both humans (Brewer and Barton, 2016) and macaques (Baumann et al., 2015). Auditory cortex can be parcellated into core, belt, and parabelt regions both in humans and macaques (Brewer and Barton, 2016, Hackett et al., 2001) based on anatomical criteria. A preference for processing of species-specific vocalizations in the left hemisphere has been suggested in both macaques (Heffner and Heffner, 1986a) and humans (Zatorre et al., 2002). A tendency for preferred responses to increasing stimulus complexity as one moves from core to belt areas also occurs across primate species (Semple and Scott, 2003).

1.5.3 Structural organization of the macaque auditory cortex

Auditory cortical areas in the macaque are located in the STP and the caudal two-thirds of the superior temporal gyrus (STG) (see Figure 1-3 b). The primary auditory cortex (PAC) lies within this region and receives projections from the MGB. Anatomical and electrophysiological studies in the macaque auditory cortex were started by Brodmann in 1909 and followed by various studies. Detailed structure-function mapping was initiated much later (Pandya and Kuypers, 1969, Merzenich and Brugge, 1973). The auditory cortex has a columnar organization (Linden and Schreiner, 2003) that is typically seen in sensory cortices. Cytoarchitectonic studies (see Figure 1-3 d) have shown that the auditory cortex in the macaque can be subdivided into (see Figure 1-3 c) a central core region surrounded by the medial and lateral belt and lateral parabelt regions oriented along the superior temporal lobe. Core, belt, and parabelt (see Figure 1-4 for a schematic overlaid on top of a T1 structural scan) have been argued to be sequential levels in the auditory processing hierarchy in the influential model of Kaas and Hackett (Kaas and Hackett, 2000).

In the auditory core, there are three auditory fields viz. A1 – primary auditory cortex, R – rostral, RT – rostral temporal. There is no clear consensus on whether RT should be considered as a field within the core, and also whether A1 and R should be distinct subfields within the core (Kaas and Hackett, 2000, Rauschecker et al., 1997).

A1 area is koniocortex since it has features of primary sensory cortex like dense myelination and granule cell proliferation in layer IV of the cortical columns (Morel et al., 1993, Jones et al., 1995, Pandya, 1995). Each core area makes reciprocal connections with ipsilateral surrounding belt areas and homotopic contralateral core areas (Kaas and Hackett, 2000). The core is characterized by sharp frequency tuning (Recanzone et al., 2000a) and tonotopic organization.

In the auditory belt regions, there are several auditory fields which are named based on their location with respect to the core areas. The auditory belt areas on the lateral side from caudal to rostral are CL - caudal lateral, ML - medial lateral, AL - anterior lateral, RTL - rostral temporal lateral, while on the medial side from caudal to rostral are CM: caudal medial, MM: medial middle, RM: rostral medial, RTM: rostral temporal medial. The medial belt shows properties of prokoniocortex like relative hypocellularity and prominence of deep cell layers while lateral belt shows properties of parakoniocortex like increased differentiation of layer III and de-emphasis of deeper layers (Morel et al., 1993). Individual fields of the belt region are defined not on the basis of cytoarchitectonics but on electrophysiological properties like tonotopy. The lateral belt areas also have reciprocal connections with the adjacent belt and parabelt regions apart from adjacent core regions. Information on medial belt regions and their connections are limited in comparison to the lateral belt because of the difficulty in studying these regions due to its anatomical location (Kaas and Hackett, 2000).

Multiple tonotopic representations are found in auditory core and belt areas (Rauschecker et al., 1995, Rauschecker et al., 1997, Kosaki et al., 1997). Low to high frequencies are represented from rostral to the caudal direction in A1. Core fields show responses to pure tones with best frequencies and narrow frequency-response curves. Lateral belt fields in general exhibit greater responses to narrow-band noise than to pure tones. The tonotopy in the belt is organized in a way that is parallel to the core regions i.e. the tonotopy borders of the core extends into belt regions. Thus, the distinctions between the different fields within the core and belt regions can be based on the tonotopic gradient reversals as the tonotopic maps reverse orientation across adjoining auditory fields.

The lateral parabelt (Kajikawa et al., 2015) has two auditory fields viz. CPB: caudal parabelt, RPB: rostral parabelt. The distinctions between parabelt fields are not well

understood and the division is based on a distinct pattern of connections with the belt (Hackett et al., 1998a). The parabelt makes reciprocal connections with contralateral homotopic parabelt areas as well as with adjoining cortical regions. The parabelt projections to core and MGB are minimal, consistent with a model based on serial projections from core to belt to parabelt (Hackett et al., 1998b). The parabelt also makes connections with adjoining cortical regions, CPB with Tpt and caudal STG while RPB with rostral STG, STS (both upper and lower banks) (Kaas and Hackett, 2000, Poremba et al., 2003). Rostral parabelt responds to white noise but not pure tones (Kosaki et al., 1997) while caudal parabelt responds to noise and pure tones over a wide range of frequencies (Kaas et al., 1999).

As one moves from core to belt and parabelt, selective responses to complex stimuli (animal call sounds) are observed. At the level of the belt, responses are rarely specific to a single species-specific vocalisation. However, at the level of the parabelt, responses are sometimes selective for particular complex sounds. This suggests that the belt is at an interim stage (Tian et al., 2001) and parabelt is at an advanced stage of information processing.

1.5.3.1 Two-stream hypothesis

Like vision, the cortical organization of auditory analysis has been suggested to be based on distinct parallel processing pathways (Rauschecker, 1998, Romanski et al., 2000, Kaas and Hackett, 1999, Cohen and Wessinger, 1999). In this scheme, sound identity is processed in a ventral ‘what’ pathway since rostral and lateral belt and parabelt show selectivity for patterns that characterize sounds (Belin et al., 2000) while sound location is processed in a dorsal ‘where’ pathway since caudal belt and parabelt show selectivity for spatial information (Tian et al., 2001, Recanzone et al., 2000b). The ventral pathway starts in the rostral belt and projects to rostral STG and frontal areas via rostral parabelt.

1.5.4 Human auditory cortex

The organizational scheme of the human auditory cortex is similar to that of macaques (see Figure 1-3). Galaburda and Sanides (1980), Hackett et al. (2001), Fullerton and Pandya (2007) and Clarke and Morosan (2012) have proposed cytoarchitectonic homologies of macaque and human auditory cortex. It consists of a central primary region core surrounded by the non-primary belt and parabelt regions. In humans, the primary auditory cortex is located in the medial portion of the

transverse gyrus of the Heschl (HG) in the superior temporal plane. Based on cytoarchitectonic criteria, HG comprising of koniocortex is subdivided into (see Figure 1-3 h) Te1.1, Te1.0, Te1.2 regions (Morosan et al., 2001) which is Koniocortex medial (KAm), Koniocortex lateral (KAlt), rostral Parakoniocortex (PaAr) as per Fullerton and Pandya (2007) (see Figure 1-3 f) that correspond to macaque core A1 and R regions. Surrounding HG is Te2.1, which is lateral parakoniocortex internal (PaAi) corresponding to macaque auditory belt cortex and Tl1 which is Prokoniocortex (ProA). Caudally these regions are circumscribed by Te2.2 and Te3 regions which is lateral parakoniocortex external (PaAe).

Neurophysiology in the human auditory cortex is at best only opportunistic based on pre-surgical studies of patients with epilepsy, unlike the systematic electrophysiological investigation that is possible in animals. However, non-invasive functional imaging like functional magnetic resonance imaging (fMRI) can be applied to both humans and animals to conduct a systematic investigation of structure-function relationships in vivo. Electrophysiology in medial HG in humans shows a tonotopic organization with higher frequencies located posterior and medially while lower frequencies located anteriorly and laterally. This arrangement is similar to that seen in core regions in the macaque. The phasic and tonic temporal response patterns exhibited by single unit recordings from the HG (Howard et al., 1996) in humans are similar to the that seen in monkeys (Recanzone, 2000).

Responses to speech stimuli occur in STG with a long latency (Steinschneider et al., 1999) suggesting that processing of complex stimulus properties occurs farther away from the primary regions. Upstream areas in auditory cortex encode basic stimulus properties while downstream areas farther from core regions encode complex stimulus properties (Mesulam, 1998). Further, the increase in the size of the window of temporal integration observed in humans (Husain et al., 2004) as one moves along the auditory ventral stream is consistent with this idea. Also, there is evidence (Zatorre et al., 2004) to show that the farther one goes along the ventral stream the more sensitive neurons are to invariant auditory features that characterize the individual auditory objects. This hierarchical organization of the auditory cortex seen in humans is consistent in macaques (Leaver and Rauschecker, 2010) where the most anterior regions on the ventral pathway represent the complex acoustic signature of auditory objects.

1.6 Key problems addressed

1.6.1 Auditory Figure-Ground segregation

The brain bases and mechanisms (Teki et al., 2011), behavioural performance (Teki et al., 2013) and the temporal dynamics (Teki et al., 2016, O'Sullivan et al., 2015) underlying auditory segregation based on temporal coherence have been studied in humans. However, to perform a systematic neurophysiological investigation an animal model is needed. Thus, this study aims to uncover the brain bases underlying the process of pre-attentive stimulus-driven auditory figure-ground segregation in rhesus macaques using functional MRI. Next, this study also aims to understand whether macaques can perceive figure in the stochastic figure-ground stimulus and describe how their figure detection performance change as a function of figure coherence. I hypothesize that if macaques are a good model of human auditory segregation then I would find BOLD activation for the processing of the figure in SFG stimulus in similar regions as identified in human fMRI studies i.e. parabelt homologues on the convexity of the STG, and macaque homologue of human intraparietal sulcus.

1.6.2 Spectral Flux processing

The brain bases underlying the processing of spectral flux has been investigated in humans using fMRI (Overath et al., 2008). So this study aims to understand spectral flux analysis at the systems level in a macaque model. Based on BOLD activity from sparse fMRI, the auditory cortical areas responsible for spectral flux analysis will be identified using a range of spectral flux values that span natural sounds. I hypothesize that if macaques are a good model of human spectral flux processing then I would find no sensitivity to spectral flux in auditory core regions, increasing BOLD response to decreasing spectral flux in CM, CL, anterior STG and right lateralized STS.

Chapter 2 Techniques and Methods

2.1 Summary

This chapter outlines the experimental methods used to analyse the neuroimaging data presented in this thesis. In the first section, the operant conditioning training methods using positive reinforcement to train head restraint and visual fixation are covered. The second section presents the basic principles behind functional magnetic resonance imaging (fMRI) – from the physics to statistical analyses.

2.2 Non-Human Primates

The imaging and behaviour data for the experiments reported here were obtained from rhesus macaques (*Macaca mulatta*). The details of the animals including their gender, age, and weight, are presented in Table 2-1 along with the type of experiments they were employed on. In 3Rs, 'Reduction' refers to methods that minimize the number of animals used in an experiment by promoting the use of well-designed and analysed experiments that are robust and reproducible. In the experiments reported here, utmost four animals were used in an experiment. Monkey M7 was trained for participating in fMRI and visual fixation task, however, it was not possible to obtain data from this animal in either experiment.

Animal ID	Gender	Age [years]	Weight [kg]	Experiment
M1	Male	11	9	Both
M2	Male	11	11	SFG
M3	Female	5	6	SFG
M4	Male	12	17	Both
M5	Male	9	16	Flux
M6	Male	9	10	Flux
M7	Male	5	8	-

Table 2-1 Summary of subjects participating in all experiments

2.2.1 Positive reinforcement training

Positive reinforcement training (PRT) methods are valuable refinement to laboratory animal management. In 3Rs, 'Refinement' refers to methods that minimise the pain, suffering and distress that may be caused to animals used in research by promoting high standards of animal welfare that results in good quality scientific data.

UK Home Office recommends the use of PRT with non-human primates (NHPs) (Prescott and Buchanan-Smith, 2007). PRT is used for neuroscience experiments to train animals in complex cognitive and motor behaviours for food or fluid rewards. In imaging and behaviour studies reported here, PRT was used to successfully train NHPs to move within/between enclosures (Veeder et al., 2009, Bloomsmith et al., 1998), sit on a primate chair, restrained using a neck plate and restrained using the surgically implanted head-post. This benefits the welfare of the animal as a result of their ability to control the environment and exercise free will (Prescott and Buchanan-Smith, 2007). PRT regimes are very labour intensive where operant behaviours need

to be consistently associated with rewards on a daily basis (Fernström et al., 2009) for long periods to achieve successful training.

2.2.2 Operant conditioning

First classical conditioning was employed to create a positive association for the animal with a clicker by pairing it with food reward during the conditioning phase. After successful conditioning clicker can now substitute for reward but this behaviour needs to be reinforced on a daily basis. Next, a similar positive association was created for the animal with primate chair, neck plate, head-post, MRI scanner by pairing any voluntary interaction with these objects with a food reward or the clicker. Here this constituted operant conditioning using PRT since I rewarded with food and liquids of the animal's liking for their voluntary behaviour. Here the animals were motivated by restricting fluid intake in the home enclosure and providing unlimited fluid and high-value food in the primate chair. However, during fMRI data acquisition only fluid rewards were employed due to the limitations of space within the scanner.

2.2.3 Head Implant

During the fMRI recording, the head of the animal was positioned with a custom-made polyetheretherketone (PEEK) head holder attached to a cranial implant. The details of the surgical procedures and the materials used for the cranial implant are discussed in (Thiele et al., 2006) while positioning procedures are discussed in (Baumann et al., 2010). All experiments were carried out in accordance with the UK Animals (scientific procedures) Act (1986), European Communities Council Directive RL 2010/63/EC and the US National Institute of Health Guidelines for the care and use of animals for experimental procedures and were performed with proper care to ensure the well-being of the monkeys. This project supports the principles of the consortium on Animal Research Reporting of In Vivo Experiments (ARRIVE).

2.2.4 Eye fixation and Head restraint

The monkeys were trained to fixate at a visual cue displayed on the screen during the fMRI imaging. This simple task has advantages both to the quality of data acquired and to the wellbeing of the animal. Firstly, this task ensured that the attention of the animal is taken away from the stimulus. Further, this task fixed the attentional set of the animal. Next, it ensured that the levels of attention on the distractor cue remained consistent across the entire session. Finally, it minimized the body movement of the animal by reducing any stress it may have. A brief description

of how eye fixation and head restraint was achieved in the animal is presented next. All the steps were achieved via operant conditioning using PRT methods. I do not have training data for macaques M1 - M6 except monkey M7. However, monkey M7 was lost before I could get experimental data from this animal.

First, the monkey was trained to sit comfortably in the purpose-built Plexiglas primate chair using PRT based operant conditioning methods. Next, the animal's trust was gained to allow it to be neck plated in the primate chair by rewarding voluntary acceptance of neck plate with a high-value food reward. This chair also contained a touch bar as well as a reward tube through which the monkey could be given a fluid of its liking upon expected behaviour. Using PRT based operant conditioning method, the animal was taught a bar release reward relationship. First, even a slight contact with the metal bar was rewarded with high-value fluid. Next, the animal was rewarded only for contact of slightly longer duration. This allows the animal to correlate the bar contact reward relationship. Further, only complete bar release was paired with reward ensuring that desired bar release reward relationship is fully established.

A cathode ray tube (CRT) monitor was placed in front of the chair at a distance of 55 cm. A visual cue in the form of a square was presented at the centre of the screen. A bar release upon contrast dimming of the visual cue resulted in a fluid reward. At the start, this change in contrast and the size (visual angle subtended at the eye of the animal) of the visual cue was very huge making it salient to the animal (see Figure 2-1). But gradually this contrast was reduced to match the monkey's just noticeable difference (JND) while the size was also reduced from 10 degrees to 0.2 degrees. This required the animal to fixate on the stimulus to detect any change in contrast (see Figure 2-1).

Next, the animal was head restrained using the previously implanted head post to different degrees gradually until full head restraint is achieved. Finally, the eye of the head restrained animal was monitored using ViewPoint EyeTracker systems (Arrington Research, USA), which recorded the gaze location of the right eye using a 220Hz monocular camera. Now the eye position was monitored and rewarded when the animal fixated on the visual cue without the need for a bar release. Thus, prior to scanning, the animals had been previously habituated to the scanner environment as well as exposed to some experimental auditory stimuli and trained for visual fixation task. The animals were scanned in awake behaving state.

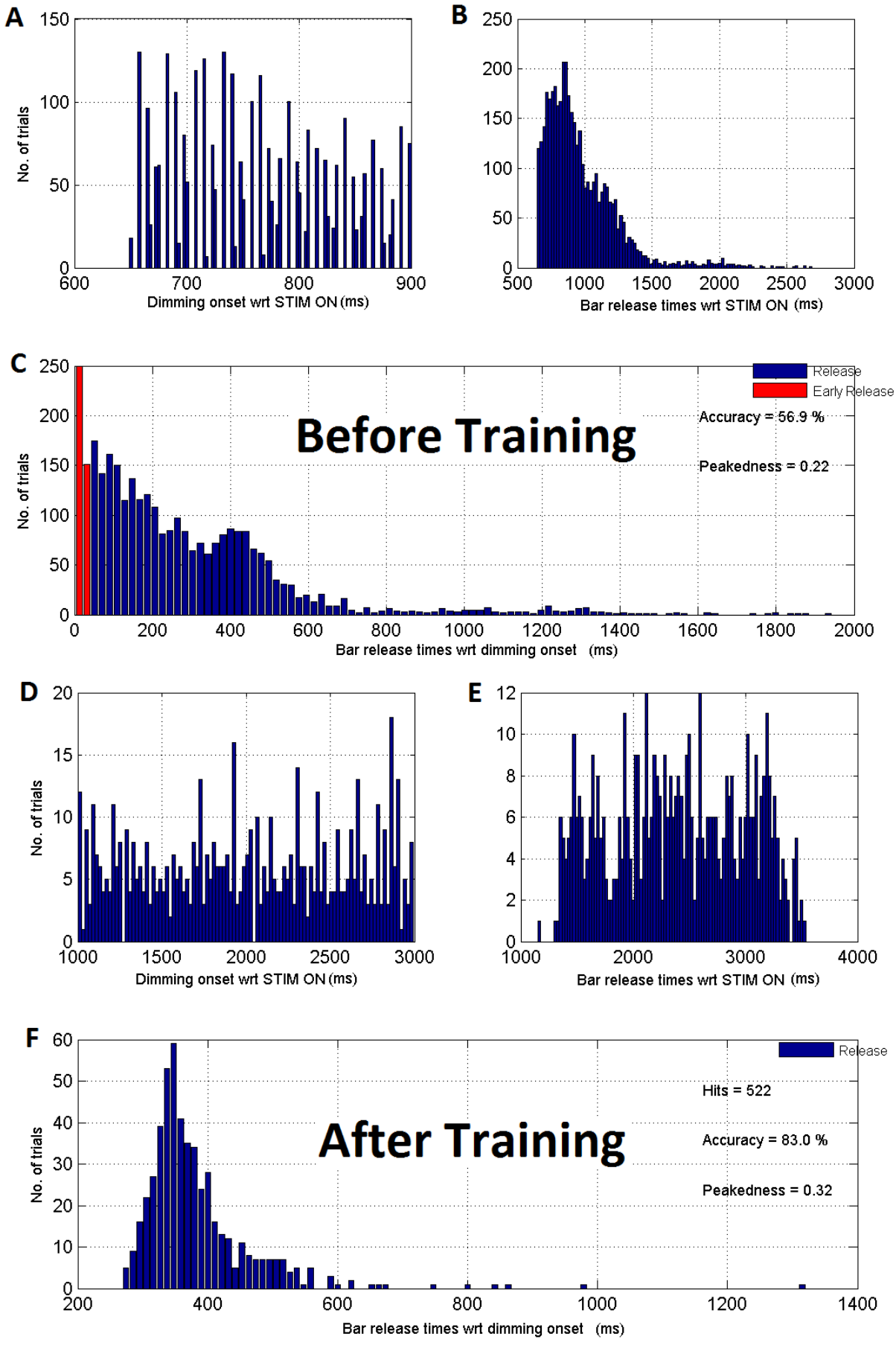

A
No. of trials
Dimming onset wrt STIM ON (ms)
B
Bar release times wrt STIM ON (ms)
C
Release
Early Release
Before Training
Accuracy = 56.9 %
Peakedness = 0.22
No. of trials
Bar release times wrt dimming onset (ms)
D
No. of trials
Dimming onset wrt STIM ON (ms)
E
Bar release times wrt STIM ON (ms)
F
Release
Hits = 522
After Training
Accuracy = 83.0 %
Peakedness = 0.32
No. of trials
Bar release times wrt dimming onset (ms)

Figure 2-1 Performance of monkey M7 at the start and end of the training on a visual cue dimming task in a bar release paradigm which is designed to train the animal on a visual fixation task. (**A**-**C**) Performance on the first day of the training. (**A**).Dimming of a very salient visual cue (size 25 x 25 degrees) occurs randomly in a narrow range of 650-900 ms after visual stimulus onset to keep the difficulty of the task low (**B**) Metal bar release times of the monkey with respect to stimulus onset (with a maximum timeout of 2 seconds) conveying that the animal has is releasing the metal bar at delay from stimulus onset. Thus the animal does not understand the task. (**C**) Metal bar release times with respect to the dimming of a salient visual cue. Here bars with red face colour signify early release and were penalized with no reward while bars with blue face colour signify release times on trials with a correct response. One can notice the considerable variance in the distribution of metal bar release times apart from the early release signifying that the animal does not understand the task. (**D**-**F**) Performance on the final day of the training (**D**) Dimming of a very subtle visual cue (size 0.2 x 0.2 degrees) occurs randomly in a wide range of 1000-3000 ms after visual stimulus onset to make the task difficult (**E**) Metal bar release times of the monkey with respect to stimulus onset conveying that the animal does not subvert the task by releasing the metal bar at constant delay from stimulus onset. (**F**) Metal bar release times with respect to the dimming of a subtle visual cue. Bars with blue face colour signify release times on correct trials. One can note that the peak of the distribution of bar release time occurs around 300 ms (response latency) after dimming of the visual cue signifying that the animal has learnt the visual fixation task. Since the visual cue is very subtle it requires that the animal fixates on the visual cue to spot the dimming and earn a reward. Thus this regime has been successful in training on visual fixation task.

2.3 Magnetic Resonance Imaging

Subatomic particles automatically pair up to spin at the same rate but in opposite directions. This principle applies not just to electrons but to nucleons (i.e. the sum of protons and neutrons) as well. Thus, atomic nuclei containing an odd number of nucleons have a net nuclear spin unlike the nuclei with even numbers of nucleons. Due to the laws of electromagnetic induction, such nuclei acquire a magnetic moment since they have both a (positive) charge and a net spin. These nuclei behave as tiny magnets and can align with a magnetic field. For example, hydrogen (H^1) and carbon (C^{13}) atoms exhibit this phenomenon.

Thus, these nuclei align with an externally applied magnetic field by precessing at a frequency proportional to the applied field strength. This alignment could either be parallel (low energy state or spin-up nuclei) seen in some nuclei or anti-parallel (high energy state or spin-down nuclei) to the external magnetic field as seen in fewer nuclei. The transitions between the spin states are accompanied by an absorption (from low to high energy state) or emission (from high to low energy state) of the difference in energy between the two spin states in the form of an electromagnetic wave in the radio frequency range. Thus, the phenomenon of nuclear magnetic resonance (NMR) was discovered in these nuclei. This is the fundamental principle underlying magnetic resonance imaging (MRI) that is used to measure properties from individual atoms. In 1940s Felix Bloch and Edward Purcell independently discovered the NMR phenomenon.

The fact that a human body contains ‘MR active’ hydrogen atoms in huge proportions (around ~60%) in the form of water in tissues allows us to use NMR method for studies of the human tissue. Because hydrogen has a solitary proton, it gives a relatively large magnetic moment as well. Both these characteristics enable the utilization of the maximum amount of available magnetization in the body.

A powerful magnetic field and radio frequency (RF) coils that can transmit and receive high energy pulses are essential to perform MR imaging. Magnetic fields are measured in Tesla (T), 1T = 10,000 Gauss. The imaging experiments reported in this thesis were conducted in a 4.7T Bruker scanner. In comparison, the earth’s magnetic field is 0.5 Gauss: four orders of magnitude less.

MR active nuclei have a net spin and absorb a photon of specific frequency given by the following equation

$$\nu = \gamma \cdot \mathrm{B}$$

where ν is the Larmor frequency in MHz, B is the strength of the external magnetic field in Tesla (T) and γ is the gyromagnetic ratio in MHz per Tesla. The value of γ for hydrogen nuclei is 42.58 MHz/T.

The absorption or emission of energy by the aligned nuclei during the transitions between the states is governed by the following equation

$$E = h \cdot \nu$$

where E is the energy difference between the states, and h is the Planck's constant ($h = 6.625 \times 10^{-34} Js^{-1}$)

There are two components to the nuclear spin magnetization viz. longitudinal and transverse components. The longitudinal component (denoted M_z) is the component of the magnetization along the z-axis i.e. in alignment to the external magnetic field. The transverse component (denoted M_{xy}) is the component of the magnetization perpendicular to the z-axis lying in the xy-plane due to the precession of the nuclei along the z-axis.

At equilibrium, the net magnetization (denoted M_0) is the same as the longitudinal component since the transverse component is zero. This equilibrium state can be perturbed by an application of an RF pulse whose energy matches the difference between the energies of the two spin states. The longitudinal component is nil when the system is saturated, but it returns to equilibrium due to the reverse transition of nuclei to the lower energy spin state. The time constant associated with this longitudinal relaxation called spin-lattice relaxation time (denoted T1), is characterized in the equation below

$$M_z(t) = M_{z,eq} \cdot \left(1 - e^{-t/T1}\right)$$

The precession of the nuclei aligns in phase upon the application of an RF pulse causing a net transverse magnetization. This component decays since the individual nuclei dephase due to short-lived molecular interactions such as collisions. This

decay occurs with a time constant, denoted T2, called the spin-spin relaxation time characterized in the equation below

$$M_{xy}(t) = M_{xy}(0) \cdot e^{-t/T2}$$

However, the effective time constant, denoted $T2^*$, associated with the decay of the transverse magnetization, is governed by both the molecular interactions ($T2$) as well as the field inhomogeneity ($T2_{inhomo}$) resulting in different rates of precession of the nuclei. The relationship between these constants is given in the equation below.

$$1/T2^* = 1/T2 + 1/T2_{inhomo}$$

MR imaging makes use of several tissue properties. The NMR signal varies as a function of proton density which is different for each tissue type. Also, the magnetization characteristics which are a function of both $T1$ and $T2^*$ and determines the rate of signal decay, differ between tissue types.

Spin-echo and Gradient-echo are MR imaging techniques that consist of applying RF pulses to a tissue at equilibrium which results in tissue-specific $T1$ and $T2$ effects. This RF pulse tilts the longitudinal component by 90º into the transverse plane. Now the transverse component begins to dephase. In the spin-echo technique, later a second RF pulse is applied which rotates the transverse component by 180º. This pulse causes the component to rephase partially and to cause a signal called the echo. This results in better detection of small inhomogeneities within tissues. In contrast to the gradient-echo technique, no second RF pulse is applied, making it more susceptible to T2* effects and hence frequently used in fMRI. The time at which the decay signal is read out with an RF receiver coil is known as ‘time-to-echo’ (TE).

MR images are obtained by changing the field strength along each dimension in a linear gradient manner. This makes the resonant frequency a function of spatial position. The MR signal obtained at the RF receiver at time TE is a complex of different frequencies and requires the use of Fourier decomposition to be analysed. The spatial frequency encoding is determined by the amplitude and duration of the gradients. A planar image is constructed on a grid in the Fourier spatial frequency domain or ‘k-space’, via two orthogonal gradients with the use of appropriate RF pulses. Along the x-axis, a read-out gradient is employed that encodes the spatial frequency, while along the y-axis, a ‘phase-encode’ gradient advances the phase.

Along the z-axis, an orthogonal gradient enables slice-selection by ensuring that only protons in a single slice become resonant and emit RF signal. Here the slice thickness is determined by the bandwidth of the RF pulse. In this spatial frequency domain, low spatial frequencies are represented in the centre while high spatial frequencies are represented in the edges. The k-space trajectory is the path traversed through k-space to acquire data. An inverse Fourier transform is applied to the recorded signal in each plane to recover the spatial characteristics of the imaged tissue. Here, Time-to-Repeat (TR) is the time between successive phase-encoding pulses.

The spatial resolution of the imaging is determined by the strength of the primary magnetic field used in the MRI scanner, where higher resolution is obtained at higher field strengths. A good spatial resolution enables a detailed visualization of the tissue and accurate localization of the activity in the brain. The resolution is characterized by the dimensions of volume element (voxel) determined as the ratio of the field of view (FOV) to the total number of voxels within the FOV.

In the experiments conducted as a part of this thesis, the parameters for the structural scans are as follows. Anatomical MR images were T1-weighted (T1w) images, consisting of a 2D magnetization-prepared rapid gradient-echo (MPRAGE) sequence with a 180° preparation pulse, TR = 2000 ms, TE = 3.74 ms, TI = 750 ms, 30° flip angle, receiver bandwidth = 50 KHz, an in-plane resolution of 0.67 x 0.67 mm^2 with a slice thickness of 0.6 mm.

2.3.1 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging has ushered in a new era of non-invasive imaging in neuroscience as it enables localization of basis for perceptual and cognitive processes with reasonable spatial precision as compared to electroencephalography (EEG), although using an indirect measure of neuronal activity. However, fMRI incurs the wrath of multiple comparisons due to the very high number of voxels sampled in a whole brain imaging at a high spatial resolution. So investigation has to be motivated by a specific hypothesis. Further, fMRI in NHPs has enabled comparison of the neurobiology of cognitive functions between humans and monkeys. Unlike positron emission tomography (PET), fMRI modality does not involve the ingestion of radioactive tracers. Further, most electrophysiological studies tend to record from layer IV in a cortical column which biases towards thalamocortical

inputs instead of assessing response properties that are different at different depths i.e. as a function of laminar depth. Though fMRI at low field strengths do not have the resolution to distinguish activity from different cortical layers, it is not biased to any one particular cortical layer. However, very high field strength fMRI can distinguish activity from different cortical layers. Next, the principles of fMRI, its neurophysiological bases, protocols for fMRI data acquisition for auditory experiments, data pre-processing and statistical analysis steps involved in the processing of fMRI data to obtain correlates of brain activity, are presented.

2.3.1.1 Echo-planar imaging

Functional MRI requires the quick acquisition of images unlike the leisurely pace of the structural MRI imaging sequences. Echo-planar imaging (EPI) enables very quick acquisition of images from the x-y plane via the application of a single excitation pulse per volume. To cover the entire plane, gradients pertaining to frequency and phase are rapidly switched. There are two subtypes: gradient-echo and spin-echo. The signal from gradient-echo is more sensitive to local field inhomogeneities (T2*) primarily caused by deoxyhaemoglobin. Thus, gradient-echo EPI is better suited for fMRI. Gradient echoes can be generated by an oscillating gradient along the read-out direction that follows a zigzag trajectory in k-space.

2.3.1.2 Physiological basis of the BOLD signal

To execute a task, the brain mobilizes localized specific sites that are part of a functional network responsible for the task. Synaptic activity in a given site is coupled with vascular response via the mediation of astrocytes. It is this vascular activity that can be detected by fMRI. The vascular response duration is a lot slower than neuronal events. So, to infer the task-related neuronal activity the time course of the recorded MRI signal needs to be compared with the expected signal.

The image contrast is termed as the BOLD signal which stands for Blood Oxygenation Level Dependent signal as it is dependent on the level of deoxyhaemoglobin in the blood vessels at a given location (Ogawa et al., 1990a, Ogawa, 2012). Seiji Ogawa demonstrated that blood oxygenation level altered T2 and T2* weighted signals (Ogawa et al., 1990b, Ogawa et al., 1990a). The foundation for this work was done by Linus Pauling, who showed that oxyhaemoglobin is diamagnetic while deoxyhaemoglobin is paramagnetic (Pauling and Coryell, 1936). Due to this difference, the NMR signal of deoxyhaemoglobin

decays faster than oxyhaemoglobin which results in magnetic susceptibility differences between haemoglobin-containing vasculature and surrounding tissue. This leads to greater dephasing of the protons and the reduction in corresponding T2* signal. Thus, neural activity is coupled to changes in BOLD signal (Ogawa et al., 1992) and the spatial distribution of these intensity changes provides the location of the activity.

2.3.1.3 Relationship of the BOLD signal to neural activity

The BOLD signal was characterized with respect to the neuronal activity in experiments that combined acquisition of BOLD signal with intracortical microelectrode recordings from the visual cortex of anaesthetized as well as awake monkeys. Logothetis et al. (2001), Logothetis (2002), Logothetis (2003) established that BOLD correlates more strongly with low-frequency components of extracellular local field potential (LFP) rather than spiking activity of local neuronal ensemble (Logothetis, 2012). The correlation between LFP and multiunit activity (MUA) is present but not strong due to inhibitory postsynaptic potentials which positively contribute to LFP but negatively to MUA. Extracellular field potentials primarily reflect inputs to, and local (~100 μm) processing of neuronal information (Logothetis and Wandell, 2004), within a region in which a signal change is detected, which includes several effects such as neuromodulation, interactions between interneurons and pyramidal cells. Further, the haemodynamic responses depend on the size of the activated populations (Logothetis, 2012).

2.3.1.4 Neural codes sensitive to BOLD signal

Given the underlying basis of BOLD signal change is an increase in the metabolic demand, BOLD is sensitive to only those neural coding schemes that increase the neuronal metabolism above the baseline. Thus, BOLD is sensitive to a neuronal coding scheme based on the average number of spikes per unit time (‘rate code’) i.e. BOLD signal increases due to an increase in the firing rate. This correlation between firing rate and the BOLD signal was confirmed using MRI/MRS studies in rats (Smith et al., 2002, Hyder et al., 2002, Logothetis and Wandell, 2004). But BOLD is neither sensitive to a neural coding scheme based on the precise timing of single spikes (‘temporal code’) nor sensitive to a coding scheme based on activation of a relatively small set of neurons that change for different stimuli (‘sparse code’). Further, BOLD is not sensitive to ‘population code’ where individual neurons have overlapping selectivities, so that many neurons respond to a given stimulus i.e. each neuron has

a distribution of responses over some set of inputs that overlap with other neurons, and responses of many neurons need to be combined to infer the value about the input stimulus.

2.3.1.5 Limitations of fMRI

First, BOLD is an indirect measure of neural activity and it lags behind by several seconds. Also, the location of BOLD signal might not coincide with the location of neurally active tissue since a large component of the BOLD signal is in venules. Next, MR acquisition sequences sensitive to BOLD are also sensitive to any magnetic field inhomogeneity. So images can be distorted, have reduced or absent signal due to susceptibility related dropout in brain regions near bone or sinuses. Further, there are some limitations of BOLD-based fMRI that arise due to the circuitry and functional organization of the brain as well as the use of inappropriate experimental protocols and these need to be considered (Logothetis, 2008). Certain pulse sequences, like Gradient Echo based EPI, are more sensitive to signals from within and around large blood vessels. This can bias the responses towards those regions that are close to macrovessels. So this bias needs to be factored in while interpreting results when using these sequences. Next, fMRI signals are very sensitive to neuromodulation effects (like arousal, attention, memory) that are slow and lead to reduced spatiotemporal resolution and specificity. So the use of sequences that has a sampling time of utmost few seconds could be optimal to address this limitation. Further, fMRI signals cannot easily differentiate between bottom-up and top-down signals. Also BOLD signal may potentially confuse between excitation and inhibition. Finally, fMRI signals are not sensitive to the size of the activated population of neurons. So the magnitude of fMRI signal cannot be used to quantify differences between brain regions or between tasks within the same region.

2.3.1.6 Haemodynamic response function

The BOLD response is characterized by the haemodynamic response function (HRF) with distinct characteristic phases (Logothetis, 2002) that captures the complex interactions between cerebral blood flow, blood volume, and blood oxygenation. The BOLD response shows an initial dip that might be due to an increase in oxygen consumption which alters the ratio of deoxyhaemoglobin to oxyhaemoglobin (Malonek and Grinvald, 1996). Following this initial dip the blood flow increases in the active regions which might represent an increase in the blood oxygenation or equivalently a decrease in the oxygen extraction fraction as demonstrated using

Positron Emission Tomography (Fox and Raichle, 1986). This increase results in a peak that is around 4-6 seconds from stimulus onset but takes up to 18 seconds after stimulus offset to return to baseline in the primary auditory cortex (Belin et al., 1999, Hall et al., 1999). Increased blood flow results in vasodilation i.e. expansion of blood vessels due to local venous blood volume, which results in a post-stimulus undershoot in the HRF (Buxton et al., 1998). The amount of change in BOLD response is typically of the order of 1.5% in the human auditory cortex (Talavage and Edmister, 2004).

2.3.2 fMRI for auditory stimulation

The primary problem in using fMRI for auditory research is the loud acoustic noise produced by the scanner. This noise is due to the switching of the gradient coils. The main source of this noise is due to the read-out gradients (Ravicz et al., 2000). Further, the coolant pump for the scanner's permanent magnet is another source of low-frequency low-intensity noise.

The acoustic properties of the scanner noise, like bandwidth, fundamental frequency, etc., depending on the mechanical resonances of the coil assemblies, on the type of imaging sequence used, and its switching frequency (Hall et al., 2000). Typically, the spectrum of the scanner noise is broadband (250 to 8000 Hz) with a peak around 1 – 2 kHz (Hall et al., 1999, Ravicz et al., 2000, Chambers et al., 2001) which overlaps with the critical frequency in the hearing range of humans and macaques. The scanner noise is always intense and exceeds 97 dB (A) (Price et al., 2001). The major route of acoustic conduction is through ear canal at low frequencies which can be attenuated by the use of earplugs and ear defenders (by about 40 dB utmost) while bone conduction plays a major role at high frequencies (Ravicz and Melcher, 2001) which is comparatively harder to attenuate. Active noise cancellation might have been helpful however these cannot tackle sound conduction via bone conduction. So, the scanner noise will be audible to the subject despite best efforts to eliminate it.

This acoustic contamination has several implications (Hall et al., 2000) for an auditory fMRI study apart from the unpleasant experience due to the loud background scanner noise. First, the noise might mask the auditory stimulus employed implying that subjects may not properly hear the stimulus. Next, if different stimuli are masked to different degrees then stimulus effects might be confounded

with hearing difficulty due to masking. Further, hearing out a stimulus against scanner noise makes this an auditory segregation task (Scheich et al., 1998). Since the scanner noise induces a BOLD response in the auditory regions of a variable magnitude (Moelker and Pattynama, 2003) whose magnitude increases non-linearly with duration of acquisition (Talavage and Edmister, 2004), the dynamic range associated with stimulus-induced BOLD response might be reduced because of the elevated baseline. Next, scanner noise precludes the use of silent baseline essential for cognitive subtraction analysis. The difference in BOLD response between the auditory condition and silent baseline devoid of scanner noise is not the same as when the baseline has scanner noise (Gaab et al., 2007). The scanner background noise results in habituation of response to stimulus whose relative magnitude varies across auditory field maps (Di Salle et al., 2001).

2.3.2.1 Auditory imaging protocols

There are a few auditory imaging protocols that aim to circumvent the scanner noise (Talavage and Hall, 2012). Quiet acquisition sequences were developed in some cases (Sander et al., 2003) while others techniques employed active noise cancellation devices (Ravicz and Melcher, 2001, Chambers et al., 2001) or presented auditory stimulus during the silent phase with volume acquisition at a time that minimally affected the BOLD response due to experimental manipulation (Belin et al., 1999, Hall et al., 1999).

The sparse imaging techniques (see Figure 2-2) address all concerns raised earlier except for the reduced temporal sampling of the BOLD response when compared over continuous acquisition paradigms. Belin et al. (1999) developed an event-related paradigm that had a silent period between successive volume acquisitions (TR = 10s) during which auditory stimulus was presented at different lags with respect to the acquisition to enable sampling of the BOLD response at various points. In a similar technique developed by Hall et al. (1999), the sampling of the BOLD response was done at a time point where the peak is expected to occur, thus, further maximizing the SNR of the BOLD signal. Since the questions raised in the experiments reported in this thesis concern with the brain bases and representation of acoustic features, peak SNR afforded by sparse temporal sampling paradigms is more needed than the benefit of accurate temporal sampling of the BOLD response afforded by continuous imaging paradigms. Thus, the sparse temporal sampling paradigm is adopted in all experiments reported here.

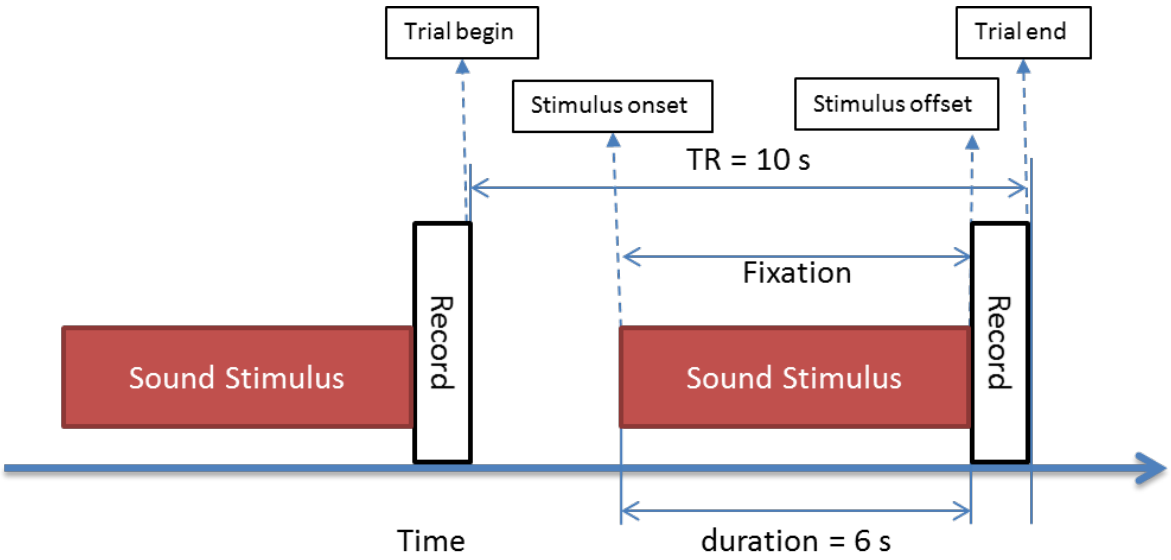


Figure 2-2 Schematic of the 'sparse temporal' fMRI design employed in this thesis. The duration of the sound stimulus is 6 seconds during which the macaque had to fixate on a visual cue. Repetition time (TR) is 10 s.

2.3.3 Image analysis

The analysis of functional MRI data involves pre-processing steps that include realigning volumes within and across sessions for head movement of the subject, reslicing, smoothening the images to improve the SNR. Experimental design modelling and statistical analysis were conducted next to yield spatial activation for each subject that was co-registered with a subject-specific structural image. All steps were applied using Statistical Parametric Mapping (SPM12) software implemented using MATLAB 2012 (MathWorks Inc.). A brief description of the underlying theoretical principles of all these steps is provided next.

Realignment

Realignment deals with compensating the head movement of the subject within a single session and realignment of volumes across separate sessions conducted on different days but from the same subject. Head movement limits the precise estimation of brain activity since the location of voxels keeps changing over time. In all the fMRI studies reported here, all animals in all sessions were head restrained, nevertheless, there was still minor head movement of the order of up to 1 mm within a session in the worst case. Even tiny movements across volumes within a session, even when not accounting for the volumes acquired on separate sessions, can contaminate the data (Friston et al., 1995) and contribute to misalignment and up to 90% of the variance can be explained (Friston et al., 1996). Thus, head movement leads to misattributed brain activation and hence a major source of a problem that needs to be addressed in the pre-processing step.

Movement artefacts are eliminated by realigning successive volumes to a common spatial reference frame, which is the first volume that was acquired in the first session with the animal. The realignment is based on least squares approach on affine rigid-body spatial transformation (3 parameters for translation and 3 parameters for rotation) that computes the movement associated with each volume (Friston et al., 1995, Andersson et al., 2001). These six parameters are used to reslice the volumes to the estimated positions as determined by the transformation (Grootoonk et al., 2000).

Normalisation

Normalisation is the process of aligning brains from different individuals on to a common anatomical space. This process is not a linear transformation as it needs to account for huge variability in the shape and size across different subjects. This process is achieved by a nonlinear estimation of spatial deformation patterns in addition to spatial transformation matrix (12 parameter affine transformation). Since the results shown in all the experiments reported here are presented as a group of single-subject analysis, the normalization step has not been performed on the data presented here. Further, co-registration of the functional and structural image is assured since the structural data is acquired alongside the functional in the same session.

Smoothing

Next, the data is smoothed using a Gaussian kernel of fixed width, typically conveyed as full-width-at-half-maximum (FWHM) that is 2-3 times the voxel size. This convolution improves the SNR by reducing the noise but more importantly, smoothing makes the data closer to the assumptions of the Gaussian that random field theory assumes which is needed for the statistical analysis of the brain activations. In this final stage of pre-processing, a Gaussian kernel of 3 mm was applied in the analysis of all experiments. The human fMRI studies typically apply 8 mm smoothing kernel as they co-register to a standard MNI template which requires local individual structural variations to be down-weighted. However, in the macaque studies presented here, the data are interpreted as a group of single-subject analyses and hence data is not co-registered to a standard space at the time of analysis.

2.3.4 Statistical analysis of fMRI data

The theory behind the statistical analysis of fMRI data is conveyed in brief here. Under the null hypothesis that there is no regionally specific effect, the signal in every voxel is assumed to have a Gaussian distribution. Through a mass-univariate approach, this hypothesis is tested at each voxel using the General Linear Model (GLM). This consisted of estimation of GLM parameters that is specified as a design matrix and evaluation of results at each voxel that yields statistical parametric maps (SPMs) of brain activity. All these steps are implemented using SPM12.

General Linear Model

GLM provides a framework for the statistical analysis of neuroimaging data, including BOLD data, using common parametric tests, like student's t-test etc. that can be conducted in each voxel. Briefly, GLM is applied at each voxel to model signal intensity as a linear combination of effects of interest, effects of no interest also known as confounds. The following equation describes GLM analysis (Friston et al., 1994)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{H\gamma} + \boldsymbol{e}$$

where $\boldsymbol{Y}$ is the vector with signal intensity values as a function of time at each voxel, $\boldsymbol{X}$ is a matrix reflecting experiment variables that are convolved with canonical haemodynamic response function, $\boldsymbol{H}$ is a matrix of confounds or regressors of no interest such as motion estimates; $\boldsymbol{\beta}$ is a vector of parameter estimates for effects of interest while $\boldsymbol{\gamma}$ is a vector of parameter estimates for effects of no interest; $\boldsymbol{e}$ is a vector of error values that are normally distributed. Both the matrices $\boldsymbol{X, H}$ are specified as the design matrix that has as many rows as the number of volumes ($\boldsymbol{N}$) and as many columns as the total number of regressors ($\boldsymbol{p}$). The parameter estimates $\widehat{\boldsymbol{\beta}}$ can be computed using ordinary least-squares (OLS) solution. An example of an OLS solution to simple GLM equation is given below

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y}$$

The parameter estimates are contrasted against each other by appropriately weighted contrast ($\boldsymbol{c}$) vectors. A $\boldsymbol{t}$ -statistic is computed as the ratio of contrast weighted parameter estimates to the estimated variance that depends on the noise variance estimate ($\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}$) computed using the error term ($\hat{\boldsymbol{\varepsilon}}$).

$$\boldsymbol{T} = \frac{\boldsymbol{c'}\widehat{\boldsymbol{\beta}}}{\sqrt{\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}\boldsymbol{c'}(\boldsymbol{X'X})^{-1}\boldsymbol{c}}} \sim \boldsymbol{t}_{\boldsymbol{N-p}}$$

$$\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}} = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\boldsymbol{N-p}}$$

Random Field Theory

Random field theory can be invoked to test for significant activation in each voxel. This assumes that, under the null hypothesis, the map of the parameter estimates

across all voxels is distributed according to a certain probability distribution i.e. a $\boldsymbol{t}$ or $\boldsymbol{F}$ distribution. Any deviation of this distribution that exceeds a pre-specified statistical threshold can be attributed to regressor of interest with a probability of $1-\alpha$, where α is the Type I error related to the false rejection of the null hypothesis.

Correction for multiple comparisons is required in the case of fMRI data due to a large number of tests conducted at each voxel. However, typical methods of such correction, like Bonferroni correction in which correction factor is arrived at by normalizing the false positive rate at each voxel with the huge number of tests, are too conservative due to a large number of voxels and hence tests. This implies that such a traditional correction would lead to missing out on true activations. Further, Bonferroni assumes independence but these tests are not truly independent due to the spatial correlation across neighbouring voxels both inherent as well as due to smoothing. Thus, an appropriate statistical framework is necessary to control the false positive rate that accounts for this spatial correlation. On the other hand, uncorrected statistics are used where there is an a priori hypothesis of activation as long as the significance threshold is stringent, for instance, $p<0.001$ (uncorrected). When there is no a priori hypothesis, typically correction for multiple comparisons based on family-wise error rate (FWE) (Logan et al., 2008, Nandy and Cordes, 2007) or false discovery rate (FDR) (Genovese et al., 2002) are to be used.

Group Analysis

In the analysis of fMRI data, there are three types of analysis: fixed-effects, random-effects and single-subject analyses. These vary in their underlying assumptions about how data from individuals is considered for group analysis. In the fixed-effects analysis, the assumption is that there is no variability between subjects for a particular effect of interest and hence any inter-subject variability is considered noise. On the other hand, the random-effects analysis assumes the inter-subject variability in activation as a random variable. This enables inference about the average behaviour across a group. In essence, fixed effects analysis shows the typical behaviour of a sample while random effects analysis shows the invariant behaviour of the population from which the sample was drawn. Typically, around 16 participants are required for robust results from the random-effects analysis. Finally, the single-subject analysis is a method that does not combine the results from each subject but compares the result across subjects for consistency. This approach is typically taken

in experiments with very few subjects like rare patient studies, primate studies. The primate studies reported in this thesis are single-subject analysis as the number of animals used are too few to warrant group level analysis. Further, it is reasonable to assume that the brain bases uncovered using a small set of subjects reflect the entire population as there is no inter-subject variability for basic processes that concern with auditory perception that does not require prior experience (like speech, music) unlike for high-level cognitive processes involving frontal cortex where this assumption is less valid.

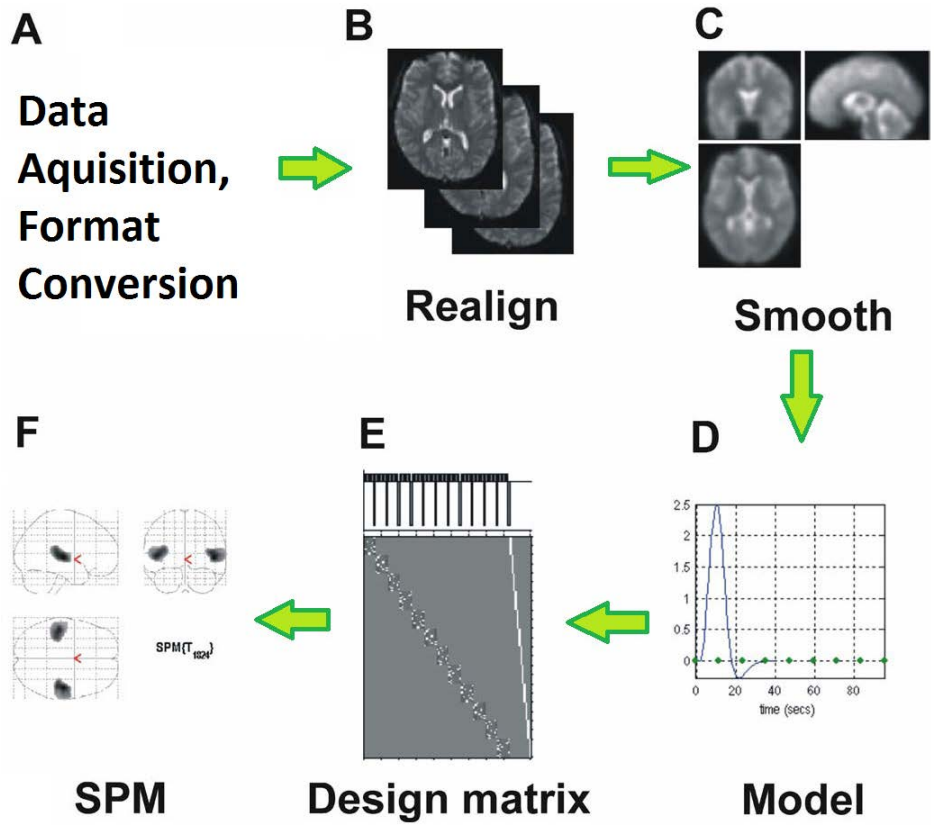


Figure 2-3 Analysis pipeline used on NHP fMRI data.(**A**) Raw data in the native scanner format is converted into the standard NifTi format for further analysis. A software package like SPM is used to perform the following steps. (**B**) Functional data volumes are then realigned to correct for head movement within and across sessions for a subject using an algorithm that minimizes variance between images. (**C**) Realigned volumes are smoothed using the Gaussian kernel of a specific size which improves signal-to-noise ratio but at the cost of a reduction in the spatial resolution. (**D**) Realigned and smoothened volumes are then analysed using a specified model, which also involves convolution with haemodynamic response function to account for the time course of cerebral blood flow in relation to the neuronal activity. (**E**) A design matrix is generated based on the general linear model, where rows correspond to volumes while columns to the total number of regressors and confounds. Parameter estimates also known as coefficients of regression for each regressor is estimated. A contrast between experimental conditions is defined as a weighted sum of parameter estimates. A t-statistic for a given contrast is derived at each voxel. (**F**) A statistical parametric map of the statistic can be plotted as 'glass brain' projections in axial, coronal, sagittal planes, or it could also be rendered on a structural image to indicate relationships of activation to brain anatomy. These activations can be thresholded based on specific significance level.

Chapter 3 Figure-Ground Segregation

3.1 Summary

A critical aspect of auditory scene analysis is the ability to extract a sound of relevance (figure) from a background of competing sounds (ground), known as segregation. In contrast to previous studies, this study employs a stimulus paradigm where the figure and the ground components overlap in the spectro-temporal space since it better approximates the challenges of segregation encountered in natural scenarios. The neural bases of pre-attentive stimulus-driven auditory segregation are investigated in rhesus macaques using functional magnetic resonance imaging while the animal performed a stimulus-irrelevant visual fixation task.

Significant activation in anterior superior temporal gyrus and posterior superior temporal sulcus (STS) is observed. These results suggest that similar to the human intraparietal sulcus (IPS), macaque posterior STS is involved in mediating pre-attentive auditory segregation identified using an identical stimulus construct as the previous human fMRI study. Further, these results also suggest that the neural bases for auditory stimulus-driven segregation in macaques are similar to that observed in humans. The data support the use of macaques as a model of human auditory scene analysis that allows us to carry out system-level characterisation and systematic neuronal specification of the system

3.2 Background

A major drawback of previous investigations of auditory streaming is that they employed simple narrowband stimuli with predictable sequences that do not reflect the complex spectral and temporal overlaps encountered in natural scenes. These were simple deterministic stimuli that either has a protective spectral band around the figure (Elhilali et al., 2009a) or differed from the background components in low-level acoustic attributes (Gutschalk et al., 2008). Other investigations have employed complex ethological stimuli like speech in noise, concurrent vowels (Alain, 2007), but these have semantic confounds that restrict their translation from human to animal studies.

In this work, a novel stimulus paradigm is employed, which is known as stochastic figure-ground (SFG) stimulus, (Teki et al., 2011) that has a stochastic variation of its elements in spectro-temporal space. The stimulus consists of a set of inharmonic pure tone elements that are temporally coherent (start and stop together), known as figure, amidst a set of randomly varying background elements in spectro-temporal space, known as ground. Spontaneous percept of a figure popping out of the ground occurs by grouping temporally coherent frequency elements that repeat across time. This stimulus design elicits a level of complexity in segregation that is typical of a natural task as the properties of the figure and the ground overlap in the feature space that varies across trials.

In human listeners, Teki et al. (2013) showed that the figure detection in these stimuli is associated with the segregation mechanism and not based on the detection of any low-level changes. They posit that these stimuli tap fundamental segregation mechanisms that might correspond to an early ‘generic’ form of figure-ground analysis relevant to a wide range of natural stimuli. They also showed that the segregation mechanisms investigated are robust but more susceptible to spectral rather than temporal perturbations of the figure components in these stimuli. Further, these stimuli have no semantic attribution or relevance to any species, thus, it allows us to test the existence of a common mechanism for the auditory segregation across species.

3.2.1 Macaque model

Auditory stream segregation has been suggested to exist across all animal species (Hulse, 2002). It is an important function of the auditory system and is crucial for

survival since natural environments typically contain multiple sound sources. This study adopted functional magnetic resonance imaging (fMRI) in macaques. This approach can bridge human fMRI studies with non-human primates (NHP) neurophysiological studies by establishing initial models for an organization that can provide prior areas for targeted neurophysiology. This method provides a direct framework for parallel modelling in NHPs based on both the blood flow data and neurophysiology to establish the neuronal basis that is not feasible in humans. Functional imaging in macaque auditory cortex is possible since the blood oxygenation level dependent (BOLD) signal has been characterized (Baumann et al., 2010) in addition to the previous syntheses of the functional cortical organization in macaques (Baumann et al., 2013).

3.2.2 Previous results

Previous modelling of auditory segregation in humans based on haemodynamic data (Teki et al., 2011) showed significant activations bilaterally in superior temporal sulcus (STS) and intraparietal sulcus (IPS), a non-auditory region, to increasing coherence. However, they did not find any significant activity in the primary auditory cortex contrary to reports from previous studies using other types of streaming paradigms. They reported that these non-primary auditory areas and non-auditory regions as the primary brain areas involved in the process of automatic, stimulus-driven figure-ground decomposition.

3.2.3 Current work

This study elucidated the brain bases underlying the process of pre-attentive stimulus-driven auditory segregation based on temporal coherence in rhesus macaques using fMRI while the naïve animals performed a stimulus irrelevant visual fixation task. Based on the previous report in humans (Teki et al., 2011), I hypothesized that bilateral activity in non-primary auditory areas would be seen i.e. parabelt in STS and activations in macaque intraparietal cortex. Thus, whole brain functional imaging was performed instead of just the auditory cortex.

3.3 Materials and methods

3.3.1 Subjects

The imaging data were obtained from scanning sessions with three rhesus macaques (*Macaca mulatta*) while behavioural data were obtained in two macaques. The details of the animals are presented in Table 2-1. Only one animal had both imaging and behavioural experiment data collected from it. The animals had been previously habituated to the scanner environment as well as exposed to some experimental auditory stimuli. Further, they had been trained to sit in a primate chair and perform a visual fixation task. A primate chair was used to position the animal in the magnet. The animals were scanned in awake behaving state. During the recording, the head of the animal was positioned with a custom-made polyetheretherketone (PEEK) head holder attached to a cranial implant.

Animal ID	Imaging	Behaviour
M1	Y	N
M2	Y	Y
M3	N	Y
M4	Y	N

Table 3-1 Summary of subjects participating in figure-ground experiments

3.3.2 Stimuli

The Stochastic Figure Ground (SFG) stimuli from Teki et al. (2011) are employed here as they model the natural scenarios closely by approximating the challenges of segregation typically encountered. It is characterised by a sequence of repeating chords (figure) occurring amidst an otherwise random background (ground).

The coherence of an SFG stimulus is the number of temporally coherent spectral elements that repeat in time. For example in the stimulus presented in Figure 3-1 A the coherence is ten as there are 10 spectral elements that are held constant from 2.0 to 4.0 s against a background of randomly varying spectral elements.

Sound stimuli were created using scripts written in MATLAB (MathWorks, Natick, USA) version 7.1 at a sample rate of 44.1 kHz and 16 bit resolution. Stochastic figure-ground stimuli were created with chords, defined as a sum of multiple pure tone elements. Chords contain a random set of pure tone components that are not harmonically related. These component frequencies were drawn randomly from a set

of 129 values equally spaced on a logarithmic scale between 179 to 7246 Hz with successive frequencies separated by 1/24th of an octave. Each chord is of 50 ms duration. The onset and offset of each tone were shaped by a 10 ms raised-cosine ramp. Each stimulus was 6 seconds long (or 120 chords) with the inter-chord interval set at 0 ms.

3.3.2.1 Stimuli for fMRI experiment

For functional imaging, stimuli were constructed using chords that made up the background consisting of a minimum of 5 to a maximum of 15 incoherent components. A stimulus with the figure was made of 10 additional coherent components that repeated for 40 chords corresponding to 2 seconds in duration located in the middle one-third of the stimulus i.e. 2 s or 40 chords after onset. To control for the increased power due to the presence of the figure in the regular stimuli, the control stimuli had an additional 10 incoherent components that changed from chord to chord and thus, did not form a figure. Thus, in the fMRI analysis, use of 'figure minus control' contrast ensures that the increase in sound intensity due to the presence of the SFG figure does not confound my inference on auditory segregation. Figure 3-1 provides a visual representation of the spectro-temporal decomposition of an SFG stimulus employed in this fMRI study – (A) contains a figure (B) control stimulus without a figure. Overall, subjects were presented with 135 exemplars of each stimulus type, or equivalently 50 per cent of all stimuli contained figure.

3.3.2.2 Stimuli for the behavioural experiment

For behavioural testing, stimuli contained 60 chords (3s in duration) and had a fixed number of components per chord (n = 15). In contrast to the imaging stimuli, extra elements were not added on top but incorporated into the existing stream of chords to remove any sound intensity cues. The coherence level of the figure was varied between 4, 6, 8, 10 and 12 components. Figure onset times were randomised between 0.3 and 2 seconds. This slight change in the stimulus design from the one employed in the imaging experiment was to keep the sound intensity constant both in the background as well as when a figure was added as the monkeys are sensitive to changes in intensity which increases the false alarm rates.

3.3.3 Stimuli Presentation in imaging paradigm

To reduce the interfering effect of the high-intensity noise generated by the MRI scanner that can acoustically mask the auditory stimuli as well as saturate the

auditory cortical areas, a ‘sparse temporal’ design is utilized where the acoustic stimulus can be presented devoid of scanner noise. Figure 3-2 illustrates the timeline for each trial. With the use of a pseudo-random sequence, each adjacent trial was ensured to have a different figure or control sound stimulus. The duration of each sound stimulus was 6 seconds. This duration is sufficient for the BOLD response in the macaque auditory cortex to reach a plateau (Baumann et al., 2010).

The sound stimuli were presented to the monkey using Cortex software (Salk Institute) at an RMS sound pressure level (SPL) of 75 dB using 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).

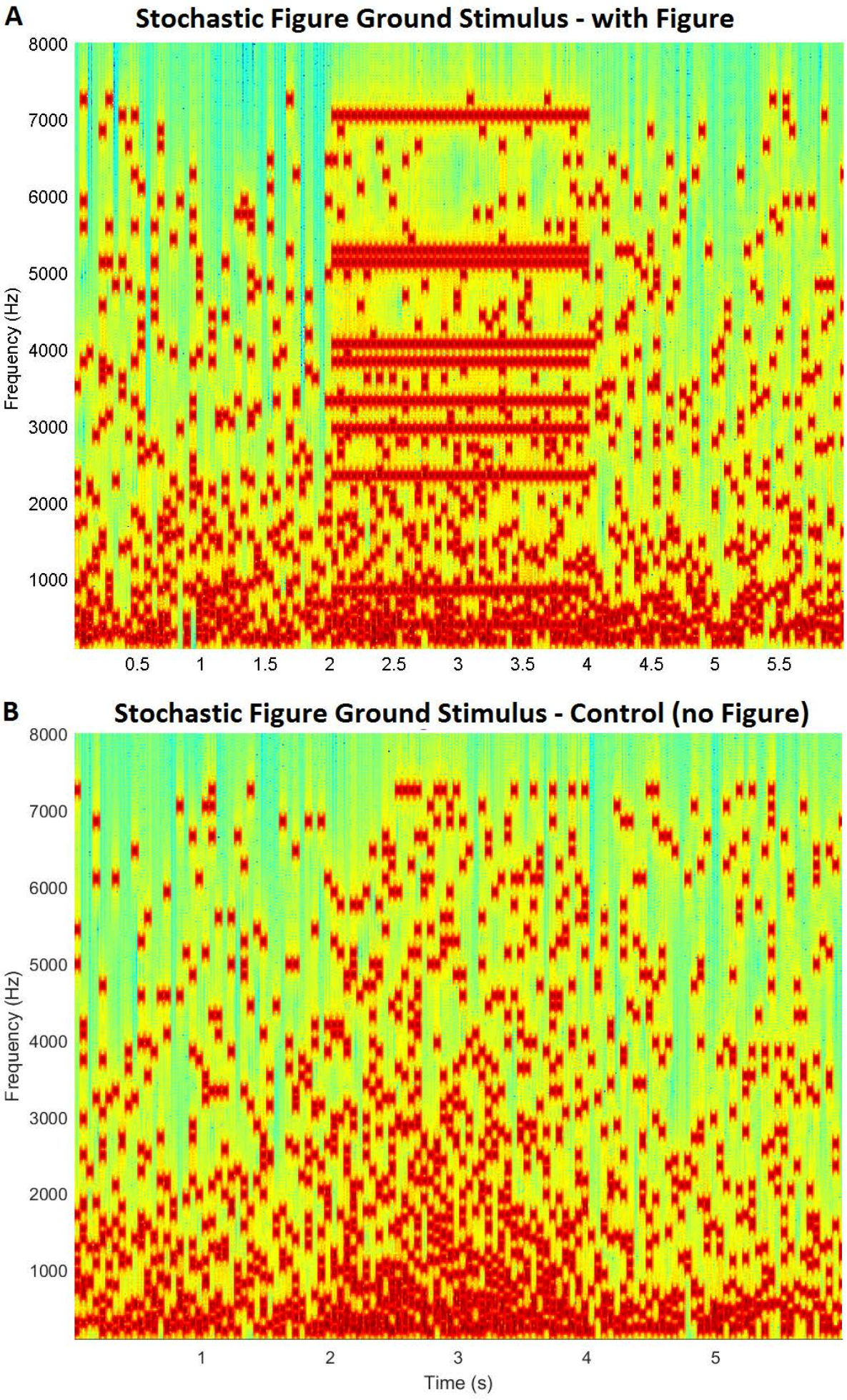

A
Stochastic Figure Ground Stimulus - with Figure
Frequency (Hz)
8000
7000
6000
5000
4000
3000
2000
1000
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
B
Stochastic Figure Ground Stimulus - Control (no Figure)
Frequency (Hz)
8000
7000
6000
5000
4000
3000
2000
1000
1
2
3
4
5
Time (s)

Figure 3-1 Spectrogram of example stimuli employed in fMRI study. (**A**) SFG stimulus showing an auditory object (or figure) embedded within a background (or ground). The ground is made up of a minimum of 5 to a maximum of 15 components chosen randomly from a frequency pool ranging from 179 to 7246 Hz. The figure with 10 coherent components repeating for 40 consecutive chords or 2s in duration is embedded within this ground from time 2.0s till 4.0s. Extraction of the figure is only possible by simultaneously grouping of elements across frequency as well as integrating such grouped elements across time. (**B**) SFG stimulus example that was used as a control that did not contain a figure. Instead of having a figure from time 2.0s to 4.0s, it has 10 additional components as the figure would have had over the same duration but incoherent in nature i.e. these additional components changed from chord to chord and thus, it did not form a coherent figure but controlled for the increased RMS of the sound stimulus.

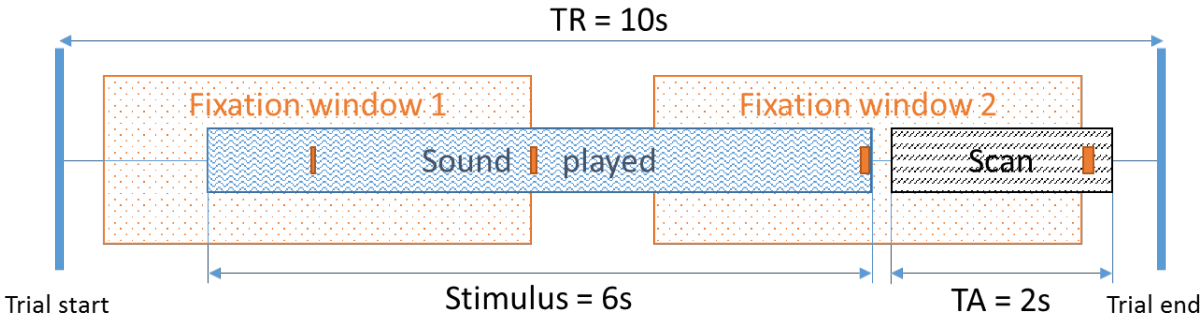


Figure 3-2 Schematic of the 'sparse temporal' fMRI design employed in this study. The duration of the sound stimulus is six seconds during which the macaque had to continuously fixate on a visual cue to be rewarded at regular intervals during the trial. Repetition time (TR) is 10.0 s while acquisition time (TA) is 2.01 s.

3.3.4 Task during imaging

As the aim of the present study was to uncover automatic stimulus-driven mechanisms that underlie auditory segregation, the monkey was required to perform a stimulus-irrelevant visual fixation on a cue during the entire time sound stimulus was presented. This simple task has certain advantages to it both to the quality of data acquired and to the wellbeing of the animal. First, this task assured that the attention of the animal is taken away from the stimulus. Next, it ensured that the levels of attention on the distractor cue remained consistent across the entire session. Further, it minimized the body movement of the animal by alleviating any stress it may have.

The eye position was monitored at 60 Hz with a tracking (camera-based with Infra-Red illumination) of the pupil using iView software (SMI, www.smivision.com, Teltow, Germany). The position, X and Y coordinates, of the pupil was communicated to Cortex software. The task was to fixate on a target (small red square) positioned at the centre of a screen, when the eye trace entered within a window of fixation (~ five degrees centred on the target) a timer started and the fixation target turned green. A continuous visual fixation (no saccades) of a pre-defined duration of 1.9 s was rewarded immediately by the delivery of a juice via a gravity-fed dispenser. If the continuous-fixation persisted for a total of 3.9 s (pre-defined) then it was further rewarded. This fixation regime is repeated twice in every trial including trials where no sound was presented.

3.3.5 fMRI Data Acquisition

Magnetic resonance imaging (MRI) was conducted in an actively shielded 4.7 Tesla vertical scanner (Bruker Biospec 47/60 VAS) dedicated to imaging in NHPs. It has an inner-bore width of 38 cm and a Bruker GA-38S gradient system from Bruker Medical, Ettlingen, Germany. Shimming was performed with the MAPSHIM algorithm (Kanayama et al., 1996) which measures B0 field inhomogeneity to apply first and second order corrections to it.

Data were acquired with parallel imaging with 2-fold GRAPPA acceleration using custom designed (www.wkscientific.com) 4-channel array receive coil. The RF transmission was achieved using a custom designed saddle coil from the same company in transmit mode. Both structural and functional data covered the whole brain. A navigator scan helped with the slice selection.

Functional MRI measurements by blood oxygenation level-dependent (BOLD) contrast consisted of single-shot gradient-recalled echo-planar imaging (GR-EPI) sequences with an in-plane resolution of 1.2 x 1.2 mm^2 and slice thickness of 1.2 mm, yielding 1.72 mm^3 voxels and a volume acquisition time (TA) of 2.011 s. Typical acquisition parameters were as follows – time to echo (TE) of 21 ms, flip angle (FA) of 90º, receiver spectral bandwidth of 200 kHz, a field of view (FOV) of 9.6 x 9.6 cm^2, with an acquisition matrix of 96 x 96.

A sparse design was employed where the acquisition of each volume was separated by a 10 s repetition time (TR) gap. This TR duration was necessary and sufficient to avoid recording the BOLD response to the gradient noise of the previous scan. The stimuli were presented during the 6 s immediately before a volume acquisition. These durations were based on previous characterisation of BOLD response time course in the auditory system of macaques (Baumann et al., 2010). On average, for every fourth volume acquisition, no stimulus was presented to obtain data for the silent baseline. In each functional imaging session of one-hour duration, 360 volumes were acquired resulting in 270 volumes for all stimuli or 135 volumes per each stimulus type (figure and control i.e. ground only) and 90 volumes for silence. In the combined sessions, there were 12, 10 and 4 sessions in subject M1, M2, and M4. The differing number of scanning sessions across the animals used in the fMRI data analysis leads to a difference in power in this single subject analysis presented here. This difference in the number of sessions was due to the non-availability of the monkey M4 for scans.

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 = 2000 ms, TE = 3.74 ms, TI = 750 ms, 30° flip angle, receiver bandwidth = 50 KHz, an in-plane resolution of 0.67 x 0.67 mm^2 with a slice thickness of 0.6 mm. These structural scans cover the same field of view as the functional scans.

3.3.6 fMRI Data Analysis

MR images were first converted from scanner's native file format into a common MINC file format, 3D for the anatomical data and 4D (x, y, z, t) for the functional data, using the Perl script pvconv.pl available online (http://pvconv.sourceforge.net/). From

MINC format, it was converted to NIfTI file format standard using MINC tools. This raw fMRI data were processed using Statistical Parametric Mapping (SPM12b; Wellcome Trust Centre for Neuroimaging) software (www.fil.ion.ucl.ac.uk/spm), using MATLAB 8 software.

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. Then, this data was spatially smoothened using a Gaussian kernel with full-width-at-half-maximum (FWHM) of 3 mm. 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 the stimulus onsets were convolved with a canonical hemodynamic 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 that is caused mainly due to scanner drift. Finally, this data was adjusted for global signal fluctuations also known as global scaling to account for differences in system responses across multiple sessions.

In a general linear model (GLM) analysis (Friston et al., 1994) of the combined sessions that included the motion parameters, the voxel-wise response estimates the regression coefficients, denoted beta. The t-values for the contrast of the different conditions versus the silent baseline were also calculated. The response to silent baseline was not explicitly modelled in the GLM and hence ‘sound minus silent baseline’ contrast looked for values of beta weights that were greater than zero.

A contrast map was generated to identify the brain areas in which activity is modulated by the presence of a figure in the stimuli. First, the functional data of the acquired volumes was projected onto the anatomical scans. Next, the response strength for the figure stimuli was contrasted with the control ground stimuli. This contrast map was calculated voxel by voxel by summing the differentially weighted (1, -1) regression coefficients (beta) of the figure and the control stimuli, henceforth referred to as ‘figure-versus-control’ contrast.

The ‘single subject inference’ was performed in these three subjects. Data were thresholded at $p<0.001$ (uncorrected for multiple comparisons across the brain) for areas where there was a prior hypothesis based on the human studies, i.e. auditory

cortex. Results from only monkey M2 survived $p<0.05$, family-wise error FWE corrected across the brain, and it showed a pattern similar to that presented here while results from other monkeys M1 and M4 did not survive corrections for multiple comparisons.

3.3.7 Probabilistic maps

The applied probabilistic maps are an estimate of functional areas of the auditory field in standard space (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 (Kaas and Hackett, 2000);(Baumann et al., 2013); (Hackett et al., 2014); (Brewer and Barton, 2016) . Isofrequency lines from mirror reversals between core and belt areas i.e. A1/R and ML/AL, were extended laterally to approximate the border between rostral and caudal parabelt. 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, it was made sure that each voxel is in at least 50% of the scanned population within the boundaries of the respective functional field.

3.3.8 Behavioural training

Both animals were naïve to the behavioural detection task during functional imaging. By means of positive reinforcement methods, a bar release was associated with a reward and thus, a relationship was established. Since animals needed to be trained in a figure detection task, a fixed target stimulus was paired via operant conditioning method. This target was a plain figure (duration: 1000ms, coherence: 10) without any incoherent ground elements. After the monkeys responded proficiently to the sound, the background components were introduced. The signal to noise ratio was gradually decreased by increasing the number of ground elements. After this phase, the sound level of the ground elements was fixed whereas the sound level of the figure was incremented to give animals an extra cue to the target. This sound intensity cue was then gradually decreased until animals could detect the figures without an intensity cue. Finally, figure coherence was systematically manipulated in order to assess the animal’s performance. Both animals trained on the task were able to perform successfully.

3.3.9 Behavioural task

To make inferences about the ability of macaques to segregate auditory figure from ground, a figure detection task was designed as a Go/No-Go paradigm (see Figure 3-3). For behavioural testing, macaques sat in a primate chair (Christ Instruments) and initiated trials by touching a bar placed in front of them. Two free-field speakers (Yamaha Monitor Speaker MS101 II), located at approximately 45 degrees to the left and right of the animal (distance: ~65cm from ear), delivered the stimuli at ~60dB 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. In case the monkeys missed to respond to a target, no reward was administered but a 3s penalty time-out was imposed. 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. Each behavioural session lasted around two hours (average number of trials per session: M2 = 1000, M3 = 873). Data were acquired, saved and analysed using MATLAB.

3.3.10 Behavioural data analysis

For behavioural data analysis, signal detection theory was applied. In total, data from 52 behavioural sessions were included in this analysis (M2: 23, M3: 29). The performance was evaluated based on d-prime, a measure of discriminability between responses to different stimuli (Nevin, 1969). Computation of d-prime values was done by using the formula below:

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

where Z is the z-transform defined as the inverse of the cumulative Gaussian distribution. Since d-prime 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-prime values across all sessions for each coherence condition was the basis for the assessment of the behavioural performance. Trials with responses below 0.4s after stimulus onset were excluded from further analysis (M2: 1.67%, M3: 1.38%). Since these trials were rejected thus not classified as false alarms. Reaction times were corrected for sound output latency. 95% confidence intervals were calculated via bootstrapping. Data were fitted with a second-order polynomial function. For statistical testing, data of both subjects were pooled. Effects of coherence were tested across sessions with repeated measures ANOVA for d-prime values, reaction times and responses variability respectively.

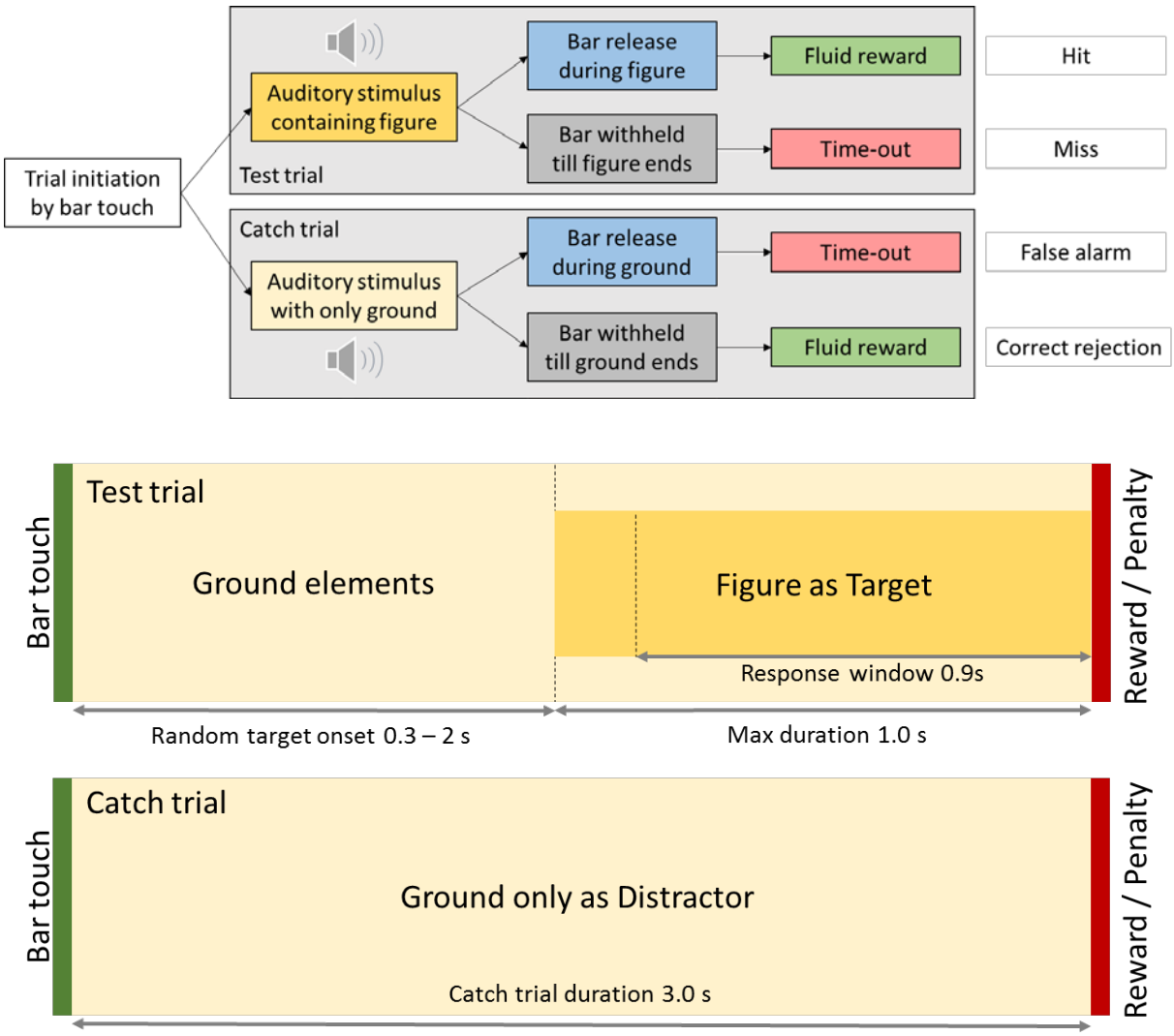


Figure 3-3 Schematic of the behavioural task. Animal initiates the trial by touching the bar which presents the auditory stimulus. In a test trial, the stimulus contains a figure as a target of maximum duration 1s with a random onset time (0.3-2.0 s) while stimulus in catch trial contains ground only (without a figure) as a distractor. Bar release during response window in test trials (hit) and bar withheld in catch trials is rewarded with fluid. A 3s time-out is applied as a penalty for bar release in catch trials (false alarm) or bar not released during response window in test trials.

3.4 Results

3.4.1 Activation to sound

The BOLD response to both figure and control sound stimuli across the entire brain was recorded. The BOLD activation associated with sound stimulation compared to silence was analysed in voxel space. Sound related activation ($p < 0.001$ uncorrected for multiple comparisons across the brain) was observed in the superior temporal plane that had a symmetrical pattern across the hemispheres in the three monkeys. Table 3-2 and Table 3-4 summarizes the percentage of voxels that are significant for sound versus silence contrast and the corresponding maximum t-statistic in each auditory ROI in the three monkeys identified using the probabilistic maps. Reddish-yellow hue in Figure 3-4 panel A shows the areas activated to sound stimulation in monkey M1, panel C and D shows activations in monkey M2 and panel B shows activations in monkey M4. This data indicates that the SFG stimulus robustly activates auditory cortex bilaterally.

3.4.2 Processing of the figure

Bluish-green hue in Figure 3-4 panel A shows activations for the processing of figure in anterior STG bilaterally in monkey M1, while panel C and panel D shows activations in left and right anterior STG respectively of monkey M2, and panel B shows activations in bilateral anterior STG of monkey M4. Figure 3-5 shows activations for processing of figure in non-auditory regions i.e. posterior STS bilaterally. The figure-versus-control contrast represents the degree of preference for the figure over control ground stimuli. Whole brain analysis found no brain region that responded more strongly during the control condition than the figure condition. Activations for figure versus control contrast from monkeys M1, M2 and M4 were rendered on top of a standard macaque brain as shown in Figure 3-6 conveying that segregation of figure occurs in the rostral-lateral belt and rostral parabelt of the monkey auditory cortex.

Table 3-3 and Table 3-5 summarizes the percentage of voxels that are significant for figure versus control contrast and the corresponding maximum t-statistic in each auditory ROI in the three monkeys identified using the probabilistic maps. In four hemispheres bilaterally across three animals, rostral parabelt (RPB) showed significant involvement in the processing of the figure. In three hemispheres bilaterally across two animals, antero-lateral belt area (AL) showed significant

involvement in the processing of the figure. In three hemispheres bilaterally across two animals, rostral temporal lateral belt area (RTL) showed significant involvement in the processing of the figure. In two hemispheres bilaterally in one animal, rostral temporal area (RT) showed significant involvement in the processing of the figure. The left caudal parabelt (CPB) in one monkey showed significant involvement in the processing of the figure. A visual summary of the regions involved in the processing of the figure is presented in Figure 3-7.

Per cent of voxels significant for sound vs silence contrast	Monkey M1		Monkey M2		Monkey M4	
	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi
A1	97.4%	94.8%	97.1%	100%	63.3%	89.2%
AL	100%	100%	100%	87.6%	100%	69.4%
CL	95.2%	77.7%	23.2%	74.5%	0%	31.1%
CM	13.1%	2.4%	1.4%	30.7%	0%	0%
CPB	100%	100%	100%	96.7%	31.2%	87.2%
ML	100%	100%	100%	98.4%	94.4%	87.4%
MM	6.0%	12.8%	88%	54.3%	82.1%	0.5%
R	63.6%	100%	97.4%	96.6%	100%	62.3%
RM	0%	23.4%	16.3%	6.4%	87.8%	0%
RPB	100%	90.8%	100%	67.1%	91.2%	69.4%
RT	63.3%	59.5%	96.3%	73.2%	100%	14.2%
RTL	100%	26.2%	100%	99.6%	89%	0%
RTM	17.3%	0%	31.6%	7%	100%	1%

Table 3-2 Summary of fMRI results in three monkeys – the proportion of voxels that are activated for sound stimulation. Percentage of voxels in each ROI that is significant ($p<0.001$ uncorrected for multiple comparisons across the brain) for sound versus silence contrast are summarized for each ROI in both hemispheres.

Per cent of voxels significant for figure vs control contrast	Monkey M1		Monkey M2		Monkey M4	
	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi
A1	0%	0%	0%	0%	0%	1.9%
AL	0%	0%	54.5%	32.6%	16.3%	0%
CL	0%	0%	0%	0%	0%	0%
CM	0%	0%	0%	0%	0%	0%
CPB	0%	0%	13.9%	3.3%	0%	0%
ML	0%	0%	3.4%	0%	0%	4.3%
MM	0%	0%	0%	0%	0%	0.4%
R	0%	0%	1%	0%	0%	0.5%
RM	0%	0%	0%	0%	0%	0%
RPB	0%	19.1%	79.2%	45.8%	12.7%	1.3%
RT	0%	0%	20.7%	8.9%	0%	0%
RTL	12.3%	0%	99.3%	82.7%	0%	0%
RTM	0%	0%	0%	4.0%	0%	0%

Table 3-3 Summary of fMRI results in three monkeys – the proportion of voxels that are activated for processing of figure. Percentage of voxels in each ROI that is significant ($p<0.001$ uncorrected for multiple comparisons across the brain) for figure versus control contrast are summarized for each ROI in both hemispheres.

Max t-statistic for Sound vs Silence contrast	Monkey M1		Monkey M2		Monkey M4	
	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi
A1	42.99	29.04	45.33	46.66	26.44	29.83
AL	31.81	22.05	30.50	33.08	19.75	14.24
CL	20.41	16.56	9.49	21.16	20.31	0.86
CM	15.56	4.88	4.69	15.30	1.00	9.61
CPB	31.49	31.17	13.80	16.27	1.65	2.35
ML	44.30	23.34	25.54	21.32	9.56	28.79
MM	6.70	7.56	38.62	26.35	13.73	25.79
R	25.62	25.40	40.99	38.70	21.98	3.29
RM	1.01	8.63	12.27	9.94	30.10	22.62
RPB	27.15	36.19	20.09	22.47	14.79	24.32
RT	10.34	8.54	28.94	25.42	24.53	5.05
RTL	21.53	6.67	27.73	21.94	18.55	2.03
RTM	4.56	2.83	14.50	10.63	21.05	8.29

Table 3-4 Summary of fMRI results in three monkeys – maximum t-statistic in each ROI in both hemispheres that is activated for sound stimulation (sound versus silence contrast). The t-statistic threshold of 3.09 is considered significant ($p<0.001$ uncorrected for multiple comparisons across the brain).

Max t-statistic for figure vs control contrast	Monkey M1		Monkey M2		Monkey M4	
	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi
A1	0.81	0.94	2.93	1.89	2.20	3.19
AL	1.98	1.88	5.00	4.01	3.78	2.26
CL	-0.65	0.23	1.91	2.53	1.92	1.77
CM	-0.55	-0.22	1.00	0.28	1.75	2.26
CPB	0.29	2.06	4.31	3.43	2.26	2.92
ML	0.82	0.23	3.55	2.29	2.62	3.42
MM	-0.99	1.14	2.16	1.46	2.61	3.13
R	1.16	1.29	3.15	2.07	1.49	3.07
RM	1.23	1.59	1.60	2.24	2.72	2.38
RPB	2.78	3.66	5.68	5.71	3.78	3.23
RT	1.43	1.73	4.07	3.49	1.33	0.24
RTL	3.29	1.90	5.99	6.44	2.48	2.05
RTM	0.91	2.04	1.99	3.21	1.30	0.67

Table 3-5 Summary of fMRI results in three monkeys – maximum t-statistic in each ROI in both hemispheres that are activated for processing of figure (figure versus control contrast). The t-statistic threshold of 3.09 is considered significant ($p<0.001$ uncorrected for multiple comparisons across the brain).

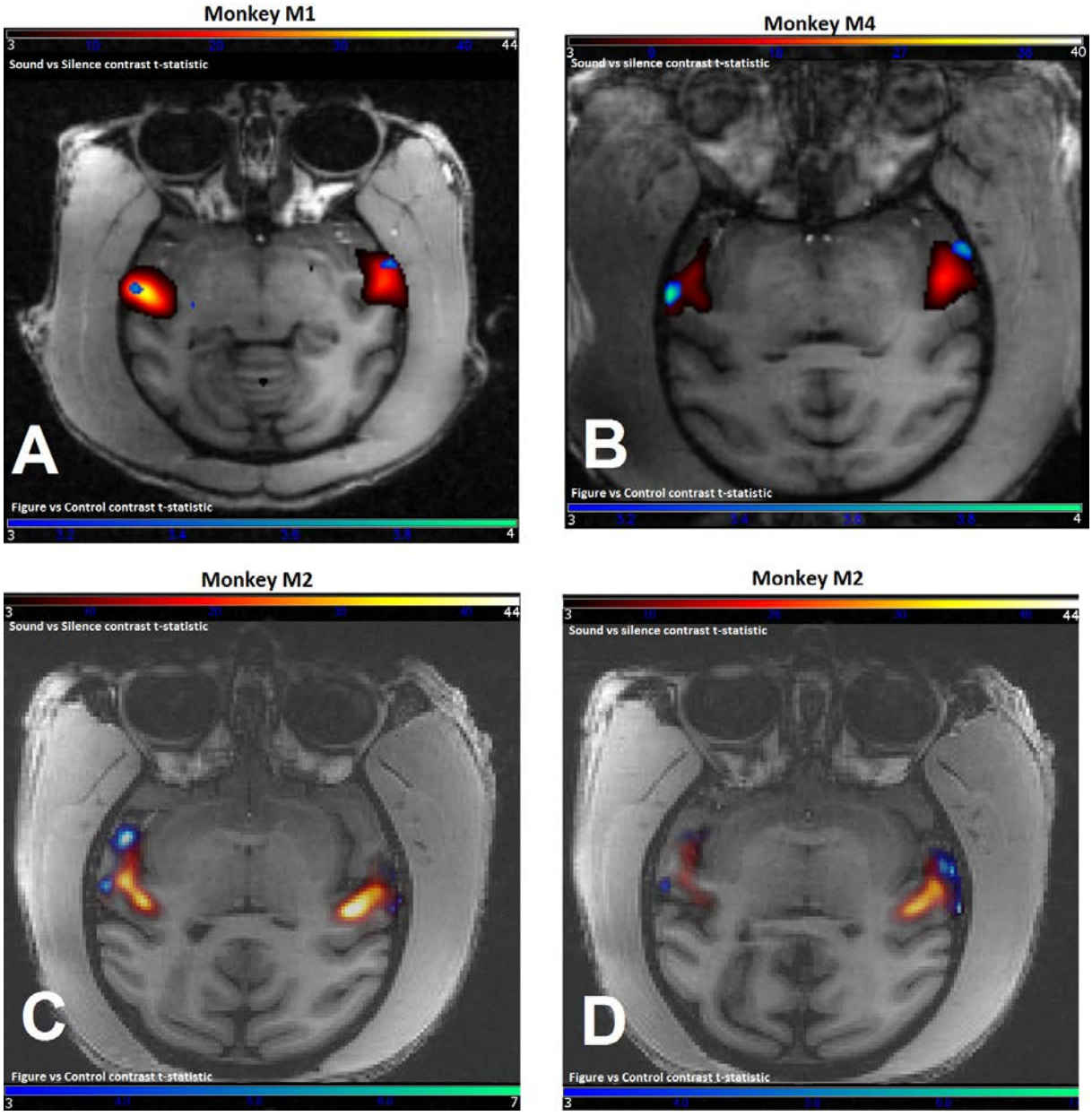


Figure 3-4 Contrast for figure vs control showing activations in STG bilaterally and sound versus silence contrast in three monkeys. Figure vs control contrast (bluish-green hue) is overlaid on top of sound over silent baseline contrast (reddish-yellow hue). Both these contrast maps are rendered on top of an axial section (T1 structural scan). The thresholds on statistical maps were kept at $p<0.001$ uncorrected for multiple comparisons across the brain. Colour scale indicates t-statistic for each contrast. Panel (**A**) shows the activations in monkey M1 with the axial plane passing through Z = -4 in this macaque. The maximum t-statistic for figure vs control contrast at left STG is 3.80 with sound vs silence t-statistic at 31.77 at that voxel while the maximum t-statistic for figure vs control contrast at right STG is 3.78 with sound vs silence t-statistic at 13.22 at that voxel. Panel (**B**) shows the activations in monkey

M4 with the axial plane passing through Z=-11 in this macaque. The maximum t-statistic for figure vs control contrast at left anterior STG is 4.51 with sound vs silence t-statistic at 3.87 at that voxel while the maximum t-statistic for figure vs control contrast at right anterior STG is 3.92 with sound vs silence t-statistic at 2.67 at that voxel. Panel (**C, D**) shows the activations in monkey M2 with the axial plane passing through Z=-3 (C) and Z=-5 (D) in this macaque. The maximum t-statistic for figure vs control contrast at left STG is 6.55 with sound vs silence t-statistic at 12.17 at that voxel while the maximum t-statistic for figure vs control contrast at right STG is 6.12 with sound vs silence t-statistic at 17.64 at that voxel.

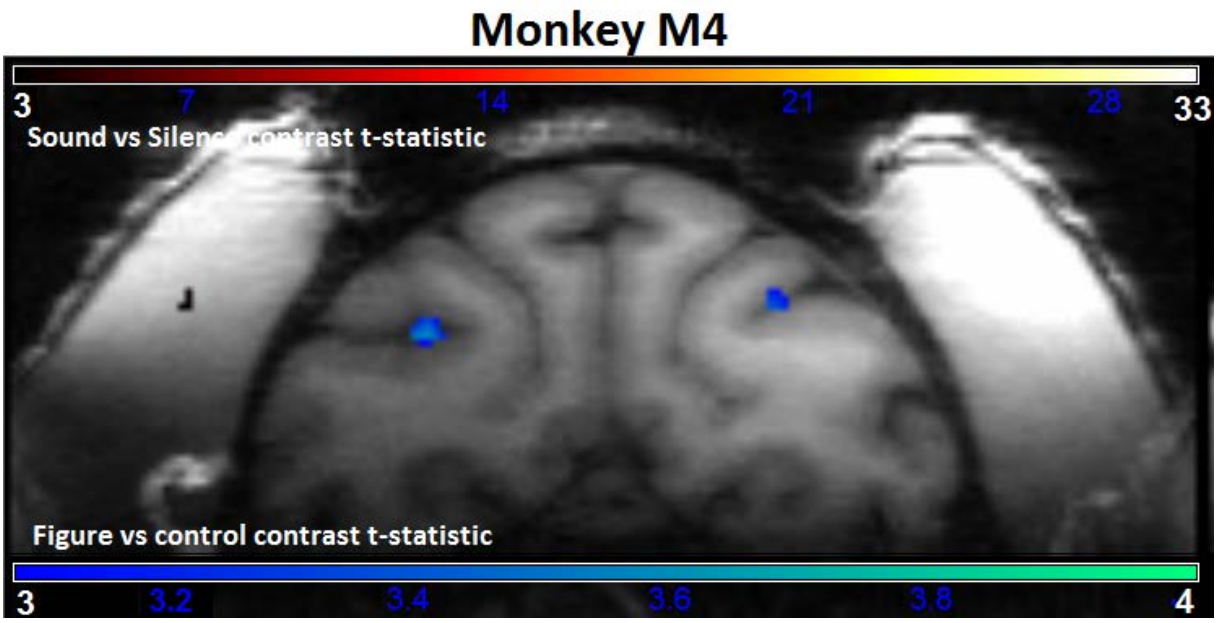


Figure 3-5 Contrast for figure vs control showing activations in non-auditory regions in monkey M4 i.e. posterior STS in the parietal lobe bilaterally. This contrast map (bluish-green colour) is rendered on top of a coronal view of the structural scan. The thresholds on statistical maps were kept at $p<0.001$ uncorrected for multiple comparisons across the brain. Colour scale indicates t-statistic for each contrast. The maximum t-statistic for figure vs control contrast at left posterior STS is 3.66 with sound vs silence t-statistic at -0.81 at that voxel while the maximum t-statistic for figure vs control contrast at right posterior STS is 3.57 with sound vs silence t-statistic at -2.38 at that voxel.

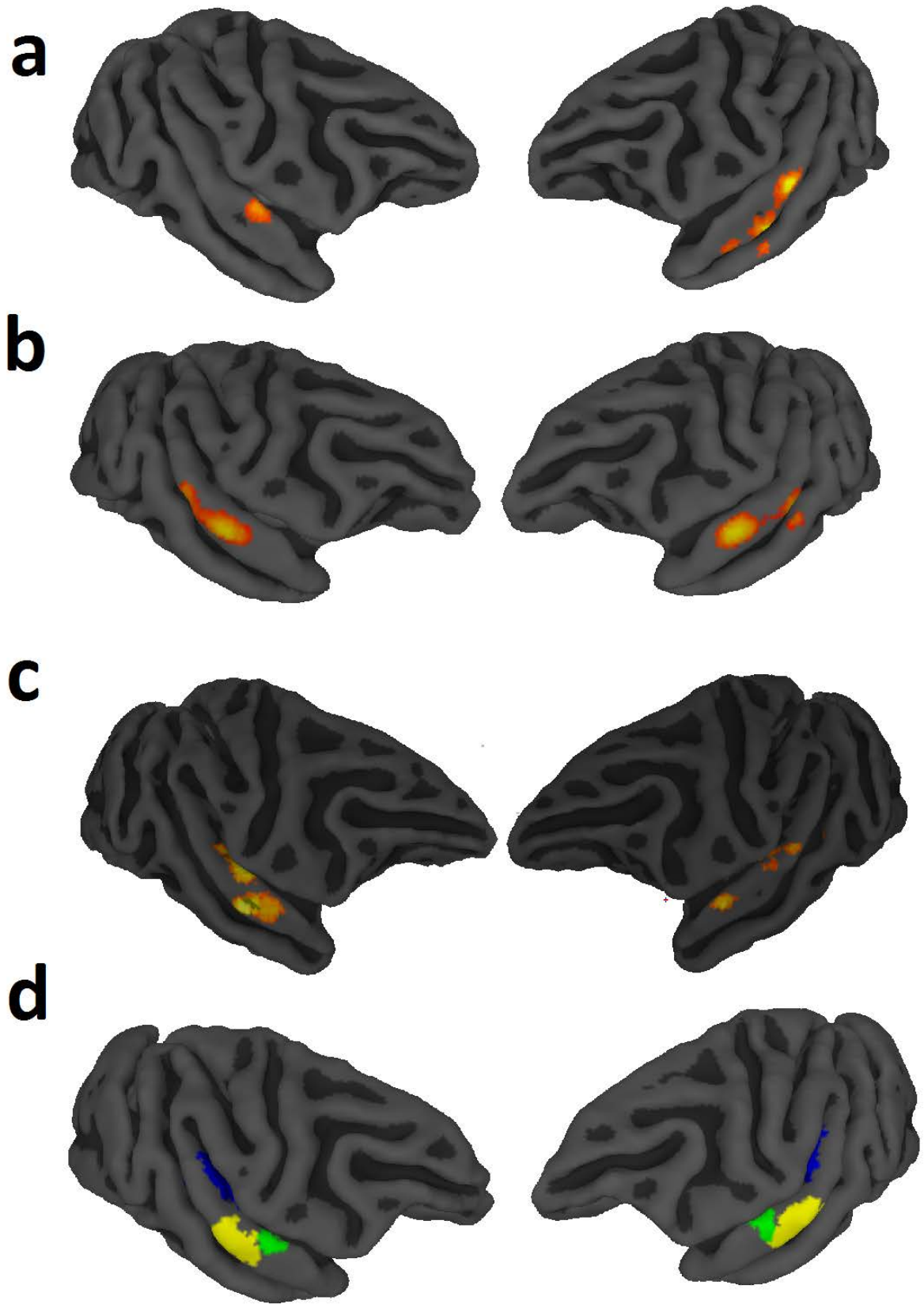

a
b
c
d

Figure 3-6 Contrast for figure vs control from three monkeys rendered on top of the standard macaque brain displayed in reddish-yellow hue is right and left hemisphere activations in anterior STG of monkey M1 (**a**), monkey M2 (**b**), and monkey M4 (**c**). (**d**) Colour-coded probabilistic maps of functional areas overlaid on standard macaque brain. Functional areas: A1 - Primary auditory cortex (blue), RPB – Rostral parabelt (yellow), RTL - Lateral rostro-temporal area (green). The activation for the segregation of figure occurs in the rostral lateral belt and rostral parabelt regions bilaterally

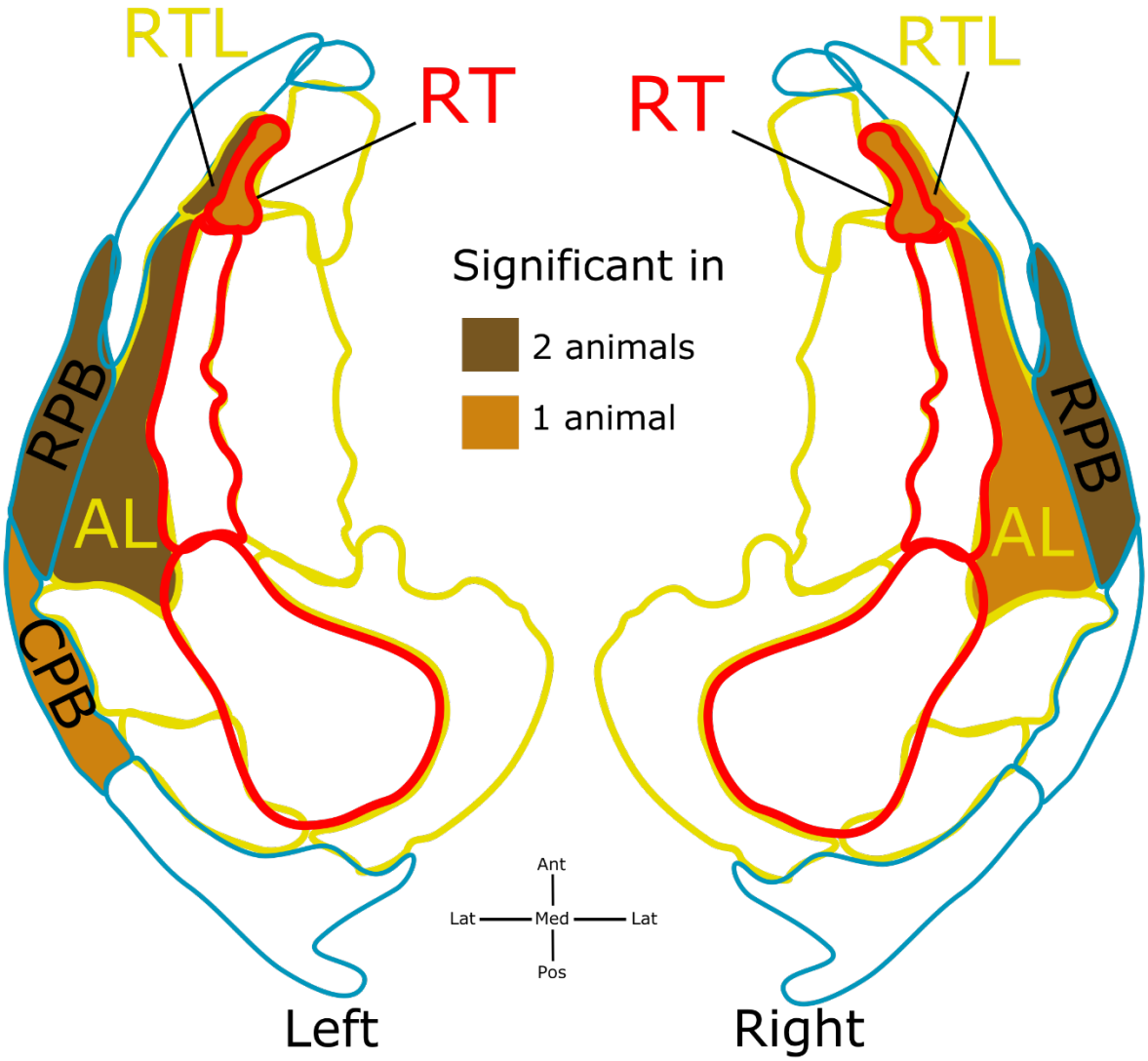


Figure 3-7 Visual summary of the fMRI results for the processing of a figure. Summary for figure versus control contrast evaluated across 3 monkeys in each ROI of auditory cortex obtained using probabilistic maps. ROIs are colour coded for the number of animals in which the result is significant (defined as 5% of voxels in an ROI have $t>3.09$ or $p<0.001$ uncorrected for multiple comparisons across the brain). The activation for the segregation of figure occurs predominantly in the rostral parabelt (RPB) areas, antero-lateral (AL) belt areas, and rostral temporal lateral (RTL) belt areas bilaterally.

3.4.3 Behavioural results

Behavioural experiments tested whether macaques can segregate auditory figure from background. Two monkeys were trained to perform an active figure detection task (see Figure 3-3). Proficiency on the task is indicated by the mean d-prime on the most salient condition i.e. condition with highest figure coherence (M2: 1.87, M3: 2.55). The reaction time (RT) distribution show a clear peak in both subjects (Figure 3-8, M2: Peak bin: 0.49 s - 0.53 s, Mean RT: 0.56 s; M3: Peak bin: 0.42 s - 0.46 s, Mean RT: 0.50 s), indicating competent segregation of auditory figures. In support of this conclusion, Figure 3-9 shows RT distribution for two coherence levels, where it is noticed that for a figure coherence of 4 (least salient) the hit rate is much lower with no clear peak in histogram while for a figure coherence of 12 (most salient) the hit rate is much higher with a clear peak in histogram

False alarm rate was at 22.08% in monkey M2 and 12.21% in monkey M3. False alarm rates were lower than hit rate across all coherence levels tested suggesting that monkeys could competently withhold responses to control stimuli without a figure. Further, a difference in false alarm rates between the two monkeys implies that the internal decision response criteria employed by the monkeys were different.

Next, repeated measures (RM-ANOVA) ANOVA was conducted on hit rate, cumulative d-prime, response latency and response variability to assess the effect of figure coherence. Suitable post-hoc tests were conducted to support the claim.

Hit rates (see Table 3-6 for monkey M2 and monkey M3) increased with figure saliency (see Figure 3-10 panel A). The main effect of figure coherence on hit rate was significant i.e. RM-ANOVA: F(1, 50) = 933.03, p = 5.26e-34, indicating that the number of coherent elements has an impact on the hit rates in both monkeys throughout sessions. Mauchly's test of sphericity on mean hit rates indicated that the assumption of sphericity was violated. $\chi^2(9) = 32.3$, p=1.7e-4. So a conservative lower bound estimate $(\varepsilon = 0.25)$ was applied as a correction for violating sphericity. One-sample Kolmogorov-Smirnov test on the mean hit rate in each coherence level of each monkey indicated that data was not normally distributed (see Table 3-7). So a non-parametric post-hoc test was used for comparison of figure detection performance across figure coherence levels. Wilcoxon signed rank test indicated a significant difference between mean hit rates across adjacent figure coherence levels (see Table 3-8).

Cumulative d-prime values increase as a function of increasing figure salience (see Figure 3-10 panel B). The main effect of figure coherence on cumulative d-prime was significant i.e. RM-ANOVA: F(4, 200) = 743.66, p = 8.13e-119, indicating that the number of coherent elements has an impact on the detection performance in both monkeys throughout sessions. Mauchly's test of sphericity on cumulative d-prime indicated that the assumption of sphericity was not violated. $\chi^2(9) = 8.28,\ p = 0.5$. One-sample Kolmogorov-Smirnov test on cumulative d-primes in each coherence level of each monkey indicated that data was not normally distributed (see Table 3-7). So a non-parametric post-hoc test was used for comparison of figure detection performance across figure coherence levels. Wilcoxon signed rank test indicated a significant difference between cumulative d-prime across adjacent figure coherence levels (see Table 3-8).

Reaction times (see Figure 3-10 panel C) decreased with increasing figure saliency. Mauchly's test of sphericity on mean reaction times indicated that the assumption of sphericity was violated. $\chi^2(9) = 78.5$, p=3.19e-13, since the variance of the differences between all combinations of reaction times (RT) are not equal. A conservative lower bound estimate $(\varepsilon = 0.25)$ was applied as a correction for violating sphericity. The main effect of figure coherence on mean RT was significant i.e. RM-ANOVA: F(1, 50) = 253.89, p = 3.12e-21, indicating that the number of coherent elements has an impact on the RT in both monkeys throughout sessions. One-sample Kolmogorov-Smirnov test on mean RT in each coherence level of each monkey indicated that data was not normally distributed (see Table 3-7). So a non-parametric post-hoc test was used for comparison of figure detection RT across figure coherence levels. Wilcoxon signed rank test indicated a significant difference between mean RT across adjacent figure coherence levels (see Table 3-8).

Response variability (see Figure 3-10 D) reduced with increasing figure saliency. It was assessed using the coefficient of variation (COV) i.e. ratio of standard deviation of the reaction time to mean of the reaction time in every session. The main effect of figure coherence on mean COV was significant i.e. RM-ANOVA: F(1, 50) = 44.64, p = 1.9e-8, indicating that the number of coherent elements has an impact on the response variability in both monkeys throughout sessions. Mauchly's test of sphericity on mean reaction times indicated that the assumption of sphericity was violated. $\chi^2(9) = 43.7$, p=1.6e-6. A conservative lower bound estimate $(\varepsilon = 0.25)$

was applied as a correction for violating sphericity. One-sample Kolmogorov-Smirnov test on mean COV in each coherence level of each monkey indicated that data was not normally distributed (see Table 3-7). So a non-parametric post-hoc test was used for comparison of response variability in figure detection times across figure coherence levels. Wilcoxon signed rank test did not indicate a significant difference between mean COV between adjacent figure coherence levels (see Table 3-8).

Behaviour in monkeys	Coherence level				
	4	6	8	10	12
	Monkey M2				
Mean hit rates	0.3445	0.5246	0.6957	0.8009	0.8639
Mean false alarm rate			0.2208		
Mean cumulative d-prime	0.3694	0.8314	1.2816	1.6144	1.8677
Mean reaction time [s]	0.5913	0.6144	0.5878	0.5539	0.5176
Mean coefficient of variation	0.2984	0.2648	0.2601	0.2430	0.2367
	Monkey M3				
Mean hit rates	0.4443	0.7047	0.8376	0.8997	0.9174
Mean false alarm rate			0.1221		
Mean cumulative d-prime	1.0242	1.7023	2.1488	2.4442	2.5522
Mean reaction time [s]	0.5950	0.5571	0.5090	0.4689	0.4283
Mean coefficient of variation	0.2531	0.2638	0.2445	0.2198	0.1849

Table 3-6 Summary of behavioural results from two monkeys M2 and M3 in the active figure detection task across all coherence conditions. Increasing hit rates as a function of increasing coherence and a low false alarm rate confirms that the figure detection performance of monkeys depends on coherence. Increasing d-prime as a function of increasing coherence confirms that the figure detection performance of monkeys depends on coherence. Decreasing reaction time as a function of increasing coherence (saliency) confirms that the figure detection performance of monkeys depends on coherence. Decreasing mean response variability as a function of increasing coherence confirms that the figure detection performance of monkeys depends on coherence.

Measure	Monkey	KS test	Figure coherence level				
			4	6	8	10	12
Hit rates	M2	Z stat	0.57	0.65	0.71	0.76	0.8
		p-value	1.9e-7	2.5e-9	3.2e-11	1e-12	6.3e-14
	M3	Z stat	0.62	0.71	0.76	0.78	0.8
		p-value	1.6e-10	8.7e-14	9.9e-16	1.6e-16	2.6e-17
Cumulative d-primes	M2	Z stat	0.49	0.66	0.81	0.87	0.93
		p-value	1.6e-5	7.9e-10	2.5e-14	1.3e-16	6e-19
	M3	Z stat	0.7	0.88	0.93	0.95	0.98
		p-value	2.2e-13	3.3e-21	3.2e-23	2.1e-24	6e-26
Mean reaction time	M2	Z stat	0.7	0.7	0.71	0.7	0.68
		p-value	8e-11	5.9e-11	3.5e-11	7.5e-11	2.3e-10
	M3	Z stat	0.7	0.68	0.68	0.67	0.66
		p-value	1.1e-13	7e-13	1.4e-12	2.4e-12	7.7e-12
Response variability	M2	Z stat	0.56	0.58	0.59	0.58	0.58
		p-value	2.9e-7	1.5e-7	1e-7	1.8e-7	1.5e-7
	M3	Z stat	0.58	0.58	0.58	0.57	0.55
		p-value	3.1e-9	1.7e-9	2.2e-9	4.2e-9	1.3e-8

Table 3-7 Test of normality on various measures for figure detection behavioural task. Using one-sample Kolmogorov-Smirnov test in each coherence level and each monkey indicated that the hit rates, cumulative d-primes, mean reaction times, response variability are not normally distributed. This implies that the post hoc tests on the measures need to employ non-parametric methods.

Measure	Monkey	Figure Coherence levels			
		4 and 6	6 and 8	8 and 10	10 and 12
Hit rates	M2	2.7e-5	2.7e-5	2.7e-5	4.2e-5
	M3	2.56e-6	4.32e-6	1.95e-5	0.03
Cumulative d-primes	M2	2.7e-5	2.7e-5	2.7e-5	4.6e-5
	M3	2.6e-6	4.3e-6	2.4e-5	0.037
Mean reaction time	M2	0.013	0.013	2.7e-5	2.7e-5
	M3	2.6e-5	1.1e-5	2.8e-6	2.6e-6
Response variability	M2	7.4e-3	n.s.	0.015	n.s.
	M3	n.s.	4.2e-3	2.1e-4	3.46e-5

Table 3-8 Post hoc test results on various measures from figure detection behavioural task testing for an effect of figure coherence. Using Wilcoxon signed rank test between adjacent coherence levels in each monkey the p-values given in the table indicated an effect of figure coherence on hit rates, d-primes, and reaction times. However, the effect of figure coherence on response variability was not clear. n.s. - not significant, p> 0.05

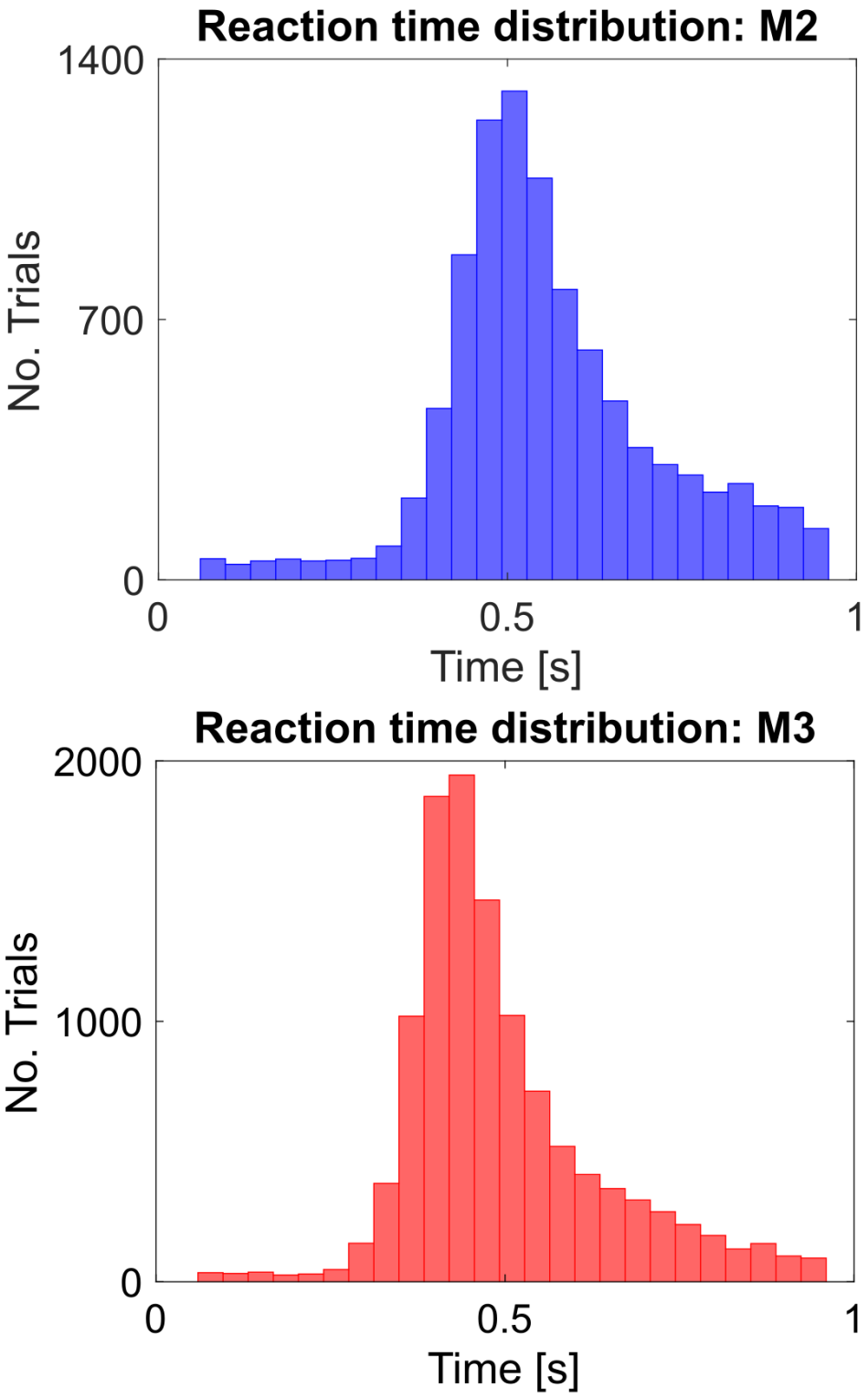


Figure 3-8 Reaction time (RT) histogram on the Go/No-Go active auditory figure detection task, averaged across all coherence conditions for monkeys M2 (blue) and M3 (red). RT data have been corrected for sound output latency. A single peak in the reaction time histogram confirms that monkeys are able to perform auditory figure-ground segregation.

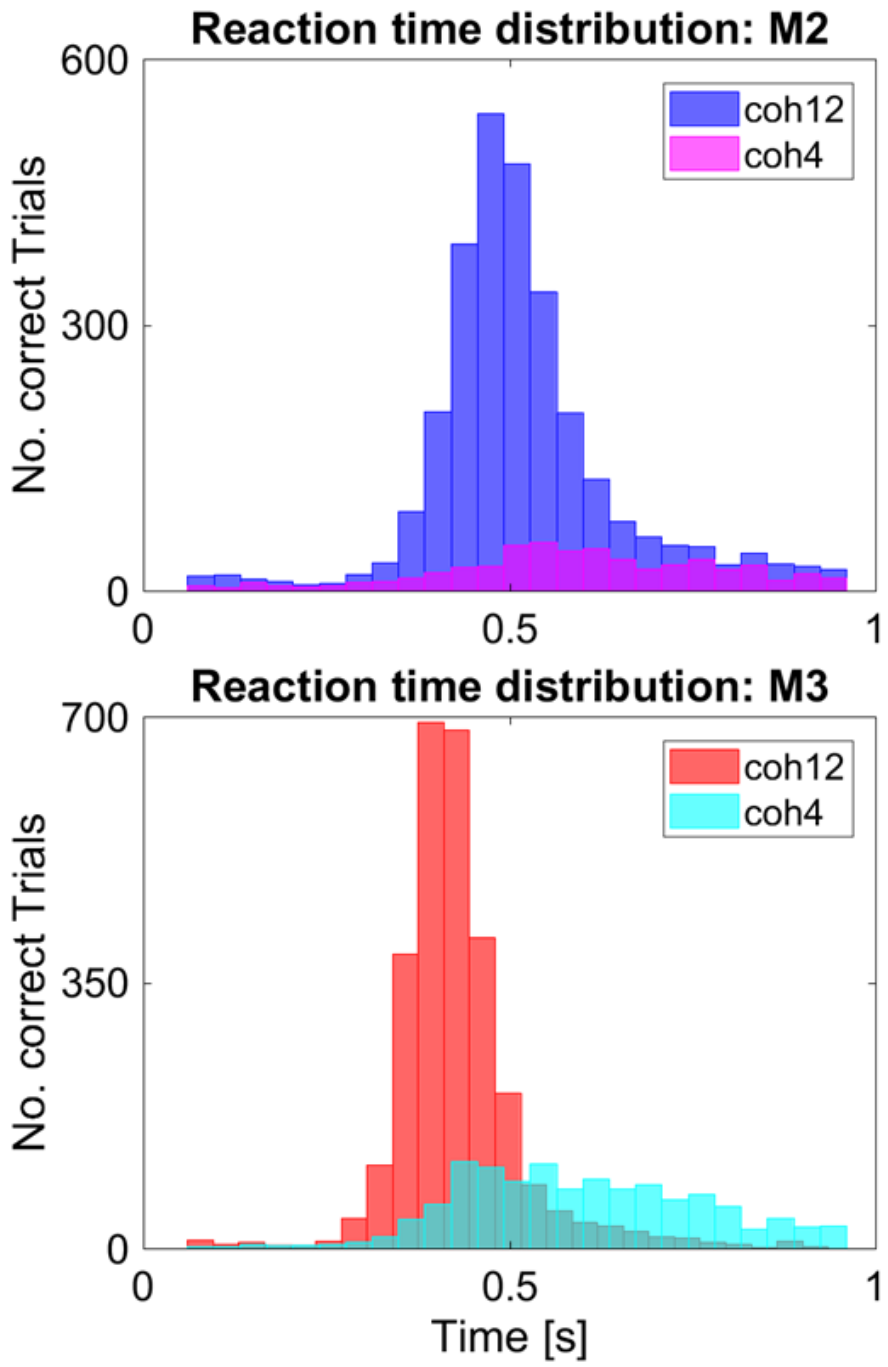


Figure 3-9 Reaction time (RT) histogram as a function of coherence on the active auditory figure detection task for the coherence of 4 (magenta/cyan) and 12 (blue/red) elements in monkeys M2 (upper panel) and M3 (lower panel). RT data have been corrected for sound output latency. Lower hit rate and broad RT distribution for a figure coherence of four elements in comparison to the coherence of 12 elements indicate difficulty in detecting figures.

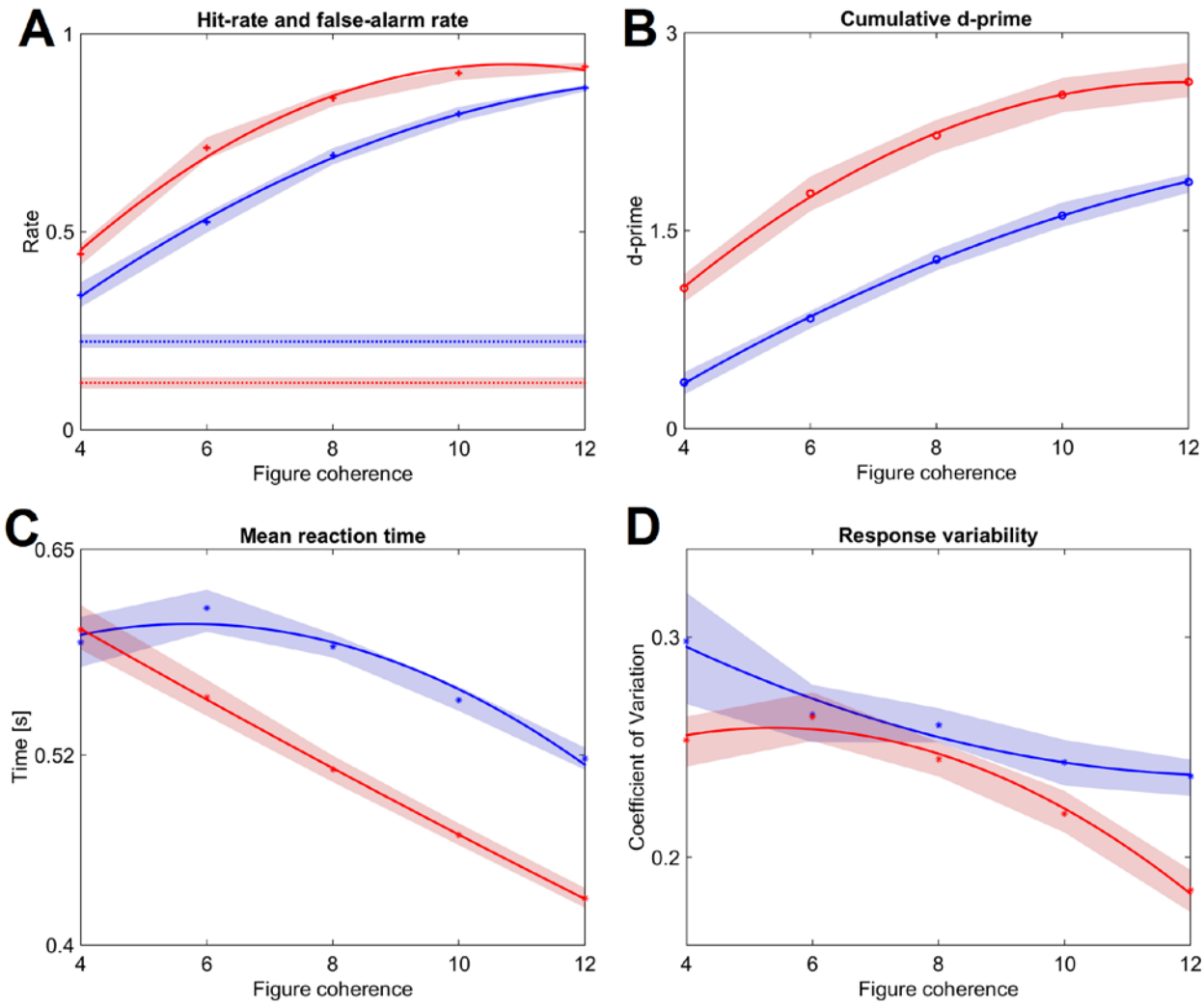


Figure 3-10 Summary of behavioural performance on the active figure detection task in two monkeys (**A**) Hit rate (solid line) and false alarm rate (dotted line) in monkeys M2 (blue) and M3 (red). Increasing hit rate as a function of increasing coherence and a low false alarm rate confirms that the figure detection performance of monkeys depends on coherence. (**B**) D-prime measure in monkeys M2 (blue) and M3 (red). Increasing d-prime as a function of increasing coherence confirms that the figure detection performance of monkeys depends on coherence. (**C**) Mean-reaction time across all coherence conditions in monkeys M2 (blue) and M3 (red). Decreasing reaction time as a function of increasing coherence (saliency) confirms that the figure detection performance of monkeys depends on coherence. (**D**) Mean response variability across all coherence conditions in monkeys M2 (blue) and M3 (red). Decreasing mean response variability as a function of increasing coherence confirms that the figure detection performance of monkeys depends on coherence.

3.5 Discussion

3.5.1 BOLD correlate of auditory segregation

The earlier human study by Teki et al. (2011) that used the same stimulus paradigm as employed in this work, implicate bilateral STS and IPS in the stimulus-driven partitioning of figure and ground components. Using fMRI, bilateral anterior STG was found to reflect the state of auditory perceptual organisation in all three monkeys. Bilateral posterior STS activation in the parietal lobe was seen in one monkey but not in the others.

3.5.2 Behavioural performance

The behavioural performance of both monkeys indicates that they can segregate auditory figures present in SFG stimulus from the background and that the segregation ability increases with increasing figure coherence which increases the saliency and signal to noise ratio. Similarly, in human listeners (Teki et al., 2013) the figure detection performance increases with increasing figure coherence. The behavioural results also indicate that the figure detection threshold i.e. a d-prime of 1, of one macaque seems to be around a coherence level of four elements while in the other macaque it is around between six and eight elements (Table 3-6). Humans can detect (i.e. d-prime of 1) a figure (a basic SFG of chord duration 50 ms) with a coherence level of four elements (Teki et al., 2013). Thus based on the results from two macaques, this limited data suggest that the threshold for figure detection is similar in humans and macaques.

Christison-Lagay and Cohen (2014) used a stimulus construct with 2-tone streaming complex with temporal coherence as proposed in Elhilali et al. (2009a), and they report behavioural results in macaques that are qualitatively similar to results seen in humans. This macaque behavioural study that inferred the effect of temporal coherence on segregation further supports the findings from my macaque behavioural experiment presented here where the auditory figure was composed of temporally coherent spectral elements.

3.5.3 Neural correlates of perceptual organization

Significant figure versus control contrast was found in bilateral anterior STG in three monkeys tested. Using probabilistic maps, these activations were found to correspond to antero-lateral (AL) belt, rostral lateral temporal (RTL) belt and rostral parabelt (RPB). The earlier human fMRI study (Teki et al., 2011) found modulation of

the BOLD signal in bilateral anterior STS that increased with increasing coherence of figures while the ECoG study in humans found activity in the convexity of STG (Griffiths, 2017). Thus parabelt homologue regions in the human auditory cortex are involved in the auditory segregation. Similarly, rostral lateral belt and rostral parabelt regions of the macaque auditory cortex are involved in the auditory segregation.

In addition, based on fMRI study in humans, Leaver and Rauschecker (2010) suggest a hierarchical organization of the antero-ventral auditory-processing stream, with the most anterior regions representing the complete acoustic signature of auditory objects. One could argue that the auditory figure in the current stimulus is similar to an auditory object by definition i.e. temporally coherent elements that repeat in time. Thus, one would expect the involvement of anterior regions of the ventral stream in the processing of figure. Since rostral lateral belt (AL, RTL) and rostral parabelt (RPB) lie along the ventral stream, the results in monkeys presented here are in good agreement with existing literature in humans. Thus, the neural correlate of perceptual organization in the auditory cortex identified in this fMRI experiment in macaques is valid across all primates.

3.5.4 Involvement of primary auditory core in segregation

Previous imaging studies in humans (Gutschalk et al., 2005, Snyder et al., 2006, Wilson et al., 2007) have identified activity in A1 that correlated with a streaming percept. However, the earlier study (Teki et al., 2011), as well as this study, found no evidence of activity in A1 for the processing of figure. It could be due to the difference in the stimulus employed in the current study against the simplistic ones used in the previous studies on auditory streaming. Further, the time required to extract the figure from the ground in the current paradigm is an order of magnitude lower than the typical build-up rate reported in the earlier studies of streaming. Thus, the absence of activity in primary auditory areas in the current paradigm when compared to earlier paradigms could be due to the underlying mechanisms of the perceptual organisation being different between these paradigms.

3.5.5 Involvement of parietal cortex in segregation

Contrary to previous studies but consistent with Cusack (2005), Teki et al. (2011) found modulation of the BOLD signal in IPS with increasing coherence of figures which also increased its perceptual segregation (Teki et al., 2013). Since the human participants in Teki et al. (2011) were not making perceptual judgments on the figure

during imaging, it is unclear the extent to which IPS reflected perceptual processing as opposed to automatic stimulus processing (Snyder et al., 2012). But Teki et al. (2011) suggested that IPS plays an automatic role in the auditory figure-ground segregation. However, this study did not find activity in the monkey IPS which has been implicated (Grefkes and Fink, 2005) in mediating object representations, binding of sensory features within and across different modalities as well as attentional selection. Instead, in one animal, significant bilateral activity for figure versus control contrast was found in posterior STS located in the parietal cortex. The sound versus silence contrast in this region (see Figure 3-5) was negative confirming that this is a non-auditory region. Given that the activity in the parietal cortex has been seen only in one animal without training in passive paradigm, evidence from imaging in animals while they perform an active figure detection paradigm may be able to ascertain the involvement of parietal cortex in monkeys. So, the role of monkey posterior parietal cortex in the perceptual organization is still uncertain.

3.5.6 Prediction of properties of neurons

Segregation of figure in this SFG stimulus requires detection of temporal coherence across a huge range of frequencies that are not harmonically related. This would require single neurons with broad multi-peaked tuning. The necessary broad tuning expected of such units is described in belt cortex of monkeys (Rauschecker and Tian, 2004). Given the parabelt cortex lies in the high level in the hierarchy (Scott et al., 2017, Kaas and Hackett, 2000, Hackett et al., 2014), one could expect broadband tuning properties of units in these areas (Kajikawa et al., 2015). Thus, this study predicts the existence of single units with broadband responses in the rostral parabelt of the monkey.

Further, based on fMRI in humans that employed natural sounds, Moerel et al. (2013) reported that the cluster representing the auditory population which has a sensitivity to multiple frequency bands with no clear harmonic relationship was populated mostly in the anterior STG. In the observations made herein, monkeys showed anterior STG (i.e. RPB) activation for the processing of figure comprising spectral elements that did not have a harmonic relationship. Thus, there is haemodynamic data in humans for the existence of neuronal units with multiple peaks without a harmonic relationship in rostral parabelt.

The fMRI data in humans (Teki et al., 2011) showed that the activity in the convexity of superior temporal sulcus in the auditory cortex varied parametrically with figure coherence, i.e. activity increased with increasing figure coherence. Since fMRI based on BOLD signal is sensitive to only a ‘rate code’ – a neuronal coding scheme that increases the baseline metabolism – I predict that the single units in macaque rostral lateral belt and parabelt convey information on the figure coherence via the average rate of neuronal spikes which increases with increasing number of simultaneously occurring frequencies but agnostic of the actual constituent frequencies.

However, using fMRI one cannot detect neuronal activity that employs coding schemes like temporal code, sparse code and population codes as they do not increase baseline metabolic demand. So I will not be able to speculate on whether these neuronal codes are employed in the detection of temporal coherence.

3.5.7 Conclusions

Given the similarity in the brain basis underlying auditory figure-ground segregation in humans and macaques and the similarity in the behavioural performance for segregation of figures from the background, one can conclude that macaques are a good model of human cortical analysis of the auditory perceptual organization.

Chapter 4 Spectral flux

4.1 Summary

Spectral flux is a key determinant of timbre, defined as the rate of change of spectral energy. In this work, the mechanisms underlying the encoding of spectral flux were explored in the auditory cortex of macaques through fMRI. Synthetic stimuli with systematic variation in the degree of energy fluctuation in the acoustic spectrum were employed for characterising the brain activation corresponding to different degrees of spectral flux. A previous study in humans that employed the above stimuli reported bilateral sensitivity to decreasing flux in belt homologues viz. planum temporale and anterior superior temporal gyrus, and right lateralized activity in parabelt homologues viz. superior temporal sulcus. Contrary to these findings, my data from three macaques suggested bilateral sensitivity to increasing flux in both core and belt auditory cortices bilaterally. The preference of the auditory core in macaque was for higher flux unlike the lack of differential sensitivity exhibited by the auditory core in humans. These findings support a functional organization of spectral flux in macaques that is different from that in humans. I speculate that these differences are related to the differences in the perception of temporal windows between the species.

4.2 Background

Spectral flux, one of the dimensions of timbre, is defined as the rate of change of spectral energy as a function of time. In the case of speech signals, phonemes have a high spectral flux while syllables have lower flux (Rosen, 1992). The degree of spectral energy fluctuation may be characterized by Pearson product-moment correlation between amplitude spectra in adjacent time frames (Caclin et al., 2005, Krumhansl, 1989). Alternatively one can interpret this spectrotemporal correlation in terms of time window by the duration of window required within which any two frames have a minimum amount of correlation. Thus, higher spectral flux implies a shorter time window while higher spectral flux implies a longer time window. This allows us to interpret spectral flux as the underlying preference of time window duration.

4.2.1 Need for synthetic stimuli

In this study, synthetic stimuli were used as these have advantages over natural sounds. First, these stimuli afford systematic alteration of its statistical properties of a specific underlying feature without changing other features. Thus, it enables one to seek the organization of this feature's processing in the brain. Next, these sounds match the acoustic complexity of sounds like speech but have no semantic attribution or relevance to any particular species. Thus, it allows one to test the existence of a common mechanism for the analysis of timbre across primates. Further, it also enables one to establish a macaque model of human cortical analysis of timbre.

4.2.2 Current work

This study aims to understand spectral flux analysis at the systems level in a macaque model. Previous modelling of spectral flux organization in humans (Overath et al., 2008) was based on haemodynamic data. Current work was aided by previous syntheses of the functional cortical organization in macaques (Baumann et al., 2013). fMRI in macaque auditory cortex is possible since the blood oxygenation level dependent (BOLD) signal has been characterized (Baumann et al., 2010). Based on BOLD activity from sparse fMRI, the auditory cortical areas responsible for spectral flux analysis will be identified using a range of spectral flux values that span natural sounds.

4.2.3 Previous results

The previous study in humans, Overath et al. (2008) reported bilateral sensitivity in planum temporale (PT) and anterior superior temporal gyrus (aSTG) to longer time

window duration (or higher spectrotemporal correlation) while also observing a significantly right lateralized activity in superior temporal sulcus (STS). Further, there was no differential sensitivity to spectral flux in core homologues in humans.

4.2.4 Hypothesis

I expected the results in macaques that are consistent with the human study i.e. I expected that in macaques I would see a preference for longer time windows in corresponding homologue regions. So, similar to the human study, I conducted an fMRI study in macaques covering only the auditory cortex. The regions analogous to human PT in macaques are Caudal Medial (CM) and Caudal Lateral (CL) located posterior to the 'protuberance', the macaque's version of Heschl's Gyrus (HG) (Semple and Scott, 2003, Jones et al., 1995, Baumann et al., 2013). The anterior STG is located laterally next to RT, while STS has a corresponding structure in monkeys. Thus, I hypothesized to observe sensitivity of the BOLD contrast to longer time window duration (or higher spectrotemporal correlation) in the CM, CL, and STGr regions apart from right lateralized activity in STS. Further, I hypothesized to see a flat or no differential sensitivity to spectral flux in the auditory core regions (A1, R regions).

4.2.5 Analysis

To characterize the spectral flux results, I employ a linear parametric contrast on the general linear model regression coefficients of various spectral flux conditions to reveal the preference of time window duration. Next, to characterize the results as a function of auditory field, I need to parcellate the auditory cortex into fields and hence I conduct tonotopy and collect anatomical MRI including T1w and T2w images. Further, to characterize the relative preference of time window duration, I fit a linear function on the general linear model regression coefficients of various spectral flux conditions.

4.3 Materials and methods

4.3.1 Subjects

The imaging data were obtained from scanning sessions with four male rhesus macaques (*Macaca mulatta*), denoted M1, M4, M5, and M6. The animals have been previously habituated to the scanner environment as well as exposed to some experimental auditory stimuli prior to scanning. Further, they had been trained to sit in a primate chair and perform a visual fixation task. A primate chair was used to position the animal in the magnet. The animals were scanned in awake behaving state.

4.3.2 Spectral flux characterisation

Spectral flux was characterised by constraining the Pearson product moment correlation (denoted r_1), henceforth termed as 'correlation', between amplitude spectra of adjoining frames. This systematic variation in the degree of fluctuation of acoustic spectral energy quantified spectral flux. The correlation was formulated (Overath et al., 2008) as given below:

Equation 4-1

$$r_1(k) = -\left(\frac{1}{s_k \cdot s_{k+1}}\right) \cdot \frac{1}{n} \sum_{j=1}^{n} \left(\left(a_{j,k} - \bar{a}_k\right) \cdot \left(a_{j,k+1} - \bar{a}_{k+1}\right) \right)$$

In Equation 4-1, r_1 is the Pearson product moment correlation between adjacent frames k and $k+1$ whose amplitude spectra is denoted as $a_{j,k}$ for the amplitude (expressed in dB) of the j^{th} frequency of n such frequency components belonging to the k^{th} frame, while $\bar{a}_k$ denotes the mean and s_k denotes the standard deviation of the amplitude spectra corresponding to the k^{th} frame.

The interpretation of the relation between spectrotemporal correlation and spectral flux is intuitive. As the correlation increases, the amplitude spectra of two adjacent frames of the sound vary to a less extent. This implies that the spectral flux of a stimulus synthesized with high correlation is low. To further illustrate this aspect, consider, for a stimulus with a correlation value of one, the spectral flux is zero since there is no change in the acoustic energy over time, while for a stimulus with a correlation value of zero, the spectral flux is highest due to the drastic changes in the

spectral energy as a function of time. However, this inverse relationship between spectral flux and spectrotemporal correlation does not hold for negative values.

Equation 4-2

$$r_n = (r_1)^n; \quad win\_len = frame\_dur \cdot \frac{\ln(r_{min})}{\ln(|r_1|)}$$

The correlation between any two frames in a stimulus is characterised by the number of frames between them and the correlation between adjacent frames. Equation 4-2 describes the correlation between two frames, denoted as r_n, as a function of the spectrotemporal correlation r_1 between adjacent frames and the temporal distance between the frames, denoted as n when the selected frame is n frames away from the reference frame. This equation (Overath et al., 2008) also determines the length of a time window (denoted win_len) required to reach a minimum level of correlation (denoted r_{min}) between any two frames within it or alternatively the correlation between farthest frames contained within the window. The window duration is a function of the correlation r_1 and r_{min} and the duration of a frame denoted $frame\_dur$. Figure 4-1 presents the relationship between the parameters across the sample values.

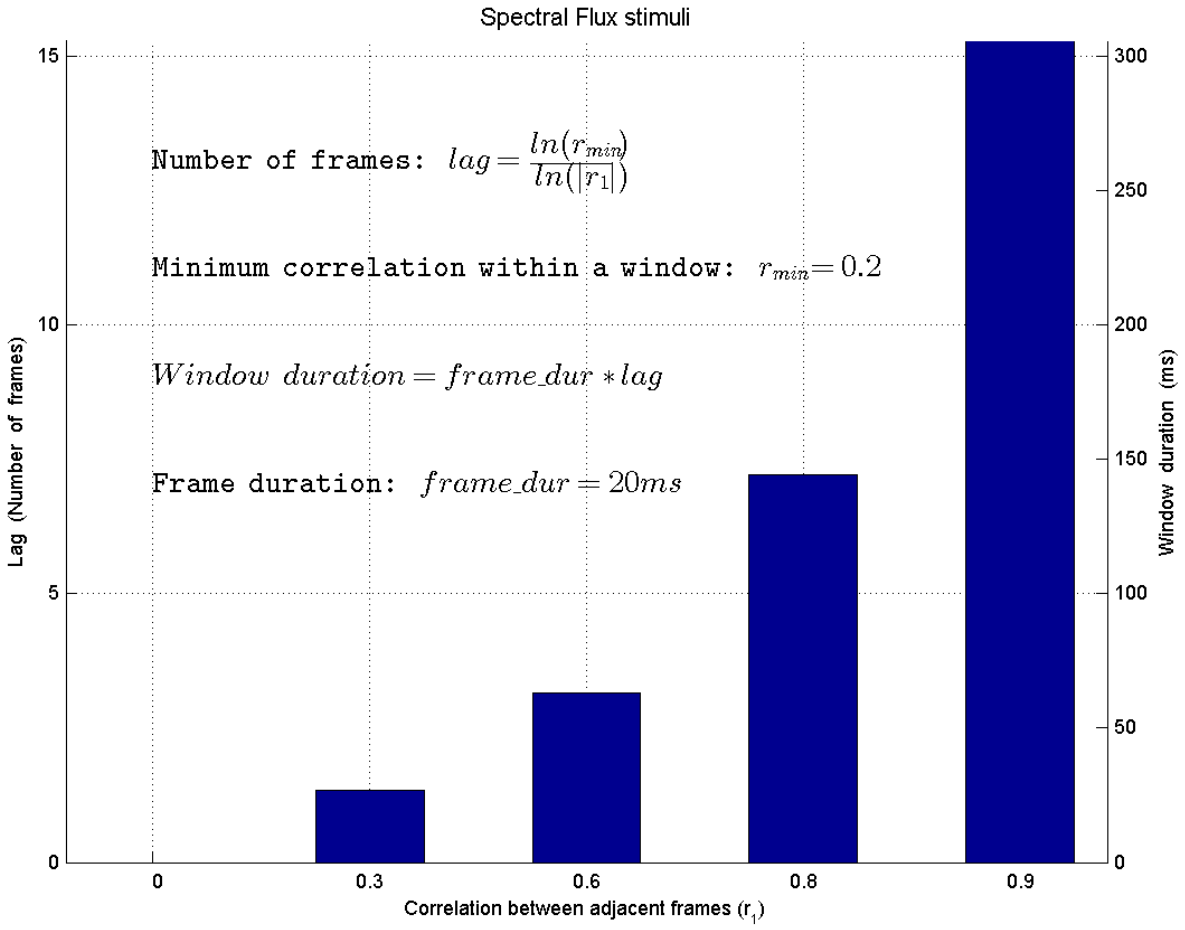


Figure 4-1 Bar plot showing the relation between correlation r_1 and the number of frames in a window (lag) with its associated duration in ms, the values of the parameters are r_{min} = 0.2, frame duration = 20 ms.

Stimulus	**I**	**II**	**III**	**IV**	**V**
Spectro-temporal Correlation (r_1)	0.0	0.3	0.6	0.8	0.9
Number of frames (lag)	1	1.35	3.15	7.2	15.3
Window duration (ms)	20	27	63	144	306
Spectral Flux (qualitative)	High		Medium		Low

Table 4-1 List of correlation values (r_1) used in the experiment, and the corresponding number of frames in a window within which any two frames must show a minimum degree of correlation (r_{min}) and associated time window duration in ms, the values of the parameters are r_{min} = 0.2, frame duration = 20 ms.

4.3.3 Spectral Flux Stimuli

Sound stimuli were created using scripts written in MATLAB (MathWorks, Natick, USA) version 7.1 at a sample rate of 44.1 kHz and 16 bit resolution. The amplitude spectrum was defined in terms of frames of 20 ms duration. Each synthetic stimulus was synthesised using 20 sinusoids (i.e. n = 20) chosen randomly from a pool of 101 logarithmically spaced frequencies between 246 and 4435 Hz. This frequency range was defined (Overath et al., 2008) to encompass the critical range of the human audiogram. The most sensitive part of the macaque audiogram is similar to the one found in humans (Jackson et al., 1999). Further, linear spline interpolation was applied to amplitude transitions between frames to avoid sudden amplitude jumps. The rise time and fall time for each sound stimulus were set at 20 ms. The mean and the standard deviation of the amplitude spectra were set at 65 dB-rms and 10 dB-rms respectively. These values were identical for each of the frequency components and for all correlation levels.

The parameters in the study by Overath et al. (2008) were chosen to encompass the timescales required for processing of phonemes (20 ms) and syllables (300 ms) (Rosen, 1992). This choice ensured that one could draw inference on the mechanisms employed for the analysis of sounds similar in complexity to speech, though not specific to it. Since I am interested in comparing the findings from macaques with the earlier human study in Overath et al. (2008), this motivated me to choose the same values for the parameters. Hence, r_{min} was set at 0.2, frame_duration was set at 20 ms. However, in my study, the amplitude spectra of given stimuli were allowed to vary with one of the five correlations, resulting in the different sampling of the spectral flux. The correlation r_1 for each stimulus was fixed as one of these five different positive values listed here: 0.0, 0.3, 0.6, 0.8, and 0.9. This was done to increase the number of times volumes were acquired for silence condition and hence a better sound versus silence contrast. Though this sampling of the correlations is different from the values chosen (0.0, 0.2, 0.4, 0.6, 0.8, 0.9) in the human study (Overath et al., 2008), both set of values sample the entire valid range of correlations possible and thus does not affect the inter-species comparison.

Table 4-1 provides the range of values of r_1 used to generate the acoustic stimuli and its corresponding window duration. Figure 4-2 provides a visual representation of the spectrotemporal decomposition of exemplars of the various spectral flux stimuli employed in this study.

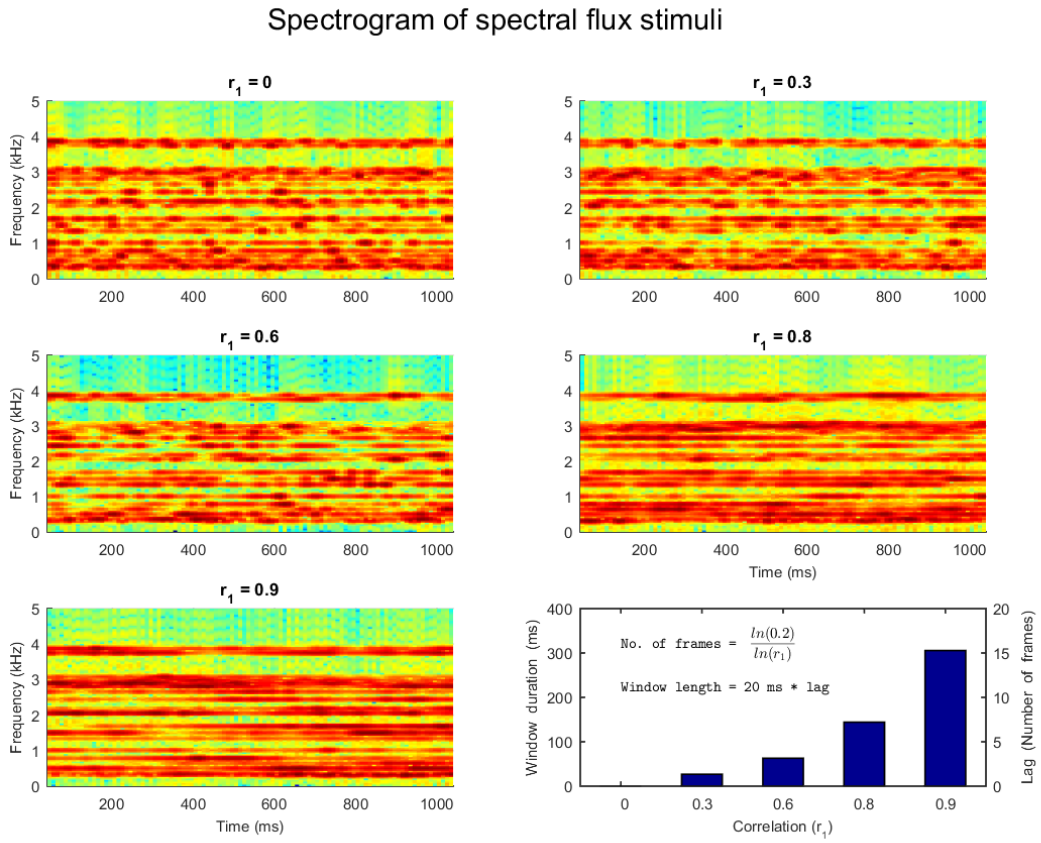


Figure 4-2 Spectrogram of exemplar stimulus from each of five different spectral flux levels employed in this study, showing the degrees of energy fluctuation from a high rate at r_1 = 0.0 to a low rate at r_1 = 0.9. The relationship between the spectrotemporal correlation r_1 and the duration of the window to achieve a minimum correlation between any two frames within it is shown in the bottom right plot (inset formulae).

4.3.4 Tonotopy Stimuli

For characterising tonotopy using the BOLD response to spectral frequencies sound stimuli were based on random-phase noise carrier with seven different pass-bands, 0.125-0.25 kHz, 0.25-0.5 kHz, 0.5-1 kHz, 1-2 kHz, 2-4 kHz, 4-8 kHz, and 8-16 kHz resulting in seven different stimuli that encompassed different spectral ranges. The carriers were amplitude modulated with a sinusoidal envelope of 90% depth at 10 Hz to achieve a robust response in the auditory system.

4.3.5 Stimulus Presentation

To record data from the auditory system that is devoid of activity due to the high-intensity noise generated by the MRI scanner, a 'sparse temporal' design is utilized. With the use of a pseudo-random sequence, each adjacent trial was ensured to have a different spectral flux sound stimulus. The duration of each sound stimulus was 6 seconds. This duration is sufficient for the BOLD response in the macaque auditory cortex to reach a plateau (Baumann et al., 2010).

The onset and offset of the stimuli were smoothed by a linear ramp of 50 ms. The sound stimuli were presented to the monkey at an RMS sound pressure level (SPL) of 75 dB using 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).

4.3.6 Task during imaging

The monkey performed visual fixation on a cue during the entire time the sound stimulus was presented. This simple task has some advantages to it, both to the quality of data acquired and to the wellbeing of the animal. First, it ensured that the levels of attention remained consistent across the entire session. Next, it minimized the body movement of the animal by alleviating any stress it may have. The eye position was monitored at 60 Hz with a tracking (camera-based with Infra-Red illumination) of the pupil using iView software (SMI, www.smivision.com, Teltow, Germany). The position, X and Y coordinates, of the pupil was communicated to the Cortex software. The task was to fixate on a target (small red square) positioned at the centre of a screen, when the eye trace entered within a window of fixation (~ five

degrees centred on the target) a timer started and the fixation target turned green. A continuous visual fixation (no saccades) of a randomly defined duration of 2-2.5 s was rewarded immediately by the delivery of a juice via a gravity-fed dispenser.

4.3.7 Data Acquisition

MRI was conducted in an actively shielded 4.7 Tesla vertical scanner (Bruker Biospec 47/60 VAS) dedicated to imaging in NHPs. It has an inner-bore width of 38 cm and a Bruker GA-38S gradient system from Bruker Medical, Ettlingen, Germany. Shimming was performed with the FASTMAP algorithm (Gruetter, 1993) which measures B0 field inhomogeneity to apply first and second order corrections to it.

Data were acquired with parallel imaging with 2-fold GRAPPA acceleration using custom designed (www.wkscientific.com) 4-channel array receive coil. The RF transmission was achieved using a custom designed saddle coil (from the same company) in transmit mode. Both structural and functional data covered the temporal lobe and aligned to the superior temporal plane (STP). A navigator scan helped with the slice selection.

Functional MRI measurements by BOLD contrast consisted of single-shot gradient-recalled echo-planar imaging (GR-EPI) sequences with an in-plane resolution of 1.2 x 1.2 mm^2 and slice thickness of 1.2 mm, yielding 1.72 mm^3 voxels and a volume acquisition time (TA) of 1.35 s. Typical acquisition parameters were as follows – time to echo (TE) of 21 ms, flip angle (FA) of 90º, receiver spectral bandwidth of 200 kHz, the field of view (FOV) of 9.6 x 9.6 cm^2, with an acquisition matrix of 96 x 96. A sparse design was employed where the acquisition of each volume was separated by a 10 s repetition time (TR) gap. This TR duration was necessary and sufficient to avoid recording the BOLD response to the gradient noise of the previous scan.

The stimuli were presented during the last six of 10 s inter-trial interval. The timing was based on previous characterisation of BOLD response time course in the auditory system of macaques (Baumann et al., 2010). For every five volume acquisitions, three volumes were acquired where no stimulus was presented to obtain data for a silent baseline. In each session of one-hour duration, 360 volumes were acquired resulting in 225 volumes for all stimuli or 45 volumes per each of 5 stimulus levels while 135 volumes correspond to silence. Data from monkey M4 was collected over five sessions (thus 225 volumes were obtained for each stimulus

level), data from monkey M5 was collected over four sessions (thus 180 volumes were obtained for each stimulus level), while data from monkey M6 was collected over six sessions (thus 270 volumes were obtained for each stimulus level), and data from monkey M1 was collected over five sessions (thus 225 volumes were obtained for each stimulus level).

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 = 2000 ms, TE = 3.74 ms, TI = 750 ms, 30° flip angle, receiver bandwidth = 50 KHz, an in-plane resolution of 0.67 x 0.67 mm^2 with a slice thickness of 0.6 mm. These structural scans cover the same field of view as the functional scans.

4.3.8 Data Analysis

MR images were first converted from scanner's native file format into a common MINC file format, 3D for the anatomical data and 4D (x, y, z, t) for the functional data, using the Perl script pvconv.pl available online (http://pvconv.sourceforge.net/). From MINC format, it was converted to NIfTI file format standard using MINC tools. This raw fMRI data were processed using Statistical Parametric Mapping (SPM12) software (www.fil.ion.ucl.ac.uk/spm), using MATLAB 7.1 software.

In the pre-processing steps, first, rigid body motion compensation was performed. Next, image volumes from multiple sessions were combined by realigning all volumes to the first volume of the first session. Then, this data was spatially smoothened using a Gaussian kernel with full-width-at-half-maximum (FWHM) of 3 mm. A standard SPM regression model was used to partition components of the BOLD response at each voxel. The five conditions, each of five different spectrotemporal correlation values were modelled as effects of interest and their stimulus onsets were convolved with a canonical hemodynamic response function. Next, the time series was high pass filtered with a cut-off of 120 s to remove low-frequency variations in the BOLD signal that is caused mainly due to scanner drift. Finally, this data was adjusted for global signal fluctuations also known as global scaling to account for differences in system responses across multiple sessions.

In a general linear model (GLM) analysis of the combined sessions that included the motion parameters, the voxel-wise response estimates the regression coefficients (denoted beta). The t-values for the contrast of the different stimuli versus the silent baseline were also calculated. The response to silent baseline was not explicitly modelled in the GLM and hence 'sound minus silent baseline' contrast looked for values of beta weights that were greater than zero. The data were masked retaining voxels with significant values for the combined stimuli versus silent baseline ($p<0.001$, uncorrected for multiple comparisons across the auditory cortex).

4.3.9 Best frequency tonotopy map

Data for the tonotopy experiment was acquired from the monkeys after data for the main spectral flux experiment was acquired. Tonotopy data using 3 frequency bands (0.5-1 kHz, 2-4 kHz, 8-16 kHz) was collected from monkey M4 over two sessions (180 volumes per frequency band in total), and from monkey M5 over one session (150 volumes per frequency band in total). No tonotopy data was collected in monkey M6. Tonotopy data using all seven frequency bands were collected from monkey M1 over seven sessions (290 volumes per frequency band in total). The higher number of bands and sessions in monkey M1 was motivated by a different project but the data was used here.

Map of preferred response to different frequency bands is known as 'best-frequency map'. This map is calculated by identifying voxel by voxel, in each animal across all voxels whose sound versus silence contrast was significant ($T>3.1$, $p<0.001$ uncorrected for multiple comparisons across the auditory cortex), which of the frequency conditions showed the highest beta i.e. regression coefficient. The resulting map represents the preferred frequency for each voxel.

4.3.10 Parcellation

To map the auditory subfields, information from tonotopy fMRI data, macro-anatomical features (cortical folding), anatomical MRI were combined. The ratio (Joly et al., 2014a) of T1w and T2w images provided an index that represented average intensities across the cortical thickness. Highest values of T1/T2 ratio indicated grey matter voxels and were used to identify the location of A1 and R fields. The boundary between A1 and R was identified via the frequency reversal occurring between these regions in the best frequency map of the tonotopy experiment since the posterior end of A1 and anterior end of R prefers high frequency while the anterior end of A1 and

the posterior end of R i.e. boundary prefers low frequency. To overcome the similarity of frequency preference between core and belt regions and the difficulty in parcellation of medial belt regions, the T1/T2 ratio is utilized to demarcate between core and belt since this ratio is high in the core regions but lower in the belt regions.

The exact method and tools used in parcellation are described here. The subject-specific parcellation of the auditory cortical subfields follows the scheme reported in Reveley et al. (2017). The original atlas was used as provided in the registered format with the population MRI primate brain template published in Seidlitz et al. (2018) and available at https://github.com/jms290/NMT. For each monkey, information from the tonotopic mapping from bold-weighted functional MRI data, macro-anatomical features (cortical folding of the lateral sulcus), anatomical MRI were combined. The lateral fissure was used to run a (local) surface-based co-registration from the NMT template to the subject-native space in order to initialize the registration then non-linear registration was further computed with alignment of the antero-posterior border between A1 and R to the first reversal from High-Low-High frequency reversal from the tonotopic mapping (Joly et al., 2014a) using 3D Slicer (ITK based registration framework, www.slicer.org). Finally, the final local lateral adjustment of the full parcellation was applied to overlap the x-coordinate of the centre of the core regions (especially A1/R) to the peak location (within the grey matter) of the T1w-bias corrected map (Joly et al., 2014a, Geyer et al., 2011, Glasser and Van Essen, 2011).

Thus, the following fields were identified in each hemisphere in each monkey M4, M5 and M1 viz. A1, AL, CL, CM, CPB, ML, R, RM, RPB, RT, RTL, RTM, RTp, STGr, and Tpt. I could not collect tonotopy data in monkey M6 and parcellation is based solely on macro anatomical features (cortical folding of the lateral sulcus) identified combined with the anatomical MRI of the animal.

4.3.11 Window duration preference

To reveal the spatial organization of window duration preference, a contrast map was generated by projecting the functional data of the acquired volumes onto the anatomical scans. Next, the response strength of the shorter time windows (or lower spectrotemporal correlation) was contrasted with the longer time windows (or higher spectrotemporal correlation). This contrast map was calculated voxel by voxel by summing the differentially weighted regression coefficients (beta) of the various

spectrotemporal correlations. The contrast maps obtained using the following weights (2, 1, 0, -1, -2) are henceforth referred to as 'linear negative parametric' contrast while those obtained using the following weights (-2, -1, 0, 1, 2) are henceforth referred to as 'linear positive parametric' contrast. The negative parametric contrast represents the degree of preference for smaller over longer time window duration (or alternatively low over high spectrotemporal correlation levels) while the positive parametric contrast represents the degree of preference for longer over shorter time window duration (or alternatively high over low spectrotemporal correlation levels).

4.3.12 Preferred window of temporal integration

The preferred window of temporal integration for a given cortical area was estimated by the slope of linear regression of the BOLD signal for different spectrotemporal correlation to the underlying time window duration. This linear regression was performed using lm() function in R software. The fitted linear functions were of the form:

$$\beta(r_1) = m * w_{min} + \mathrm{c}$$

where β is the regression coefficient averaged across those voxels within an ROI whose sound versus silence contrast is significant (T>3.1, p<0.001 uncorrected for multiple comparisons across the auditory cortex), r_1 is the spectrotemporal correlation, w_{min} is the duration of time window showing a minimum correlation r_{min}, m is the slope of the linear regression and c is the y-intercept of the fit.

4.4 Results

4.4.1 Tonotopy

In the tonotopy experiment, the BOLD response was recorded across the entire auditory cortex to sound stimuli with bandpass noise with different pass-bands. These tonotopy stimuli were presented to three macaques (M4, M5, and M1) undergoing fMRI. Tonotopy data was not collected in monkey M6. The BOLD activation associated with sound stimulation was analysed in voxel space. Sound related activation ($p < 0.001$ uncorrected for multiple comparisons across the auditory cortex) was observed in the superior temporal plane that had a symmetrical pattern across the hemispheres. Figure 4-3 panel A shows the areas activated to sound stimulation using colour coded regions in monkey M4, panel B for monkey M5, and panel C for monkey M1. Here the different colours of the 'best frequency map' represent the frequency preference of the voxels which are responsive to sound i.e. sound versus silence contrast was statistically significant ($T>3.1$, $p<0.001$ uncorrected for multiple comparisons across the auditory cortex).

Best-frequency maps showed well-established mirror symmetric high-low-high frequency gradients across the auditory core and belt regions bilaterally (Kosaki et al., 1997, Merzenich and Brugge, 1973, Morel et al., 1993, Rauschecker et al., 1997, Bendor and Wang, 2008, Baumann et al., 2015, Baumann et al., 2013, Joly et al., 2014a). Parcellation of the auditory cortex in macaques into various regions of interest (ROI) was achieved using a combination of best-frequency maps from tonotopy experiments and high-resolution T1 and T2 images (Joly et al., 2014a).

4.4.2 Activation to sound

In the main experiment on spectral flux, the BOLD response was recorded across the entire auditory cortex to sound stimuli with five different spectrotemporal correlations. These stimuli corresponding to varying degrees of spectral flux were presented to four macaques undergoing fMRI. The BOLD activation associated with sound stimulation was analysed in voxel space. Sound related activation ($p < 0.001$ uncorrected for multiple comparisons across the auditory cortex) was observed in the superior temporal plane that had a symmetrical pattern across the hemispheres. Figure 4-4 shows the areas activated to sound stimulation using reddish-yellow hue in monkey M4, M5 and M6 while Figure 4-5 shows in monkey M1. This data indicates that the synthetic spectral flux stimulus robustly activates auditory cortex bilaterally.

4.4.3 Window duration preference

The contrast maps for monkey M4, M5, M6 and monkey M1 are shown in Figure 4-4 and Figure 4-5 respectively as bluish-green hue (linear negative parametric contrast) or greenish hue (linear positive parametric contrast) overlaid on auditory activation in reddish-yellow hue. The monkeys M4, M5, and M6 show a negative parametric effect in the auditory cortex bilaterally though a positive parametric effect is seen only in monkey M1 anterior STG. So BOLD increases with decreasing time windows or BOLD is highest for shorter time windows in the auditory cortex of three monkeys. However, this relationship seen in macaques is opposite to that seen in humans, where BOLD increased with increasing time windows or BOLD was highest for longer time windows (Overath et al., 2008).

4.4.4 ROI based analysis

Using MarsBaR toolbox (version 0.44) (Brett et al., 2002), the sound versus silence contrast and linear negative parametric contrast within each ROI (estimated earlier) was averaged across all voxels whose sound versus silence contrast is significant ($T>3.1$, $p<0.001$ uncorrected for multiple comparisons across the auditory cortex). Figure 4-6 visualizes this data as a function of ROIs in three monkeys. The number of voxels that survived statistical thresholding in each ROI of four animals is listed in Table 4-2. Table 4-3 and Table 4-4 provides the beta and significance values for sound versus silent baseline contrast and linear negative parametric contrast across various ROIs in monkeys M4, M5, and M6 respectively while Table 4-5 provides the same in monkey M1. Because the number of comparisons across ROIs (30 per animal) was much less compared to the number of voxels, this justified the use of appropriate ROI-level statistical threshold at $p<0.05$, corrected for multiple comparisons across the auditory ROIs in a given animal.

Across most ROIs, a linear negative parametric effect of BOLD contrast with time window duration was seen including core, belt, and parabelt regions bilaterally in monkey M4, M5, and M6 (3 animals, 6 hemispheres) except in monkey M1 where a positive parametric effect of BOLD contrast with time window duration was seen in right lateralized anterior STG but not in core, belt, parabelt regions. In core fields, a statistically significant negative parametric effect of the BOLD signal with time window was observed in A1: 4 hemispheres bilaterally (2 animals) and R: 5 hemispheres bilaterally (3 animals). In belt regions, a statistically significant negative parametric effect was observed in AL: 3 hemispheres bilaterally (2 animals), CM: 3

hemispheres bilaterally (3 animals), RM: 3 hemispheres bilaterally (2 animals) and left ML of monkey M4. However, in parabelt regions (RPB and CPB), this negative parametric effect was not significant. There was no significant parametric effect seen in STS in any of the four animals.

4.4.5 Preferred window of temporal integration

Table 4-6 summarizes the slope of the linear regression against time window duration and its corresponding significance level from both hemispheres in monkey M4, M5, and M6 respectively across those ROIs where there are voxels whose sound versus silence contrast is significant ($T>3.1$, $p<0.001$ uncorrected for multiple comparisons across the auditory cortex). From the curve fit of BOLD contrast with time window duration using linear regression in each ROI where sound vs silence contrast was significant, I analysed the slope of the fitted line that conveys the degree of the relative preference towards different time window durations. Across three monkeys M4, M5, M6 (a total of 6 hemispheres), a negative slope was noticed in most auditory regions bilaterally. This implied that most cortical areas relatively prefer a shorter window over a longer window. The slope averaged across the hemispheres of three animals was as follows - in the core regions (A1, R): -5.79e-3; in the belt regions (CM, AL, ML, RTL, RTM, RM): -3.72e-3; in parabelt (RPB, CPB): -1.3e-3. Thus, the slope was steepest in the core regions. Next, this negative slope reduced, despite staying negative, as one moved from core to belt regions. Finally, the slope was closest to being flat in parabelt regions. This implied that the duration of the preferred window of temporal integration in a given cortical region widened as one progressed from core to belt and parabelt regions.

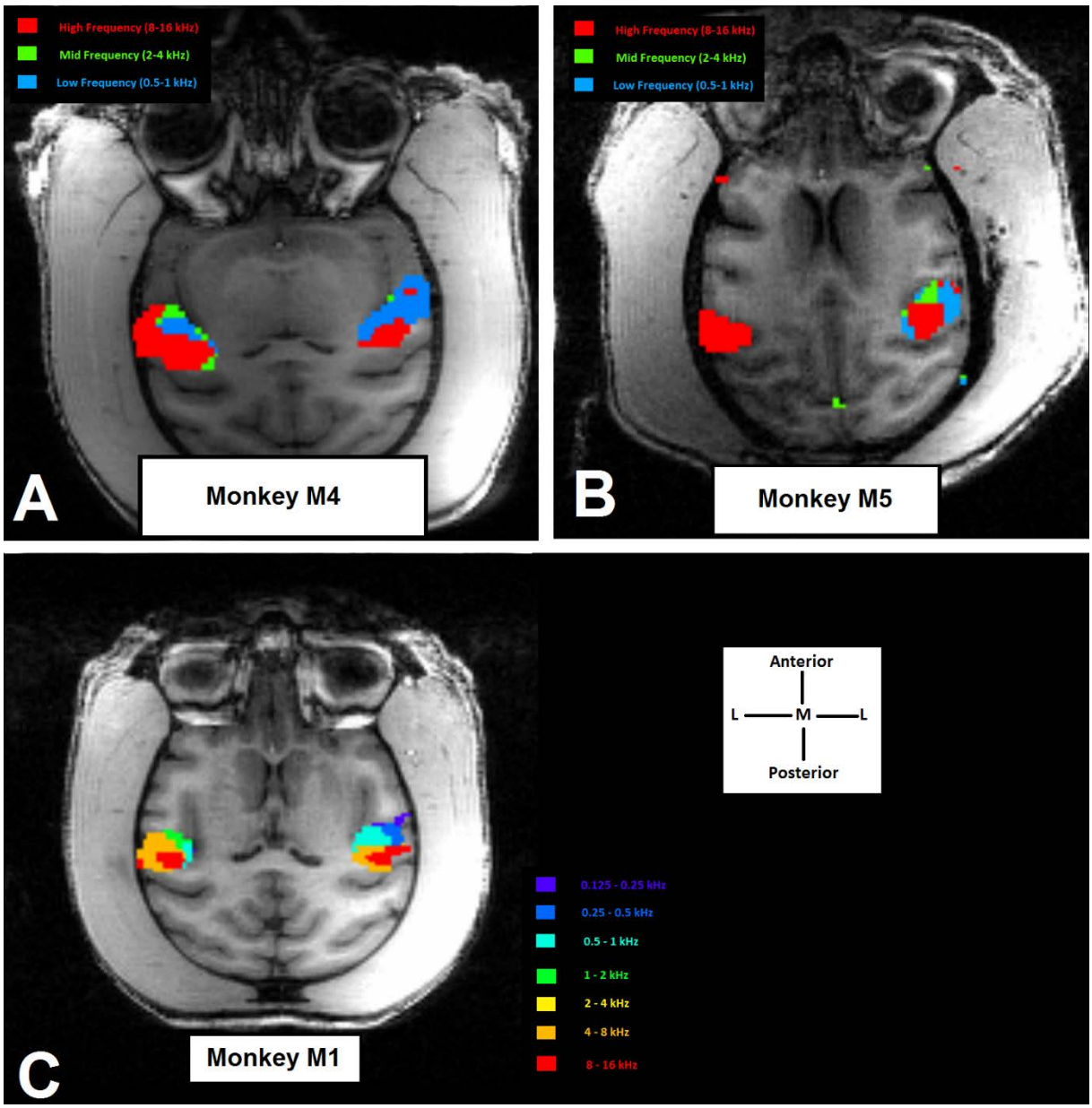


Figure 4-3 Best frequency tonotopy map in three monkeys. This best frequency map is rendered on top of an axial section of the T1 structural scan. The map is thresholded on sound versus silence contrast T>3.1, p<0.001 uncorrected for multiple comparisons across the auditory cortex. The frequency reversals present in this tonotopy maps along with myelination maps are used in the parcellation of the auditory cortex into auditory core, belt and parabelt areas. (**A**) Monkey M4's best frequency map colour coded as blue for 0.5-1 kHz, green for 2-4 kHz, red for 8-16 kHz. (**B**) Monkey M5's best frequency map colour coded as blue for 0.5-1 kHz, green for 2-4 kHz, red for 8-16 kHz. (**C**) Monkey M1's best frequency map colour coded as violet for 0.125-0.25 kHz, blue for 0.25-0.5 kHz, cyan for 0.5-1 kHz, green for 1-2 kHz, yellow for 2-4 kHz, orange for 4-8 kHz, red for 8-16 kHz. Ant – Anterior, M – Medial, L – Lateral, Pos – Posterior.

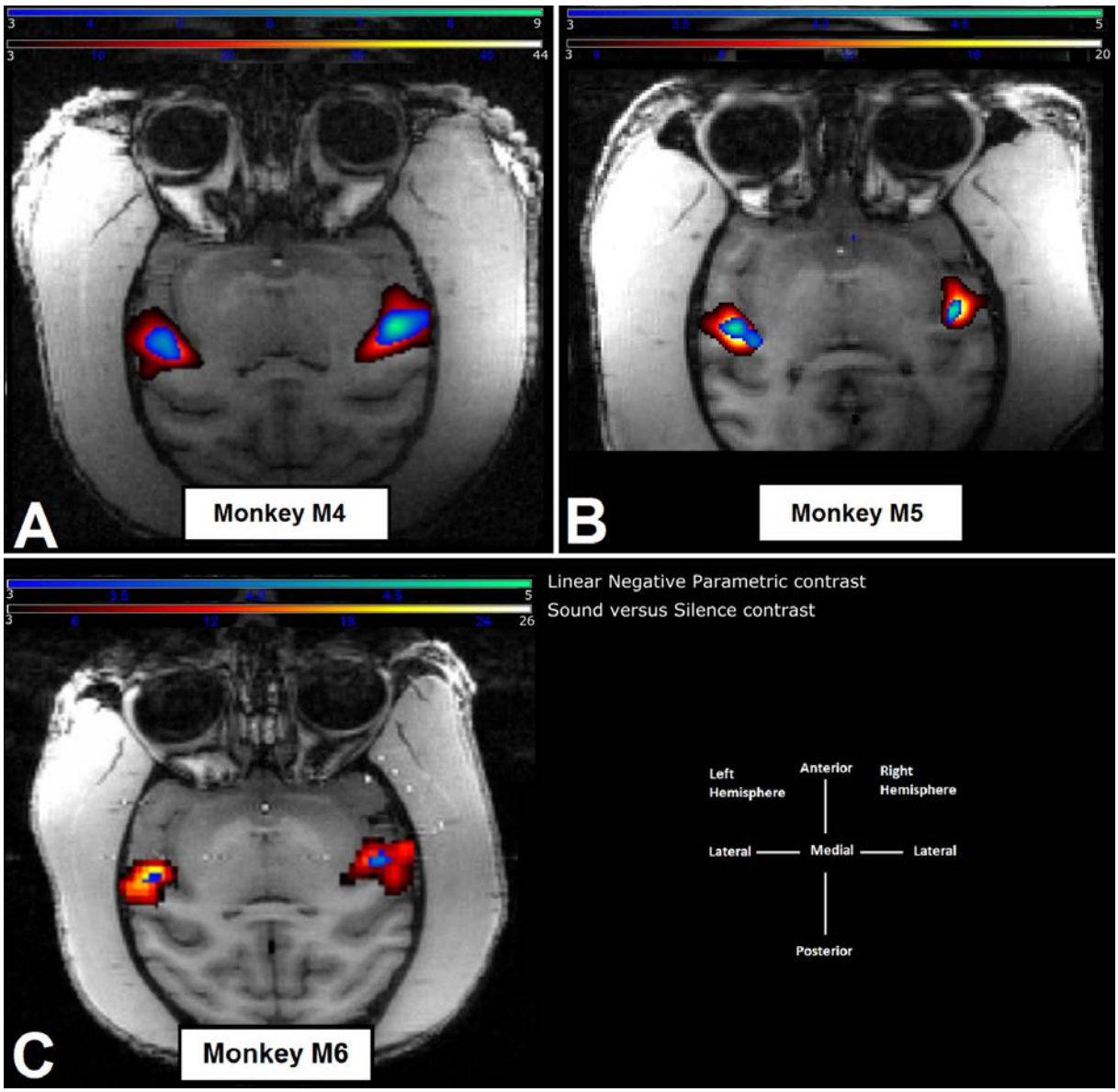


Figure 4-4 Contrast for the negative parametric effect of time window duration and sound versus silence contrast in three monkeys. Linear negative parametric contrast (bluish-green hue) is overlaid on sound minus silent baseline contrast (reddish-yellow hue). Both these contrast maps are rendered on top of an axial section passing through auditory cortex (T1 structural scan) in the Superior Temporal Plane in (**A**) Monkey M4, (**B**) Monkey M5 and (**C**) Monkey M6. The thresholds on statistical maps were kept at $T > 3.1$ or $p<0.001$ uncorrected for multiple comparisons across the auditory cortex. Sound versus silence contrast shows that this synthetic stimulus employed in this study robustly activates most auditory cortical areas bilaterally. Negative parametric contrast (implies BOLD decreases with increasing time windows) is seen in the auditory core and medial belt regions bilaterally.

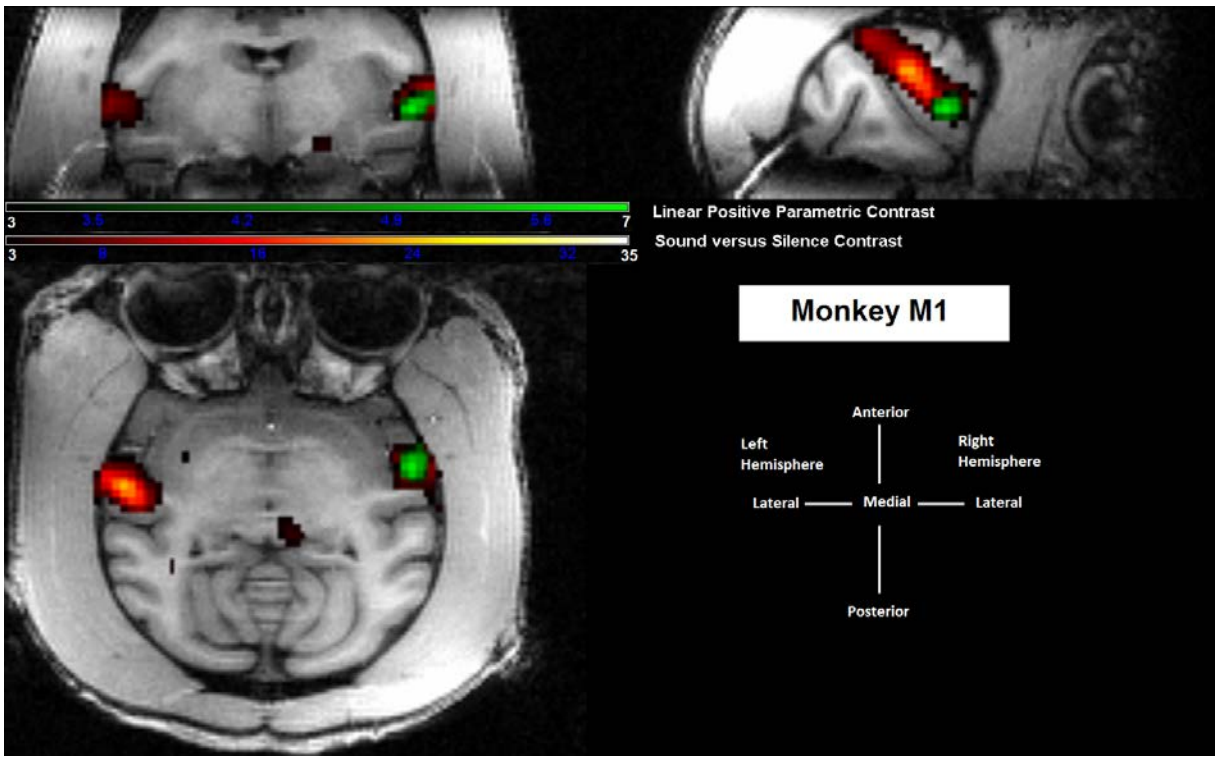


Figure 4-5 Linear positive parametric contrast of BOLD with time window duration overlaid on sound versus silence contrast in monkey M1. Linear positive parametric contrast (green hue) is overlaid on sound minus silent baseline contrast (reddish-yellow hue). Both these contrast maps are rendered on top of an axial section (T1 structural scan). The thresholds on statistical maps were kept at $T>3.1$ or $p<0.001$ uncorrected for multiple comparisons across the auditory cortex. The linear positive parametric contrast is seen in anterior STG but not seen in the auditory cortex.

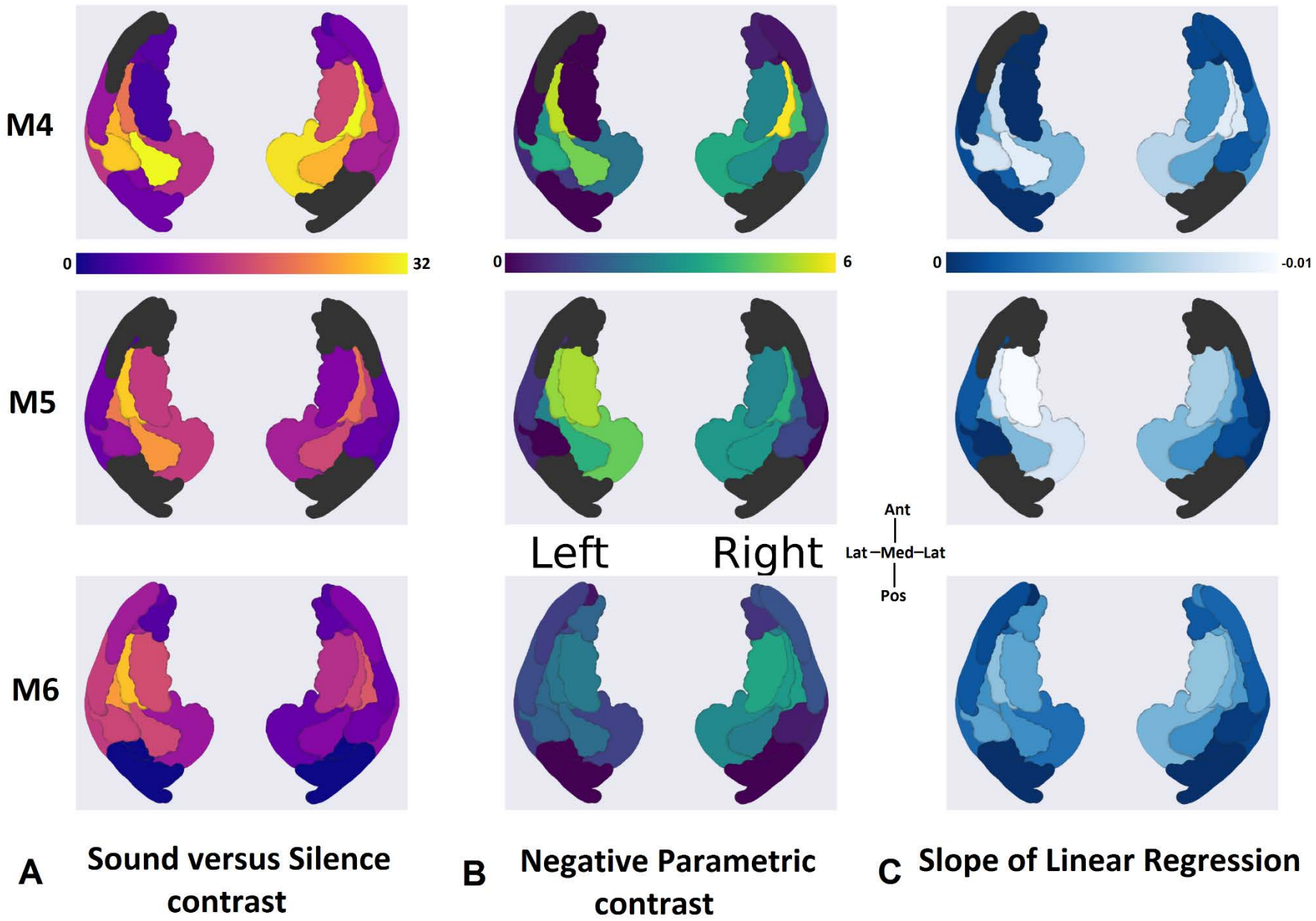

M4
M5
M6
0
32
0
6
0
-0.01
Left
Right
Ant
Lat–Med–Lat
Pos
A
Sound versus Silence contrast
B
Negative Parametric contrast
C
Slope of Linear Regression

Figure 4-6 Visual representation of sound versus silence contrast, negative parametric contrast and slope of linear regression across various ROIs of three macaques. The auditory region of interest are colour-coded individually for sound minus silent baseline (reddish yellow) and linear negative parametric contrast (green) and the slope of the linear regression (blue) in each hemisphere of monkey M4, M5, and M6. (**A**) Sound versus silence contrast panel shows that this synthetic spectral flux stimulus employed in this study robustly activates most auditory cortical areas bilaterally. (**B**) Negative parametric contrast panel shows that BOLD decreases with increasing time window duration in the auditory core and medial belt regions bilaterally. (**C**) The slope of the linear regression panel shows that it is negative in most regions bilaterally implying preference to short time windows. Next, the slope is highest in the auditory core and medial belt regions bilaterally signifying these areas most strongly prefers short windows. Further, the slope reduces (less shallow despite staying negative) in lateral belt regions signifying a relative shift in preference towards longer time windows. Finally, the slope is closest to being flat in parabelt regions bilaterally signifying a no differential preference to time window duration.

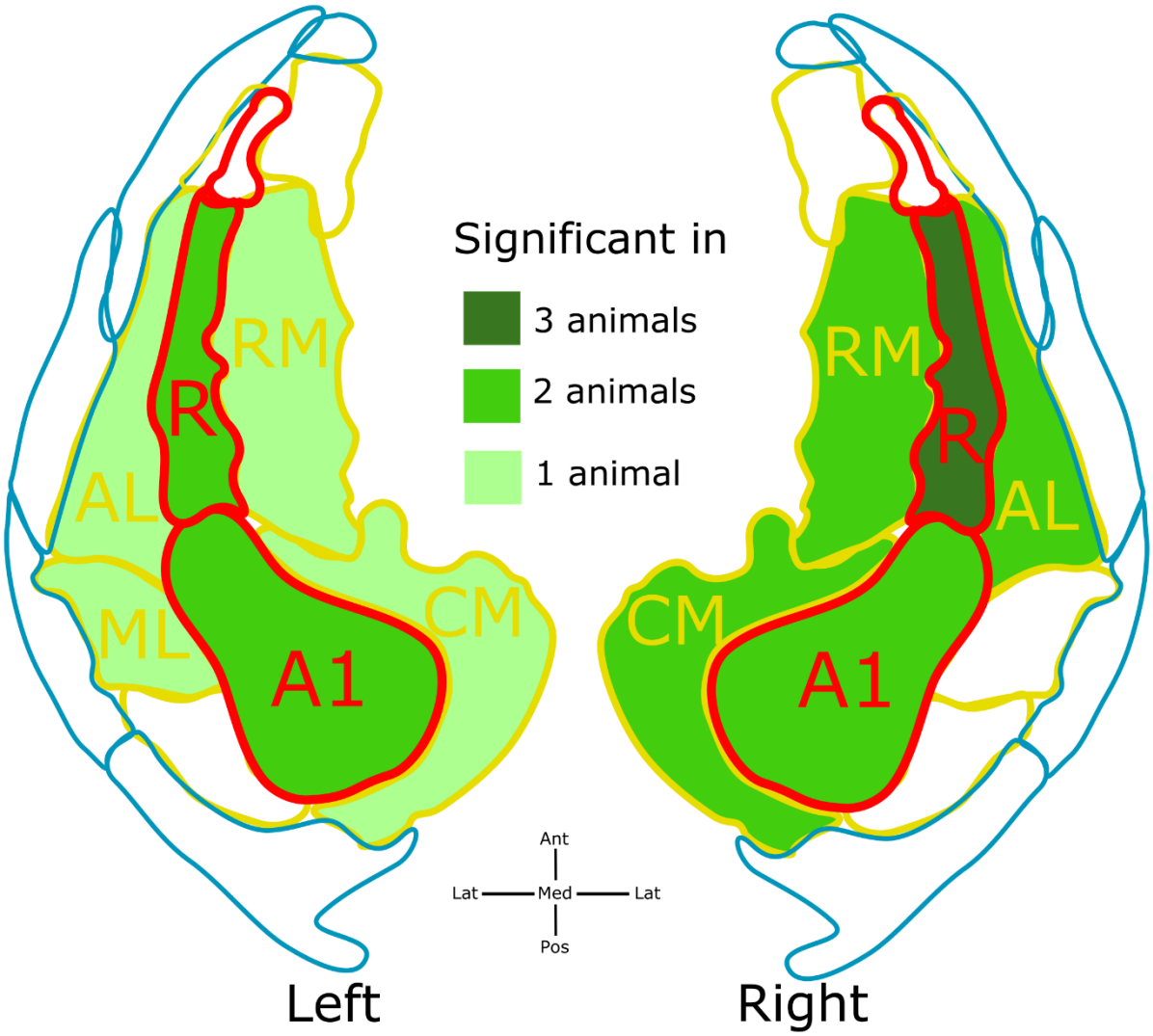


Figure 4-7 Visual summary of the statistical significance of linear negative parametric contrast results from three monkeys. Summary of statistics on linear negative parametric contrast evaluated across three monkeys (monkey M4, M5 and M6) in each ROI across the auditory cortex in left and right hemisphere. ROIs are colour coded for the number of animals in which the negative parametric contrast is significant. The negative parametric contrast signifies that BOLD signal is highest for short time windows in the auditory core and medial belt regions.

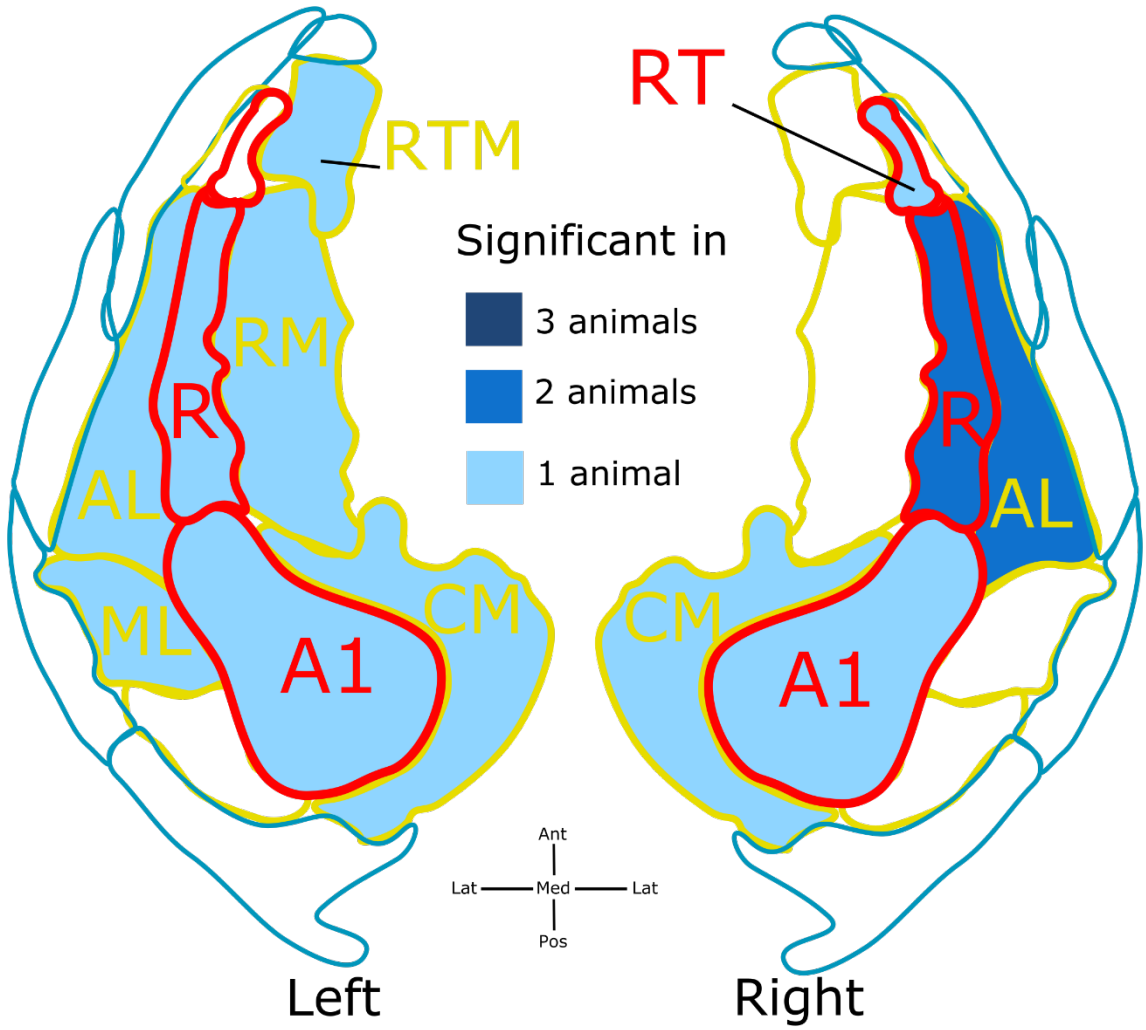


Figure 4-8 Visual summary of the statistical significance of the slope of the linear regression results from three monkeys. Summary for the slope of the linear regression evaluated across 3 monkeys (monkey M4, M5 and M6) in each ROI across the auditory cortex in left and right hemisphere. ROIs are colour coded for the number of animals in which the result is significant. The slope is not significant in parabelt regions signifying a lack of differential sensitivity for time window duration opposite to that seen in humans that preferred long time windows.

No of voxels	Monkey M4		Monkey M5		Monkey M6		Monkey M1	
	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi	Left Hemi	Right Hemi
A1	42	24	21	10	53	26	145	180
AL	51	50	40	20	135	124	119	155
CL	8	0	0	0	0	0	4	15
CM	14	24	8	8	41	13	43	14
CPB	53	18	16	1	121	76	142	152
ML	24	8	12	3	29	11	38	38
R	23	29	23	24	25	22	21	15
RM	1	27	6	7	71	59	51	5
RPB	28	22	14	13	28	3	69	20
RT	7	6	0	0	23	22	10	8
RTL	5	17	1	0	15	12	1	1
RTM	1	7	0	0	1	2	9	3
RTp	0	2	0	0	2	1	2	0
STGr	0	17	0	0	69	30	1	0
Tpt	4	0	0	0	0	0	2	26

Table 4-2 Number of voxels in each ROI of monkeys that survive statistical threshold at a single voxel level of $p<0.001$, uncorrected for multiple comparisons across the auditory cortex

ROI	Monkey M4				Monkey M5				Monkey M6			
	Left hemi		Right hemi		Left hemi		Right hemi		Left hemi		Right hemi	
	beta	p	beta	p	beta	p	beta	p	beta	p	beta	p
A1	50	<1e-7	29.62	<1e-7	24.56	<1e-7	12.12	<1e-7	17.18	<1e-7	8.95	<1e-7
AL	30.85	<1e-7	35.65	<1e-7	24.11	<1e-7	11.89	<1e-7	29.13	<1e-7	15.47	<1e-7
CL	12.27	<1e-7	-	-	-	-	-	-	-	-	-	-
CM	22.95	<1e-7	41.82	<1e-7	19.2	<1e-7	13.5	<1e-7	15.16	<1e-7	7.71	8e-6
CPB	18.09	<1e-7	14.69	<1e-7	8	<1e-7	5.48	5e-4	19.13	<1e-7	9.08	<1e-7
ML	47.53	<1e-7	17.98	<1e-7	12.1	<1e-7	3.6	2e-6	26.54	<1e-7	7.9	<1e-7
R	22.53	<1e-7	42.39	<1e-7	35.42	<1e-7	18.85	<1e-7	46.84	<1e-7	17.78	<1e-7
RM	4.8	0.024	17.31	<1e-7	21.7	<1e-7	12.98	<1e-7	22.27	<1e-7	15.17	<1e-7
RPB	12.8	<1e-7	13.22	<1e-7	9.01	<1e-7	6.64	5e-6	10.71	<1e-7	4.58	2e-6
RT	9.03	6e-6	16.01	<1e-7	-	-	-	-	16.08	<1e-7	11.3	<1e-7
RTL	9.14	7e-6	16.4	<1e-7	6.78	1e-3	-	-	16.4	<1e-7	9.84	<1e-7
RTM	6.67	1e-3	10.92	<1e-7	-	-	-	-	7.97	1e-3	9.32	6e-5
RTp	-	-	10.25	3e-3	-	-	-	-	12.3	3e-6	7.34	2e-3
STGr	-	-	8.84	<1e-7	-	-	-	-	10.28	<1e-7	7.92	<1e-7
Tpt	8.49	<1e-7	-	-	-	-	-	-	-	-	-	-

Table 4-3 Sound minus silent baseline contrast details from various fields in the auditory cortex of three monkeys. Beta and significance value is given for ‘sound versus silence’ contrast across all the ROIs in monkeys M4, M5, and M6. The beta is averaged across those voxels whose t-statistic for ‘sound versus silence’ meets $p<0.001$ uncorrected for multiple comparisons across the auditory cortex. In some ROIs, where the values are omitted, no voxels within this ROI survived this threshold at a single voxel level. The significance value (p-value) is corrected for multiple comparisons across ROIs within each monkey. This data indicates that this synthetic stimulus employed in this study robustly activates most auditory cortical areas bilaterally.

ROI	Monkey M4				Monkey M5				Monkey M6			
	Left hemi		Right hemi		Left hemi		Right hemi		Left hemi		Right hemi	
	beta	p	beta	p	beta	p	beta	p	beta	p	beta	p
A1	6.81	3e-5	3.47	0.04	4.15	1e-3	2.53	0.01	2.41	n.s.	2.73	n.s.
AL	3.63	0.03	6.52	3e-4	3.15	n.s.	1.58	n.s.	2.52	n.s.	3.02	0.01
CL	1.58	n.s.	-	-	-	-	-	-	-	-	-	-
CM	4.08	n.s.	5.22	4e-3	6.28	4e-5	3.95	0.01	1.82	n.s.	3.78	n.s.
CPB	0.97	n.s.	3.02	n.s.	0.81	n.s.	-0.38	n.s.	1.52	n.s.	0.52	n.s.
ML	6.45	3e-3	1.38	n.s.	-0.36	n.s.	0.77	n.s.	3.37	n.s.	0.56	n.s.
R	6.08	1e-6	6.71	1e-7	6.65	5e-6	3.72	1e-3	4.17	n.s.	3.81	0.02
RM	-1.87	n.s.	3.03	n.s.	8.09	2e-6	4.42	0.05	3.46	n.s.	4.27	5e-3
RPB	0.12	n.s.	1.48	n.s.	1.18	n.s.	0.34	n.s.	1.05	n.s.	1.2	n.s.
RT	-0.91	n.s.	0.44	n.s.	-	-	-	-	1.9	n.s.	3.38	n.s.
RTL	-1.38	n.s.	0.09	n.s.	1.27	n.s.	-	-	1.03	n.s.	3.27	n.s.
RTM	-1.81	n.s.	0.91	n.s.	-	-	-	-	3.39	n.s.	1.53	n.s.
RTp	-	-	1.32	n.s.	-	-	-	-	0.64	n.s.	2.5	n.s.
STGr	-	-	0.45	n.s.	-	-	-	-	1.07	n.s.	1.54	n.s.
Tpt	-1.16	n.s.	-	-	-	-	-	-	-	-	-	-

Table 4-4 Linear negative parametric contrast details from various fields in the auditory cortex of three monkeys. Beta and significance value is given for ‘linear negative parametric’ contrast across all the ROIs in monkeys M4, M5, and M6. The beta is averaged across those voxels whose t-statistic for ‘sound versus silence’ meets $p<0.001$ uncorrected for multiple comparisons across the auditory cortex. In some ROIs, where the details are omitted, no voxels within this ROI survived this threshold at a single voxel level. The significance value (p-value) is corrected for multiple comparisons across ROIs within each monkey. (n.s. – not significant i.e. $p > 0.05$). This data indicates that BOLD decreases with increasing time window duration in most auditory cortical areas in these monkeys. This relationship seen in monkeys is opposite to that seen in humans.

Contrast across ROIs	Left Hemisphere				Right Hemisphere			
	Sound vs Silence		Linear Negative Parametric		Sound vs Silence		Linear Negative Parametric	
	beta	p	beta	p	beta	p	beta	p
A1	25.01	< 1e-7	0.63	n.s.	20.91	< 1e-7	-0.02	n.s.
AL	25.78	< 1e-7	-0.49	n.s.	17.72	< 1e-7	-3.78	n.s.
CL	11.32	1e-6	0.83	n.s.	18.3	< 1e-7	-0.72	n.s.
CM	17.47	< 1e-7	-0.38	n.s.	11.56	< 1e-7	-0.72	n.s.
CPB	14.47	< 1e-7	-0.5	n.s.	15.16	< 1e-7	-1.59	n.s.
ML	21.55	< 1e-7	1.02	n.s.	32.43	< 1e-7	-0.68	n.s.
R	34.07	< 1e-7	-0.46	n.s.	9.47	< 1e-7	-1.44	n.s.
RM	13.67	< 1e-7	-0.7	n.s.	5.11	0.004	-0.3	n.s.
RPB	12.39	< 1e-7	-1.04	n.s.	8.17	< 1e-7	-3.28	n.s.
RT	5.39	2e-4	-1	n.s.	6.29	3e-5	-0.29	n.s.
RTL	4.27	0.016	-0.43	n.s.	5.8	0.01	-3.88	n.s.
RTM	5.65	5e-5	-1.01	n.s.	4.58	0.007	-1	n.s.
RTp	6.85	1e-5	-1.61	n.s.	-	-	-	-
STGr	4.11	0.009	-0.24	n.s.	-	-	-	-
Tpt	6.51	5e-4	0.56	n.s.	18.71	< 1e-7	-1.6	n.s.

Table 4-5 Contrasts from various fields in auditory cortex of monkey M1. Beta and p-value (corrected for multiple comparisons across ROIs) for ‘sound versus silence’ contrast as well as linear negative parametric contrast averaged across those voxels whose t-statistic for ‘sound versus silence’ meets $p<0.001$ uncorrected for multiple comparisons. n.s. – not significant, $p > 0.05$. Though the sound minus silent baseline contrast is significant in all auditory areas, the linear negative parametric contrast is not significant. Thus the null hypothesis of no relationship between BOLD and spectral flux in this monkey cannot be rejected.

ROI	Monkey M4				Monkey M5				Monkey M6			
	Left hemi		Right hemi		Left hemi		Right hemi		Left hemi		Right hemi	
	m	p	m	p	m	p	m	p	m	p	m	p
A1	-9e-3	0.03	-5e-3	n.s.	-6e-3	n.s.	-3e-3	n.s.	-2e-3	n.s.	-4e-3	0.05
AL	-5e-3	0.05	-9e-3	0.04	-4e-3	n.s.	-1e-3	n.s.	-2e-3	n.s.	-4e-3	0.01
CL	-3e-3	n.s.	-	-	-	-	-	-	-	-	-	-
CM	-6e-3	n.s.	-7e-3	n.s.	-9e-3	0.01	-5e-3	0.04	-2e-3	n.s.	-4e-3	n.s.
CPB	-2e-3	n.s.	-4e-3	n.s.	-2e-3	n.s.	1e-3	n.s.	-2e-3	n.s.	-5e-4	n.s.
ML	-9e-3	0.01	-2e-3	n.s.	-9e-4	n.s.	-1e-3	n.s.	-4e-3	n.s.	-1e-3	n.s.
R	-9e-3	0.05	-9e-3	0.02	-9e-3	n.s.	-4e-3	n.s.	-5e-3	n.s.	-5e-3	0.01
RM	2e-3	n.s.	-4e-3	n.s.	-0.01	n.s.	-4e-3	n.s.	-5e-3	0.04	-5e-3	n.s.
RPB	-5e-4	n.s.	-1e-3	n.s.	-2e-3	n.s.	-1e-3	n.s.	-1e-3	n.s.	-1e-3	n.s.
RT	6e-4	n.s.	-2e-3	n.s.	-	-	-	-	-4e-3	n.s.	-4e-3	0.01
RTL	7e-4	n.s.	-1e-3	n.s.	-2e-3	n.s.	-	-	-3e-3	n.s.	-3e-3	n.s.
RTM	2e-3	n.s.	-2e-3	n.s.	-	-	-	-	-6e-3	0.05	-3e-3	n.s.
RTp	-	-	-3e-3	n.s.	-	-	-	-	-3e-3	n.s.	-3e-3	n.s.
STGr	-	-	-4e-5	n.s.	-	-	-	-	-2e-3	n.s.	-2e-3	n.s.
Tpt	2e-3	n.s.	-	-	-	-	-	-	-	-	-	-

Table 4-6 Slope of a straight line fitted on beta from various fields in auditory cortex of three monkeys. A straight line is fit on the betas as a function of window duration using linear regression and the slope of this line is determined along with its significance value. The beta used for each condition / correlation is averaged across those voxels whose t-statistic for ‘sound versus silence’ meets $p<0.001$ uncorrected for multiple comparisons across the auditory cortex. n.s. – not significant, $p>0.05$. A negative slope is noticed in most areas. Further, this slope reduces as one moves from core to belt to parabelt regions

4.5 Discussion

This work examined the network underlying the processing of spectral flux in the macaque auditory cortex. The synthetic stimuli employed were similar in complexity to speech yet devoid of semantic confounds that enable us to infer the underlying mechanisms in the processing of spectral flux in macaques and compare it with results from humans. The stimuli had systematic variation in the spectrotemporal correlation that is independent of bandwidth and this systematic manipulation enabled us to seek the organization in the processing of spectral flux in the auditory cortical areas. This spectrotemporal correlation r_1 is inversely proportional to spectral flux but directly proportional to the duration of the temporal window (see Figure 4-1). Henceforth, I will refer to time window duration instead of spectrotemporal correlation.

This experiment investigated the differences in the BOLD signal as a function of the time window duration, in the auditory core, belt and parabelt regions. Using the same synthetic stimuli (see Figure 4-2) as employed in the previous study (Overath et al., 2008), bilateral sensitivity to the decreasing time windows was observed in the core, and belt regions of monkeys M4, M5 and M6. However, it is surprising that this negative parametric effect seen in the core regions of the three animals is not evident in monkey M1. Given that the parametric effect in the auditory cortex of this monkey M1 is not statistically significant bilaterally, one cannot reject the null hypothesis of flat or a no-relationship in the auditory cortex of this animal.

Further, a right-lateralized activity in anterior STG that increased with increasing time windows was observed in monkey M1 (see Figure 4-5) which is not observed in the other three animals. The activation of these anterior areas of STP, as measured by sound versus silence contrast, in the other three animals have been unsuccessful unlike in this animal M1. Thus the signal acquired from these anterior regions is much better in this animal M1 but not sufficient enough to elicit robust activation in the other animals. This could be due to optimal scan settings in this animal M1 like the position of saturation slice, placement of the receiver coils, etc. which might have been suboptimal in the other animals.

Thus, the parametric effect seen in auditory cortex of monkey M1 is not statistically significant while the positive parametric effect seen in anterior STG is not replicated in the other three animals. This makes a case for drawing conclusions based on monkeys M4, M5, M6 and excluding data from monkey M1.

In studies based on single subject inference, when the result from one of the subjects differ from the rest, it can be considered as an outlier and the results from the majority of subjects can be taken as the outcome of the experiment. In this instance, the result from monkey M1 must be considered an outlier as the pattern seen in this animal is not seen in any of the other three animals.

4.5.1 Sensitivity to decreasing time window

The earlier human study (Overath et al., 2008), employing these same synthetic stimuli, reported bilateral sensitivity to increasing time windows in PT and anterior STG while significantly right lateralized activity in STS. If the spectral flux processing in macaque were similar to humans, then one would expect to observe BOLD increase with the increasing time windows, in other words, one would expect to see a flat i.e. no-relationship in the core region, gradually increasing to show a positive parametric effect in the belt and parabelt regions. However, this experiment finds a negative parametric effect (i.e. BOLD increased with decreasing time windows) in the core regions that reduce though staying negative as one moves to the belt and parabelt regions. In essence, the current data show a difference in the relationship between BOLD and spectral flux in the auditory cortex of macaques and humans. So this difference in sensitivity for spectral flux between humans and macaques may be related to differences in the preferred window of temporal integration.

4.5.2 Temporal duration preference

The representation of sounds of different time scales within the auditory cortex is investigated. Using BOLD as a measure of local ensemble activity, a preference for shorter temporal windows is observed in all auditory regions consistent across both hemispheres of monkey M4, M5 and M6. Though this is consistent, there is a slight relative shift in the preference towards longer temporal windows as one progresses from core to belt and then to parabelt regions. This is consistent with Rauschecker et al. (1995) that suggested a hierarchical organization in the processing of communication sounds with lateral areas of the monkey auditory cortex preferring complex stimuli.

Overath et al. (2008) reported a preference to longer time windows only in auditory association cortex (AAC) and right STS in humans. They reported no preference for any specific time window in bilateral HG. Thus, the preference for temporal window

duration between core and belt homologue regions in humans does not match the preference seen in macaques found using identical synthetic stimuli.

4.5.3 Activity in auditory core

I assessed whether differences in BOLD activation for these synthetic stimuli could be a physiological way of discriminating between the auditory core and belt areas. Since these auditory regions have a similar tonotopic organization, a recent method to delineate core from belt proposed using a measure of cortical myelination (Joly et al., 2014a). Brewer and Barton (2016) proposed to use the combination of tonotopy and periodotopy for the delineation of auditory field maps. A robust method could aid in the further refinement of the functional cortical organization of the auditory cortex.

Though the earlier human study (Overath et al., 2008) reported that core homologues in HG showed a lack of differential sensitivity to spectral flux while belt homologues showed sensitivity to decreasing spectral flux, the current macaque study reported a strong differential sensitivity in the core, belt and parabelt regions bilaterally. Thus, the use of this synthetic stimulus does not allow delineation of core versus belt regions in monkeys.

4.5.4 Lateralization

Understanding the extent of lateralization of function within auditory cortex is important for establishing a macaque model of human auditory object analysis. Zatorre and Gandour (2008) support a hemispheric specialization to process sounds like speech in humans. Zatorre et al. (2002) proposed spectro-temporal trade-off theory of cortical functional asymmetry in which it is hypothesized that the left hemisphere is specialized for temporal processing like speech while the right hemisphere is specialized for spectral processing like music in humans. Poeppel (2003) proposed ‘asymmetric sampling in time’ (AST) hypothesis where the left hemisphere preferentially extracts auditory information at short temporal windows while the right hemisphere extracts information at long temporal windows in humans. In a human study using speech sounds, Obleser et al. (2008) reported lateralization in STS for temporal and spectral processing in left and right hemispheres respectively. The human study employing this synthetic stimulus by Overath et al. (2008) suggested a dual hierarchical organization and lateralization in the representation of temporal windows in AAC. The present data from three macaques does not show lateralization in the processing of these stimuli.

4.5.5 Behaviour

I do not have behavioural results to show that macaques can detect spectrotemporal correlation which was manipulated in this study. However, the human study (Overath et al., 2008) conducted psychophysical tests to show that humans can detect spectrotemporal correlation. So obtaining this psychophysical results from macaques to show that they can discriminate between spectrotemporal correlations in this synthetic stimulus is the next step in this investigation.

4.5.6 Conclusions

Here is the summary of the results so far. Macaques show sensitivity to decreasing time window in postero-medial AC while humans show sensitivity to increasing time window in antero-lateral AC. Strongest differential sensitivity was seen in macaque auditory core. But there was no differential sensitivity in human auditory core homologues. No lateralization was seen in macaques. But humans showed sensitivity in right lateralized STS.

Thus, one cannot assume that the macaque model of the human auditory cortex is the correct representation of cortical organization given the differences in the functional organization of spectral flux between macaques and humans shown in the current study. It is at best a useful approximate model of human auditory cortex with some differences.

Chapter 5 General Discussion

5.1 Summary

In this work, I aim to develop a primate model of human cortical analysis of auditory objects. To this end, I investigate how good rhesus macaques are as a model of human cortical analysis of auditory objects. Specifically, I focussed on two aspects of an auditory object - auditory segregation and timbral analysis.

Chapter 3 concerned with inferring the cortical network underlying auditory figure-ground segregation in monkeys and comparing it with the network known in humans. Here, I presented behavioural in macaques that showed a similar performance to humans. Next, I presented fMRI data in macaques showing the similarity of the brain basis underlying perceptual organization in monkeys with humans. So macaques seem to be a good model of human cortical analysis of auditory segregation. This study has already provided spatial priors for targeted neurophysiological experiments in monkeys.

Chapter 4 concerned with inferring the functional organization of processing of spectral flux, a dimension of timbre, and comparing it with the functional organization known in humans. Here, I presented fMRI data showing the dissimilarity in the sensitivity and functional organization of spectral flux between macaques and humans. However, I also report an outlier result in one monkey which calls for caution while drawing conclusions from this study. Further, I do not have behavioural data in macaques to show that they can discriminate between the different levels in this synthetic spectral flux stimuli. Given that I employ a synthetic stimulus that does not have ethological significance to the monkeys, spectral flux detection and discrimination psychophysical results from monkeys and comparison with results in humans would help bolster the study.

So given this dissimilarity in timbral processing, macaque auditory cortex does not seem to be an accurate model of human cortical analysis of auditory objects despite the similarity in cortical organization viz. a hierarchical structure with similar tonotopic mapping, and similarity in the network underlying auditory segregation between the monkeys and humans.

5.2 Perceptual Organization

Chapter 3 examined the network underlying the process of pre-attentive, stimulus-driven auditory figure-ground decomposition in macaques. To do this, I employed complex synthetic stimuli called 'Stochastic Figure Ground' (Teki et al., 2011) with temporally coherent spectral components ('figure') which can only be perceived by the process of binding (spectral grouping of components that repeat in time). These stimuli are different in two main ways from those employed in traditional streaming paradigms. Firstly, the rich spectro-temporal overlaps between figure and background segments in SFG stimuli capture the complexity of natural acoustic scenes unlike the simple deterministic streaming sequences employed previously (Pressnitzer et al., 2008, Elhilali et al., 2009a). Secondly, the time required to segregate figure from the background is less in SFG stimuli, as they don't demonstrate the build-up rate typically seen with simple, streaming sequences (Micheyl et al., 2007a, Gutschalk et al., 2008). These differences imply that the mechanisms that mediate segregation with SFG stimuli might be different from those with more simplistic stimuli (Teki et al., 2011). Further, these stimuli do not have semantic confounds and are equally relevant to macaques and humans. Thus, they allow us to seek common behavioural and neural mechanisms for auditory segregation across humans and monkeys.

In this study of macaques, I outlined the brain bases underlying figure-ground segregation due to temporal coherence involving bilateral regions of the rostral lateral belt and rostral parabelt in the auditory cortex of macaques. This result is already guiding targeted and efficient neurophysiology in macaques by providing spatial priors.

5.2.1 Macaque model

Previous studies of auditory streaming suggest that Japanese monkeys (*Macaca fuscata*) perceptually segregate tone sequences in a similar way to humans (Izumi, 2002). Further, studies in rhesus macaques (*Macaca mulatta*) have demonstrated 'auditory induction' (a process that follows the rules of auditory scene analysis where the auditory system restores an occluded sound of interest) in a manner akin to humans (Petkov et al., 2003, Petkov et al., 2007). Investigating the effect of temporal coherence on segregation, Christison-Lagay and Cohen (2014) used a stimulus construct with 2-tone streaming complex with temporal coherence as proposed in

Elhilali et al. (2009a), and report behavioural results in rhesus macaques (*Macaca mulatta*) that are qualitatively similar to results seen in humans. Thus, the evidence suggests that monkeys perform stream segregation in a way similar to humans.

Furthermore, Fishman and colleagues (Fishman et al., 2001, Fishman et al., 2004) studied the neural correlates of stream segregation by presenting a simple streaming sequence (alternating tone pattern ABAB..) and recording multiunit activity in the primary auditory cortex (PAC) of crab-eating macaques (*Macaca fascicularis*). Their results suggest a parallel between this neural activity and the previous human psychoacoustics results. Single unit responses recorded in the PAC of awake rhesus macaques (*Macaca mulatta*), using the same alternating-tone pattern, also demonstrated the features of stream formation seen in humans (Micheyl et al., 2005).

I have shown that rhesus macaques (*Macaca mulatta*) are able to segregate the figures present in stochastic figure-ground stimuli. Further, the auditory figure detection performance in macaques increased with figure coherence in a way similar to that seen in humans (Teki et al., 2013). Next, I have shown that the cortical network underlying perceptual organization in macaques involves the auditory rostral lateral belt and rostral parabelt bilaterally. Such a network in macaques, involving rostral parabelt regions, would be consistent with the activity found on the convexity of STG in humans using fMRI (Teki et al., 2011), MEG (Teki et al., 2016) and ECoG (Griffiths, 2017).

In the light of these converging pieces of evidence, I conclude that macaques are suited for system-level characterisation of auditory segregation based on temporal coherence and that they can be used as an animal model to understand brain mechanisms underlying auditory scene analysis in humans.

5.2.2 Models of stream segregation

Previous animal neurophysiological work on stimulus-driven auditory segregation suggests the involvement of primary auditory cortex (Micheyl et al., 2007a, Fishman et al., 2004, Fishman et al., 2001). Stimulus-driven segregation is thought to be mediated by neurons in the auditory cortex via their basic response properties – namely, frequency selectivity, forward suppression, and adaptation – such that distinct neuronal populations are activated for the constituent elements. This ‘population separation’ (PS) model of auditory stream segregation postulates that

segregated streams are perceived whenever sounds evoke spatially segregated responses along any of the dimensions of responses in the auditory cortex (e.g. tonotopy, virtual pitch, temporal and spectral modulation rates), and this was thought to be both necessary and sufficient for segregation to occur.

However, this model cannot account for the integration seen in temporally coherent streams (when the streamed elements are presented simultaneously as in Elhilali et al. (2009a)) despite the spatial separation of neuronal responses that code for the constituent elements of the stream. Thus, it was concluded that spatial separation of neuronal responses is necessary but not a sufficient condition for stream segregation. As a result, the ‘spatiotemporal’ model of segregation has been proposed which suggests that binding of components is at least partly dependent on their temporal coherence (Shamma et al., 2011). In other words, components that are temporally coherent are grouped together while incoherent components are perceived as belonging to separate sources.

Recent neurophysiological experiments (Fishman et al., 2017), using an ABA paradigm, tested whether population separation model can explain the integration seen in synchronous tones, by recording responses from macaque A1. A greater tonotopic separation was seen for alternating tones (which are perceived as separate streams) when compared to synchronous tones (which are perceived as a single stream). The authors suggest that population separation model remains a viable model of stream segregation.

However, stream integration can be seen even with synchronous tones that are much farther apart than the maximum frequency separation tested in this study (6 semitones). In cases where the frequency separation between synchronous tones is much greater than 12 semitones, separate tonotopic activations still evoke an integrated percept, questioning the validity of the PS model for sounds encountered in natural scenarios. Further, studies employing SFG stimuli have shown that spatial separation of neuronal responses is neither a necessary nor a sufficient condition for segregation to occur.

Previous human fMRI work also suggests the involvement of IPS, a specific non-auditory region (Cusack, 2005), in a ‘multilevel model’ by determining the final perceptual organization by mediating the combination of information across regions

competing to organize the auditory scene that is based on different representations. Using SFG stimuli, previous studies (Teki et al., 2011, Teki et al., 2016) have reported IPS activity in humans and suggested a bottom-up role in segregation. However, more evidence is needed to ascertain the involvement of monkey posterior parietal cortex in the perceptual organization based on temporal coherence.

5.2.3 Temporal coherence model

Acoustic features that start and stop at roughly the same time segregate together. Further acoustic features that co-vary in time are grouped together. So temporal coherence mechanisms need to track the evolution of acoustic features that have common onset and offset. The first stage of the 'temporal coherence model' of segregation computes the multidimensional depiction of various sound attributes including pitch, location, and the spectrogram of the acoustic input. The second stage computes pair-wise correlations between all channels of the cortical representation. The cochlea performs the spectral decomposition of the acoustic signal into different frequency bands while primary core regions in the cortex perform the temporal integration in each frequency band (Shamma et al., 2011, Elhilali et al., 2009a, Chi et al., 2005, Elhilali and Shamma, 2008). For the second stage, the temporal coherence detectors need to exhibit conjunctive effects i.e. they only respond if several specific features are present together but not in isolation. Such conjunctive effects enable selectivity for complex spectrotemporal features expected of natural sounds as well as that modelled by SFG stimulus. However, spectrotemporal receptive fields (STRF) of auditory cortical neurons have not previously described such conjunctive effects since conjunction might require capturing the effect of a cortical network of neurons instead of just a single neuron.

Since traditional STRF do not capture the fact that the response of a cortical neuron results from the complex nonlinear network in which it is embedded, Harper et al. (2016) modelled the cortical receptive field by fitting a feedforward network of 1–7 nonlinearly-interacting lower-order model neurons to cortical responses to natural sounds. This network receptive field captured non-linear functional characteristics in auditory cortical neurons like conjunctive feature selectivity. They found that 67% of multi-feature neurons recorded from primary auditory cortex of ferrets exhibited conjunctive effects. However, Elhilali et al. (2009a) searched for but did not find any evidence of temporal coherence detectors in ferret A1. This could be because, in addition to conjunctive effects, temporal coherence requires computation of

synchronicity of onsets. Further longer windows of temporal integration as observed in secondary auditory cortex are required to track the evolution of acoustic features. Thus temporal coherence detectors should exist in secondary auditory cortical areas. The findings I present here would predict the existence of single neurons in rostral lateral belt and parabelt that encode temporal coherence across frequencies in complex sounds which do not necessarily have a harmonic relationship amongst their spectral components.

In essence, the neurons in rostral lateral belt and parabelt would act as temporal coherence detectors that compute temporal coherence amongst multiple inputs each arising from upstream auditory regions that code for the various features.

5.2.4 Role of dual-stream hypothesis in stream segregation

The dual-stream hypothesis (Rauschecker and Tian, 2000) suggests that auditory object and spatial processing occur in separate and parallel pathways: the ventral pathway for ‘what’ and the dorsal pathway for ‘where’ processing. Auditory object extraction occurs in the ventral pathway which starts in auditory core areas that indirectly project to the ventrolateral prefrontal cortex via anterior lateral belt areas. Similarly, auditory spatial processing occurs in the dorsal pathway which projects from core areas via caudal areas to the dorsolateral prefrontal cortex. Under this hierarchical processing model, as one goes further along the pathway the representations become distant and more abstract from the stimulus, with higher cortical areas representing perceptual features, and finally object or category-specific representations. Consistent with this framework is the fact that the bilateral rostral parabelt (ventrally located) is activated for the processing of auditory figure (Leaver and Rauschecker, 2010).

However, activation of the parietal cortex (Teki et al., 2011, Teki et al., 2016) is thought to play a role in the perceptual organization of auditory stimuli, which lies in the dorsal pathway. This suggests that the dual-stream model of hierarchical processing might be too simplistic, and there might be overlap between the two streams during the creation of consistent perceptual representations of the auditory world. For instance, spatial information, processed in the dorsal stream, can act as a grouping cue (McDonald and Alain, 2005) for auditory segregation processed in the ventral stream.

5.2.5 Future directions

5.2.5.1 Temporal coherence detectors

I hypothesize the existence of single units whose firing rate increase with the number of simultaneous frequency components of the sound. I expect that these units would be agnostic of the exact constituent frequencies but are only concerned with the temporal coherence (i.e. has common onset and co-varies in time) among the incident frequencies. Thus these neurons should respond even to sounds with components that do not have a harmonic relationship but have simultaneous onset. At the population level, using ultra-high field (7T) fMRI in humans, Moerel et al. (2013) reported the presence of voxels in the rostral STG that respond to sounds that do not have a harmonic relationship. However, at the neuronal level, single units with such properties are yet to be reported from neurophysiological studies in monkeys. Note that Feng and Wang (2017) report single units that respond to sounds with harmonically-related components in the auditory core areas of monkeys but they did not find units that respond to sounds that do not have a harmonic relationship. I hypothesize that these temporal coherence detectors might exist in the rostral lateral belt and rostral parabelt in the auditory cortex of macaques.

To test this prediction, electrophysiological experiments could be performed to record single unit activity from the rostral lateral belt and rostral parabelt of a macaque, and search for units that respond to broadband stimuli containing multiple frequency components without a harmonic relationship but more importantly have simultaneous onset and temporally coherent, like the figures in SFG stimuli. Furthermore, an increase in the number of temporally coherent components of an SFG figure should elicit an increase in the firing rate from such temporal coherence detectors, making the auditory figures more salient to the animal. Further, the neuronal projections from and to (i.e. inputs and outputs) such temporal coherence detectors need to be characterised.

5.2.5.2 Role of top-down attention

There is a suggestion (Treisman and Gelade, 1980) in visual neuroscience that visual stimulus features (colour, shape, movement, orientation) register automatically without the need for top-down attention. However, focussed attention on any one of the visual features of a visual object binds together all these features into a coherent object is required to bind all the features of the visual object. Similarly, there is a

suggestion in auditory neuroscience (Shamma et al., 2011) that top-down attention might aid in auditory perceptual organization i.e. when attention is directed to a particular auditory feature (e.g. timbre) then it aids in binding together of all temporally coherent auditory features (like pitch, timbre, location) of that auditory source and this helps in segregating it from the incoherent features of other auditory sources that may be occurring at the same time. So it is assumed that top-down attention operates on objects be it auditory or visual.

Recent neurophysiological experiments (Lu et al., 2017) in macaques have tested the temporal coherence model using the ABA paradigm. They have demonstrated that responses and sensitivity are enhanced by synchronous stimuli and suppressed by alternating tones, but only when the animal pays attention to the stimulus. O'Sullivan et al. (2015) reported evidence for pre-attentive computations of temporal coherence that was enhanced by active analysis of the auditory scene. So these studies suggest that top-down attention is an important factor in the perceptual organization of sounds based on temporal coherence and that it has a modulatory effect on segregation.

In studies by Teki et al. (2011) and Teki et al. (2016), human participants were engaged in an irrelevant task, while monkeys in my study were distracted by an irrelevant non-auditory task. This suggested that figure-ground segregation related to bottom-up stimulus-driven mechanisms and that top-down attention is not essential. However, it is also possible that this “unattended” auditory stimulus might be processed by left-over attentional resources due to a lower load of an irrelevant task.

To concretely address whether top-down attention is essential for segregation to occur, macaque neurophysiological recordings could be made from temporal coherence detectors that are presented with SFG stimulus but are distracted with a visual task. Responses to SFG stimuli during trials where the visual task difficulty is high can be compared against trials where the visual task difficulty is low to highlight whether attention is necessary for figure-ground segregation to occur. Further, the role of top-down attention in figure-ground segregation could be inferred using neurophysiological recordings from temporal coherence detectors in macaques that are presented with the SFG stimuli and are responding whether a figure was present in the SFG stimulus. Successful trials (figure present and detected by a lever press) could be compared with false-negative trials (figure present, but missed) to elucidate

the differential role of top-down mechanisms of segregation (after discounting any motor response-related activity). I predict that neural activity would be enhanced in this contrast of hit versus miss trials. Similarly, false-negative trials (figure present, but missed) could be compared with true-negative trials (figure not present, and so no lever press) to explore bottom-up processes underlying segregation.

5.2.5.3 Interaction of top-down versus bottom-up mechanisms

One hypothesis on the role of top-down attention and bottom-up segregation is that top-down expectancies on stream constituents are shaped by bottom-up evidence on stream segregation which is in-turn furthered by the top-down attentional mechanisms that might aid in the binding of stream components. Thus, a recursive iterative relationship between top-down processes of attention and bottom-up mechanisms of segregation may be involved in achieving stream segregation (Riecke et al., 2015).

Based on my fMRI findings in macaques with a passive presentation of SFG stimuli, I suggest that rostral lateral belt and parabelt are the bottom-up mechanisms of segregation providing evidence on stream constituents based on temporal coherence. The prefrontal cortex (Fritz et al., 2010) engages top-down attention that can aid in stream segregation by providing expectancies on constituents of a stream. I speculate that posterior STS in the parietal cortex of macaques might mediate between prefrontal cortex and sensory regions. In support of multilevel model of perceptual organization (Cusack, 2005) and in line with proposed automatic bottom-up role (Teki et al., 2011, Teki et al., 2016) for human IPS in segregation, I propose that posterior STS in the parietal lobe of macaques might determine the final perceptual organization by integrating information from different regions each of which is competing for organization based on different features (temporal coherence, frequency, pitch, timbre, and spatial location). So future work should, therefore, explore this hypothesis and provide evidence for the involvement of posterior STS in the parietal lobe of macaques in figure-ground segregation.

5.3 Spectral flux processing

Chapter 4 examined the network underlying the encoding of spectral flux in the macaque auditory cortex. The synthetic stimuli used in my experiments are similar in complexity to speech yet devoid of semantic confounds. This enables us to infer the underlying mechanisms in the processing of spectral flux in macaques and compare it with results from humans. My results show a difference between macaques and humans in the relationship between BOLD and spectral flux in auditory cortex. These findings support a functional organization of spectral flux in macaques that is different from that in humans. However, since there is no behavioural data in macaques to show that they are able to discriminate between the different levels of spectral flux in this synthetic stimuli, it would be the next step in this investigation of timbral processing in a macaque model.

5.3.1 Neuronal code for spectrotemporal correlation

Natan et al. (2017) recorded from single units in primary auditory cortex (A1) of awake rats to spectral flux stimuli but did not find changes in mean population firing rate of units to systematic manipulation of spectrotemporal correlation though instead found changes in the gain of the stimulus-response. They suggested that this gain control mechanism might normalize the responses that are sensitive to the BOLD signal. Since BOLD signal may average out the heterogeneous changes in the neuronal spiking responses (Logothetis and Wandell, 2004), they suggested this as a possible reason for observing lack of differential sensitivity to spectral flux stimuli in bilateral HG in humans (Overath et al., 2008). Further, Natan et al. (2017) suggested that though in A1, some neurons might be inhibited while others excited leading to no significant change in population activity, areas downstream to A1 might convert the heterogeneous changes in firing rate of A1 neurons to an increase in firing activity and thus sensitive to the BOLD signal.

However, I find differential sensitivity to spectral flux stimuli in macaque A1 using sparse fMRI. This suggests that the neuronal coding scheme based on gain control observed in rat A1 might not be applicable to monkeys and also humans due to evolutionary proximity. So this difference in the functional organization for spectral flux between humans and macaques may be related to differences in the preferred window of temporal integration.

5.3.2 Preference for temporal integration window

I speculate that the differences in the sensitivity for spectral flux in humans and macaques are related to the differences in the perception of temporal windows between species. A number of previous studies (in humans and macaques) have examined the mechanisms for analysis of time windows (see Table 1-2). Spectral flux can also be interpreted as the duration of analysis window that is required to reach a minimum spectrotemporal correlation between any two frames within it.

In the experiment presented in Chapter 4, performed in macaques, I found a preference for shorter time windows in the auditory core and postero-medial belt regions, though there is a relative shift in preference towards long time windows as one moves towards antero-lateral belt and parabelt regions.

Using the same stimulus as employed in Chapter 4 (window duration ranging from 20 ms to 306 ms), and employing a continuous-acquisition fMRI in humans, Overath et al. (2008) report that only auditory association cortex (AAC) and right STS show a preference to longer time windows. The same study also finds no preference for any specific time window in bilateral HG nor any preference for the shorter time windows anywhere on the STP or STG. This lack of preference towards short time windows in postero-medial auditory cortex is in contrast with the majority of studies in humans (see Table 1-2). However, the preference towards long time windows in antero-lateral auditory cortex was in agreement with the majority of human studies.

So this difference in sensitivity for spectral flux between humans and macaques shown in Chapter 4 may be related to differences in the preferred window of temporal integration, despite the underlying similarity in the anatomical organization of time window processing (see Table 1-2).

Scott et al. (2011) recorded responses from single units in auditory cortex of awake macaques to sinusoidally amplitude-modulated tones. They (see Fig. 9) reported a continuous increase in onset latencies along the rostro-caudal axis of the auditory cortex. They also analysed synchronized spike discharges that suggest that the window of temporal integration is longer in R than in A1. They argue that the functional organization of the auditory cortex is best understood in terms of temporal processing.

Bendor and Wang (2008) have reported recordings from single units in A1, R and RT regions of the auditory cortex of awake marmosets. They compare responses to pure tones and sinusoidally amplitude-modulated tones (rates: 4 to 2048 Hz, i.e. window of 0.5 to 250 ms). They show that neurons in R and RT regions exhibit poorer stimulus synchronization to AM tones when compared to neurons in A1 region. They also find that neurons in R and RT regions have longer minimum latencies than those in A1. They propose a model of spectral and temporal processing sub-pathways within the 'what' pathway in primates. In this model, in the temporal processing sub-pathway, A1 has the smallest temporal integration window and the duration of this window continues to increase as one moves to R and RT along this pathway (caudal-rostral axis). Their findings of preference for shorter windows in A1 and longer windows in R and RT from electrophysiological studies in marmosets are in line with findings from neurophysiology and fMRI studies in macaques and humans (see Table 1-2).

Macaque neurophysiology studies have confirmed that onset latencies increase rostrally from A1 along the ventral stream (Kusmierek and Rauschecker, 2009, Kikuchi et al., 2010, Scott et al., 2011). However, this widening of the window of temporal integration seen in the auditory cortex actually begins in the auditory periphery (Wang et al., 2008). Though auditory-nerve fibers faithfully represent fine structures of complex sounds further along the auditory pathway, the upper limit of temporal representation of sounds gradually decreases due to temporal integration of converging inputs from one station to the next, from the cochlear nucleus to inferior colliculus and then to MGB and finally auditory cortex.

Bendor and Wang (2008) suggest that an increase in the window of temporal integration can decrease stimulus synchronization, creating a temporal-to-rate coding transformation for temporal information of the audio signal. In line with this observation, Bendor and Wang (2007) have previously suggested that R and RT are involved in transforming the temporal representation of the acoustic signal's envelope in A1 into a rate code that no longer relies on spike timing. However, Scott et al. (2011) observed that fewer neurons in R showed a significant rate response than in A1 which does not support the suggestion that R is involved in transforming a temporal code to a rate based code.

5.3.3 Species-specific differences

Spectral flux dimension corresponds to an interaction between temporal and spectral dimensions of sound. Further, spectral flux can be interpreted as the duration of the time window required to achieve a minimum spectrotemporal correlation. Thus species-specific differences in the processing in the temporal domain become relevant to understanding the differences seen in the processing of spectral flux.

There is behavioural evidence for differences between humans and macaques in the auditory temporal analysis. Clack (1966) and O'Connor et al. (1999) report that the rates of temporal integration for pure tones in rhesus macaques are slightly slower than in humans. The difference limen for detecting changes in tone duration is higher in several macaque species than in humans as reported in Sinnott et al. (1987). Further, Sinnott and Brown (1993) report that pure-tone frequency-difference limens decrease more rapidly with increasing frequency in macaques than in humans. Hopp et al. (1992) compared the perception of humans and Japanese macaques (*Macaca fuscata*) for detection of changes in the temporal position of pitch peak along a synthetic coo continuum. They reported that the sensitivity to detect spectral peak position is higher in humans than in monkeys. Similarly Sinnott and Brown (1997) compared the perception of humans and monkeys (*Macaca fuscata*) using phonemic /ra-la/ continuum. They reported that humans were more sensitive than monkeys to temporal variations while monkeys were more sensitive than humans to spectral variations in this synthetic continua.

Zarco et al. (2009) have compared psychometric performance in rhesus macaques and humans on a motor timing task. Their data show that monkeys are more accurate at shorter intervals (450 ms) than at longer intervals (900 ms) when cues are auditory in nature. Humans are as accurate at shorter intervals (450 ms) as they are at longer intervals (1 s) in auditory cued tasks. However, in a direct comparison between species, the authors find that humans are more accurate at longer intervals than monkeys but accuracy is similar at shorter intervals. This suggests a greater preference for longer intervals in humans than in monkeys. Furthermore, this study supports primate species differences in auditory rhythm processing.

O'Connor et al. (2011) have compared sensitivity to detect sinusoidally varying amplitude modulation (AM) between rhesus macaques and humans. In humans, they found peak sensitivity to detection AM that ranged between 10-30 Hz modulation

rates depending on tone duration while for macaques the peak sensitivity ranged between 40-100 Hz. Further, they reported a greater sensitivity in humans over macaques, for detection of amplitude-modulated noise at lower modulation rates (< 10 Hz). This data suggests a greater preference for slower temporal rates in humans. This study also supports species differences in temporal processing among primates.

More recent work has investigated the cortical encoding of natural sounds using fMRI in humans and made a comparison with macaques using identical stimuli and identical modelling. The temporal modulation function in macaques shows a preference for faster modulation rates with its peak greater than 30 Hz (Erb et al., 2019) while the preference in humans is for slower modulation rates centred at 3 to 4 Hz (Santoro et al., 2014). These studies support differences in preference for temporal modulation rate between primates.

Scott et al. (2011) recorded from single neurons in the auditory cortex of awake macaques in response to amplitude-modulated tones. They reported (see Fig 14D) more neurons in core regions (A1, and R areas) preferentially synchronize to slower modulation rates (<10 Hz) than to faster modulation rates (>50 Hz). Further based on very limited data, they reported that neurons in belt areas (CM, MM, and ML areas) favour a more restricted temporal receptive field with a peak at 5 Hz modulation rate. A similar result from core areas was reported by Bendor and Wang (2008) who recorded from single neurons in awake marmoset monkeys in response to amplitude-modulated tones. They reported (see Fig 14C) temporal best modulation frequency that peaked at 8.1 Hz in A1, 6.5 Hz in R, and 5.7 Hz in RT areas.

It is surprising that the results (Scott et al., 2011) from single neurons in core areas of macaques showing preference to slower modulation rates are not in line with behaviour results of O'Connor et al. (2011) that employed same stimulus showing preference to faster modulation rates. The results from single neurons in the belt could suffer from sampling bias as data from these areas was limited unlike data from core regions. However, the behaviour is a manifestation of coordinated interaction between multiple areas each a result of neuronal ensembles and not just a single neuron. So future research should reconcile this discrepancy between results from single neuron neurophysiology and behaviour.

Further, the results (Scott et al., 2011) from single neuron recordings in macaques also differed with the results from modelling of the BOLD data (Erb et al., 2019) that employed natural sounds instead of synthetic sounds. Since the topographic organization for temporal modulation obtained using synthetic AM noise (Baumann et al., 2015) and natural sounds (Erb et al., 2019) are very similar in the auditory cortex of macaques, differences due to the stimulus can be rejected as a possible reason for the observed discrepancy. BOLD signal reflects the input and local processing of information in a neuronal ensemble (Logothetis and Wandell, 2004) and BOLD sensitivity is reflective of the overall preference of an ensemble rather than a single neuron. So future research should reconcile this discrepancy between results from single neuron neurophysiology and behaviour as well as fMRI results.

Thus, there is behavioural data (O'Connor et al., 2011) and BOLD data (Erb et al., 2019) to support an increased preference towards slower rates (~3 Hz) or longer time windows in humans over monkeys. Modulation at 3–10 Hz seems critical for processing of spoken syllables and speech intelligibility (Luo and Poeppel, 2007). This suggests a possible reason for observing increased preference for longer time windows in humans. Cohen et al. (2007) reported high variance between macaque vocalisation categories at high temporal modulation frequencies between 5 to 20 Hz which are very relevant for categorisation of vocalisations. Joly et al. (2012) had reported that certain macaque vocalisations have very high temporal modulation rates compared to human speech. These studies suggest a possible reason for observing increased sensitivity to shorter time windows in monkeys. Thus this need to process speech in humans and vocalisations in macaques might account for the differences in the sensitivity for temporal processing rates between humans and monkeys. Thus, the tuning of the auditory cortex to syllabic rate (i.e. a long time window) might be unique to humans and possibly an outcome of divergent evolution in humans alongside the development of speech.

A syllabic rate of 3-10 Hz implies a temporal analysis window of 100 to 300 ms which corresponds to a spectro-temporal correlation of 0.8 to 0.9 (see Table 4-1). Thus the increased sensitivity towards longer time windows (higher correlation levels) observed in humans in the spectral flux experiment might be due to species differences in the preferred window of temporal integration. Figure 5-1 shows a possible way to reconcile the differences in the functional organization of spectral flux between humans and macaques by factoring the increased preference for slower

syllabic rate (i.e. longer time windows) in humans. Thus an increased preference for long time windows in humans could make the negative parametric effect seen in macaques seem to be a positive parametric effect in humans. This could be the reason for observing a positive parametric effect in humans while a negative parametric effect is seen in monkeys.

To summarize, the work I have presented in chapter 4 supports species differences between humans and monkeys in the processing of spectral flux. However, it remains to be investigated whether the species differences between macaques and humans that are found in the analysis of spectral flux are unique to humans or just one instance of a general disparity between humans and other species.

5.3.4 Future directions

The current experiment was limited to functional imaging in macaques and no behaviour was quantified which is important since this synthetic stimulus is not ethological relevant. Using this synthetic stimulus, the ability to detect spectral flux (just noticeable difference at correlation $r_1 = 1$) and the difference limen at different spectrotemporal correlations need to be characterized in macaques as well as humans. The behavioural results could provide further evidence on the interspecies differences in the processing of spectral flux.

Future work should characterize the spectral flux of these different macaque vocalizations, in terms of spectrotemporal correlation of a running analysis window of a specific duration. This would provide evidence as to whether a particular correlation dominates the macaque vocalizations, and how this relates to the spectral flux preference seen in macaques.

In this thesis, I have described an anatomical basis of the spectral flux processing in monkeys, with evidence of both core and belt region involvement. This spatial prior can be utilized for future targeted neurophysiological experiments in macaques. I predict that the multiunit spiking activity in response to a synthetic spectral flux stimulus would be most pronounced for correlation $r_1 = 0$ in core (A1, R) regions more than belt (CM, RM, AL) regions.

Human
Macaque (default-primate)
** Significant positive parametric effect
(e.g. ant STG, parabelt homologues)
** Significant negative parametric effect
(e.g. A1, R, CM)
Increased preference for
long-time windows in humans
(evolutionary adaption to speech)
Response
Window length
Cortical area preference for
Long windows / slow rates
(Antero-lateral)
Short windows / fast rates
(Postero-medial)

Figure 5-1 Interpretation of spectral flux results in macaques and humans. The top panels represent postero-medial auditory cortex while bottom panels represent antero-lateral auditory cortex. Macaques (shown in blue lines) exhibit a significant negative parametric effect in the auditory core and medial belt regions i.e. BOLD is highest for short time windows in postero-medial auditory regions, while there is no significant parametric effect in parabelt. However, macaque antero-lateral auditory regions have more preference for longer time windows than postero-medial auditory regions. Humans (shown in red lines) show a positive parametric effect in ant STG, parabelt homologues i.e. BOLD is highest for long time windows in antero-lateral auditory regions, while there is no significant parametric effect in core homologues. However, human postero-medial auditory regions have more preference for shorter time windows than antero-lateral auditory regions. Thus, the anatomical organization of time window processing is similar across both humans and macaques, i.e. postero-medial auditory cortex prefers short time windows while antero-lateral auditory cortex prefers long time windows. The macaque result is overlaid on the human result (shown in blue dotted lines) to summarize the difference in the functional organization of spectral flux between macaques (default-primate) and humans. This apparent difference in sensitivity for spectral flux between species can be reconciled by factoring in the increased preference (shown in green arrows) for long time windows in humans over monkeys. This increased sensitivity to long time windows or slow rates (~3 Hz) in humans could be due to an evolutionary adaptation to speech.

5.4 Role of cortical synaptic synchrony

Chapter 3 concerned with the perceptual organization based on temporal coherence. Auditory segregation due to temporal coherence requires single units that can detect the common onset of auditory features and track their co-variation in time. Common-onset detection concerns with detecting coincidence which is efficient when there is high synchrony of synaptic inputs to such detectors (Grande et al., 2004).

The synaptic influence of multiple neurons converging onto a single neuron in a cortical network is much stronger if these input neurons fire at or near synchrony (Abeles, 1982, Abeles, 1991). Thus synchronous firing, implicated in segregation based on temporal coherence, is present as a neural code (Aertsen et al., 1994) in the brain.

Chapter 4 concerned with the processing of spectral flux. Since spectral flux is a dynamic aspect of timbre, spectral flux processing concerned with the mapping of preference for the duration of the window of temporal integration across the auditory cortex. Temporal integration does not require high synchrony amongst the inputs to such single units rather a low degree of synchrony is sufficient for its computation.

Thus, pyramidal cells in the auditory cortex can act as either coincidence detectors or temporal integrators depending on the degree of synchrony among the synaptic inputs - both thalamocortical inputs and intracortical ones (Rudolph and Destexhe, 2003). High input synchrony leads to efficient computation of coincidence detection while a low input synchrony facilitates computation of temporal integration (Eggermont, 2007). Thus level of cortical synaptic synchrony plays a significant role in the cortical analysis of auditory objects.

Further, neurons, over which either coincidence detection or temporal integration is performed, may belong to clusters that exhibit stronger correlation within themselves than the usual level of correlation seen (Eggermont, 2006). Correlation can ensure efficient propagation of information (Kistler and Gerstner, 2002) apart from increasing signal to noise ratio in the STRF (Eggermont, 2006) as well as preserving the temporal precision of neural firing in further downstream areas (Kimpo et al., 2003).

5.5 The macaque as a model of auditory object processing

Animal models allow single neuron and intracellular recordings, terminal procedures like retrograde and anterograde staining methods (Hackett et al., 2014), apart from destructive lesioning (Fritz et al., 2005) none of these are suitable for humans but are important to advance our understanding of the functional organization of the brain. Further, an animal model that has a similar underlying substrate and mechanism to humans is advantageous to understand the mechanism of auditory perception in normal and abnormal human listening.

Of all possible mammals (e.g. rat, mice, cat, and ferret), macaques are more suited as an animal model given their phylogenetic proximity to humans with the exception of chimpanzees which are not available for invasive research. But how good are macaques as an animal model for systems neuroscience research? Functional MRI in awake behaving macaques has enabled comparison of the neurobiology of cognitive functions with humans in an unprecedented way.

Primates have forward facing eyes enabling increased depth perception and visual acuity and thus results in their increased reliance on vision. The early visual areas and motion area are suggested to be conserved in humans (Orban et al., 2004). Tsao et al. (2008) suggest a closer anatomical correspondence between macaque and human face-processing systems. Thus, in visual neuroscience, the visual system of macaque is chosen over other mammals for the study of visual perception due to its similarity with humans. However, with more detailed investigations a number of differences in the visual cortical organization are also emerging (Tootell et al., 1997, Tootell and Taylor, 1995, Tootell et al., 1995, Jacobs and Deegan, 1997).

Primates have a precision grip as they have opposable thumbs and nails instead of claws. Sensorimotor systems of primates constitute levels of increasing size and complexity. Prosimians, monkeys, apes, and humans group to form four grade shifts with each primate level characterized by a more elaborate sensorimotor system (Kaas, 2004). Despite the increase in complexity, the motor system of macaques and humans have structural similarities, comparable topographical relations, architecture and regional receptor distribution patterns all of which support the notion that there are homologous regions in the motor cortex (Roland and Zilles, 1996, Zilles et al., 1995) including primary motor cortex (M1), premotor cortex (PM), supplementary

motor area (SMA) and caudal cingulate motor area (CMAc). Thus macaques are used as an animal model in motor neuroscience research.

However, there are structural and functional differences, despite some similarities, between macaques and humans in the organization of the intraparietal cortex (Grefkes and Fink, 2005) implicated, amongst many things, in the integration of multimodal information for constructing the spatial representation of the external world. These differences possibly reflect the differences in the evolution of the dorsal visual stream. Thus there has been a divergence in the evolution of the intraparietal cortex compared to the motor and visual cortex whose function has remained relatively conserved. Similarly, Donahue et al. (2018) report that the amount of gray matter and white matter in the prefrontal cortex in humans is disproportionately larger than in non-human primates. This suggests a divergence in the evolution of prefrontal cortex.

Coming to the auditory system, the hearing range in macaques (Jackson et al., 1999) is similar to that in humans. Further macaques are able to possibly distinguish conspecifics and particular individuals based on the call structure. Macaques are excellent sound source localizers similar to humans in their spatial acuity. There is evidence that macaques perceive pitch in a similar way to humans (Joly et al., 2014b). So the auditory abilities of macaques are increasingly indicated as being very close to that of humans.

The similarity of auditory cortical organization in terms of core, belt and parabelt, between macaques and humans is fairly established (Baumann et al., 2013). Further, the general consensus is that the organization of the macaque auditory cortex is better established than in humans (Moerel et al., 2014). But how similar is the functional organization of the auditory cortex of macaques and humans? Macaques show some evidence of specialisation for processing of vocalisations in general and conspecific vocalisation in particular similar to how humans show specialisation for processing of speech.

Responses to speech stimuli occur in STG with a long latency (Steinschneider et al., 1999) suggesting that processing of complex stimulus properties occurs farther away from the primary regions. Upstream areas in auditory cortex encode basic stimulus properties while downstream areas further from core regions encode complex

stimulus properties (Mesulam, 1998). Consistent with this idea, there is an increase in the size of the window of temporal integration observed in humans as one moves along the auditory ventral and dorsal streams (Husain et al., 2004). Also, there is evidence that the further one goes along the ventral stream, the more it is sensitive to invariant auditory features, which characterize the individual auditory objects (Zatorre et al., 2004). This hierarchical organization of the auditory cortex seen in humans is consistent in macaques (Leaver and Rauschecker, 2010) where most anterior regions on the ventral pathway represent a complete acoustic signature of auditory objects.

My study is the first and the only investigation to show evidence of capabilities of auditory segregation based on temporal coherence in any non-human animal and the similarity of its underlying brain basis with humans. Though there is anecdotal evidence that most animals are able to extract sounds of interest from complex acoustic scenes, there are no investigations to show the behavioural performance and brain basis underlying auditory segregation based on temporal coherence in a non-primate species. So the question of whether a species ‘lower’ than macaques is better suited as an animal model is still open. Although a ‘lower’ species might be able to perform auditory segregation, it is less likely that they employ similar brain mechanisms to humans. In this regard, I show that macaques are particularly suited as an animal model of human auditory segregation.

My study is also the first to show differences in the functional organization of processing of spectral flux, one of the dimensions of timbre, between humans and non-human primates. Though I do not have behavioural results to show that macaques can perceive manipulations in synthetic stimulus employed in this study, the human study employed psychophysical criteria for participant inclusion that showed humans can detect spectrotemporal correlation. This difference in the functional organization of the spectral flux processing between macaques and humans is surprising given the phylogenetic proximity between the species. However, when spectral flux is interpreted in terms of a temporal window of integration, it highlights the similarity in the anatomical organization of time window processing apart from providing clues on the reason for this disparity in terms of possible specialization of the human auditory cortex for processing of speech with an increased preference for longer time windows.

5.6 Conclusions

The macaque can be argued to be a good model for human cortical analysis, based on the existence of a core, belt, and parabelt areas in the superior temporal plane, and in the adjacent cortex of the lateral superior temporal gyrus. In this thesis, I have demonstrated that, at the level of auditory scene analysis, the macaque shows homology with humans for the auditory cortical mechanisms underlying segregation based on temporal coherence. However, at the level of timbral analysis, the macaque does not show such a clear homology to humans. This emphasises the need for exercising caution when moving from simple cues like frequency to timbral cues associated with object identity.

One cannot assume that the macaque model of auditory cortex is a good representation of human cortical organization given the differences in the functional organization of spectral flux between macaques and humans. It is at best a useful model of human auditory cortex with some differences.

REFERENCES

ABELES, M. 1982. *Local cortical circuits: An electrophysiological study*, Springer Science & Business Media.

ABELES, M. 1991. *Corticonics: Neural circuits of the cerebral cortex*, Cambridge University Press.

AERTSEN, A., ERB, M. & PALM, G. 1994. Dynamics of Functional Coupling in the Cerebral-Cortex - an Attempt at a Model-Based Interpretation. *Physica D-Nonlinear Phenomena,* 75**,** 103-128.

ALAIN, C. 2007. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear Res,* 229**,** 225-36.

ALAIS, D., BLAKE, R. & LEE, S. H. 1998. Visual features that vary together over time group together over space. *Nat Neurosci,* 1**,** 160-4.

AMERICAN-STANDARDS-ASSOCIATION & ACOUSTICAL-SOCIETY-OF-AMERICA 1960. *American standard acoustical terminology:(including mechanical shock and vibration)*, American Standards Association.

ANDERSSON, J. L., HUTTON, C., ASHBURNER, J., TURNER, R. & FRISTON, K. 2001. Modeling geometric deformations in EPI time series. *Neuroimage,* 13**,** 903-19.

BARTON, B., VENEZIA, J. H., SABERI, K., HICKOK, G. & BREWER, A. A. 2012. Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc Natl Acad Sci U S A,* 109**,** 20738-43.

BAUMANN, S., GRIFFITHS, T. D., REES, A., HUNTER, D., SUN, L. & THIELE, A. 2010. Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. *Neuroimage,* 50**,** 1099-108.

BAUMANN, S., JOLY, O., REES, A., PETKOV, C. I., SUN, L., THIELE, A. & GRIFFITHS, T. D. 2015. The topography of frequency and time representation in primate auditory cortices. *Elife,* 4.

BAUMANN, S., PETKOV, C. I. & GRIFFITHS, T. D. 2013. A unified framework for the organization of the primate auditory cortex. *Front Syst Neurosci,* 7**,** 11.

BEE, M. A. 2015. Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. *Int J Psychophysiol,* 95**,** 216-37.

BEE, M. A. & KLUMP, G. M. 2004. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol,* 92**,** 1088-104.

BEECHER, M. D., PETERSEN, M. R., ZOLOTH, S. R., MOODY, D. B. & STEBBINS, W. C. 1979. Perception of conspecific vocalizations by Japanese macaques. Evidence for selective attention and neural lateralization. *Brain Behav Evol,* 16**,** 443-60.

BELIN, P., ZATORRE, R. J., HOGE, R., EVANS, A. C. & PIKE, B. 1999. Event-related fMRI of the auditory cortex. *Neuroimage,* 10**,** 417-29.

BELIN, P., ZATORRE, R. J., LAFAILLE, P., AHAD, P. & PIKE, B. 2000. Voice-selective areas in human auditory cortex. *Nature,* 403**,** 309-12.

BELIN, P., ZILBOVICIUS, M., CROZIER, S., THIVARD, L., FONTAINE, A., MASURE, M. C. & SAMSON, Y. 1998. Lateralization of speech and auditory temporal processing. *J Cogn Neurosci,* 10**,** 536-40.

BENDOR, D. & WANG, X. 2007. Differential neural coding of acoustic flutter within primate auditory cortex. *Nat Neurosci,* 10**,** 763-71.

BENDOR, D. & 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.

BIZLEY, J. K. & COHEN, Y. E. 2013. The what, where and how of auditory-object perception. *Nature Reviews Neuroscience,* 14**,** 693-707.

BIZLEY, J. K., NODAL, F. R., NELKEN, I. & KING, A. J. 2005. Functional organization of ferret auditory cortex. *Cereb Cortex,* 15**,** 1637-53.

BLAKE, R. & LEE, S. H. 2005. The role of temporal structure in human vision. *Behav Cogn Neurosci Rev,* 4**,** 21-42.

BLOOMSMITH, M. A., STONE, A. M. & LAULE, G. E. 1998. Positive reinforcement training to enhance the voluntary movement of group-housed chimpanzees within their enclosures. *Zoo Biology,* 17**,** 333-341.

BOEMIO, A., FROMM, S., BRAUN, A. & POEPPEL, D. 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci,* 8**,** 389-95.

BREGMAN, A. S. 1990. *Auditory scene analysis: The perceptual organization of sound*, MIT press.

BREGMAN, A. S. & CAMPBELL, J. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *J Exp Psychol,* 89**,** 244-9.

BREGMAN, A. S. & PINKER, S. 1978. Auditory streaming and the building of timbre. *Can J Psychol,* 32**,** 19-31.

BRETT, M., ANTON, J.-L., VALABREGUE, R. & POLINE, J.-B. Region of interest analysis using an SPM toolbox. 8th international conference on functional mapping of the human brain, 2002. Sendai, 497.

BREWER, A. A. & BARTON, B. 2016. Maps of the Auditory Cortex. *Annu Rev Neurosci,* 39**,** 385-407.

BROADBENT, D. E. & LADEFOGED, P. 1959. Auditory perception of temporal order. *J Acoust Soc Am,* 31**,** 1539-1539.

BROSCH, M. & SCHREINER, C. E. 1997. Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol,* 77**,** 923-43.

BURGOYNE, J. A. & MCADAMS, S. A meta-analysis of timbre perception using nonlinear extensions to CLASCAL. International Symposium on Computer Music Modeling and Retrieval, 2007. Springer, 181-202.

BUXTON, R. B., WONG, E. C. & FRANK, L. R. 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med,* 39**,** 855-64.

CACLIN, A., MCADAMS, S., SMITH, B. K. & WINSBERG, S. 2005. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America,* 118**,** 471-482.

CHAMBERS, J., AKEROYD, M. A., SUMMERFIELD, A. Q. & PALMER, A. R. 2001. Active control of the volume acquisition noise in functional magnetic resonance imaging: method and psychoacoustical evaluation. *J Acoust Soc Am,* 110**,** 3041-54.

CHENEY, D. L. & SEYFARTH, R. M. 1990. *How monkeys see the world: Inside the mind of another species*, University of Chicago Press.

CHERRY, E. C. 1953. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am,* 25**,** 975-979.

CHI, T., RU, P. & SHAMMA, S. A. 2005. Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am,* 118**,** 887-906.

CHRISTISON-LAGAY, K. L. & COHEN, Y. E. 2014. Behavioral correlates of auditory streaming in rhesus macaques. *Hear Res,* 309**,** 17-25.

CIORBA, A., BIANCHINI, C., PELUCCHI, S. & PASTORE, A. 2012. The impact of hearing loss on the quality of life of elderly adults. *Clin Interv Aging,* 7**,** 159-63.

CLACK, T. D. 1966. Effect of signal duration on the auditory sensitivity of humans and monkeys (Macaca mulatta). *J Acoust Soc Am,* 40**,** 1140-6.

CLARKE, S. & MOROSAN, P. 2012. Architecture, connectivity, and transmitter receptors of human auditory cortex. *The human auditory cortex.* Springer.

COHEN, Y. E., THEUNISSEN, F., RUSS, B. E. & GILL, P. 2007. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol,* 97**,** 1470-84.

COHEN, Y. E. & WESSINGER, C. M. 1999. Who goes there? *Neuron,* 24**,** 769-71.

CUSACK, R. 2005. The intraparietal sulcus and perceptual organization. *J Cogn Neurosci,* 17**,** 641-51.

CUSACK, R. & ROBERTS, B. 2000. Effects of differences in timbre on sequential grouping. *Percept Psychophys,* 62**,** 1112-20.

DALLOS, P. 1992. The active cochlea. *J Neurosci,* 12**,** 4575-85.

DARWIN, C. J. & CARLYON, R. P. 1995. Hearing. Academic San Diego, CA.

DARWIN, C. J. & HUKIN, R. W. 1999. Auditory objects of attention: the role of interaural time differences. *J Exp Psychol Hum Percept Perform,* 25**,** 617-29.

DARWIN, C. J., HUKIN, R. W. & AL-KHATIB, B. Y. 1995. Grouping in pitch perception: evidence for sequential constraints. *J Acoust Soc Am,* 98**,** 880-5.

DI SALLE, F., FORMISANO, E., SEIFRITZ, E., LINDEN, D. E., SCHEFFLER, K., SAULINO, C., TEDESCHI, G., ZANELLA, F. E., PEPINO, A., GOEBEL, R. & MARCIANO, E. 2001. Functional fields in human auditory cortex revealed by time-resolved fMRI without interference of EPI noise. *Neuroimage,* 13**,** 328-38.

DONAHUE, C. J., GLASSER, M. F., PREUSS, T. M., RILLING, J. K. & VAN ESSEN, D. C. 2018. Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. *Proc Natl Acad Sci U S A,* 115**,** E5183-E5192.

DUBNO, J. R., DIRKS, D. D. & MORGAN, D. E. 1984. Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am,* 76**,** 87-96.

DYKSTRA, A. R. & GUTSCHALK, A. 2013. Time is of the essence for auditory scene analysis. *Elife,* 2**,** e01136.

EGGERMONT, J. J. 2006. Properties of correlated neural activity clusters in cat auditory cortex resemble those of neural assemblies. *J Neurophysiol,* 96**,** 746-64.

EGGERMONT, J. J. 2007. Correlated neural activity as the driving force for functional changes in auditory cortex. *Hear Res,* 229**,** 69-80.

ELHILALI, M., MA, L., MICHEYL, C., OXENHAM, A. J. & SHAMMA, S. A. 2009a. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron,* 61**,** 317-29.

ELHILALI, M. & SHAMMA, S. A. 2008. A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J Acoust Soc Am,* 124**,** 3751-71.

ELHILALI, M., XIANG, J., SHAMMA, S. A. & 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.

ELLIOTT, T. M., HAMILTON, L. S. & THEUNISSEN, F. E. 2013. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J Acoust Soc Am,* 133**,** 389-404.

ERB, J., ARMENDARIZ, M., DE MARTINO, F., GOEBEL, R., VANDUFFEL, W. & FORMISANO, E. 2019. Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex. *Cereb Cortex,* 29**,** 3636-3650.

FAY, R. R. 1992. Analytic listening by the goldfish. *Hear Res,* 59**,** 101-7.

FAY, R. R. 1998. Auditory stream segregation in goldfish (Carassius auratus). *Hear Res,* 120**,** 69-76.

FENG, L. & WANG, X. 2017. Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proc Natl Acad Sci U S A,* 114**,** E840-E848.

FERNSTRÖM, A. L., FREDLUND, H., SPÅNGBERG, M. & WESTLUND, K. 2009. Positive reinforcement training in rhesus macaques—training progress as a result of training frequency. *Am J Primatol,* 71**,** 373-379.

FISHMAN, Y. I., AREZZO, J. C. & STEINSCHNEIDER, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am,* 116**,** 1656-70.

FISHMAN, Y. I., KIM, M. & STEINSCHNEIDER, M. 2017. A Crucial Test of the Population Separation Model of Auditory Stream Segregation in Macaque Primary Auditory Cortex. *J Neurosci,* 37**,** 10645-10655.

FISHMAN, Y. I., RESER, D. H., AREZZO, J. C. & STEINSCHNEIDER, M. 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res,* 151**,** 167-187.

FORMISANO, E., KIM, D. S., DI SALLE, F., VAN DE MOORTELE, P. F., UGURBIL, K. & GOEBEL, R. 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron,* 40**,** 859-69.

FOX, P. T. & RAICHLE, M. E. 1986. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc Natl Acad Sci U S A,* 83**,** 1140-4.

FRISTON, K. J., ASHBURNER, J., FRITH, C. D., POLINE, J. B., HEATHER, J. D. & FRACKOWIAK, R. S. J. 1995. Spatial registration and normalization of images. *Human Brain Mapping,* 3**,** 165-189.

FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D. & FRACKOWIAK, R. S. 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum brain mapp,* 2**,** 189-210.

FRISTON, K. J., WILLIAMS, S., HOWARD, R., FRACKOWIAK, R. S. & TURNER, R. 1996. Movement-related effects in fMRI time-series. *Magn Reson Med,* 35**,** 346-55.

FRITZ, J., MISHKIN, M. & SAUNDERS, R. C. 2005. In search of an auditory engram. *Proc Natl Acad Sci U S A,* 102**,** 9359-64.

FRITZ, J. B., DAVID, S. V., RADTKE-SCHULLER, S., YIN, P. & SHAMMA, S. A. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci,* 13**,** 1011-9.

FULLERTON, B. C. & PANDYA, D. N. 2007. Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *J Comp Neurol,* 504**,** 470-98.

GAAB, N., GABRIELI, J. D. & GLOVER, G. H. 2007. Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum Brain Mapp,* 28**,** 721-32.

GALABURDA, A. & SANIDES, F. 1980. Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol,* 190**,** 597-610.

GALABURDA, A. M. & PANDYA, D. N. 1983. The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey. *J Comp Neurol,* 221**,** 169-84.

GANNON, P. J., HOLLOWAY, R. L., BROADFIELD, D. C. & BRAUN, A. R. 1998. Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science,* 279**,** 220-2.

GENOVESE, C. R., LAZAR, N. A. & NICHOLS, T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage,* 15**,** 870-8.

GEYER, S., WEISS, M., REIMANN, K., LOHMANN, G. & TURNER, R. 2011. Microstructural Parcellation of the Human Cerebral Cortex - From Brodmann's Post-Mortem Map to in vivo Mapping with High-Field Magnetic Resonance Imaging. *Front Hum Neurosci,* 5**,** 19.

GHAZANFAR, A. A. & HAUSER, M. D. 1999. The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cogn Sci,* 3**,** 377-384.

GIRAUD, A. L., KLEINSCHMIDT, A., POEPPEL, D., LUND, T. E., FRACKOWIAK, R. S. & LAUFS, H. 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron,* 56**,** 1127-34.

GIRAUD, A. L., LORENZI, C., ASHBURNER, J., WABLE, J., JOHNSRUDE, I., FRACKOWIAK, R. & KLEINSCHMIDT, A. 2000. Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol,* 84**,** 1588-98.

GLASSER, M. F. & VAN ESSEN, D. C. 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci,* 31**,** 11597-616.

GRANDE, L. A., KINNEY, G. A., MIRACLE, G. L. & SPAIN, W. J. 2004. Dynamic influences on coincidence detection in neocortical pyramidal neurons. *J Neurosci,* 24**,** 1839-51.

GREFKES, C. & FINK, G. R. 2005. The functional organization of the intraparietal sulcus in humans and monkeys. *J Anat,* 207**,** 3-17.

GREY, J. M. 1977. Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am,* 61**,** 1270-7.

GREY, J. M. & GORDON, J. W. 1978. Perceptual Effects of Spectral Modifications on Musical Timbres. *Journal of the Acoustical Society of America,* 63**,** 1493-1500.

GRIFFITHS, T., MICHEYL, C. & OVERATH, T. 2012. Auditory Object Analysis. *In:* POEPPEL, D., OVERATH, T., POPPER, A. N. & FAY, R. R. (eds.) *The Human Auditory Cortex.* Springer New York.

GRIFFITHS, T. D. & WARREN, J. D. 2004. What is an auditory object? *Nature Reviews Neuroscience,* 5**,** 887-892.

GRIFFITHS, T. M. C. I. C. P. D. P. G. S. K. S. T. 2017. Figure-ground analysis based on the sequential grouping of spectral patterns. *In:* (ARO), A. F. R. I. O. (ed.) *ARO Annual MidWinter Meeting* Baltimore.

GRIMAULT, N., BACON, S. P. & MICHEYL, C. 2002. Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am,* 111**,** 1340-8.

GROOTOONK, S., HUTTON, C., ASHBURNER, J., HOWSEMAN, A. M., JOSEPHS, O., REES, G., FRISTON, K. J. & TURNER, R. 2000. Characterization and correction of interpolation effects in the realignment of fMRI time series. *Neuroimage,* 11**,** 49-57.

GRUETTER, R. 1993. Automatic, localized in Vivo adjustment of all first- and second-order shim coils. *Magn Reson Med,* 29**,** 804-811.

GUTSCHALK, A., MICHEYL, C., MELCHER, J. R., RUPP, A., SCHERG, M. & OXENHAM, A. J. 2005. Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci,* 25**,** 5382-8.

GUTSCHALK, A., MICHEYL, C. & OXENHAM, A. J. 2008. 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**,** 133-46.

HACKETT, T. A., DE LA MOTHE, L. A., CAMALIER, C. R., FALCHIER, A., LAKATOS, P., KAJIKAWA, Y. & SCHROEDER, C. E. 2014. Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. *Front Neurosci,* 8**,** 72.

HACKETT, T. A., PREUSS, T. M. & 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., STEPNIEWSKA, I. & KAAS, J. H. 1998a. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol,* 394**,** 475-95.

HACKETT, T. A., STEPNIEWSKA, I. & KAAS, J. H. 1998b. Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol,* 400**,** 271-86.

HALL, D. A., HAGGARD, M. P., AKEROYD, M. A., PALMER, A. R., SUMMERFIELD, A. Q., ELLIOTT, M. R., GURNEY, E. M. & BOWTELL, R. W. 1999. "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp,* 7**,** 213-23.

HALL, D. A., SUMMERFIELD, A. Q., GONCALVES, M. S., FOSTER, J. R., PALMER, A. R. & BOWTELL, R. W. 2000. Time-course of the auditory BOLD response to scanner noise. *Magn Reson Med,* 43**,** 601-6.

HALL, J. W., HAGGARD, M. P. & FERNANDES, M. A. 1984. Detection in noise by spectro-temporal pattern analysis. *J Acoust Soc Am,* 76**,** 50-6.

HARPER, N. S., SCHOPPE, O., WILLMORE, B. D., CUI, Z., SCHNUPP, J. W. & KING, A. J. 2016. Network Receptive Field Modeling Reveals Extensive Integration and Multi-feature Selectivity in Auditory Cortical Neurons. *PLoS Comput Biol,* 12**,** e1005113.

HAUSER, M. 1998. Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Anim Behav,* 55**,** 1647-58.

HAUSER, M. D. 1996. *The evolution of communication*.

HEFFNER, H. E. & HEFFNER, R. S. 1986a. Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *J Neurophysiol,* 56**,** 683-701.

HEFFNER, H. E. & HEFFNER, R. S. 1986b. Hearing loss in Japanese macaques following bilateral auditory cortex lesions. *J Neurophysiol,* 55**,** 256-71.

HEFFNER, H. E. & HEFFNER, R. S. 1990. Effect of bilateral auditory cortex lesions on absolute thresholds in Japanese macaques. *J Neurophysiol,* 64**,** 191-205.

HERDENER, M., ESPOSITO, F., SCHEFFLER, K., SCHNEIDER, P., LOGOTHETIS, N. K., ULUDAG, K. & KAYSER, C. 2013. Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex,* 49**,** 2822-33.

HILL, K. T., BISHOP, C. W., YADAV, D. & MILLER, L. M. 2011. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci,* 12**,** 85.

HOPP, S. L., SINNOTT, J. M., OWREN, M. J. & PETERSEN, M. R. 1992. Differential sensitivity of Japanese macaques (Macaca fuscata) and humans (Homo sapiens) to peak position along a synthetic coo call continuum. *J Comp Psychol,* 106**,** 128-36.

HOWARD, M. A., 3RD, VOLKOV, I. O., ABBAS, P. J., DAMASIO, H., OLLENDIECK, M. C. & GRANNER, M. A. 1996. A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res,* 724**,** 260-4.

HULSE, S. H. 2002. Auditory scene analysis in animal communication. *Advances in the Study of Behavior, Vol 31,* 31**,** 163-200.

HULSE, S. H., MACDOUGALL-SHACKLETON, S. A. & WISNIEWSKI, A. B. 1997. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris). *J Comp Psychol,* 111**,** 3-13.

HUSAIN, F. T., TAGAMETS, M. A., FROMM, S. J., BRAUN, A. R. & HORWITZ, B. 2004. Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *Neuroimage,* 21**,** 1701-20.

HYDER, F., ROTHMAN, D. L. & SHULMAN, R. G. 2002. Total neuroenergetics support localized brain activity: implications for the interpretation of fMRI. *Proc Natl Acad Sci U S A,* 99**,** 10771-6.

IVERSON, P. 1995. Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *J Exp Psychol Hum Percept Perform,* 21**,** 751-63.

IVERSON, P. & KRUMHANSL, C. L. 1993. Isolating the dynamic attributes of musical timbre. *J Acoust Soc Am,* 94**,** 2595-603.

IZUMI, A. 2002. Auditory stream segregation in Japanese monkeys. *Cognition,* 82**,** B113-22.

JACKSON, L. L., HEFFNER, R. S. & HEFFNER, H. E. 1999. Free-field audiogram of the Japanese macaque (Macaca fuscata). *J Acoust Soc Am,* 106**,** 3017-23.

JACOBS, G. H. & DEEGAN, J. F., 2ND 1997. Spectral sensitivity of macaque monkeys measured with ERG flicker photometry. *Vis Neurosci,* 14**,** 921-8.

JAMISON, H. L., WATKINS, K. E., BISHOP, D. V. & MATTHEWS, P. M. 2006. Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb Cortex,* 16**,** 1266-75.

JOLY, O., BAUMANN, S., BALEZEAU, F., THIELE, A. & GRIFFITHS, T. D. 2014a. Merging functional and structural properties of the monkey auditory cortex. *Front Neurosci,* 8**,** 198.

JOLY, O., BAUMANN, S., POIRIER, C., PATTERSON, R. D., THIELE, A. & GRIFFITHS, T. D. 2014b. A perceptual pitch boundary in a non-human primate. *Front Psychol,* 5**,** 998.

JOLY, O., RAMUS, F., PRESSNITZER, D., VANDUFFEL, W. & ORBAN, G. A. 2012. Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cereb Cortex,* 22**,** 838-53.

JONES, E. G., DELL'ANNA, M. E., MOLINARI, M., RAUSELL, E. & HASHIKAWA, T. 1995. Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J Comp Neurol,* 362**,** 153-70.

KAAS, J. H. 2004. Evolution of somatosensory and motor cortex in primates. *Anat Rec A Discov Mol Cell Evol Biol,* 281**,** 1148-56.

KAAS, J. H. & HACKETT, T. A. 1999. 'What' and 'where' processing in auditory cortex. *Nat Neurosci,* 2**,** 1045-7.

KAAS, J. H. & HACKETT, T. A. 2000. Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A,* 97**,** 11793-9.

KAAS, J. H., HACKETT, T. A. & TRAMO, M. J. 1999. Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol,* 9**,** 164-70.

KAJIKAWA, Y., FREY, S., ROSS, D., FALCHIER, A., HACKETT, T. A. & 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**,** 4140-50.

KANAYAMA, S., KUHARA, S. & SATOH, K. 1996. In vivo rapid magnetic field measurement and shimming using single scan differential phase mapping. *Magn Reson Med,* 36**,** 637-42.

KANT, I. 1929. Critique of pure reason. *Cambridge: Cambridge University Press;(1781/translated 1999)*.

KANWAL, J. S., MEDVEDEV, A. V. & MICHEYL, C. 2003. Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network,* 14**,** 413-35.

KIDD, G., JR., MASON, C. R., DELIWALA, P. S., WOODS, W. S. & COLBURN, H. S. 1994. Reducing informational masking by sound segregation. *J Acoust Soc Am,* 95**,** 3475-80.

KIKUCHI, Y., HORWITZ, B. & MISHKIN, M. 2010. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci,* 30**,** 13021-30.

KIMPO, R. R., THEUNISSEN, F. E. & DOUPE, A. J. 2003. Propagation of correlated activity through multiple stages of a neural circuit. *J Neurosci,* 23**,** 5750-61.

KISTLER, W. M. & GERSTNER, W. 2002. Stable propagation of activity pulses in populations of spiking neurons. *Neural Comput,* 14**,** 987-97.

KONDO, H. M. & KASHINO, M. 2009. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J Neurosci,* 29**,** 12695-701.

KOSAKI, H., HASHIKAWA, T., HE, J. & JONES, E. G. 1997. Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol,* 386**,** 304-16.

KRIMPHOFF, J., MCADAMS, S., WINSBERG, S., PETIT, H., BAKCHINE, S., DUBOIS, B., LAURENT, B., MONTAGNE, B., TOUCHON, J. & ROBERT, P. 1994. Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification. *J. de Physique,* 4**,** 625-628.

KRUMHANSL, C. L. 1989. Why Is Musical Timbre So Hard to Understand. *Structure and Perception of Electroacoustic Sound and Music,* 846**,** 43-53.

KUSMIEREK, P. & RAUSCHECKER, J. P. 2009. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol,* 102**,** 1606-22.

LAKATOS, S. 2000. A common perceptual space for harmonic and percussive timbres. *Percept Psychophys,* 62**,** 1426-39.

LEAVER, A. M. & RAUSCHECKER, J. P. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci,* 30**,** 7604-12.

LICKLIDER, J. C. R. 1951. Basic correlates of the auditory stimulus.

LIN, F. R., THORPE, R., GORDON-SALANT, S. & FERRUCCI, L. 2011. Hearing loss prevalence and risk factors among older adults in the United States. *J Gerontol A Biol Sci Med Sci,* 66**,** 582-90.

LINDEN, J. F. & SCHREINER, C. E. 2003. Columnar transformations in auditory cortex? A comparison to visual and somatosensory cortices. *Cereb Cortex,* 13**,** 83-9.

LIPP, R., KITTERICK, P., SUMMERFIELD, Q., BAILEY, P. J. & PAUL-JORDANOV, I. 2010. Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia,* 48**,** 1417-25.

LOGAN, B. R., GELIAZKOVA, M. P. & ROWE, D. B. 2008. An evaluation of spatial thresholding techniques in fMRI analysis. *Hum Brain Mapp,* 29**,** 1379-89.

LOGOTHETIS, N. K. 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci,* 357**,** 1003-37.

LOGOTHETIS, N. K. 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci,* 23**,** 3963-71.

LOGOTHETIS, N. K. 2008. What we can do and what we cannot do with fMRI. *Nature,* 453**,** 869-878.

LOGOTHETIS, N. K. 2012. Intracortical recordings and fMRI: an attempt to study operational modules and networks simultaneously. *Neuroimage,* 62**,** 962-9.

LOGOTHETIS, N. K., PAULS, J., AUGATH, M., TRINATH, T. & OELTERMANN, A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature,* 412**,** 150-7.

LOGOTHETIS, N. K. & WANDELL, B. A. 2004. Interpreting the BOLD signal. *Annu Rev Physiol,* 66**,** 735-69.

LU, K., XU, Y., YIN, P., OXENHAM, A. J., FRITZ, J. B. & SHAMMA, S. A. 2017. Temporal coherence structure rapidly shapes neuronal interactions. *Nat Commun,* 8**,** 13900.

LUO, H. & POEPPEL, D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron,* 54**,** 1001-10.

LYON, R. & SHAMMA, S. 1996. Auditory representations of timbre and pitch. *Auditory computation.* Springer.

MA, L., MICHEYL, C., YIN, P., OXENHAM, A. J. & SHAMMA, S. A. 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). *J Comp Psychol,* 124**,** 317-30.

MACDOUGALL-SHACKLETON, S. A., HULSE, S. H., GENTNER, T. Q. & WHITE, W. 1998. Auditory scene analysis by European starlings (Sturnus vulgaris): perceptual segregation of tone sequences. *J Acoust Soc Am,* 103**,** 3581-7.

MALONEK, D. & GRINVALD, A. 1996. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science,* 272**,** 551-4.

MAY, B., MOODY, D. B. & STEBBINS, W. C. 1989. Categorical perception of conspecific communication sounds by Japanese macaques, Macaca fuscata. *J Acoust Soc Am,* 85**,** 837-47.

MCADAMS, S. 1999. Perspectives on the contribution of timbre to musical structure. *Computer Music Journal,* 23**,** 85-102.

MCADAMS, S. & CUNIBLE, J. C. 1992. Perception of timbral analogies. *Philos Trans R Soc Lond B Biol Sci,* 336**,** 383-9.

MCADAMS, S. & GIORDANO, B. L. 2009. The perception of musical timbre. *The Oxford handbook of music psychology***,** 72-80.

MCADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G. & KRIMPHOFF, J. 1995. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res,* 58**,** 177-92.

MCDONALD, K. L. & ALAIN, C. 2005. Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J Acoust Soc Am,* 118**,** 1593-604.

MENON, V., LEVITIN, D. J., SMITH, B. K., LEMBKE, A., KRASNOW, B. D., GLAZER, D., GLOVER, G. H. & MCADAMS, S. 2002. Neural correlates of timbre change in harmonic sounds. *Neuroimage,* 17**,** 1742-54.

MERZENICH, M. M. & BRUGGE, J. F. 1973. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res,* 50**,** 275-96.

MESULAM, M. M. 1998. From sensation to cognition. *Brain,* 121 (Pt 6)**,** 1013-52.

MICHEYL, C., CARLYON, R. P., GUTSCHALK, A., MELCHER, J. R., OXENHAM, A. J., RAUSCHECKER, J. P., TIAN, B. & COURTENAY WILSON, E. 2007a. The role of auditory cortex in the formation of auditory streams. *Hear Res,* 229**,** 116-31.

MICHEYL, C., SHAMMA, S. A. & OXENHAM, A. J. 2007b. Hearing out repeating elements in randomly varying multitone sequences: a case of streaming? *Hearing–From Sensory Processing to Perception.* Springer.

MICHEYL, C., TIAN, B., CARLYON, R. P. & RAUSCHECKER, J. P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron,* 48**,** 139-48.

MIRBAGHERI, M., AKRAM, S. & SHAMMA, S. An auditory inspired multimodal framework for speech enhancement. Thirteenth Annual Conference of the International Speech Communication Association, 2012.

MOELKER, A. & PATTYNAMA, P. M. 2003. Acoustic noise concerns in functional magnetic resonance imaging. *Hum Brain Mapp,* 20**,** 123-41.

MOEREL, M., DE MARTINO, F. & FORMISANO, E. 2014. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci,* 8**,** 225.

MOEREL, M., DE MARTINO, F., SANTORO, R., UGURBIL, K., GOEBEL, R., YACOUB, E. & FORMISANO, E. 2013. Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *J Neurosci,* 33**,** 11888-98.

MOORE, B. C., GLASBERG, B. R. & PETERS, R. W. 1986. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J Acoust Soc Am,* 80**,** 479-83.

MOORE, B. C. J. & GOCKEL, H. 2002. Factors influencing sequential stream segregation. *Acta Acustica United with Acustica,* 88**,** 320-333.

MOREL, A., GARRAGHTY, P. E. & KAAS, J. H. 1993. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol,* 335**,** 437-59.

MOROSAN, P., RADEMACHER, J., SCHLEICHER, A., AMUNTS, K., SCHORMANN, T. & ZILLES, K. 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage,* 13**,** 684-701.

NANDY, R. & CORDES, D. 2007. A semi-parametric approach to estimate the family-wise error rate in fMRI using resting-state data. *Neuroimage,* 34**,** 1562-76.

NATAN, R. G., CARRUTHERS, I. M., MWILAMBWE-TSHILOBO, L. & GEFFEN, M. N. 2017. Gain Control in the Auditory Cortex Evoked by Changing Temporal Correlation of Sounds. *Cereb Cortex,* 27**,** 2385-2402.

NEVIN, J. A. 1969. SIGNAL DETECTION THEORY AND OPERANT BEHAVIOR: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. *J Exp Anal of Behav,* 12**,** 475-480.

NITYANANDA, V. & BEE, M. A. 2011. Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PLoS One,* 6**,** e21191.

NODA, T., KANZAKI, R. & TAKAHASHI, H. 2013. Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. *PLoS One,* 8**,** e83544.

O'CONNOR, K. N., BARRUEL, P., HAJALILOU, R. & SUTTER, M. L. 1999. Auditory temporal integration in the rhesus macaque (Macaca mulatta). *J Acoust Soc Am,* 106**,** 954-65.

O'CONNOR, K. N., JOHNSON, J. S., NIWA, M., NORIEGA, N. C., MARSHALL, E. A. & SUTTER, M. L. 2011. Amplitude modulation detection as a function of modulation frequency and stimulus duration: comparisons between macaques and humans. *Hear Res,* 277**,** 37-43.

O'SULLIVAN, J. A., SHAMMA, S. A. & LALOR, E. C. 2015. Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. *J Neurosci,* 35**,** 7256-63.

OBLESER, J., EISNER, F. & KOTZ, S. A. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci,* 28**,** 8116-23.

OGAWA, S. 2012. Finding the BOLD effect in brain images. *Neuroimage,* 62**,** 608-9.

OGAWA, S., LEE, T. M., KAY, A. R. & TANK, D. W. 1990a. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A,* 87**,** 9868-72.

OGAWA, S., LEE, T. M., NAYAK, A. S. & GLYNN, P. 1990b. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med,* 14**,** 68-78.

OGAWA, S., TANK, D. W., MENON, R., ELLERMANN, J. M., KIM, S. G., MERKLE, H. & UGURBIL, K. 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A,* 89**,** 5951-5.

ORBAN, G. A., VAN ESSEN, D. & VANDUFFEL, W. 2004. Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn Sci,* 8**,** 315-24.

OVERATH, T., KUMAR, S., VON KRIEGSTEIN, K. & GRIFFITHS, T. D. 2008. Encoding of spectral correlation over time in auditory cortex. *J Neurosci,* 28**,** 13268-73.

PANDYA, D. N. 1995. Anatomy of the auditory cortex. *Rev Neurol (Paris),* 151**,** 486-94.

PANDYA, D. N. & KUYPERS, H. G. 1969. Cortico-cortical connections in the rhesus monkey. *Brain Res,* 13**,** 13-36.

PAPEZ, J. W. 1929. *Comparative neurology: a manual and text for the study of the nervous system of vertebrates*, Thomas Y. Crowell Company.

PAULING, L. & CORYELL, C. D. 1936. The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc Natl Acad Sci U S A,* 22**,** 210-6.

PETKOV, C. I., O'CONNOR, K. N. & SUTTER, M. L. 2003. Illusory sound perception in macaque monkeys. *J Neurosci,* 23**,** 9155-61.

PETKOV, C. I., O'CONNOR, K. N. & SUTTER, M. L. 2007. Encoding of illusory continuity in primary auditory cortex. *Neuron,* 54**,** 153-65.

POEPPEL, D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun,* 41**,** 245-255.

POREMBA, A., SAUNDERS, R. C., CRANE, A. M., COOK, M., SOKOLOFF, L. & MISHKIN, M. 2003. Functional mapping of the primate auditory system. *Science,* 299**,** 568-72.

PRESCOTT, M. J. & BUCHANAN-SMITH, H. M. 2007. Training laboratory-housed non-human primates, part I: a UK survey. *Animal Welfare,* 16**,** 21-36.

PRESSNITZER, D., SAYLES, M., MICHEYL, C. & WINTER, I. M. 2008. Perceptual organization of sound begins in the auditory periphery. *Curr Biol,* 18**,** 1124-8.

PRICE, D. L., DE WILDE, J. P., PAPADAKI, A. M., CURRAN, J. S. & KITNEY, R. I. 2001. Investigation of acoustic noise on 15 MRI scanners from 0.2 T to 3 T. *J Magn Reson Imaging,* 13**,** 288-93.

RAUSCHECKER, J. P. 1998. Parallel processing in the auditory cortex of primates. *Audiol Neurootol,* 3**,** 86-103.

RAUSCHECKER, J. P. & TIAN, B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A,* 97**,** 11800-6.

RAUSCHECKER, J. P. & TIAN, B. 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol,* 91**,** 2578-89.

RAUSCHECKER, J. P., TIAN, B. & HAUSER, M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science,* 268**,** 111-4.

RAUSCHECKER, J. P., TIAN, B., PONS, T. & MISHKIN, M. 1997. Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol,* 382**,** 89-103.

RAVICZ, M. E. & MELCHER, J. R. 2001. Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am,* 109**,** 216-31.

RAVICZ, M. E., MELCHER, J. R. & KIANG, N. Y. 2000. Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am,* 108**,** 1683-96.

RECANZONE, G. H. 2000. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hear Res,* 150**,** 104-18.

RECANZONE, G. H., GUARD, D. C. & PHAN, M. L. 2000a. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol,* 83**,** 2315-31.

RECANZONE, G. H., GUARD, D. C., PHAN, M. L. & SU, T. K. 2000b. Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol,* 83**,** 2723-39.

REVELEY, C., GRUSLYS, A., YE, F. Q., GLEN, D., SAMAHA, J., B, E. R., SAAD, Z., A, K. S., LEOPOLD, D. A. & SALEEM, K. S. 2017. Three-Dimensional Digital Template Atlas of the Macaque Brain. *Cereb Cortex,* 27**,** 4463-4477.

RIECKE, L., SACK, A. T. & SCHROEDER, C. E. 2015. Endogenous Delta/Theta Sound-Brain Phase Entrainment Accelerates the Buildup of Auditory Streaming. *Curr Biol,* 25**,** 3196-201.

ROBERTS, B., GLASBERG, B. R. & MOORE, B. C. 2002. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am,* 112**,** 2074-85.

ROLAND, P. E. & ZILLES, K. 1996. Functions and structures of the motor cortices in humans. *Curr Opin Neurobiol,* 6**,** 773-81.

ROMANSKI, L. M., TIAN, B., FRITZ, J. B., MISHKIN, M., GOLDMAN-RAKIC, P. S. & RAUSCHECKER, J. P. 2000. Reply to "What', 'where' and 'how' in auditory cortex'. *Nat Neurosci,* 3**,** 966.

ROSEN, S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci,* 336**,** 367-73.

RUDOLPH, M. & DESTEXHE, A. 2003. Tuning neocortical pyramidal neurons between integrators and coincidence detectors. *J Comput Neurosci,* 14**,** 239-51.

SALEEM, K. S. & LOGOTHETIS, N. K. 2012. *A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates*, Academic Press.

SAMSON, S., ZATORRE, R. J. & RAMSAY, J. O. 2002. Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain,* 125**,** 511-23.

SANDER, K., BRECHMANN, A. & SCHEICH, H. 2003. Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Res Brain Res Protoc,* 11**,** 81-91.

SANTORO, R., MOEREL, M., DE MARTINO, F., GOEBEL, R., UGURBIL, K., YACOUB, E. & FORMISANO, E. 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol,* 10**,** e1003412.

SCHEICH, H., BAUMGART, F., GASCHLER-MARKEFSKI, B., TEGELER, C., TEMPELMANN, C., HEINZE, H. J., SCHINDLER, F. & STILLER, D. 1998. Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. *Eur J Neurosci,* 10**,** 803-9.

SCHNUPP, J., NELKEN, I. & KING, A. 2011. *Auditory neuroscience: Making sense of sound*, MIT Press.

SCHOLES, C., PALMER, A. R. & SUMNER, C. J. 2015. Stream segregation in the anesthetized auditory cortex. *Hear Res,* 328**,** 48-58.

SCHÖNWIESNER, M., RÜBSAMEN, R. & VON CRAMON, D. Y. 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci,* 22**,** 1521-1528.

SCOTT, B. H., LECCESE, P. A., SALEEM, K. S., KIKUCHI, Y., MULLARKEY, M. P., FUKUSHIMA, M., MISHKIN, M. & SAUNDERS, R. C. 2017. Intrinsic Connections of the Core Auditory Cortical Regions and Rostral Supratemporal Plane in the Macaque Monkey. *Cereb Cortex,* 27**,** 809-840.

SCOTT, B. H., MALONE, B. J. & SEMPLE, M. N. 2011. Transformation of temporal processing across auditory cortex of awake macaques. *J Neurophysiol,* 105**,** 712-30.

SEIDLITZ, J., SPONHEIM, C., GLEN, D., YE, F. Q., SALEEM, K. S., LEOPOLD, D. A., UNGERLEIDER, L. & MESSINGER, A. 2018. A population MRI brain template and analysis tools for the macaque. *Neuroimage,* 170**,** 121-131.

SEMPLE, M. N. & SCOTT, B. H. 2003. Cortical mechanisms in hearing. *Curr Opin Neurobiol,* 13**,** 167-73.

SHAMMA, S. A., ELHILALI, M. & MICHEYL, C. 2011. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci,* 34**,** 114-23.

SHAMMA, S. A. & MICHEYL, C. 2010. Behind the scenes of auditory perception. *Curr Opin Neurobiol,* 20**,** 361-6.

SINGH, P. G. 1987. Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre? *J Acoust Soc Am,* 82**,** 886-99.

SINGH, P. G. & BREGMAN, A. S. 1997. The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *J Acoust Soc Am,* 102**,** 1943-52.

SINNOTT, J. M. & BROWN, C. H. 1993. Effects of varying signal duration on pure-tone frequency discrimination in humans and monkeys. *J Acoust Soc Am,* 93**,** 1541-6.

SINNOTT, J. M. & BROWN, C. H. 1997. Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *J Acoust Soc Am,* 102**,** 588-602.

SINNOTT, J. M., OWREN, M. J. & PETERSEN, M. R. 1987. Auditory duration discrimination in Old World monkeys (Macaca, Cercopithecus) and humans. *J Acoust Soc Am,* 82**,** 465-70.

SMITH, A. J., BLUMENFELD, H., BEHAR, K. L., ROTHMAN, D. L., SHULMAN, R. G. & HYDER, F. 2002. Cerebral energetics and spiking frequency: the neurophysiological basis of fMRI. *Proc Natl Acad Sci U S A,* 99**,** 10765-70.

SNYDER, J. S., ALAIN, C. & PICTON, T. W. 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci,* 18**,** 1-13.

SNYDER, J. S., GREGG, M. K., WEINTRAUB, D. M. & ALAIN, C. 2012. Attention, awareness, and the perception of auditory scenes. *Front Psychol,* 3**,** 15.

STAINSBY, T. H., FULLGRABE, C., FLANAGAN, H. J., WALDMAN, S. K. & MOORE, B. C. 2011. Sequential streaming due to manipulation of interaural time differences. *J Acoust Soc Am,* 130**,** 904-14.

STEINSCHNEIDER, M., VOLKOV, I. O., NOH, M. D., GARELL, P. C. & HOWARD, M. A., 3RD 1999. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol,* 82**,** 2346-57.

SUSSMAN, E. S., HORVATH, J., WINKLER, I. & ORR, M. 2007. The role of attention in the formation of auditory streams. *Percept Psychophys,* 69**,** 136-52.

TALAVAGE, T. M. & EDMISTER, W. B. 2004. Nonlinearity of FMRI responses in human auditory cortex. *Hum Brain Mapp,* 22**,** 216-28.

TALAVAGE, T. M. & HALL, D. A. 2012. How challenges in auditory fMRI led to general advancements for the field. *Neuroimage,* 62**,** 641-7.

TEKI, S., BARASCUD, N., PICARD, S., PAYNE, C., GRIFFITHS, T. D. & CHAIT, M. 2016. Neural Correlates of Auditory Figure-Ground

Segregation Based on Temporal Coherence. *Cereb Cortex,* 26**,** 3669-80.

TEKI, S., CHAIT, M., KUMAR, S., SHAMMA, S. & GRIFFITHS, T. D. 2013. Segregation of complex acoustic scenes based on temporal coherence. *Elife,* 2**,** e00699.

TEKI, S., CHAIT, M., KUMAR, S., VON KRIEGSTEIN, K. & GRIFFITHS, T. D. 2011. Brain bases for auditory stimulus-driven figure-ground segregation. *J Neurosci,* 31**,** 164-71.

THIELE, A., DELICATO, L. S., ROBERTS, M. J. & 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.

TIAN, B., RESER, D., DURHAM, A., KUSTOV, A. & RAUSCHECKER, J. P. 2001. Functional specialization in rhesus monkey auditory cortex. *Science,* 292**,** 290-3.

TOOTELL, R. B., MENDOLA, J. D., HADJIKHANI, N. K., LEDDEN, P. J., LIU, A. K., REPPAS, J. B., SERENO, M. I. & DALE, A. M. 1997. Functional analysis of V3A and related areas in human visual cortex. *J Neurosci,* 17**,** 7060-78.

TOOTELL, R. B., REPPAS, J. B., KWONG, K. K., MALACH, R., BORN, R. T., BRADY, T. J., ROSEN, B. R. & BELLIVEAU, J. W. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci,* 15**,** 3215-30.

TOOTELL, R. B. & TAYLOR, J. B. 1995. Anatomical evidence for MT and additional cortical visual areas in humans. *Cereb Cortex,* 5**,** 39-55.

TREISMAN, A. M. & GELADE, G. 1980. A feature-integration theory of attention. *Cogn Psychol,* 12**,** 97-136.

TSAO, D. Y., MOELLER, S. & FREIWALD, W. A. 2008. Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A,* 105**,** 19514-9.

VAN NOORDEN, L. P. A. S. 1975. Temporal coherence in the perception of tone sequences.

VEEDER, C. L., BLOOMSMITH, M. A., MCMILLAN, J. L., PERLMAN, J. E. & MARTIN, A. L. 2009. Positive reinforcement training to enhance the voluntary movement of group-housed sooty mangabeys (Cercocebus atys atys). *J Am Assoc Lab Anim Sci,* 48**,** 192-5.

VLIEGEN, J. & OXENHAM, A. J. 1999. Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am,* 105**,** 339-46.

VON BEKESY, G. 1970. Travelling waves as frequency analysers in the cochlea. *Nature,* 225**,** 1207-9.

WANG, X. 2000. On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A,* 97**,** 11843-9.

WANG, X., LU, T., BENDOR, D. & BARTLETT, E. 2008. Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience,* 154**,** 294-303.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M. & WARREN, R. P. 1969. Auditory sequence: confusion of patterns other than speech or music. *Science,* 164**,** 586-7.

WESSEL, D. L. 1979. Timbre space as a musical control structure. *Comput Music J***,** 45-52.

WIEGAND, K. & GUTSCHALK, A. 2012. Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *Neuroimage,* 61**,** 62-9.

WILSON, E. C., MELCHER, J. R., MICHEYL, C., GUTSCHALK, A. & OXENHAM, A. J. 2007. Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol,* 97**,** 2230-8.

WINGFIELD, A., TUN, P. A. & MCCOY, S. L. 2005. Hearing loss in older adulthood - What it is and how it interacts with cognitive

performance. *Current Directions in Psychological Science,* 14**,** 144-148.

ZARCO, W., MERCHANT, H., PRADO, L. & MENDEZ, J. C. 2009. Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol,* 102**,** 3191-202.

ZATORRE, R. J. & BELIN, P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex,* 11**,** 946-53.

ZATORRE, R. J., BELIN, P. & PENHUNE, V. B. 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci,* 6**,** 37-46.

ZATORRE, R. J., BOUFFARD, M. & BELIN, P. 2004. Sensitivity to auditory object features in human temporal neocortex. *J Neurosci,* 24**,** 3637-42.

ZATORRE, R. J. & GANDOUR, J. T. 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci,* 363**,** 1087-104.

ZILLES, K., SCHLAUG, G., MATELLI, M., LUPPINO, G., SCHLEICHER, A., QU, M., DABRINGHAUS, A., SEITZ, R. & ROLAND, P. E. 1995. Mapping of human and macaque sensorimotor areas by integrating architectonic, transmitter receptor, MRI and PET data. *J Anat,* 187 (Pt 3)**,** 515-37.

ZOLOTH, S. & GREEN, S. 1979. Monkey vocalizations and human speech: parallels in perception? *Brain Behav Evol,* 16**,** 430-42.

APPENDIX A

Papers arising from this thesis

- Felix Schneider*, Pradeep Dheerendra*, Fabien Balezeau, Michael Ortiz-Rios, Yukiko Kikuchi, Christopher I Petkov, Alexander Thiele, Timothy D Griffiths, "Auditory figure-ground analysis in rostral belt and parabelt of macaque monkey", Scientific Reports, vol. 8, pp. 17948, 2018 (* equal first authors)

Poster abstracts arising from this thesis

- Pradeep Dheerendra, Olivier Joly, Simon Baumann, Alexander Thiele, Timothy D Griffiths, "Measurement of the BOLD response to acoustic spectral flux in the macaque superior temporal plane", SfN Annual Meeting, Nov 2014, Washington DC
- Pradeep Dheerendra, Fabien Balezeau, Sukhbinder Kumar, Andrew Blamire, Alexander Thiele, Timothy D Griffiths, "Primate BOLD data demonstrating fundamental bases for auditory figure-ground analysis", SfN Annual Meeting, Nov 2016, San Diego
- Felix Schneider, Pradeep Dheerendra, Fabien Balezeau, Alexander Thiele, Timothy D Griffiths, "Primate behavioural and functional-imaging model for auditory figure-ground segregation", SfN Annual Meeting, Nov 2017, Washington DC
- Pradeep Dheerendra, Simon Baumann, Oliver Joly, Fabien Balezeau, Alexander Thiele, Timothy D Griffiths, "BOLD response to auditory object properties in the monkey auditory cortex", British Society of Audiology, Sept 2018, Newcastle

Publications arising from the studentship

- Colline Poirier, Simon Baumann, Pradeep Dheerendra, Olivier Joly, David Hunter, Fabien Balezeau, Li Sun, Adrian Rees, Chris I Petkov, Alexander Thiele, Timothy D Griffiths, "Auditory motion specific mechanisms in the monkey brain", PLOS Biology, vol. 15 (5), 2017
- Pradeep Dheerendra, Nicholas M Lynch, Joseph Crutwell, Mark O Cunningham, Tom V Smulders, "In-vitro characterization of gamma

oscillations in the hippocampal formation of the domestic chick", European Journal of Neuroscience, vol. 48(8), pp.2807-15, 2018

- Pradeep Dheerendra, Felipe De Carvalho, Thomas Hall, Andrew Jackson, "The spatial distribution of spike-related slow potentials in macaque primary motor cortex", BNA Festival of Neuroscience, Apr 2015, Edinburgh
- Pradeep Dheerendra, Nick Lynch, Mark O Cunningham, Tom V Smulders, "Receptor pharmacology of gamma oscillations induced in the avian hippocampal formation in vitro", SfN Annual Meeting, Oct 2015, Chicago

Author Contributions

Prof Alex Thiele, Prof Chris Petkov, Dr Yukiko Kikuchi performed surgeries for head implant in the monkeys. Fabien Balezeau helped in fMRI data acquisition in all 3 experiments. Alwin Gieselmann was involved in training monkeys on fixation task.

Experiment – Figure Ground

I designed the experiment, acquired the fMRI data from the monkeys, analysed all experimental fMRI data. I also created the stimulus employed in the monkey behaviour experiment, analysed the behavioural data and co-wrote the paper. Michael Ortiz-Rios was involved in the generation of Figure 3-6 Contrast for figure vs control from three monkeys rendered on top of the standard macaque brain. Felix Schneider trained two monkeys on active figure detection task, designed the behavioural paradigm, analysed the behavioural data and co-wrote the paper.

Experiment – Spectral Flux

I designed the experiment, acquired the fMRI data from the monkeys, analysed all experimental data and co-wrote the paper under preparation. Simon Baumann was involved in designing the experiment, in the fMRI data acquisition, and co-wrote the paper under preparation.

Experiment – Tonotopy

I designed the tonotopy experiment, acquired the fMRI data from the monkeys, analysed all experimental data. Simon Baumann was involved in designing the experiment. Olivier Joly performed the parcellation of the macaque auditory cortex based on tonotopy and T1, T2 data collected by me.

APPENDIX B

Tonotopy stimulus

This script synthesizes audio signals corresponding to a specified passband and saves the output as a wav file.

MATLAB Script

```
% clean slate at the start of execution
clear all; close all; clc;

%% user-defined configuration

% set the list of bandwidth for conditions required in kHz
list_conditions = ...
    {
  % set [ start_freq stop_freq ] % in kHz
        [0.125 0.25];
        [0.25 0.5];
        [0.5 1];
        [1 2];
        [2 4];
        [4 8];
        [8 16];
    };

% set the duration of the stimulus % in sec
tot_dur = 6;

% set sampling frequency in Hz
samp_rate = 48000;

% set flag to enable amplitude modulation
flag_am = 1;

% set the amplitude modulation frequency in Hz
am_freq = 10;

% set the amplitude modulation depth in percentage
am_depth = 90;

% set the output path to save files
out_path = 'D:/work/tonotopy/stimuli/test/';

% set verbosity level
verbose = 0;

%% initializations

% compute the time indices
time_vec = (0:1/samp_rate:tot_dur-1/samp_rate);


% compute the length of the stimulus
len = length(time_vec);
```

```
% compute frequency resolution
fres = samp_rate/len;

% set the number of stimuli types/conditions required
num_conditions = length(list_conditions);

%% processing

% iterate for each condition to generate the stimuli
for cnd = 1:num_conditions

    if verbose >= 1
    disp (['Processing: ' num2str(cnd) ' of ' num2str(num_conditions)]);
    end

    % extract the pass band frequency characteristics
    freq_range = list_conditions{cnd};

    %% random-phase noise signal synthesis

    % initiate magnitude vector
    mag_val = ones(1,len/2);

    % compute the index corresponding to lower cut-off frequency
    low_fq_indx = floor(freq_range(1)*1000/fres);

    % eliminate frequencies below lower cut-off
    mag_val(1:low_fq_indx) = 0;

    % compute the index corresponding to upper cut-off frequency
    high_fq_indx = floor(freq_range(2)*1000/fres);

    % eliminate frequencies above upper cut-off
    mag_val(high_fq_indx:end) = 0;

    % compute the magnitude vector with symmetry property
    mag_vec = [mag_val fliplr(mag_val)];

    % initialize random phase vector
    phase_rnd = 2*pi*rand(1,len/2-1);

    % compute the phase vector - ensure DC & Nyquist are real and
    % rest exhibits complex conjugate symmetry property
    phase_vec = [0 phase_rnd 0 -1*fliplr(phase_rnd)];

    % compute the frequency spectrum
    fq_domain = exp(1i * phase_vec) .* mag_vec;

    % transform the synthesized frequency spectrum into time signal
    time_domain = ifft(fq_domain);

    % error check and alert on complex time domain signal
    if real(time_domain)==0
        disp('Warning.. non real signal generated')
        time_domain = real(time_domain);
    end

    %% end game
```

```
    % perform amplitude modulation
    if flag_am==1
        signal = time_domain .* (1 + am_depth *sin(2*pi*am_freq*time_vec));
    elseif flag_am==0
        signal = time_domain;
    end

    % set the RMS level to 0.1
    signal = 0.1 * signal/std(signal);

    % shape onset and offset
    signal = window_adsr(signal, samp_rate, 50);

    %% save the output

    % output file name with/out AM
    if flag_am==1
        am_fn = ['_am_' num2str(am_freq) 'Hz' ];
    elseif flag_am==0
        am_fn = [];
    end

    % generate output file name
    out_file_name = [out_path '/' 'noise_' num2str(freq_range(1)) '_' ...
        num2str(freq_range(2)) 'kHz' am_fn '.wav'];

    % ensure a valid output path
    if ~(exist(out_path,'dir'))
        if verbose >= 2
            disp(['Creating output folder: ' out_path]);
        end
        mkdir(out_path);
    end

    h_file = fopen(out_file_name,'w');

    % error handling
    disp ' ';
    if h_file>0
        err = fclose(h_file);
        if ~err
            wavwrite(signal, samp_rate, out_file_name);
        end
        if verbose >= 1
            disp('Done. Output saved as');
        end
    else
        disp('Aborting: Output file is open in another application');
    end
    disp(['File: ' out_file_name]);

end
```

Spectral flux stimulus

This script synthesizes audio signals corresponding to a specified value of spectral flux and finally saves the output as a wav file.

MATLAB Script

```
% clean slate at the start of execution
clear all; close all; clc;

% set the path to the output results
out_path = './';

% set the total duration of each stimulus
stimulus_duration = 6; % in seconds

% set the spectral flux value
spec_flux_r1 = 0.9;

% set the exemplar number
exmplr = 1;

disp([ 'spec_flux_r1: ' num2str(spec_flux_r1) ]);


%% default configuration settings

% set sampling frequency of output audio
samp_rate = 48000; % in Hz

% set the number of frequency elements to be used in the stimulus
num_freq_elements = 20;

% set the duration of each frame in seconds
frame_duration = 20e-3;

% set tolerance limits
tol = 0.01;

% set the mean power of amplitude spectrum in dB
mean_amp_spec_pow = 65; % in dB

% set the std deviation of power of amplitude spectrum in dB
std_amp_spec_pow = 10; % in dB

% set debug flag
flag_dbg = 1;
flag_dbg_0 = 0;



%% initializations

% compute the number of frames in the stimulus
num_frames = stimulus_duration/frame_duration;

% initialize array to store the deviation factors for entire stimulus
```

```
sdev_factors = zeros(num_freq_elements, num_frames);


%% computation of amplitude deviation factors

% use of array allows generation of stimuli for continuous fMRI acquisition
% form a array of spectral flux values
frame_r1_vals = repmat(spec_flux_r1, num_frames, 1);

% generate random deviation factors for amplitude spectrum
std_fac = randn(num_freq_elements, 1);
% assign std dev factors values for first frame
sdev_factors(:,1) = std_fac;

% store current guess for dev factor as reference for next iteration
prev_fr_std_facs = std_fac;

if flag_dbg
    % initialize array of number of iterations required to discover devfacs
    num_iters = ones(num_frames, 1);
end

% iterate for all frames
for fr_no = 2:num_frames

    % select r value for current frame
    target_r1 = frame_r1_vals(fr_no);

    if flag_dbg
        % initialize iteration number
        iter_no = 0;
    end

    if target_r1==1
        % avoid iterative search when correlation is 1
        cur_fr_std_facs = prev_fr_std_facs;

    else
        %% iterative search for a dev fac that fits the correl. requirement

        % set the initial error to non-conformance of criteria
        error = 2 * tol;

        % formulate a positive definite matrix using the current target r1
        corr_mat = [1 target_r1; target_r1 1;];

        % perform cholesky decomposition
        upper_triangle = chol(corr_mat);

        % extract the transformation vector
        transform_vec = upper_triangle(:,2);

        % iterate until convergence is reached
        while(error > tol)

            if flag_dbg
                % increment iteration number
                iter_no = iter_no + 1;
            end
```

```
            % avoid direct guessing of dev facs as it takes far too many
            % iterations to converge esp TRUE for high values of r1

            % form a random vector but it is NOT the dev facs
            rand_vec = randn(num_freq_elements, 1);

            % compute deviation factors for amplitude spectrum
            guess_std_facs = [prev_fr_std_facs rand_vec] * transform_vec;

            % compute the correlation for the current guess
            r_curr_guess = corr([prev_fr_std_facs guess_std_facs] * ...
                std_amp_spec_pow + mean_amp_spec_pow);

            % compute the error in r for the current guess
            error = abs(r_curr_guess(2) - target_r1);

        end

        % retain the latest guess as answer for the current frame
        cur_fr_std_facs = guess_std_facs;

    end

    if flag_dbg
        % store number of iterations taken to find the array
        num_iters(fr_no) = iter_no;
    end

    % store the current frame stdev factors
    sdev_factors(:, fr_no) = cur_fr_std_facs;

    % store current frame's dev factor as reference for next iteration
    prev_fr_std_facs = cur_fr_std_facs;

end

% compute the array of power of amplitude spectrum values in db
raw_amp_spec_power = mean_amp_spec_pow + sdev_factors * std_amp_spec_pow;

if flag_dbg_0
    % initialize array to store the deviation factors for entire stimulus
    r_check = zeros(num_frames, 1);

    % need to set first value as it is ill defined
    r_check(1) = frame_r1_vals(1);

    % iterate for every frame
    for fr_no = 2:num_frames

        % compute the correlation for the current guess
        r_check(fr_no) = corr(raw_amp_spec_power(:,fr_no-1),
raw_amp_spec_power(:,fr_no));
    end
end

if flag_dbg_0
    % plot the histogram of the iterations it takes to discover dev factors
    hist(num_iters, 100);
end
```

```
%% selection of frequency elements

% compute the pool of frequencies to choose from randomly
freq_pool = 440 * 2 .^ ((-20:80)/24); % in Hz

% get the number of frequencies in the pool
freq_pool_len = length(freq_pool);

% get a list of random indices for freq selection
frq_ind = randperm(freq_pool_len, num_freq_elements);

% select a random subset of frequencies from the pool rearranged in
increasing order
freq_vec = freq_pool(sort(frq_ind));



%% generate sound stimuli

% get the number of samples per frame
samples_per_frame = round(frame_duration * samp_rate);

% get the number of frames required in the stimuli
num_frames = size(raw_amp_spec_power, 2);

% get the number of frequency elements
num_freqs = length(freq_vec);

% compute the vector of positions where values are known
frame_pos = [1:num_frames];

% compute the positions at which interpolation is to be performed
inter_pos = [1:1/samples_per_frame:num_frames];

% get the total number of samples in the stimulus
tot_samps = (num_frames-1) * samples_per_frame + 1;

% compute vector of time instances
time = [0:tot_samps-1]/samp_rate;

% initialize a vector for storing stimulus
stimulus = zeros(1, tot_samps);



%% construction of the raw sound stimulus

% for every freq element
for fq = 1:num_freqs

    % extract the vector of amp power at current frequency
    sparse_pow_vec_fq = raw_amp_spec_power(fq, :);

    % interpolation for values at intermediate position
    pow_vec_fq = interp1(frame_pos,sparse_pow_vec_fq,inter_pos,'spline');

    % ensure the values are normalized to avoid clipping
```

```
    pow_vec_fq = pow_vec_fq - max(pow_vec_fq);

    % convert from dB to actuals amplitude value
    amp_vec_fq = 10.^(pow_vec_fq/20);

    % retain the ampltiude values
    amp_spectrum(fq,:) = amp_vec_fq;

    % convert amplitude values to actual sinusoid component
    sin_comp = amp_vec_fq .* sin(2 * pi * freq_vec(fq) * time + 2 * pi *
rand);

    % integrate all sinusoid components
    stimulus = stimulus + sin_comp;

end

% format variable into time series
stimulus = stimulus.';

% normalize stimulus vector to avoid clipping
stimulus = 0.1 * stimulus / std(stimulus);

% apply a window
stimulus = window_adsr(stimulus, samp_rate, 50);

% generate the file name for the output audio stimulus file
out_name = sprintf('%s/corr_%d_ex_%d.wav',out_path, 10*spec_flux_r1,
exmplr);

% ensure output directory exists
% out_path = [pwd '/output/'];
if ~(exist(out_path,'dir'))
    mkdir(out_path);
end

% write the wav file
audiowrite(out_name, stimulus, samp_rate);
```

Stochastic Figure Ground stimulus

This script configures the program that synthesizes stochastic figure-ground stimulus in a three-segment format with the figure in the middle segment and ground throughout the 3 segments. If the figure is not requested then the middle segment still has additional components that figure would have had but without the coherence. This function generates stochastic figure-ground with

(1) chords having a random number of frequency components in each
(2) a fixed number of extra components in each chord as requested
(3) coherence among the extra components is decided based on whether a figure was requested by the user

MATLAB wrapper script

```
clear; close; clc

%% user defined configurations

% set whether figure is present or not
% 1 - figure present; 0 - figure absent
flag_fig_present = 1;

% set flag to enable amplitude modulation (AM) of either figure/ground
%  0: none   1: AM on fig  -1: AM on gnd
flag_modulation = 0;

% set the number of chords in each of 3 segments
num_chords = 20;

% set the number of coherent components in figure - middle segment
cc_figure = 6;

% set the number of exemplars required
num_exemplars = 2;

% set path to output directory
out_path = './stimuli/';



%% default user configurations

% set flag whether to play stimulus
flag_play_op = 0;

% set flag whether to save output file
flag_save_op = 1;

% set file name prefix
fn_prefix = 'sfg_';
```

```
% set file extension type for output
file_extn = '.wav';

% set verbosity level
verbose = 0; % 0-errors 1-regular


%% default settings

% duration of each chord in ms
dur_chord = 50;

% minimum number of components
cmp_min = 5;

% maximum number of components
cmp_max = 15;

% sampling rate in Hz
samp_rate = 44100;


%% initializations

% number of chords in the first ground segment
init_L = num_chords;

% number of coherence components in the first ground segment
init_C = 0;

% number of chords in middle (figure/ground) segment
mid_L = num_chords;

% number of chords in last ground segment
fin_L = num_chords;

% number of coherence components in last ground segment
fin_C = 0;


%% processing

% number of chords in the first, second and third segments respectively
n_chords_seg = [init_L mid_L fin_L];

% duration of a chord in each segment
dur_chord_seg = [dur_chord dur_chord dur_chord];

% number of coherent components in each segment
n_coh_seg = [init_C cc_figure fin_C];

% compute the number of segments
num_segments = length(n_chords_seg);

% iterate for each exemplar
for exmpl = 1:num_exemplars

    if verbose >= 1
        disp(['Exemplar: ' num2str(exmpl) ' of ' num2str(num_exemplars)]);
```

```
    end

    % initialize output variable
    all_seg = [];

    % for each segment
    for k = 1:num_segments

        % generate sfg stimulus for current segment
        cur_seg = generate_sfg_signal(flag_fig_present, n_chords_seg(k),
            dur_chord_seg(k), cmp_min, cmp_max, n_coh_seg(k), samp_rate,
            flag_modulation);

        % concatenate the segments into output variable
        all_seg = [all_seg; cur_seg];

    end

    % play output signal
    if flag_play_op == 1
        sound(all_seg, samp_rate);
    end

    %% save output as wave file
    if flag_save_op == 1

        % create output path if not present
        if ~(exist(out_path,'dir'))
            mkdir(out_path);
        end

        % file name prefix for figure present vs absent
        if flag_fig_present == 1
            fn2_prefix = 'fig1_';
        elseif flag_fig_present == 0
            fn2_prefix = 'fig0_';
        end

        % file name suffix for modulation type
        switch(flag_modulation)
            case 0
                fn_suffix = '';
            case 1
                fn_suffix = 'AMf_';
            case -1
                fn_suffix = 'AMg_';
        end

        % compute the file name
        out_file_name = [out_path '/' fn_prefix fn2_prefix fn_suffix 'ex_'
num2str(exmpl) file_extn];

        disp ' ';
        h_file = fopen(out_file_name,'w');

        if h_file>0
            err = fclose(h_file);
            if ~err
                wavwrite(all_seg, samp_rate, out_file_name);
            end
            if verbose >= 1
```

```
                disp('Done. Output saved as');
            end
        else
            disp('Aborting: Output file is open in another application');
        end
        disp(['File: ' out_file_name]);

    end
    disp ' ';

end
```

MATLAB function

```
function signal = generate_sfg_signal(flag_fig_present, n_chords,
dur_chord, n_comps_min, n_comps_max, n_extra_comps, samp_rate,
flag_modulation)

% Inputs:
%   flag_fig_present - flag conveys if figure is requested 1 - yes 0 - no
%   n_chords   - total number of chords requested
%   dur_chord  - duration of each chord
%   n_comps_min - minimum number of components per chord
%   n_comps_max - maximum number of components per chord
%   n_extra_comps - number of extra components requested in each chord
%   samp_rate  - sampling rate of the segment
%   flag_modulation - flag conveys if modulation is requested & which
%       aspect of the stimulus i.e. figure or ground has to modulated
%       Amplitude Modulation - tremolo; Frequency Modulation: vibrato
%        0: no modulation
%        1: AM on figure
%       -1: AM on ground
%
% Outputs:
%   signal  - output with all chords as requested

%% default settings

if nargin < 1
    % set flag to convey whether a figure is requested
    flag_fig_present = 1;
end

if nargin < 2
    % set total number of chords in this segment
    n_chords = 20;
end

if nargin < 3
    % set duration of each chord in ms
    dur_chord = 50;
end

if nargin < 4
    % set minimum number of components per chord
    n_comps_min = 5;
end

if nargin < 5
    % set maximum number of components per chord
```

```
    n_comps_max = 15;
end

if nargin < 6
    % set number of extra components requested in each chord
    n_extra_comps = 6;
end

if nargin < 7
    % set sampling rate in Hz
    samp_rate = 44100;
end

if nargin < 8
    % set flag to enable amplitude modulation (AM) of either figure/ground
    %  0: none   1: AM on fig  -1: AM on gnd
    flag_modulation = -1;
end


%% default values

% set modulation index
mod_indx = 1;

% set modulation frequency
modulator_frq = 10; % in Hz

% set flag to enable debug code
flag_dbg = 0; % disables randomization to help debugging

if n_comps_min > n_comps_max
    disp('Error: Input config - n_comps_min > n_comps_max');
end


%% initializations

% Component frequencies were randomly drawn from a set of 129 values
% equally spaced on a logarithmic scale between 179 and 7246 Hz
% note: Successive frequencies are separated by 1/24th an octave
freq_pool = 440 * 2 .^((-31:97)/24);

% set duration for onset and offset ramp
onset_duration = 10; % in ms

% create time vector - dur_chord is in ms
t_vec = [0: 1/samp_rate: dur_chord/1000 - 1/samp_rate].';

% when figure is requested, coherence among extra components is ensured
if flag_fig_present == 1

    if flag_dbg == 0
        % create random list of indices
        rnd_indx = randperm(length(freq_pool));
    else
        rnd_indx = length(freq_pool):-1:1;
    end

    % pick coherent components for all chords in this segment
```

```
    coherent_frqs = freq_pool(rnd_indx(1:n_extra_comps));

    % create non-coherent frequency pool
    non_coherent_pool = freq_pool(rnd_indx(n_extra_comps+1:end));

end

% compute the maximum number of components possible per chord
tot_max_comp = n_comps_max + n_extra_comps;

% compute the total duration of the signal
sig_dur = n_chords * dur_chord/1000;

% create time vector - sig_dur is in s
time_vector = [0: 1/samp_rate: sig_dur - 1/samp_rate].';


%% processing

% initialize variables
figure_signal = [];
ground_signal = []; % length = (dur_chord * samp_rate * n_chords, 1);

% iterate for each chord in this segment
for ch = 1:n_chords

    % initialize accumulator
    fig_comps = zeros(dur_chord/1000 * samp_rate, 1);
    gnd_comps = zeros(dur_chord/1000 * samp_rate, 1);

    %% extra components corresponding to figure

    % when fig is not requested, coherence among extra comp is eliminated
    if flag_fig_present == 0

        if flag_dbg == 0
            % create random list of indices
            rnd_indx = randperm(length(freq_pool));
        else
            % create predictable list of indices for debugging
            rnd_indx = length(freq_pool):-1:1;
        end

        % pick coherent components for all chords in this segment
        coherent_frqs = freq_pool(rnd_indx(1:n_extra_comps));

        % create non-coherent frequency pool
        non_coherent_pool = freq_pool(rnd_indx(n_extra_comps+1:end));

    end
    % ELSE: when a fig is requested, coherence among extra comp is ensured

    % iterate for each component in the chosen freq pool
    for k = 1:length(coherent_frqs)

        % extract current component frequency
        curr_comp_frq = coherent_frqs(k);

        % create current component with normalization
```

```
        curr_comp = 0.2 / tot_max_comp * sin(2*pi* curr_comp_frq * t_vec);

        % accumulate current component into a new row
        fig_comps = [fig_comps curr_comp];

    end

    %% random components corresponding to the ground

    % num of components in current chord = random no b/w n_c_min & n_c_max
    n_c = n_comps_min + round( rand * (n_comps_max - n_comps_min) );

    if flag_dbg == 0
        % create random list of indices
        rnd_indx = randperm(length(non_coherent_pool));
    else
        % create predictable list of indices for debugging
        rnd_indx = 1:length(non_coherent_pool);
    end

    % pick the non coherent components at random from the non coh frq pool
    non_coherent_frqs = non_coherent_pool(rnd_indx(1:n_c));

    % iterate for each component in the chosen freq pool
    for k = 1:length(non_coherent_frqs)

        % extract current component frequency
        curr_comp_frq = non_coherent_frqs(k);

        % create current component with normalization
        curr_comp = 0.2 / tot_max_comp * sin(2*pi* curr_comp_frq * t_vec);

        % accumulate current component into a new row
        gnd_comps = [gnd_comps curr_comp];

    end


    %% end processing for current chord

    % temporal summate components to generate a single chord
    fig_chord = sum(fig_comps, 2);
    gnd_chord = sum(gnd_comps, 2);

    % shape sound onset and offset with a raised-cosine ramp
    fig_chord = window_adsr(fig_chord, samp_rate, onset_duration);
    gnd_chord = window_adsr(gnd_chord, samp_rate, onset_duration);

    % store current chord
    figure_signal = [figure_signal; fig_chord];
    ground_signal = [ground_signal; gnd_chord];

end


%% Apply Amplitude modulation on figure or ground as requested

% apply amplitude modulation AM / tremolo effect
if 1 == abs(flag_modulation)
```

```
    % on figure chords
    if (flag_modulation == 1)

        % apply tremolo effect to figure chords
        figure_signal = (1+ mod_indx * sin(2*pi* modulator_frq *
time_vector)).* figure_signal ;
    end

    % on ground chords
    if (flag_modulation == -1)

        % apply tremolo effect to ground chords
        ground_signal = (1+ mod_indx * sin(2*pi* modulator_frq *
time_vector)).* ground_signal ;
    end

end

%% output stage

% compute output signal by putting together both figure and ground signals
signal = ground_signal + figure_signal;


end
```

Window Shaping

This function applies a raised cosine window at onset and offset of the data. This is used to shape attack and decay characteristics of stimuli.

MATLAB function

```
function out_data = window_adsr(data, samp_rate, window_ms)

% Notes:
% Hanning window a.k.a raised cosine window
%  h(theta) = 0.5*(1-cos(theta))    theta = 0 to 2*pi
%  h(n) = 0.5*(1-cos(2*pi*n/(N-1)))     n = 0 to N-1
%
% Inputs:
%   data      - input data vector assumes a single column
%   samp_rate - sampling rate in Hz
%   window_ms - attack or decay duration in ms
%
% Outputs:
%   out_data  - output data always with row-time & column-ch format

%% default values, error checks and handling

if nargin < 2
    % set sampling rate in Hz
    samp_rate = 44100;
end

if nargin < 3
    % set attack, decay duration in ms
    window_ms = 10;
end

if nargin < 1
    % error handling
    disp('Error: no valid input');
    return;
end


%% initializations

% extract the number of channels and length of the input data
if size(data, 1) < size(data, 2)
    n_ch = size(data, 1);

    dat_len = size(data, 2);

    % intialize output data
    out_data = data.';
else
    n_ch = size(data, 2);

    dat_len = size(data, 1);

    % initialize output data
    out_data = data;
```

```
end

% warn the user if in case they are using this function incorrectly
if n_ch > 2
    disp('Warning: unsupported data')
end

% compute length of the window needed
win_len = round(window_ms / 1000 * samp_rate);

% compute the Hann or raised cosine window function
hann_win = ( 1 - cos(linspace(0, 2*pi, 2*win_len).') )/2;


%% processing

% iterate for each channel
for ch = 1:n_ch

    % onset window is first half of the Hann window
    onset_win = hann_win(1:win_len);

    % apply onset window - attack characteristics
    out_data(1:win_len, ch) = out_data(1:win_len, ch) .* onset_win;

    % compute the start index of the window
    start_indx = dat_len - win_len + 1;

    % offset window is second half of Hann window
    offset_win = hann_win(win_len+1:end);

    % apply offset window - decay characteristics
    out_data(start_indx:dat_len, ch) = out_data(start_indx:dat_len, ch) .*
offset_win;

end

end
```

Tonotopy fMRI experiment Cortex code

```
/*************************************************************
// This Cortex Timing file is used for sparse data acquisition
*************************************************************/
#include "css_inc.h"
#include "extern_var.h"

// global variables
#define NumRew    _int0
#define REWTIME   _int1
#define x_move    _int2
#define y_move    _int3
#define step_sz   _int4

#define FIXWIN_SZ _float0

// Parameters of the task, to be adjusted
#define PRE_FIX         400
#define FIX_DURATION    3900
#define REWARD_DURATION 700
#define REWARD_DELAY    1900
#define SOUND_DURATION  6000
#define SOUND_DELAY     950
#define SCAN_DELAY      2100

#define SOUND_STOP  ((SOUND_DELAY) + (SOUND_DURATION))

#define TIMER_PRE_FIX   1
#define TIMER_FIX_DUR   2
#define TIMER_SND_DEL   3
#define TIMER_SND_END   4
#define TIMER_REW_DEL   5
#define TIMER_REW_DUR   6
#define TIMER_SCN_DEL   7

#define OFFSET          10
#define  TIMER_PRE_FIX_2    TIMER_PRE_FIX + OFFSET
#define  TIMER_FIX_DUR_2    TIMER_FIX_DUR + OFFSET
#define  TIMER_REW_DEL_2    TIMER_REW_DEL + OFFSET
#define  TIMER_REW_DUR_2    TIMER_REW_DUR + OFFSET

// Colors used
// background color
#define BCK_COL      GcolorABS(0,100,100,100); Gflush(1);
// fix spot color red when fixating
#define DOES_FIX_COL GcolorABS(2,255,  0,  0); Gflush(1);
// fix spot color blue when waiting for fixation
#define WAIT_FIX_COL GcolorABS(2,128, 64,  0); Gflush(1);

char snd_file_name[50];

// Variables with scope of whole file
int esc_flag   = 0;
int rew_monkey = 0;
int endfix_flag = 0;
int reward_given = 0;
int reward2_success = 0;
int scan_flag = 0;
int sound_played = 0;
int sound_stop = 0;
```

```
int cond_no;
int repeat_num;
int res;
char type[8];
int offset;

// function declaration
int give_reward();
int KEY_action();

/*****************************************/
/*********   Start > main < here   **********/
/*****************************************/
main()
{
    /* initialization routines */

    // Update user display with current value of parameters
    Mprintf(1,"rewards: %d rewtime: %d", NumRew, REWTIME);
    Mprintf(2,"fixtime: %d fixwin: %.2f", FIX_DURATION, FIXWIN_SZ);
    Mprintf(3,"step: %d", step_sz);
    Mprintf(4,"Start up");
    Mprintf(5,"NEW TRIAL");

    // instruct cortex to start collecting data
    collect_data(ON);

    // start storing eye movement data
    put_eye_data_in_buf(ON);

    // store the timestamp of start of eye data
    encode(START_EYE_DATA);

    // extract current condition number
    cond_no = BLOCKget_cond_num();

    // extract current block/repeat number
    repeat_num = get_repeat_num();

    // output single byte of data to scanner to reset the state
    res = DEVoutp(0, 1, 0);

    // clear EOG display window
    clear_eog();

    // set to default size if undefined
    if(FIXWIN_SZ == 0)
        FIXWIN_SZ = 5;

    // specify the nature of fixation window and size
    set_fixwin_params(0, FIXWIN_SZ, FIXWIN_SZ);

    // Move fixation window to a specified item position in a given screen
    ITEM_POSbind_fixspot(TEST0, 0);

    // show path of eye movement
    display_eye_path(ON);

    // make it invisible
```

```
SCREENdraw_fixwin(WHITE);

// draw reference visual angle sizes
SCREENdraw_box_on_eog(0, 0, 20, 20, BLACK);
SCREENdraw_box_on_eog(0, 0, 10, 10, LIGHTGRAY);

// render background colour
BCK_COL
Gflush(1);

// No Fix Spot for this task!
display_fixspot(0);

// Select value based on condition number
switch(cond_no)
{
    case 1:     type = "0_5_1"; break;
    case 2:     type = "2_4";   break;
    case 3:     type = "8_16";  break;
    default:    type = "sil";   break; // ensures silence plays
}

// Random exemplars based on trial number
sprintf(snd_file_name, "tonotopy\\20151005\\pass%skHz10Hz.wav",type);
// update user
Mprintf(3,"file: %s",snd_file_name);

// load sound file and assign a index
SOUNDload(cond_no, snd_file_name);

// set mixer volume - left and right, for sound of given index
SOUNDvol(31, 31, cond_no);

/***** Trial starts - 1st round of fixation *****/

endfix_flag = 0;
reward_given = 0;
sound_played = 0;
reward2_success = 0;

// prefixation routine
Mprintf(4,"PRE FIX");
MS_TIMERset(TIMER_PRE_FIX, PRE_FIX);

// when waiting for prefix to complete
while (MS_TIMERcheck(TIMER_PRE_FIX) > 0)
{
    KEY_action();
    Gon_off(TEST0, ON);
    WAIT_FIX_COL
    SCREENdraw_fixwin (BLUE);
}

Mprintf(4,"FIXATION");
// set all timers
MS_TIMERset(TIMER_FIX_DUR, FIX_DURATION);
MS_TIMERset(TIMER_SND_DEL, SOUND_DELAY);
MS_TIMERset(TIMER_REW_DEL, REWARD_DELAY);
MS_TIMERset(TIMER_SND_END, SOUND_STOP);

while( MS_TIMERcheck(TIMER_FIX_DUR) > 0 )
```

```
{
    // perform any keyboard actions
    KEY_action();

    // when fixation is in progress
    if (endfix_flag == 0)
    {
        // check for escape character
        if(KEY_action() == 1)
        {
            Mprintf(5,"BREAK TRIAL");
            break;
        }

        // 1st test of eye fixation
        if (get_fixation_state() == 1)
        {
            // when monkey is fixating
            Mprintf(4,"FIXATED!!");
            DOES_FIX_COL
            SCREENdraw_fixwin (RED);
            rew_monkey = 1;
        }

        set_timer(100); // wait for small gap
        while (timer_expired() == 0); // kill time

        // 2nd test of eye fixation
        if (get_fixation_state() == 0)
        {
            // give monkey a second chance to account for any noise

            set_timer(200); // wait for small gap
            while (timer_expired() == 0); // kill time

            // 3rd test of eye fixation
            if (get_fixation_state() == 0)
            {
                // when monkey is still not fixating
                Mprintf(4,"Fixation break");
                Gon_off(TEST0, OFF);
                WAIT_FIX_COL
                SCREENdraw_fixwin (BLACK);
                rew_monkey = 0;
                endfix_flag = 1;
            }
        }

        // when it is time to reward monkey
        if ( (MS_TIMERcheck(TIMER_REW_DEL)==0) && (reward_given == 0) )
        {
            Mprintf(4,"REWARD 1");

            REWTIME = 40;
            give_reward();

            encode(103);

            reward_given = 1;
        }
    }
```

```
    // when it is time, playout the sound
    if (MS_TIMERcheck(TIMER_SND_DEL) == 0 && sound_played == 0)
    {
        SOUNDprep(cond_no);
        SOUNDstart(cond_no);
        sound_played = 1;
    }
}

// Reward time
Mprintf(4,"REWARD 2");
MS_TIMERset(TIMER_REW_DUR, REWARD_DURATION);

REWTIME = 50;
if (give_reward() == 0)
{
    encode (5);
}
else
{
    encode(104);
    reward2_success = 1;
}

Gon_off(TEST0, OFF);
WAIT_FIX_COL
SCREENdraw_fixwin (WHITE);

// kill reward time if any is remaining
while (MS_TIMERcheck(TIMER_REW_DUR) > 0)
{
    KEY_action();
}

/***** Half time - 2nd round of fixation *****/

endfix_flag = 0;
reward_given = 0;
scan_flag = 0;
sound_stop = 0;

// prefixation routine
Mprintf(4,"PRE FIX-2");
MS_TIMERset(TIMER_PRE_FIX_2, PRE_FIX);

// when waiting for prefix to complete
while (MS_TIMERcheck(TIMER_PRE_FIX_2) > 0)
{
    KEY_action();
    Gon_off(TEST0, ON);
    WAIT_FIX_COL
    SCREENdraw_fixwin (BLUE);
}

Mprintf(4,"FIXATION-2");
// set all timers
MS_TIMERset(TIMER_FIX_DUR_2, FIX_DURATION);
MS_TIMERset(TIMER_SCN_DEL, SCAN_DELAY);
MS_TIMERset(TIMER_REW_DEL_2, REWARD_DELAY);
```

```
while( MS_TIMERcheck(TIMER_FIX_DUR_2) > 0 )
{
    // perform any keyboard actions
    KEY_action();

    // when it is time, stop the playout of sound
    if ( (MS_TIMERcheck(TIMER_SND_END) == 0) && (sound_stop == 0) )
    {
        SOUNDstop(cond_no);
        sound_stop = 1;
    }

    // when fixation is in progress
    if (endfix_flag == 0)
    {
        // check for escape character
        if(KEY_action() == 1)
        {
            Mprintf(5,"BREAK TRIAL");
            break;
        }

        // 1st test of eye fixation
        if (get_fixation_state() == 1)
        {
            // when monkey is fixating
            Mprintf(4,"FIXATED!!");
            DOES_FIX_COL
            SCREENdraw_fixwin (RED);
            rew_monkey = 1;
        }

        set_timer(100); // wait for small gap
        while (timer_expired() == 0); // kill time

        // 2nd test of eye fixation
        if (get_fixation_state() == 0)
        {
            // give monkey a second chance to account for any noise

            set_timer(200); // wait for small gap
            while (timer_expired() == 0); // kill time

            // 3rd test of eye fixation
            if (get_fixation_state() == 0)
            {
                // when monkey is still not fixating
                Mprintf(4,"Fixation break");
                Gon_off(TEST0, OFF);
                WAIT_FIX_COL
                SCREENdraw_fixwin (BLACK);
                rew_monkey = 0;
                endfix_flag = 1;
            }
        }

        // when it is time to reward monkey
        if ( (MS_TIMERcheck(TIMER_REW_DEL_2)==0)&&(reward_given == 0) )
        {
            Mprintf(4,"REWARD 3");
```

```
            if (reward2_success == 1)
                REWTIME = 70;
            else
                REWTIME = 50;

            give_reward();

            encode(105);

            reward_given = 1;
        }
    }

    // when it is time, scan
    if ( (MS_TIMERcheck(TIMER_SCN_DEL) == 0) && (scan_flag == 0) )
    {
        // send trigger to perform a scan
        res = DEVoutp(0, 1, 2);

        if (res == 2)
        {
            // update user
            Mprintf(4, "Volume aqc");

            // encode this event in the log file
            encode(200);
        }
        else
        {
            // update user
            Mprintf(4, "Error triggering scanner");
        }

        set_timer(100); // wait for small gap
        while(timer_expired() == 0); // kill time

        // signal the scanner to stop scanning
        res = DEVoutp(0, 1, 0);

        // update flag since scan was performed for the current trial
        scan_flag = 1;

    }
}

// Reward time
Mprintf(4,"REWARD 4");
MS_TIMERset(TIMER_REW_DUR_2, REWARD_DURATION);

if (reward2_success == 1)
    REWTIME = 100;
else
    REWTIME = 80;

if (give_reward() == 0)
{
    encode (5);
}
else
{
    encode(106);
```

```
    }

    Gon_off(TEST0, OFF);
    WAIT_FIX_COL
    SCREENdraw_fixwin (WHITE);

    // kill reward time if any is remaining
    while (MS_TIMERcheck(TIMER_REW_DUR_2) > 0)
    {
        KEY_action();
    }

    /*  End of Trial process */

    // clean up at end of trail
    end_trial();

    // stop storing eye movement data, and EPP data
    put_eye_data_in_buf(0);
    put_epp_data_in_buf(0);

    // store the timestamp of end of eye data
    encode(END_EYE_DATA);

    // do not show path of eye movement
    display_eye_path(OFF);

    // stop collecing data
    collect_data(OFF);

     return;
}  // main

 /* This function sets reward duration and gives reward */
int give_reward()
{
    if(rew_monkey == 1)
    {
        set_ms_reward_duration(REWTIME);    // Windows version
        reward();
        NumRew = NumRew + 1;
        return 1;
    }
    return 0;
}
```

Spectral flux fMRI experiment Cortex code

```
#include "css_inc.h"
#include "extern_var.h"

// global variables
#define NumRew    _int0
#define REWTIME   _int1
#define x_move    _int2
#define y_move    _int3
#define step_sz   _int4

#define FIXWIN_SZ _float0

// Parameters of the task, to be adjusted
#define PRE_FIX     6000
#define FIX         2500
#define SCAN_DELAY   150
#define REWARD      1500

// Colors used
#define BCK_COL      GcolorABS(0,100,100,100); // background color
#define DOES_FIX_COL GcolorABS(2,255,  0,  0); // fix spot color red when
fixating
#define WAIT_FIX_COL GcolorABS(2,128, 64,  0); // fix spot color blue when
waiting for fixation

// Variables with scope over entire file
int esc_flag   = 0;
int go_flag    = 1;
int endfix_flag = 0;
int scan_flag  = 1;
int rew_monkey = 0;
int cond_no;

int repeat_num;
int res;
int corr;

char snd_file_name[50];

// function declaration
int give_reward();
int KEY_action();
int perform_scan();

/*****************************************/
/*********   Start > main < here   **********/
/*****************************************/
main()
{
    /* initialization routines */

    // Update user display with current value of parameters
    Mprintf(1,"rewards: %d rewtime: %d", NumRew, REWTIME);
    Mprintf(2,"fixtime: %d fixwin: %.2f", FIX, FIXWIN_SZ);
    Mprintf(3,"step: %d", step_sz);
    Mprintf(4,"Start up");
    Mprintf(5,"NEW TRIAL");

    // instruct cortex to start collecting data
```

```
collect_data(ON);

// start storing eye movement data
put_eye_data_in_buf(ON);

// store the timestamp of start of eye data
encode(START_EYE_DATA);

// extract current condition number
cond_no = BLOCKget_cond_num();

// extract current block/repeat number
repeat_num = get_repeat_num();

// output single byte of data to scanner to reset the state
res = DEVoutp(0, 1, 0);

// clear EOG display window
clear_eog();

// set to default size if undefined
if(FIXWIN_SZ == 0)
    FIXWIN_SZ = 5;

// specify the nature of fixation window and size
set_fixwin_params(0, FIXWIN_SZ, FIXWIN_SZ);

// Move fixation window to a specified item position in a given screen
ITEM_POSbind_fixspot(TEST0, 0);

// show path of eye movement
display_eye_path(ON);

// make it invisible
SCREENdraw_fixwin(WHITE);

// draw reference visual angle sizes
SCREENdraw_box_on_eog(0, 0, 20, 20, BLACK);
SCREENdraw_box_on_eog(0, 0, 10, 10, LIGHTGRAY);

// render background colour
BCK_COL
Gflush(1);

// No Fix Spot for this task!
display_fixspot(0);

// GmoveABS(TEST0,x_move,y_move);
// ITEM_POSbind_fixspot(TEST0, 1);

// Select value based on condition number
switch(cond_no)
{
    case 1: corr = 0;   break;
    case 2: corr = 3;   break;
    case 3: corr = 6;   break;
    case 4: corr = 8;   break;
    case 5: corr = 9;   break;
    default: corr = 11; break; // ensures silence since no file exist
}
```

```
 // Random exemplars based on trial number
sprintf(snd_file_name,"SF_rnd_ex\\corr_%d_ex_%d.wav",corr,repeat_num+1);

 // update user
 Mprintf(3, "file: %s", snd_file_name);

 // load sound file and assign a index
 SOUNDload(cond_no, snd_file_name);

 // set mixer volume - left and right, for sound of given index
 SOUNDvol(31, 31, cond_no);


 /* Start Trial */
 endfix_flag = 0;
 scan_flag = 1;

 Mprintf(4,"PRE FIX");

 MS_TIMERset(5, PRE_FIX);

 // Prepare sound file for playout
 SOUNDprep(cond_no);
 // Playout the sound with given index
 SOUNDstart(cond_no);

 // perform while sound is playing
 while (MS_TIMERcheck(5) > 0)
 {
     KEY_action();
     if (MS_TIMERcheck(5) == 600)
     {
         Gon_off(TEST0, ON);
         WAIT_FIX_COL
         Gflush(1);
         SCREENdraw_fixwin(BLUE);
         Mprintf(4,"WAIT FIXATION");
     }
 }

 MS_TIMERset(3, SCAN_DELAY);
 MS_TIMERset(1, FIX);

 while( MS_TIMERcheck(1) > 0 )
 {
   KEY_action();
   if (endfix_flag == 0)
   {
     if(KEY_action() == 1)
     {
         Mprintf(5,"BREAK TRIAL");
         go_flag = 0;
         break;
     }

     Mprintf(4,"FIXATION!");
     DOES_FIX_COL
     Gflush(1);
     SCREENdraw_fixwin(RED);
```

```
    go_flag = 0;

    rew_monkey = 1;
    if (MS_TIMERcheck(3) == 0 && scan_flag == 1)    //Chris
    {
        SOUNDstop(cond_no);

        res = perform_scan();
    }

    if (get_fixation_state() == 0) //simon: test
    {
        set_timer(100); // was 10
        while (timer_expired() == 0);  // try to exclude noise
        {
            if (get_fixation_state() == 0) //simon: test
            {
                Mprintf(4,"Fixation break");
                Mprintf(5,"BREAK TRIAL");
                WAIT_FIX_COL
                Gflush(1);
                SCREENdraw_fixwin(BLACK);
                rew_monkey = 0;
                go_flag = 1;
                endfix_flag = 1;
                //Gon_off(TEST0, OFF);
                //Gflush(1);
                SCREENdraw_fixwin(WHITE);
                //break;
            }
        }
    }
    else if (KEY_action() == 1)
    {
        Mprintf(5,"BREAK TRIAL");
        rew_monkey = 0;
        break;
    }
  }
  else if (MS_TIMERcheck(3) == 0 && scan_flag == 1)
  {
     SOUNDstop(cond_no);

     res = perform_scan();
  }
}

// Trial done (however ...)
Gon_off(TEST0, OFF);
Gflush(1);
SCREENdraw_fixwin(WHITE);

if(esc_flag != 1)
{
    set_timer(REWARD);

    while(timer_expired() == 0)
    {
        if( give_reward() == 0 )
        {
            Mprintf(4,"NO REWARD");
```

```
                if(KEY_action() == 1)
                {
                    go_flag = 0;
                    break;
                }
            }
            else
            {
                Mprintf(4,"REWARD");
                if(KEY_action() == 1)
                {
                    break;
                }
            }
        }
    }

    /*  End of Trial process */

    // clean up at end of trail
    end_trial();

    // stop storing eye movement data, and EPP data
    put_eye_data_in_buf(0);
    put_epp_data_in_buf(0);

    // store the timestamp of end of eye data
    encode(END_EYE_DATA);

    // do not show path of eye movement
    display_eye_path(OFF);

    // stop collecing data
    collect_data(OFF);

    return;
}

/* This function sends trigger to scanner: return 1 if successful else 0 */
int perform_scan()
{
    // send trigger to scan to perform a scan
    res = DEVoutp(0, 1, 2);

    if (res == 2)
    {
        // update user
        Mprintf(4, "Volume aqc");

        // encode this event in the log file
        encode(100);

        // wait for 100 ms
        set_timer(100);
        while(timer_expired() == 0);

        // signal the scanner to stop scanning
        res = DEVoutp(0, 1, 0);

        // update flag since scan was performed in the current trial
        scan_flag = 0;
```

```
        // return successful
        return 1;
    }
    else
    {
        // update user
        Mprintf(4, "Error triggering scanner");
        // return failure
        return 0;
    }
}

/* This function sets reward duration and gives reward */
int give_reward()
{
    if(rew_monkey == 1)
    {
             if(NumRew <  250)  { REWTIME = 120; }
        else if(NumRew <  500)  { REWTIME = 140; }
        else if(NumRew <  750)  { REWTIME = 160; }
        else if(NumRew < 1000)  { REWTIME = 180; }
        else if(NumRew < 1250)  { REWTIME = 200; }
        else if(NumRew < 1500)  { REWTIME = 220; }
        else                    { REWTIME = 250; }

        set_ms_reward_duration(REWTIME);    // windows version

        // provide 3 reward pulses that are 200 ms apart each
        reward(); // first pulse

        set_timer(200);
        while(timer_expired() == 0);
        reward(); // second pulse

        set_timer(200);
        while(timer_expired() == 0);
        reward(); // third pulse

        // update counter on the number of rewards given
        NumRew = NumRew + 1;
        // return successful
        return 1;
    }
    else    {
        // return failure
        return 0;
    }
}
```

Stochastic Figure Ground fMRI experiment Cortex code

```
/*************************************************************
// This Cortex Timing file is used for sparse data acquisition
// with constant eye fixation behaviour
*************************************************************/

#include "css_inc.h"
#include "extern_var.h"

// global variables
#define NumRew    _int0
#define REWTIME   _int1
#define x_move    _int2
#define y_move    _int3
#define step_sz   _int4

#define FIXWIN_SZ _float0

// Parameters of the task, to be adjusted
#define PRE_FIX         400
#define FIX_DURATION    3900
#define REWARD_DURATION 700
#define REWARD_DELAY    1900
#define SOUND_DURATION  6000
#define SOUND_DELAY     950
#define SCAN_DELAY      2100

#define SOUND_STOP  ((SOUND_DELAY) + (SOUND_DURATION))

#define TIMER_PRE_FIX   1
#define TIMER_FIX_DUR   2
#define TIMER_SND_DEL   3
#define TIMER_SND_END   4
#define TIMER_REW_DEL   5
#define TIMER_REW_DUR   6
#define TIMER_SCN_DEL   7

#define OFFSET          10
#define  TIMER_PRE_FIX_2    TIMER_PRE_FIX + OFFSET
#define  TIMER_FIX_DUR_2    TIMER_FIX_DUR + OFFSET
#define  TIMER_REW_DEL_2    TIMER_REW_DEL + OFFSET
#define  TIMER_REW_DUR_2    TIMER_REW_DUR + OFFSET

// Colors used
#define BCK_COL      GcolorABS(0,100,100,100); Gflush(1);   // background
color
#define DOES_FIX_COL GcolorABS(2,255,  0,  0); Gflush(1);   // fix spot
color red when fixating
#define WAIT_FIX_COL GcolorABS(2,128, 64,  0); Gflush(1);   // fix spot
color blue when waiting for fixation

char snd_file_name[50];

// Variables with scope of whole file
int esc_flag   = 0;
int rew_monkey = 0;
int endfix_flag = 0;
int reward_given = 0;
int reward2_success = 0;
int scan_flag = 0;
```

```
int sound_played = 0;
int sound_stop = 0;

int cond_no;
int repeat_num;
int res;
char type[4];
int offset;

// function declaration
int give_reward();
int KEY_action();

/*****************************************/
/*********   Start > main < here   **********/
/*****************************************/
main()
{
    /* initialization routines */

    // Update user display with current value of parameters
    Mprintf(1,"rewards: %d rewtime: %d", NumRew, REWTIME);
    Mprintf(2,"fixtime: %d fixwin: %.2f", FIX_DURATION, FIXWIN_SZ);
    Mprintf(3,"step: %d", step_sz);
    Mprintf(4,"Start up");
    Mprintf(5,"NEW TRIAL");

    // instruct cortex to start collecting data
    collect_data(ON);

    // start storing eye movement data
    put_eye_data_in_buf(ON);

    // store the timestamp of start of eye data
    encode(START_EYE_DATA);

    // extract current condition number
    cond_no = BLOCKget_cond_num();

    // extract current block/repeat number
    repeat_num = get_repeat_num();

    // output single byte of data to scanner to reset the state
    res = DEVoutp(0, 1, 0);

    // clear EOG display window
    clear_eog();

    // set to default size if undefined
    if(FIXWIN_SZ == 0)
        FIXWIN_SZ = 5;

    // specify the nature of fixation window and size
    set_fixwin_params(0, FIXWIN_SZ, FIXWIN_SZ);

    // Move fixation window to a specified item position in a given screen
    ITEM_POSbind_fixspot(TEST0, 0);

    // show path of eye movement
    display_eye_path(ON);
```

```
    // make it invisible
    SCREENdraw_fixwin(WHITE);

    // draw reference visual angle sizes
    SCREENdraw_box_on_eog(0, 0, 20, 20, BLACK);
    SCREENdraw_box_on_eog(0, 0, 10, 10, LIGHTGRAY);

    // render background colour
    BCK_COL
    Gflush(1);

    // No Fix Spot for this task!
    display_fixspot(0);

    // Select value based on condition number
    switch(cond_no)
    {
        case 1:     type = "fig";   offset =  0;    break;
        case 2:     type = "fig";   offset = 45;    break;
        case 3:     type = "fig";   offset = 90;    break;
        case 4:     type = "gnd";   offset = 0;     break;
        case 5:     type = "gnd";   offset = 45;    break;
        case 6:     type = "gnd";   offset = 90;    break;
        default:    type = "sil";   break; // ensures silence since no file
}

    // Random exemplars based on trial number
    sprintf(snd_file_name, "sfg_stimuli_nc10\\sfg_%s_ex_%d.wav",type,
(repeat_num+1+offset));

    // update user
    Mprintf(3,"file: %s",snd_file_name);

    // load sound file and assign a index
    SOUNDload(cond_no, snd_file_name);

    // set mixer volume - left and right, for sound of given index
    SOUNDvol(31, 31, cond_no);

    /***** Trial starts - 1st round of fixation *****/

    endfix_flag = 0;
    reward_given = 0;
    sound_played = 0;
    reward2_success = 0;

    // prefixation routine
    Mprintf(4,"PRE FIX");
    MS_TIMERset(TIMER_PRE_FIX, PRE_FIX);

    // when waiting for prefix to complete
    while (MS_TIMERcheck(TIMER_PRE_FIX) > 0)
    {
        KEY_action();
        Gon_off(TEST0, ON);
        WAIT_FIX_COL
        SCREENdraw_fixwin (BLUE);
    }
```

```
    Mprintf(4,"FIXATION");
    // set all timers
    MS_TIMERset(TIMER_FIX_DUR, FIX_DURATION);
    MS_TIMERset(TIMER_SND_DEL, SOUND_DELAY);
    MS_TIMERset(TIMER_REW_DEL, REWARD_DELAY);
    MS_TIMERset(TIMER_SND_END, SOUND_STOP);

    while( MS_TIMERcheck(TIMER_FIX_DUR) > 0 )
    {
        // perform any keyboard actions
        KEY_action();

        // when fixation is in progress
        if (endfix_flag == 0)
        {
            // check for escape character
            if(KEY_action() == 1)
            {
                Mprintf(5,"BREAK TRIAL");
                break;
            }

            // 1st test of eye fixation
            if (get_fixation_state() == 1)
            {
                // when monkey is fixating
                Mprintf(4,"FIXATED!!");
                DOES_FIX_COL
                SCREENdraw_fixwin (RED);
                rew_monkey = 1;
            }

            set_timer(100); // wait for small gap
            while (timer_expired() == 0); // kill time

            // 2nd test of eye fixation
            if (get_fixation_state() == 0)
            {
                // give monkey a second chance to account for any noise

                set_timer(200); // wait for small gap
                while (timer_expired() == 0); // kill time

                // 3rd test of eye fixation
                if (get_fixation_state() == 0)
                {
                    // when monkey is still not fixating
                    Mprintf(4,"Fixation break");
                    Gon_off(TEST0, OFF);
                    WAIT_FIX_COL
                    SCREENdraw_fixwin (BLACK);
                    rew_monkey = 0;
                    endfix_flag = 1;
                }
            }

            // when it is time to reward monkey
            if ( (MS_TIMERcheck(TIMER_REW_DEL) == 0) && (reward_given == 0)
)
            {
                Mprintf(4,"REWARD 1");
```

```
            REWTIME = 40;
            give_reward();

            encode(103);

            reward_given = 1;
        }
    }

    // when it is time, playout the sound
    if (MS_TIMERcheck(TIMER_SND_DEL) == 0 && sound_played == 0)
    {
        // play sound file once and stop
        // SOUNDprep(cond_no); SOUNDstart(cond_no, 0);
        SOUNDplay(cond_no, 0);
        sound_played = 1;
    }
}

// Reward time
Mprintf(4,"REWARD 2");
MS_TIMERset(TIMER_REW_DUR, REWARD_DURATION);

REWTIME = 50;
if (give_reward() == 0)
{
    encode (5);
}
else
{
    encode(104);
    reward2_success = 1;
}

Gon_off(TEST0, OFF);
WAIT_FIX_COL
SCREENdraw_fixwin (WHITE);

// kill reward time if any is remaining
while (MS_TIMERcheck(TIMER_REW_DUR) > 0)
{
    KEY_action();
}

/***** Half time - 2nd round of fixation *****/

endfix_flag = 0;
reward_given = 0;
scan_flag = 0;
sound_stop = 0;

// prefixation routine
Mprintf(4,"PRE FIX-2");
MS_TIMERset(TIMER_PRE_FIX_2, PRE_FIX);

// when waiting for prefix to complete
while (MS_TIMERcheck(TIMER_PRE_FIX_2) > 0)
{
    KEY_action();
    Gon_off(TEST0, ON);
```

```
    WAIT_FIX_COL
    SCREENdraw_fixwin (BLUE);
}

Mprintf(4,"FIXATION-2");
// set all timers
MS_TIMERset(TIMER_FIX_DUR_2, FIX_DURATION);
MS_TIMERset(TIMER_SCN_DEL, SCAN_DELAY);
MS_TIMERset(TIMER_REW_DEL_2, REWARD_DELAY);

while( MS_TIMERcheck(TIMER_FIX_DUR_2) > 0 )
{
    // perform any keyboard actions
    KEY_action();

    // when it is time, stop the playout of sound
    if ( (MS_TIMERcheck(TIMER_SND_END) == 0) && (sound_stop == 0) )
    {
        SOUNDstop(cond_no);
        sound_stop = 1;
    }

    // when fixation is in progress
    if (endfix_flag == 0)
    {
        // check for escape character
        if(KEY_action() == 1)
        {
            Mprintf(5,"BREAK TRIAL");
            break;
        }

        // 1st test of eye fixation
        if (get_fixation_state() == 1)
        {
            // when monkey is fixating
            Mprintf(4,"FIXATED!!");
            DOES_FIX_COL
            SCREENdraw_fixwin (RED);
            rew_monkey = 1;
        }

        set_timer(100); // wait for small gap
        while (timer_expired() == 0); // kill time

        // 2nd test of eye fixation
        if (get_fixation_state() == 0)
        {
            // give monkey a second chance to account for any noise

            set_timer(200); // wait for small gap
            while (timer_expired() == 0); // kill time

            // 3rd test of eye fixation
            if (get_fixation_state() == 0)
            {
                // when monkey is still not fixating
                Mprintf(4,"Fixation break");
                Gon_off(TEST0, OFF);
                WAIT_FIX_COL
                SCREENdraw_fixwin (BLACK);
```

```
                rew_monkey = 0;
                endfix_flag = 1;
            }
        }

        // when it is time to reward monkey
        if ( (MS_TIMERcheck(TIMER_REW_DEL_2)==0) && (reward_given==0) )
        {
            Mprintf(4,"REWARD 3");

            if (reward2_success == 1)
                REWTIME = 70;
            else
                REWTIME = 50;

            give_reward();

            encode(105);

            reward_given = 1;
        }
    }

    // when it is time, scan
    if ( (MS_TIMERcheck(TIMER_SCN_DEL) == 0) && (scan_flag == 0) )
    {
        // send trigger to perform a scan
        res = DEVoutp(0, 1, 2);

        if (res == 2)
        {
            // update user
            Mprintf(4, "Volume aqc");

            // encode this event in the log file
            encode(200);
        }
        else
        {
            // update user
            Mprintf(4, "Error triggering scanner");
        }

        set_timer(100); // wait for small gap
        while(timer_expired() == 0); // kill time

        // signal the scanner to stop scanning
        res = DEVoutp(0, 1, 0);

        // update flag since scan was performed for the current trial
        scan_flag = 1;

    }
}

// Reward time
Mprintf(4,"REWARD 4");
MS_TIMERset(TIMER_REW_DUR_2, REWARD_DURATION);

if (reward2_success == 1)
```

```
        REWTIME = 100;
    else
        REWTIME = 80;

    if (give_reward() == 0)
    {
        encode (5);
    }
    else
    {
        encode(106);
    }

    Gon_off(TEST0, OFF);
    WAIT_FIX_COL
    SCREENdraw_fixwin (WHITE);

    // kill reward time if any is remaining
    while (MS_TIMERcheck(TIMER_REW_DUR_2) > 0)
    {
        KEY_action();
    }

    /*  End of Trial process */

    // clean up at end of trail
    end_trial();

    // stop storing eye movement data, and EPP data
    put_eye_data_in_buf(0);
    put_epp_data_in_buf(0);

    // store the timestamp of end of eye data
    encode(END_EYE_DATA);

    // do not show path of eye movement
    display_eye_path(OFF);

    // stop collecing data
    collect_data(OFF);

     return;
}  // main


/* This function sets reward duration and gives reward */
int give_reward()
{
    if(rew_monkey == 1)
    {
        set_ms_reward_duration(REWTIME);    // Windows version

        reward();

        NumRew = NumRew + 1;

        return 1;
    }
        return 0;
}
```

```
/*****************************************/
/****  Functions used in the main block  ****/
/*****************************************/

// This function reads out the key pressed and initialises necessary action
// This is adjusted for windows cortex. It won't work with DOS cortex!

int KEY_action()
{
    int xoffset_step = 25;
    int yoffset_step = 25;

    if(KeyPressed() == 1)
    {
        Mprintf(5,"Key was hit");

        if (step_sz == 0)
        {
            step_sz = 4; // Set this as default value for steps
        }

        switch(GetAKey())
        {
            // move eye position

             case VK_SHIFT:     // make y-offset smaller
                YOFFSET = YOFFSET - yoffset_step;
                set_EOGoffset_y(YOFFSET);
                break;

             case VK_TAB:  // make y-offset bigger
                YOFFSET = YOFFSET + yoffset_step;
                set_EOGoffset_y(YOFFSET);
                break;

            case VK_BACK:    // make x-offset bigger
                XOFFSET = XOFFSET + xoffset_step;
                set_EOGoffset_x(XOFFSET);
                break;

            case VK_RETURN:  // make x-offset smaller
                XOFFSET = XOFFSET - xoffset_step;
                set_EOGoffset_x(XOFFSET);
                break;

            // set size of fixation window

            case VK_ADD:
                FIXWIN_SZ = FIXWIN_SZ + 0.25;
                set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
                break;

            case VK_SUBTRACT:
                FIXWIN_SZ = FIXWIN_SZ - 0.25;
                set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
                break;

            case VK_F1:
                FIXWIN_SZ = 1.0;
                set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
```

```
    break;

case VK_F2:
    FIXWIN_SZ = 2.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F3:
    FIXWIN_SZ = 3.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F4:
    FIXWIN_SZ = 4.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F5:
    FIXWIN_SZ = 5.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F6:
    FIXWIN_SZ = 6.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F7:
    FIXWIN_SZ = 7.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F8:
    FIXWIN_SZ = 8.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

case VK_F9:
    FIXWIN_SZ = 9.0;
    set_fixwin_params (0, FIXWIN_SZ, FIXWIN_SZ);
    break;

// Move target position

case VK_LEFT:
    x_move   = -step_sz;
    y_move   = 0;
    break;

case VK_RIGHT:
    x_move   = step_sz;
    y_move   = 0;
    break;

case VK_UP:
    x_move   = 0;
    y_move   = step_sz;
    break;

case VK_DOWN:
    x_move   = 0;
```

```
                y_move   = -step_sz;
                break;

            case VK_HOME:
                x_move   = 0;
                y_move   = 0;
                break;

            // set step size for movement

            case VK_NUMPAD1:
                step_sz =  1; break;// NUMPAD KEY 1

            case VK_NUMPAD2:
                step_sz =  2; break;// NUMPAD KEY 2

            case VK_NUMPAD3:
                step_sz =  3; break;// NUMPAD KEY 3

            case VK_NUMPAD4:
                step_sz =  4; break;// NUMPAD KEY 4

            case VK_NUMPAD5:
                step_sz =  5; break; // NUMPAD KEY 5

            case VK_NUMPAD6:
                step_sz =  6; break;// NUMPAD KEY 6

            case VK_NUMPAD7:
                step_sz =  7; break;// NUMPAD KEY 7

            case VK_NUMPAD8:
                step_sz =  8; break; // NUMPAD KEY 8

            case VK_NUMPAD9:
                step_sz =  9; break;// NUMPAD KEY 9

            case VK_NUMPAD0:
                step_sz = 10; break;// NUMPAD KEY 0

            // give extra reward
            case VK_SPACE:      // SPACE BAR
                reward(); break;

            // finish experiment

            case VK_ESCAPE:
                esc_flag = 1;
                BLOCKset_next(-1, -1);
                Mprintf(5,"User: ESCAPE without resuming !");
                break;
        }
    }
    return esc_flag;
}
```