FUNCTIONAL NEUROIMAGING OF VENTRAL AND DORSAL STREAM PATHWAYS IN THE MACAQUE AUDITORY SYSTEM

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Michael Ortiz Rios
aus Rio Piedras, Puerto Rico, U.S.A.

November - 2015

Tag der mündlichen Prüfung: Februar 16, 2016.

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel

Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Prof. Dr. Nikos K. Logothetis

2. Berichterstatter: Prof. Dr. Hans-Otto Karnath

1. Prüfungskommission: Prof. Josef P. Rauschecker

2. Prüfungskommission: PD Dr. Steffen Hage

Declaration:

I hereby declare that I have produced the work entitled “Functional neuroimaging of ventral and dorsal stream pathways in the macaque auditory system”, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den ....

Date Signature

Dedicated to my beloved family

Midea Malena, Arlan Gael and Micah Mael

“Como no estás experimentado en las cosas del mundo, todas las cosas que tienen algo de dificultad te parecen imposibles. Confía en el tiempo, que suele dar dulces salidas a muchas amargas dificultades"

— Miguel de Cervantes Saavedra

ACKNOWLEDGEMENTS

The journey through my PhD was shared with many friends and colleagues that one-way or another somehow contributed directly or indirectly to improve myself as a person, as a father and/or as scientist.

First of all, I wish to thank my wife, my best friend, my love Midea Malena and my son Arlan Gael whom gave the real purpose of life. Thank you both for always standing up and supporting me during the most hectic period of my PhD and for withstanding those empty nights at the dinner table. I will like also to thank my parents for teaching me how to be perseverant, to followed my dreams and for given me the basis to form myself as a person. Also, Lizannete Velez thank you for shaping me as a person during my early carrier times and for always been there as a real friend.

I will like to give special thanks to my mentor Josef P. Rauschecker for his constant support during my PhD and for always believing in me as a scientist. Also special thanks to my second advisor, Nikos K. Logothetis for his constant support during my PhD, for allowing me to execute my science at the "edge" and for teaching me how to stand up for a strong and solid scientific basis. I will like to thank the rest of my committee members Matthias H. Munk and Hans-Otto Karnath for all their constructive comments and suggestions. I will like to also thank my adopted supervisor Georgios A. Keliris for his constant advise on quantification of fMRI data and for the many suggestions that always improved my work.

Special thanks to my colleague and best friend, Frederico A.C. Azevedo for his technical advice in many aspects of my work and for always been with me in all the up and downs of my PhD life. I am really grateful to count with such a great colleague and friend. I also would like to give special thanks to Pawel Kusmierek for being a really meticulous reviewer, always improving my writing and for his technical advise in programming and acoustics. I will like to also thank Iain DeWitt and Denis Archakov for helping me with data acquisition during my period in Georgetown University. Thanks to David Balla, Thomas Steudel and Mirko Lindig for helping during data acquisition in Tubingen. Many thanks also to Vishal Kapoor, Leonardo Azevedo, Michael C. Schmid and Xin Yu for making comments in previous versions of my manuscripts and for the many useful conversations about science.

CONTENTS

ABSTRACT

One fundamental goal of the brain is to predict sensory events in the environment in order to spatially direct actions. In vision, the ability to identify and locate objects depends on two cortical pathways: a ventral "what" stream supporting object recognition and a dorsal "where" stream supporting object localization. While this hierarchical model received strong support in vision, in audition the analogues functional roles have remained rather elusive, particularly for the dorsal "where" stream. Thus, the objective of this thesis was to explore the functional roles of auditory ventral and dorsal stream pathways in the macaque brain. We first explored the representational structure of natural sounds in early regions of the ventral pathway utilizing single-unit electrophysiology. We then used functional magnetic resonance imaging (fMRI) to map the representation of natural sounds along the ventral pathway including regions outside auditory cortex. Finally, using high-field fMRI we examined the functional representation of acoustic space in auditory cortical regions. Overall, our work confirms the role of the ventral stream in decoding sound identity and extends the evidence suggesting that vocalizations carry information that is represented outside auditory cortex. Moreover, our work in the dorsal stream also confirms the role of a posterior dorsal cortical region specialized in processing spatial information and reconciles competitive theories of spatial coding in auditory cortex. However, our space work also indicates a fundamental difference in the representation format for acoustic space in auditory cortex as compared to visual cortex. Taken together, our work confirms the functional roles of the ventral and dorsal streams and suggests incorporating subcortical level processes in the cortical model for a more integrated framework of acoustic processing in the primate brain.

1 INTRODUCTION

1.1 "What" and "Where" processing streams

In 1969, clinical studies in humans with lesions of occipital cortex led to the finding of functionally specialized cortical brain regions for recognizing and locating visual objects. In particular, behavioural studies in humans with lesions to the ventral temporal cortex suggested that these regions were involved in the recognition of objects while dorsal parietal regions were more important for localization of objects in space (Newcombe, 1969). In 1983, behavioural and lesion studies combined with anatomical tract-tracing techniques in monkeys led to the proposal of a ventral "what" pathway involved in the recognition of objects and a dorsal "where" pathway involved in sound localization (Mishkin et al., 1983).

In 1999, a similar proposal was adapted to the auditory domain (Rauschecker et al., 1997; Rauschecker, 1998; Romanski et al., 1999; Rauschecker and Tian, 2001). This proposal was based on the electrophysiological tuning of neurons and on differential anatomical projections of anterior-ventral regions and posterior-dorsal regions of auditory cortex to segregated functional domains in prefrontal cortex (Romanski et al., 1999). However, in contrast to the original dual-pathway proposal in vision, the analogues functional roles proposed in the auditory domain were based on neuronal tuning properties (Tian et al., 2001) rather than on causal relationships of cortical lesions on behaviour (Mishkin et al., 1983). Up to date, after more than a decade of research, neuroscientists have provided evidence supporting this proposal while others have challenged this notion in the auditory system (Belin and Zatorre, 2000; Bizley and

Cohen, 2013; Middlebrooks, 2002). Consequently, the proposed functional roles of these two pathways in audition have remained elusive given the evidence is mixed.

1.2 Ventral Stream: Functional and anatomical organization

While most of the data supporting a dual-pathway in audition is based on non-invasive functional imaging studies in humans (Binder et al., 2000; Maeder et al., 2001; Alain et al., 2001; Arnott et al., 2004; Leaver and Rauschecker, 2010; Chevillet et al., 2011), evidence based on fMRI in the macaque monkey is very limited. Still, single-unit studies in monkeys together with anatomical tracing techniques provided the foundation for understanding the functional organization of auditory cortex in the framework of the dual-stream hypothesis.

Previous neurophysiological studies of macaque auditory cortex identified three auditory cortical fields, primary (A1), rostral (R) and rostrotemporal (RT). These fields are functionally characterized by their mirror-symmetric tonotopic maps of the cochlea partition (see Kaas and Hackett, 2000a for review). Anatomically, they are characterized by a dense expression of cytochrome oxidase and parvalbumin in middle cortical layers, typical of primary sensory cortices. Together they make up the core auditory cortex, which is bordered by belt areas with similar tonotopic organization and a tertiary processing stage further referred as parabelt cortex. The rostral belt, parabelt and the anterior superior temporal gyrus (STG) send afferent projections into ventrolateral, polar, orbital and medial regions of the prefrontal cortex (PFC) (Jones and Powell, 1970; Hackett et al., 1999; Romanski et al., 1999; Kaas and Hackett, 2000b; Cavada et al., 2000; Hackett, 2011; Yeterian et al., 2012), and together these regional projections form the ventral cortical stream in audition.

Functionally, neurons in belt areas are known to respond more prominently to narrow-band noise rather than to simple pure tones as it is typically observed in core auditory regions (Kusmierek and Rauschecker, 2009 and Rauschecker et al., 1995). Hence, it is hypothesized that along the medial-lateral axis of auditory cortex spectral information gets integrated to represent broader spectral bandwidth. Similarly, along the posterior-anterior axis (e.g. most rostral areas) neurons do not longer show synchronized temporal discharges to temporal rate but rather integrate temporal information into a rate like code (Bendor and Wang, 2007; Hackett, 2011). Temporal integration of acoustic features is necessary for auditory stimulus identification (Rauschecker and Tian, 2000;

(Tian et al., 2013) and combined with increasing spectral integration (Rauschecker et al., 1995) it provides the equivalent to a gradual increase in receptive field complexity along the ventral stream pathway (Desimone et al., 1984; Connor et al., 2007).

Furthermore, neurons in the anterior lateral belt (area AL) were found to be more selective to monkey vocalizations than neurons in more posterior lateral belt areas (Tian et al., 2001). Anatomically, area AL projects to regions further ventral in the anterior STG and to ventrolateral PFC (Romanski et al., 1999). Thus, it is thought that along the anteriolateral axis spectrotemporal information gets integrated to form a complete acoustic "scene" or "auditory object" (Griffiths and Warren, 2004; Nudds, 2010). Such higher-level processes are important for decoding sound meaning and for invariant representations of acoustic features. These functions are generally thought to belong to regions in PFC as originally suggested in the visual domain (Riesenhuber and Poggio, 1999; Romanski and Goldman-Rakic, 2002). However, it remains an open question whether or not neurons at earlier stages of the auditory ventral stream could give rise to higher-level representations of complex natural sounds.

1.2.1 Representation of vocalizations in the macaque

Similarly to studies of the ventral stream in vision and the use of faces as primordial stimuli to probe the neural mechanisms of face identity, in audition the use of species-specific vocalizations provided valuable insight into the neural mechanisms involved in the representation of voice identity. In macaques, vocalizations carry information about the identity of individuals including gender. Moreover, vocalizations also carry information about emotional states of individuals, analogous to human vocal sounds including speech (Cheney and Seyfarth, 1990; Ghazanfar and Hauser, 1999; Yovel and Belin, 2013).

Most of the evidence for vocalization coding in the ventral pathway came from neurophysiological studies that characterized the physiological properties of the anterior STG (Russ et al., 2008; Cohen et al., 2009; Kikuchi et al., 2010, 2014; Perrodin et al., 2011; Fukushima et al., 2014). Several other studies have also examined the representation of vocalizations in the macaque using neuroimaging techniques (Poremba et al., 2003; Petkov et al., 2008; Joly et al., 2012b). A highlight among these studies was the first fMRI study by Petkov et al. (2008), which found an area sensitive to the voice of individuals in the anterior STG region of the macaque. Such imaging

studies in monkeys offered the possibility to bridge the methodological gap between neuroimaging studies in humans and single-unit studies in monkeys in the search for homologies between human speech and monkey vocalizations.

Previous comparative approaches focused on identifying the anatomical networks involved in the processing of speech in humans and vocalizations in non-human primates (Petrides and Pandya, 2009). As the use of functional MRI in monkeys became increasingly more available it provided researchers with a complementary method to study common neuronal networks in the evolution of speech and language (Rauschecker, 2012). Recently, a comparative study in humans and monkeys compared the activations of vocalizations and speech in both species and found similar activations in the STG (Joly et al., 2012a). Moreover, the authors reported that monkeys listening to either vocalizations and/or humans speech showed activation of the orbitofrontal cortex (Joly et al., 2012a, 2012b). Given the fact that the ventral pathway continues into orbitofrontal and ventrolateral PFC from anterior STG regions (Romanski et al., 1999; Cohen et al., 2007), this finding is particularly interesting.

In humans, the cortical processing of speech has been traditionally attributed to the dorsal stream (Wernicke, 1970; Dronkers and Baldo, 2010), but more recent meta-data studies have challenged this notion showing primary involvement of the ventral stream pathway (DeWitt and Rauschecker, 2012). On the other hand, a recent proposal suggested the dorsal stream might also play a role in processing temporal sequences of speech sounds (Rauschecker & Scott, 2009). Specifically, this hypothesis predicts that articulatory sequences from prefrontal and/or premotor cortices might be transmitted into the inferior parietal lobule (IPL) and posterior superior temporal (pST) regions where, presumably, the incoming sensorial speech signal could be compared with the executed motor sequences. Along the same lines, the existence of mirror neurons in premotor cortex area (F5) of the macaque monkey suggests that premotor areas might also be involved in the execution and perception of vocal actions (Rizzolatti et al., 1996). Based on these findings, it was suggested that mirror neurons represent a primitive system from which language circuits could have emerged in humans (Rizzolatti and Arbib, 1998). But it remains an open question whether the representation of vocalizations in the macaque brain also carries information about the motor actions, which are necessary to produce them, as it was shown in humans listening to speech and music (Wilson et al., 2004; Leaver et al., 2009; Rauschecker, 2012). Thus, besides the STG regions, activation of frontal, premotor and parietal regions might also be expected

when monkeys are presented with conspecific vocalization sounds, which might carry information about internal motor representations.

1.3 Dorsal Stream: Functional and anatomical organization

As opposed to the ventral pathway projecting to anterio-ventral regions, the dorsal pathway projects to posterio-dorsal regions of auditory cortex, intraparietal cortex and dorsal PFC (Romanski et al., 1999; Lewis and Van Essen, 2000). The classical functional role of the posterior-dorsal pathway relates to the processing of sound source location and motion perception (Tian et al., 2001; Recanzone, 2000; Maeder et al., 2001a; Warren et al., 2002). In monkeys, the dorsal "where" stream begins in the posterior STG regions and includes caudolateral (CL) and caudomedial (CM) belt areas. Anatomically, both of this areas show darker expression of parvalbumin as compared to other belt regions (Kaas and Hackett, 2000b). Functionally, evidence from single-unit studies indicates that some of these regions might be tonotopically organized (Rauschecker et al., 1997) . However, the functional role of caudal belt regions is more related to spatial processing. In particular, area CL was shown to contain neurons with sharp spatial tuning and to be more selective to spatial position than other areas of the lateral belt (Tian et al., 2001; Woods et al., 2006; Miller and Recanzone, 2009). More recently, electrophysiological studies confirmed the role of area CL in spatial processing and provided new evidence indicating that area CM is faster (in terms of latency) than other cortical fields (Camalier et al., 2012) while even more recent studies emphasised the posterior areas differences by showing that CM is also more temporally precise than CL (Kusmierek and Rauschecker, 2014). While these recent data suggested that the dorsal pathway might be composed of out multiple pathways emerging from posterior areas (Kusmierek and Rauschecker, 2014), the role of CL/CM, including other auditory cortical fields in spatial processing remains a matter of debate.

The notion of a posterior-dorsal region specialized in processing spatial information remains conjectural given that many other single-unit studies in monkeys (Woods et al., 2006; Miller and Recanzone, 2009; Werner-Reiss and Groh, 2008), cats (Middlebrooks et al., 1994; Stecker and Middlebrooks, 2003) and ferrets (Nelken et al., 2008) showed neurons across auditory cortical fields broadly tuned to contralateral space rather than localized in posterior regions in the form of sharply tuned neurons (Tian et al., 2001). While a large number of human neuroimaging studies support the notion of a posterior-

dorsal system for auditory space, only a handful of single-unit studies in macaques exist exploring spatial tuning in macaque posterior regions (Tian et al., 2001; Woods et al., 2006; Miller and Recanzone, 2009; Kusmierek and Rauschecker, 2014).

A critical point to notice is the fact that most of the single-unit studies in macaques do not plot spatial tuning curves, except the study by Woods et al. (2006), which showed neurons broadly tuned to contralateral space. Given functional analogues roles in spatial processing of the posterior-dorsal pathways in vision and audition, it remains highly controversial whether neurons in posterior-dorsal regions of auditory cortex code spatial position with a narrow tune response as in the visual system. This issue is important in light of the possible spatial presentations that could be observed at the population level. Narrow or sharp tuning to spatial position leads to a place code, which may (or may not) be arranged in a topographical map of space as in visual cortex. Alternatively, broad tuning across hemispheres might represent space based on a distributed rate code or hemifield map.

1.3.1 Spatial representations in auditory cortex

The controversy between broad and sharp tuning is a question that relates to how auditory cortical regions could possibly represent spatial position in auditory cortex. If the assumption of analogous functional roles between audition and vision are taken into account as originally proposed by the dual-pathway in audition (Rauschecker and Tian, 2000) one could expect narrow tuning at the cortical level for acoustic spatial position alike in visual cortex for retinal space (Wandell et al., 2007).

If narrow tuning exists in caudal regions and is arranged sequentially in cortical space one could expect a space map. However, the search for spatial maps in the auditory system of the mammalian brain was not fruitful (McAlpine and Grothe, 2003a). Instead, all evidence indicates that auditory space is not represented in the form of a spatial topography or maps but by a contralateral hemifield representation based on neurons widely distributed across auditory cortex (Grothe et al., 2010; Stecker et al., 2005; Werner-Reiss and Groh, 2008). Up to date, no map of acoustic space was ever found in any auditory structure of any mammalian species ever studied with neuronal recordings and/or optical imaging technology. Interestingly however, a map of auditory space does exist in the superior colliculi: multisensory nuclei involved in integrating different sensorial representations of space into a coherent reference frame (King et al.,

1994). Thus, it remains an unresolved mystery whether or not the representation of space is really analogous to the representation of visual space at the cortical level in primates.

1.4 Aim of projects

The purpose of this thesis was to study the functional roles of the dual-pathways in the auditory system focusing on auditory cortical networks. In the following sections, we present three studies conducted to explore how auditory cortical areas represent sound identity and acoustic space. First, we explored the representational structure of complex sounds within auditory cortical fields using population analyses of single-units. Secondly, we explored the representation of vocalizations within and outside auditory cortex using functional magnetic resonance imaging. Finally, we explored the representation of acoustic space in auditory cortex of the macaque using a combination of analytical methods applied to the BOLD signal. The overall implication of this work suggest a fundamental differences in spatial representations between vision and audition and suggest a revised framework that could incorporate subcortical and cortical processing of auditory information in primates.

2 SOUND-IDENTITY PROCESSING IN EARLY AREAS OF THE AUDITORY VENTRAL STREAM IN THE MACAQUE

2.1 Motivation

The anterior ventral stream in audition is thought to be involved in processing sensory information in a hierarchical fashion similarly to the visual ventral stream (Riesenhuber and Poggio, 1999). Thus the size and complexity of receptive fields is predicted to gradually increase along the ventral pathway (Riesenhuber and Poggio, 1999). Thus, an early increase in sensitivity for more complex acoustic features should be present already at relatively early processing stages near primary auditory cortex. For this study we examined how stimulus classes belonging to two main categories, simple and complex sounds, were represented by the population activity patterns in early areas of the auditory ventral stream using representational similarity analyses (Kriegeskorte et al., 2008; Nili et al., 2014).

2.2 Methods

Single-unit and multi-unit recordings from auditory cortical areas were obtained from four awake macaque rhesus monkeys performing a behavioral go/no-go auditory discrimination task. In this task monkeys were trained to release a bar for an infrequent (~10-15%) auditory target (broad-band noise) that was paired with a juice reward. The task served to maintain a constant level of attention to all other stimuli in the sound library. The sound library consisted of simple (i.e. pure tones and narrow-band noise) and complex sounds (i.e. monkey vocalizations and environmental sounds). All analyses were performed at the population level (n neurons = 585) and the responses of the neural population to each sound stimulus were analysed using representational dissimilarity analyses and hierarchical clustering techniques.

In brief, mean firing rate from discrete temporal windows of the peri-stimulus time histograms were calculated. First, the mean response of each neuron to the set of stimuli was aligned in a vector and normalized by subtracting the mean response of the neuron from the vector and dividing it by its Euclidean length. The response normalization for single neurons cancels the bias due to different ranges of firing rates in different neurons. Second, for any pair of stimuli, we calculated the Pearson correlation coefficient (R) between the patterns of responses evoked by the stimuli in the neural population. Multidimensional scaling and hierarchical clustering were employed on the response pattern dissimilarities to visualize the overall structure of the neural responses. The distance or dissimilarity between two stimuli was quantified by $1-R$. A dissimilarity matrix based on neural distance ($1-R$) was a substrate for some clustering methods employed to reveal the response patterns to sound stimuli, while the normalize response patterns were used for K-means clustering.

2.3 Results

Dissimilarity analyses and hierarchical clustering techniques revealed that simple sounds such as pure tones and band pass noise were clustered mainly by sound frequency. The structure for complex sounds formed a distinguishable tree structure far apart from the responses to pure tones and band-pass noise. We further quantified these results using k-means clustering and showed that if a smaller number of discrete clusters were imposed onto the data, the response patterns still reproduced the original stimulus structure with reasonable accuracy.

Using the same analyses for discrete temporal windows we found an emerging pattern of differences between responses across cortical regions. In the caudal region the response patterns to each stimuli were classified poorly, but quickly. Classification success in the middle region showed marginal performance but slightly later than the caudal region. In the rostral region classification evolved slowly: after 20 ms it could perform as good as the middle and caudal regions. Interestingly however, at approximately 60-80 ms past stimulus onset, the rostral region clearly surpassed the classification performance of any other cortical region, reaching ~90% accuracy.

We further explored the degree of classification accuracy for the rostral region and tested whether it reflected the acoustic features of stimuli. First, we applied the same classification methods as for the neural population to multiple acoustic features extracted from the stimuli. Classification based on acoustic features showed comparable results to the classification performance base on neural population in caudal and middle regions. However, classification performance from the rostral region, within 80-160 ms past stimulus, surpassed not only classifications obtained from other cortical regions, but also those extracted from acoustic features of the stimuli.

2.4 Conclusions

In this study we provided evidence showing that stimulus identity, a feature attributed to the anterior ventral stream, could be found at the level of the rostral and rostromedial areas (R/RM). Based on relatively slow development of high classification accuracy and on much higher classification accuracy than that based on acoustic features, we hypothesized about the role of top-down influences of higher regions in the process. Methodologically, this study also demonstrates the advantage of using clustering analyses with short temporal windows yielding more information than more conventional type of analyses.

3 FUNCTIONAL MRI OF THE VOCALIZATION-PROCESSING NETWORK IN THE MACAQUE BRAIN

3.1 Motivation

Species-specific vocalizations in non-human primates convey important information about affective/emotional states as well as information pertaining to the identity of objects and individuals (Ghazanfar and Hauser, 1999). While single-unit studies in the macaque monkey had focused on the neural properties of neurons in the superior temporal gyrus (STG) much less is known about the representation of vocalization networks outside the STG. Here, we examined the global representation of vocalizations using fMRI and compared their activation with other sound categories. We explained our results based on the anatomical networks projecting from the observed active regions in the STG to prefrontal and parietal cortices.

3.2 Methods

We used whole-brain fMRI in awake-behaving macaques to image the BOLD responses to auditory stimuli from three categories: macaque monkey vocalizations,

environmental sounds and scrambled versions of the same monkey vocalizations. The scrambled versions of the monkey vocalizations served as acoustic control given that those sounds preserved the local spectrotemporal features of the original monkey calls. During scanning sessions, monkeys performed a go/no-go auditory discrimination task in which they were required to response with a saccade to a sound target presented randomly within 2 sec after data acquisition. Each trial lasted for about 12 sec: 8 sec of stimulus presentation, 2 sec of data acquisition and a random interval of 2 sec maximum. For each sound condition, we used a general linear model to estimate the hemodynamic response function (HRF) to each stimulus. The HRF was convolved with on-parameter gamma distribution estimate and a square-wave stimulus function. For each condition and contrast, the obtained t-values were used to calculate a mean weighted laterality index (*LIwm*). The indices were based on laterality curves computed from a mean sample ($x_1 - x_n$) that used the t-value threshold as a weighting factor w_i for each data point *x*:

$$LIwm = \sum_{i=1}^{n} Wi * Xi \,/\, \sum_{i=1}^{n} Wi$$

This index yields a single value between − 1 and 1 indicating right- or left-sided hemisphere dominance (Wilke and Lidzba, 2007).

3.3 Results

Using a horizontal 3-T scanner without enhancing or contrast agents we were able to activate the ascending auditory pathways using all three sound categories as regressors of interest. Regions that were significantly activated included: cochlear nucleus, inferior colliculi, medial geniculate body, primary auditory cortex and the STG.

We also compared the activations evoked by monkey vocalizations against those evoked by environmental sounds and/or scrambled monkey vocalizations and found consistent activations in the anterior STG, specifically in AL, rostrotemporolateral (RTL) and rostrotemporal pole (RTp) areas of both monkeys. The laterality index showed a bias towards the right anterior STG patches (e.g. RTL and RTp).

Interestingly, when we compared monkey calls against environmental sounds we found activation patches in regions of the inferior parietal lobule (IPL), specifically in areas PF/PFG.

3.4 Conclusions

Our functional imaging results demonstrate the validity of using complex sounds for mapping cortical and subcortical auditory structures in the macaque brain (Rauschecker et al., 1995; Rauschecker 1998; Poremba et al., 2003). We also substantiate previous fMRI results showing a right-hemispheric bias in the representation of complex sounds in macaque auditory cortex (Joly et al., 2012; Petkov et al., 2008). More importantly, our results extend previous evidence showing increased sensitivity to monkey vocalizations in anterior STG regions (Poremba et al., 2004; Petkov et al., 2008; Kikuchi et al., 2010; Fukushima et al., 2014b). Taken together, our results indicate that vocalizations are processed along the ventral auditory pathway and involve a chain of interconnected regions in the anterior STG and PFC for the recognition and categorization of complex sounds (Rauschecker, 2012).

4 WIDESPREAD AND OPPONENT FMRI SIGNALS REPRESENT SOUND LOCATION IN MACAQUE AUDITORY CORTEX

4.1 Motivation

How acoustic space is represented in primate auditory cortex remains a highly debated argument in the field of auditory neuroscience (Zatorre et al., 2002; Belin and Zatorre, 2000; Middlebrooks, 2002). Two theoretical frameworks proposed different forms of coding for spatial position of sound sources in auditory cortex. While some single-unit studies in monkeys suggested that acoustic space is represented by specialized and localized cortical regions in posterio-dorsal auditory areas (Tian et al., 2001; Woods et al., 2006; Miller and Recanzone, 2009; Kusmierek and Rauschecker, 2014), other studies in monkeys (Werner-Reiss and Groh, 2008; Woods et al., 2006; Miller and Recanzone, 2009), cats (Middlebrooks et al., 1994; Stecker and Middlebrooks, 2003) and ferrets (Nodal et al., 2012; Nelken et al., 2008) indicated that neurons across auditory cortical areas responded broadly to spatial position.

The notion of distributed coding for space (Stecker et al., 2005; Stecker and Middlebrooks, 2003; Magezi and Krumbholz, 2010; Salminen et al., 2009) is in accordance with a subcortical framework (McAlpine et al., 2001; Grothe et al., 2010) which indicates that acoustic space in mammals is coded by an opponent-channel mechanism based on contralateral inhibition (Grothe, 2003; McAlpine et al., 2001; Grothe and Pecka, 2014). While strong evidence exists across mammalian species supporting this model, the majority of neuroimaging research in humans supports a different perspective which suggest a local and specialized cortical region for space in posterior auditory cortex (Maeder et al., 2001b; Baumgart et al., 1999; Warren et al., 2002; Griffiths et al., 1996). However, functional imaging data in the macaque monkey mapping the auditory spatial domain is not available. For this work we investigated how auditory space is presented across auditory cortical areas of each hemisphere of the macaque monkey utilizing high-resolution fMRI.

4.2 Methods

Binaural sound recordings were used to create a virtual acoustic space from which spatial sounds containing all individual spatial cues (interaural time delays (ITDs), interaural level differences (ILDs) and spectral cues) were played during experiments. High field imaging (0.75 x 0.75 x 2 mm resolution) was performed in anesthetized and awake monkeys to image the activation patterns to spatial sounds.

Before spatial mapping, we used phase-encoding methods to identify tonotopy maps and to use their mirror-reversals to delineate areal boundaries in auditory cortex. The coherence of the fMRI time series at the stimulus presentation cycle was used to measure the strength of the BOLD response amplitude in each voxel. Coherence measures the ratio of the amplitude at the fundamental frequency to the signal variance, ranging between 0 and 1. The measure of coherence used was:

$$C(f_0) = A(f_0)/(\sum_{f=f_0-\frac{\Delta f}{2}}^{f_0+\frac{\Delta f}{2}} A\,(f)^2)^{\frac{1}{2}}$$

where f_0 was the stimulus frequency, $A(f_0)$ the amplitude of the signal at that frequency, $A(f)$ the amplitude of the harmonic term at the voxel temporal frequency f and Δf the bandwidth of frequencies in cycles/scan around the fundamental frequency f_0. For all tonotopy stimuli f_0 corresponds to twelve cycles (12/1200 sec = 0.01 Hz) and

Δf corresponds to the frequencies around the fundamental excluding the second and third harmonics. Any given coherent voxel exceeding threshold was used to plot the phase peak to each stimulus condition (e.g. frequency range). The same method was used to map the spatial domain in auditory cortex.

Furthermore, we used general linear modelling to obtain the BOLD responses to each spatial sector. To determine the degree of contralaterality between responses across hemifields we calculated hemispheric differences with a mean weighted laterality index (*LIwm*) as described above (see section 3.2). In addition, spatial tuning curves were calculated from the spatial spread of the BOLD responses for each auditory cortical region using circular statistics. Finally, multivariate pattern dissimilarity was used to compare spatial representations across cortical regions with a hemifield code model.

4.3 Results

Using phase-encoding methods we mapped the tonotopic organization in auditory cortex, identified mirror-reversal frequency maps and delineated areal boundaries between regions. For the spatial domain we found a broad hemifield-tuning response and a lack of topographical organization at the millimeter scale in each auditory cortical region.

Using a general linear model of the BOLD response we found that the representation of space was dominated by distributed hemifield representation of positive and negative BOLD signals across the cerebral hemispheres. The positive BOLD responses showed a maximum amplitude and spatial spread for contralateral sectors (e.g. near the ear ~ ± 90° - 120°). On average, each cortical field of the same hemisphere showed similar broad hemifield tuning curves.

Contrast between the responses to each hemifield sector also showed a robust contralateral bias in auditory cortex and inferior colliculi. Given the profound roles that inhibition plays in the medial superior olive (e.g. at the brainstem level) for coding ITDs we conducted control experiments in which we tested the effects of removing ITD cues from the original spatial sounds. These control experiments suggested that the suppression effects (either in the form of small patches of positive BOLD and/or negative BOLD responses) were due to inhibitory inter-hemispheric processes brought

by ITD cues. In addition, the lack of suppression caused by the removal of ITD cues particularly affected the right hemisphere response necessary for contralateral tuning.

Finally, our multivariate pattern dissimilarity analyses showed that BOLD signals in right posterior STG were greatly modulated by spatial sounds showing that this region contains more spatial information than any other field of auditory cortex, and codes full acoustic azimuthal space similarly to a hemifield rate code.

4.4 Conclusions

In summary, our results showed that the representation of auditory space relies on distributed as well as specialized mechanisms of cortical processing. The functional representation of auditory space was biased towards contralateral space and dependent on the suppression effects provided by ITD cues. Suppression was most pronounced in the right posterior region, allowing this region to modulate its activity to a greater degree than any other field of auditory cortex and thus to code full acoustic space. Such functional specialization of the right posterior region suggests similar cortical machinery between humans and monkeys for processing auditory space. Taken together, our results reconcile seemingly contradictory views of auditory space coding (Belin and Zatorre, 2000) showing that (1) the representation of space follows a broadly-tuned hemifield code (at least on the millimeter scale used in this fMRI study), but (2) that this representation generates higher sensitivity for spatial location in posterior-dorsal auditory cortex, as commonly seen in spatial studies of primate auditory cortex (Tian et al., 2001; Miller and Recanzone, 2009; Griffiths et al., 1996; Warren et al., 2002).

5 DISCUSSION

5.1 Higher level representations in early "what" stream

Population analyses of single-units (Kuśmierek et al., 2012) and functional imaging experiments (Ortiz-Rios et al., 2015) confirmed the role of an anterior ventral stream involved in processing complex natural sounds (Rauschecker and Tian, 2000). Our results also extended the evidence implicating an anterior STG region in coding species-specific vocalizations (Poremba et al., 2004; Petkov et al., 2008; Kikuchi et al., 2010; Fukushima et al., 2014) and are in accordance with previous single-unit results showing increased selectivity for vocalizations in area AL (Tian et al., 2001). Furthermore, our single-unit studies of the rostral region also found increased sensitivity to other sound categories (Kuśmierek et al., 2012). In addition, our fMRI results showed that even when control stimuli, which preserved the low-level spectrotemporal acoustic information were used, the anterior STG sensitivity for vocalizations still held. Taken, together our studies using both single-unit electrophysiology and fMRI were consistent with the notion of hierarchical processing along the ventral stream pathway.

Hierarchical processing along the ventral stream is thought to emerge from neural ensembles whose selectivity for more complex acoustic features increases, as shown in anterior regions of the STG (Poremba et al. 2004; Petkov et al. 2008; Kikuchi et al. 2010). However, our single-unit data indicated that higher-level representations for stimulus-identity were already present at much earlier cortical regions (e.g. area R and

RM). These results suggest that at the level of the rostral region the representation may shift from feature-based towards object-based.

This transformation is relevant to consider in comparison to the representation of faces in the inferior temporal cortex (e.g. along the visual ventral stream). In the inferior temporal (IT) cortex an increase in receptive field size leads to invariant representation of 3D objects (Logothetis et al., 1994) and faces (Freiwald and Tsao, 2010). Such responses in anterior temporal regions are essential for building invariant object representations, which could be robust to changes to individual features as in higher-level visual areas. Given that both visual and auditory ventral pathways project anteriorly to the temporal pole and subsequently to the hippocampal formation it might be worthwhile to investigate how multimodal representation of objects might get integrated in space and time in the temporal lobe (Howard and Eichenbaum, 2014).

In our fMRI experiments using complex sounds we observed activations in higher-level visual areas, such as the middle temporal (MT) and IT areas. These regions are known to be involved in the processing of visual motion (Maunsell and Van Essen, 1983) and face identity, respectively (Tsao et al., 2006; Ku et al., 2011). Their activation by environmental sounds raises the possibility that multisensory processing of dynamic audio-visual stimuli, such as facial expressions which naturally occur in conjunction with vocalizations and/or motion of the face (Furl et al., 2012; Polosecki et al., 2013; Perrodin et al., 2014) might recruit recurrent neural networks between superior and inferior temporal cortices.

While most studies of higher-level visual cortex focused on the use of static faces, as stimuli to prompt the neural mechanism of object-identity, more recent research has begun to explore the use of dynamic motion in faces (Polosecki et al., 2013; Furl et al., 2012). Such stimuli are more comparable to complex sound "objects" given that sounds by nature are never static and their representation might involve recurrent networks for a time-evolving acoustic scene (Sak et al., 2014). Thus the classical comparison between static faces and vocalizations might not be the most adequate for the functional analogy between ventral stream function in vision and audition. Our analyses of the neural population patterns showed a gradual increase in categorical structure in the representation of natural sounds across time at the level of auditory cortex (Kuśmierek et al., 2012). However, such higher-level representations for natural sounds are typically associated with prefrontal function, and we hypothesized about possible role of top-

down modulation arising from prefrontal or anterior temporal regions in sound categorization.

In terms of the auditory ventral pathways, the ventrolateral PFC is known to be involved in higher-level integrative processes for the cognitive control of vocalizations and the interpretation of semantic content in vocalizations (Romanski and Averbeck, 2009). The activation patches we found in PFC with our fMRI studies could be due to categorical or affective information imbedded in the vocalizations. Neurons with vocalization-sensitivity are found along with face-sensitive neurons in the vlPFC (Romanski et al., 2005), allowing these regions to integrate vocalizations with the corresponding facial gestures (Romanski and Goldman-Rakic, 2002; Cohen et al., 2007; Diehl and Romanski, 2014). Thus, the processing of vocalizations in the macaque encompasses a network of areas, which involve the superior temporal gyrus and prefrontal cortices. Moreover, the involvement of prefrontal regions in cognitive control of gestural actions in vocalizations might require sensorimotor coupling between regions involved in the perception and production of active vocal behaviour.

5.2 Sensorimotor representations in vocal perception

Ventrolateral PFC along with the ventral premotor cortex (vPMC) is implicated in the cognitive control and initiation of vocalizations in the macaque monkey (Hage and Nieder, 2013). In our fMRI studies utilizing vocalizations we found an engagement of both vlPFC and vPMC (in particular area 44) in the processing of monkey vocalizations. These results are in accordance with previous fMRI studies in humans reporting premotor activation during the presentation of speech sounds (Wilson et al. 2004). It could be speculated that the vPMC regions are the source of an efference copy signal (Kauramäki et al., 2010; Rauschecker and Scott, 2009), which might be, for example, responsible for the suppression of auditory cortex during self-initiated vocalizations (Eliades and Wang, 2003).

Area 44 of the vPMC in the macaque monkey controls the orofacial musculature of the face (Petrides and Pandya, 2009) and is strongly linked via the superior longitudinal fasciculus (SLF III) to the PF/PFG region in parietal cortex (Rozzi et al. 2006; Seltzer and Pandya, 1978ab). When we compared vocalizations against environmental sounds

we observed differential activation of the PF/PFG region of the right hemisphere. The PF/PFG in the macaque is thought to be a homologue to the human supramarginal gyrus (Petrides and Pandya, 2009) an area which was linked to language perception and receptive aphasia (Dronkers and Baldo, 2010).

Previous electrophysiological studies in area 44 of the macaque found neurons sensitive to orofacial actions and somatomotor stimulation of the face (Hyvärinen and Shelepin, 1979; Leinonen et al. 1979). Thus, it is conceivable that the cortical representation of vocalizations may also include information related to the articulatory sequences and actions required to utter them as similarly suggested in humans listening to speech sounds (Wilson et al. 2004; Aboitiz, 2012). Thus, it is reasonable to speculate that premotor and parietal regions may contain an internal representation of the temporal and motor sequences necessary to utter them during vocalizations, as suggested by Rauschecker and Scott (2009). While parietal cortex, specifically regions inside the intraparietal sulcus (IPS), are known to receive auditory projections (Lewis and VanEssen, 2000) and contain neurons that respond to auditory and multimodal stimuli (Stricanne et al., 1996; Grunewald et al. 1999; Cohen and Andersen 2000, Cohen, 2009), it is not known how areas PF/PFG of the IPL respond to auditory or naturalistic stimuli such as vocalizations. Today, it remains a challenge to study sensorimotor interactions among distal cortical regions and especially challenging to train macaque monkeys to vocalize on command. However, a new study by Hage et al. (2013) opens a new window for experimental studies in which macaques could be trained to vocalize (Fukushima et al., 2014a) during chronic neurophysiological preparations and/or even inside an MRI scanner. Future experiments along these lines might reveal the overall network dynamics involved in the production and execution of vocalizations in the macaque monkey and provide the basis for further comparative studies with humans.

5.3 Widespread representation of acoustic space in auditory cortex

In humans and monkeys, it is generally accepted that auditory spatial processes belong to a dorsal “where” pathway (Romanski et al., 1999; Rauschecker and Tian, 2000) which includes specialized cortical areas CL/CM of the posterior STG (Rauschecker and Tian, 2000; Tian et al., 2001). However, how the posterior regions including other auditory cortical areas code auditory space has remained a matter of controversy in functional imaging studies in humans and single-unit studies in monkeys (Zatorre et al.,

2002; Romanski et al., 1999; Middlebrooks, 2002). The general debate is surrounded by whether or not auditory space is represented locally in specialized cortical areas in posterior regions or distributed throughout the superior temporal plane. Our functional mapping studies demonstrated that the functional representation of auditory space in the macaque as measured by functional MRI is distributed throughout auditory cortex. This finding is in accordance with previous cortical lesion studies (Jenkins and Masterton, 1982; Heffner and Masterton, 1975; Nodal et al., 2012), optical imaging results (Nelken et al., 2008) and single-unit studies in cats and monkeys (Middlebrooks et al., 1994; Miller and Recanzone, 2009; Woods et al., 2006; Stecker et al., 2005) showing neurons responding broadly across auditory cortical fields. While previous single-unit studies of the macaque area CL indicated the existence of cells with narrowly tuned responses to sound location (Tian et al., 2001), we found essentially no difference in the spatial tuning curves across cortical regions. This finding indicated that voxels in and out of posterior regions, on average, shared common spatial tuning curves (Stecker and Middlebrooks, 2003). Given that fMRI samples (at the millimeter level) the average signal across a large neuronal population it might be be possible that narrow tuning cells may exist intermixed in the cortex.

On the other hand, if we consider the original suggestion of a posterior specialized region composed of sharply tuned neurons; it might be conceivable that if such cells are arranged in compact format (e.g. in place code), a map of space could exist. We tested this notion by implementing a modified version of the phase-mapping techniques (Barton et al., 2012) typically used for mapping the retinotopic organization in visual cortex (Wandell et al., 2007). When we used this method to map the frequency organization we were able to obtain tonotopic maps as previously shown with neurophysiological and functional imaging techniques in the macaque auditory cortex (Merzenich and Brugge, 1973; Rauschecker et al., 1997; Petkov et al., 2006). However, for the spatial domain two broadly tuned responses (one in each hemisphere) indicated a contralateral hemifield representation for auditory space.

While previous neuroimaging studies in humans presenting spatial sounds or moving virtual sounds in space showed different results in relation to the degree of contralaterality (Werner-Reiss and Groh, 2008; Zatorre et al., 2002; Krumbholz et al., 2007), single-unit studies in cats and monkeys invariably show a contralateral bias in the firing rate of cortical neurons (Woods et al., 2006; Miller and Recanzone, 2009; Werner-Reiss and Groh, 2008; Stecker and Middlebrooks, 2003). When we further

investigated these results by looking at each spatial sector we observed an opponent pattern of positive and negative BOLD responses across the cerebral hemispheres. This finding indicated that a suppression effect for ipsilateral sound sources had caused the overall amplitude shift towards a contralateral representation. When we compared equidistant hemifield sectors we observed a strong contralateral bias in the BOLD response contrast including the inferior colliculi.

Given the inhibitory roles which ITD cues exert in coding space at the subcortical level (Grothe, 2003), we explored the effects of subtracting ITD cues from the original spatial sounds (e.g. sounds based only on ILD and spectral cues) in the overall BOLD responses at the cortical level. Such manipulation of spatial cues removed the negative BOLD responses, but instead generated a facilitation of the response in the right hemisphere. The removal of the suppression effect due to the removal of ITD from the spatial sounds affected contralateral tuning in the representation of space at the cortical level. Lateralization indices further confirmed these results by showing near-zero values around the midline, which indicated that ITD cues were essential in shifting the hemifield representation across the hemispheres. Furthermore, our results are in accordance with the inhibitory roles of ITD cues in subcortical and cortical system of mammals, where at least in humans, right hemisphere lesions show severe deficits in sound localization (Bisiach et al., 1984; Spierer et al., 2009) and motion discrimination (Griffiths et al., 1996) for sounds based solely on ITD cues (Spierer et al., 2009).

We also noticed a selective suppression of the right posterior STG region for sound sources in the ipsilateral side, despite the overall positive BOLD response. The concentric BOLD modulation indicated that spatial information was greatly emphasized and deemphasized in anterior and posterior regions of AC. Such functional dynamics allowed in particular the right posterior region to segregate the responses similarly to their actual position. In humans, the posterior right region is known to be sensitive to spatial auditory motion (Baumgart et al., 1999; Krumbholz et al., 2007) and our spatial coding results support this finding (Werner-Reiss and Groh, 2008; Salminen et al., 2009). In general, our dissimilarity analyses also provided support to the hypothesis of a specialized posterior region in macaque auditory cortex (Tian et al., 2001; Rauschecker and Tian, 2000) and coincides with previous single-unit studies in monkeys (Tian et al., 2001; Miller and Recanzone, 2009; Woods et al., 2006; Kusmierek and Rauschecker, 2014) by showing that posterior regions carried more spatial information than primary

cortical regions in agreement with a posterior-dorsal "where" auditory pathway (Romanski et al., 1999; Rauschecker and Tian, 2000).

5.4 Conclusions

The functional analogy between visual and auditory streams has provided a platform for understanding the neural mechanisms of higher-level processing of acoustic information in cortex. More than two decades of research in the cortical systems of primates has lead neuroscientist to either support or reject the cortical centric view of auditory processing. The fact that the processing of complex natural sounds involves a large number of widespread cortical and subcortical regions across the macaque brain challenges the notion of a purely cortical and unimodal sensory model. Instead such evidence calls for a more comprehensive view of auditory processing to a model that incorporates cortical and subcortical processing of acoustic information.

The necessity to incorporate the subcortical processes in the overall cortical model is more sounding for ITD coding given the complexity of interactions that takes place at the brainstem level. While it was originally thought that ITDs (the most salient spatial cues) were coded by a topographical arrangement of coincidence detectors as in the barn owl (Jeffress, 1948; Knudsen and Konishi, 1978), more than a decade of research in multiple mammalian species has revealed a different mechanism (Grothe et al., 2010; McAlpine and Grothe, 2003b), one in which ITDs are coded by an opponent hemifield code based on contralateral inhibition (McAlpine et al., 2001; Grothe, 2003; Grothe et al., 2010; McAlpine and Grothe, 2003b). Our spatial mapping results here support this view at the cortical level by showing an opponent representation across the cerebral hemispheres rather than a local spatial topography.

Thus, in comparison to the visual dorsal stream, the format of representation in auditory cortex is fundamentally different from the spatial topography of the retina in visual cortical regions (Wandell et al., 2007). The lack of topographic organization, together with our results showing positive and negative BOLD responses across hemispheres, strongly support the opponent-channel model (Stecker et al., 2005) for auditory space in the macaque, rather than a place code format as generally seen in retinotopic space maps in visual cortex. Although similar functional cortical organization for the representation of visual and auditory space was proposed more than a decade ago (Rauschecker and Tian, 2000), the functional analogy is weakened by the consideration of the auditory feature analogous to "where" (position of the sensory periphery) in

vision. In the auditory system, position in the sensory periphery is based on sound frequency rather than spatial position as in vision. Thus the functional analogy between visual and auditory spatial representations in cortex of primates may not reach that far.

REFERENCES

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., and Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20738–43. doi:10.1073/pnas.1213381109.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H.-J., and Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature* 400, 724–726. Available at: http://dx.doi.org/10.1038/23390.

Belin, P., and Zatorre, R. J. (2000). "What", "where" and "how" in auditory cortex. *Nat. Neurosci.* 3, 965–966. doi:10.1038/79890.

Bendor, D., and Wang, X. (2007). Differential neural coding of acoustic flutter within primate auditory cortex. *Nat. Neurosci.* 10, 763–771. doi:10.1038/nn1888.

Bisiach, E., Cornacchia, L., Sterzi, R., and Vallar, G. (1984). Disorders of perceived auditory lateralization after lesions of the right hemisphere. *Brain* 107 (Pt 1, 37–52.

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat Rev Neurosci* 14, 693–707. doi:10.1038/nrn3565.

Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., de la Mothe, L. a., and Hackett, T. a. (2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl. Acad. Sci.* 109, 18168–18173. doi:10.1073/pnas.1206387109.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., and Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb. Cortex* 10, 220–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10731218 [Accessed October 15, 2014].

Cohen, Y. E., Theunissen, F., Russ, B. E., and Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol.* 97, 1470–84. doi:10.1152/jn.00769.2006.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U. S. A.* 109, E505–14. doi:10.1073/pnas.1113427109.

Dronkers, N. F., and Baldo, J. V. (2010). "Language: Aphasia," in *Encyclopedia of Neuroscience*, 343–348. doi:10.1016/B978-008045046-9.01876-3.

Eliades, S. J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89, 2194–207. doi:10.1152/jn.00627.2002.

Freiwald, W. A., and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–51. doi:10.1126/science.1194908.

Fukushima, M., Saunders, R. C., Fujii, N., Averbeck, B. B., and Mishkin, M. (2014a). Modeling vocalization with ECoG cortical activity recorded during vocal production in the macaque monkey. in *36th Annual International Conference of the IEEE Engineering in Medicince and Biology Society*, 6794–6797. doi:10.1109/EMBC.2014.6945188.

Fukushima, M., Saunders, R. C., Leopold, D. A., Mishkin, M., and Averbeck, B. B. (2014b). Differential coding of conspecific vocalizations in the ventral auditory cortical stream. *J. Neurosci.* 34, 4665–76. doi:10.1523/JNEUROSCI.3969-13.2014.

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., and Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J. Neurosci.* 32, 15952–62. doi:10.1523/JNEUROSCI.1992-12.2012.

Ghazanfar, A., and Hauser, M. (1999). The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cogn. Sci.* 3, 377–384. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10498928 [Accessed October 14, 2014].

Griffiths, T. D., Rees, A., Witton, C., Shakir, R. A., Henning, G. B., and Green, G. G. (1996). Evidence for a sound movement area in the human cerebral cortex. *Nature* 383, 425–427. doi:10.1038/383425a0.

Griffiths, T. D., and Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892. doi:10.1038/nrn1538.

Grothe, B. (2003). New roles for synaptic inhibition in sound localization. *Nat. Rev. Neurosci.* 4, 540–550. doi:10.1038/nrn1136.

Grothe, B., and Pecka, M. (2014). The natural history of sound localization in mammals – a story of neuronal inhibition. *Front. Neural Circuits* 8. doi:10.3389/fncir.2014.00116.

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012. doi:10.1152/physrev.00026.2009.

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–46. doi:10.1016/j.heares.2010.01.011.

Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817, 45–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9889315 [Accessed October 15, 2014].

Heffner, H., and Masterton, B. (1975). Contribution of auditory cortex to sound localization in the monkey (Macaca mulatta). *J. Neurophysiol.* 38, 1340–1358.

Howard, M. W., and Eichenbaum, H. (2014). Time and space in the hippocampus. *Brain Res.*, 1–14. doi:10.1016/j.brainres.2014.10.069.

Jeffress, L.A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35–39. doi:10.1037/h0061495.

Jenkins, W. M., and Masterton, R. B. (1982). Sound localization: effects of unilateral lesions in central auditory pathways. *J. Neurophysiol.* 47, 987–1016.

Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., and Orban, G. a (2012a). Processing of vocalizations in humans and monkeys: A comparative fMRI study. *Neuroimage* 62, 1376–1389. doi:10.1016/j.neuroimage.2012.05.070.

Joly, O., Ramus, F., Pressnitzer, D., Vanduffel, W., and Orban, G. A. (2012b). Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cereb. Cortex* 22, 838–53. doi:10.1093/cercor/bhr150.

Jones, E. G., and Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4992433 [Accessed October 15, 2014].

Kaas, J. H., and Hackett, T. A. (2000a). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11793–11799. doi:10.1073/pnas.97.22.11793.

Kaas, J. H., and Hackett, T. A. (2000b). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11793–9. doi:10.1073/pnas.97.22.11793.

Kikuchi, Y., Horwitz, B., and Mishkin, M. (2010). Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J. Neurosci.* 30, 13021–30. doi:10.1523/JNEUROSCI.2267-10.2010.

King, A. J., Moore, D. R., and Hutchings, M. E. (1994). Topographic representation of auditory space in the superior colliculus of adult ferrets after monaural deafening in infancy. *J. Neurophysiol.* 71, 182–194.

Knudsen, E. I., and Konishi, M. (1978). A neural map of auditory space in the owl. *Science* 200, 795–797. doi:10.1126/science.644324.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4. doi:10.3389/neuro.06.004.2008.

Krumbholz, K., Hewson-Stoate, N., and Schönwiesner, M. (2007). Cortical response to auditory motion suggests an asymmetry in the reliance on inter-hemispheric connections between the left and right auditory cortices. *J. Neurophysiol.* 97, 1649–1655. doi:10.1152/jn.00560.2006.

Ku, S.-P., Tolias, A. S., Logothetis, N. K., and Goense, J. (2011). fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70, 352–62. doi:10.1016/j.neuron.2011.02.048.

Kuśmierek, P., Ortiz, M., and Rauschecker, J. P. (2012). Sound-identity processing in early areas of the auditory ventral stream in the macaque. *J. Neurophysiol.* 107, 1123–41. doi:10.1152/jn.00793.2011.

Kusmierek, P., and Rauschecker, J. P. (2014). Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol.* 111, 1671–85. doi:10.1152/jn.00436.2013.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., and Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *J. Neurosci.* 29, 2477–85. doi:10.1523/JNEUROSCI.4921-08.2009.

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11058227 [Accessed October 15, 2014].

Logothetis, N. K., Pauls, J., Bülthoff, H. H., and Poggio, T. (1994). View-dependent object recognition by monkeys. *Curr. Biol.* 4, 401–414. doi:10.1016/S0960-9822(00)00089-0.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001a). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–16. doi:10.1006/nimg.2001.0888.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001b). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–816. doi:10.1006/nimg.2001.0888.

Magezi, D. A., and Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J. Neurophysiol.* 104, 1997–2007. doi:10.1152/jn.00424.2009.

Maunsell, J. H., and Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* 49, 1127–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6864242 [Accessed October 15, 2014].

McAlpine, D., and Grothe, B. (2003a). Sound localization and delay lines - Do mammals fit the model? *Trends Neurosci.* 26, 347–350. doi:10.1016/S0166-2236(03)00140-1.

McAlpine, D., and Grothe, B. (2003b). Sound localization and delay lines - Do mammals fit the model? *Trends Neurosci.* 26, 347–350. doi:10.1016/S0166-2236(03)00140-1.

McAlpine, D., Jiang, D., and Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.* 4, 396–401. doi:10.1038/86049.

Merzenich, M. M., and Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* 50, 275–296. doi:10.1016/0006-8993(73)90731-2.

Middlebrooks, J. C. (2002). Auditory space processing: here, there or everywhere? *Nat. Neurosci.* 5, 824–826. doi:10.1038/nn0902-824.

Middlebrooks, J. C., Clock, A. E., Xu, L., and Green, D. M. (1994). A panoramic code for sound location by cortical neurons. *Science* 264, 842–844. doi:10.1126/science.8171339.

Miller, L. M., and Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5931–5935. doi:10.1073/pnas.0901023106.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417. doi:10.1016/0166-2236(83)90190-X.

Newcombe, F. (1969). Missile wounds of the brain: A study of psycho- logical deficits. Oxford: Oxford University Press.

Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J., and Schnupp, J. W. H. (2008). Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *J. Neurophysiol.* 99, 1928–1941. doi:10.1152/jn.00469.2007.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* 10. doi:10.1371/journal.pcbi.1003553.

Nodal, F. R., Bajo, V. M., and King, A. J. (2012). Plasticity of spatial hearing: behavioural effects of cortical inactivation. *J. Physiol.* 590, 3965–3986. doi:10.1113/jphysiol.2011.222828.

Nudds, M. (2010). What Are Auditory Objects? *Rev. Philos. Psychol.* 1, 105–122. doi:10.1007/s13164-009-0003-6.

Ortiz-Rios, M., Kuśmierek, P., DeWitt, I., Archakov, D., Azevedo, F. A. C., Sams, M., Jääskeläinen, I. P., Keliris, G. A., and Rauschecker, J. P. (2015). Functional MRI of the vocalization-processing network in the macaque brain. *Front. Neurosci.* 9, 113. doi:10.3389/fnins.2015.00113.

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2014). Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J. Neurosci.* 34, 2524–37. doi:10.1523/JNEUROSCI.2805-13.2014.

Petkov, C. I., Kayser, C., Augath, M., and Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4, 1213–1226. doi:10.1371/journal.pbio.0040215.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–74. doi:10.1038/nn2043.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170.

Polosecki, P., Moeller, S., Schweers, N., Romanski, L. M., Tsao, D. Y., and Freiwald, W. A. (2013). Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J. Neurosci.* 33, 11768–73. doi:10.1523/JNEUROSCI.5402-11.2013.

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427, 448–451. doi:10.1038/nature02268.

Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., and Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science* 299, 568–72. doi:10.1126/science.1078900.

Rauschecker, J. P. Parallel processing in the auditory cortex of primates. *Audiol. Neurootol.* 3, 86–103. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9575379 [Accessed October 14, 2014].

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* 4, 7. doi:10.3389/fnevo.2012.00007.

Rauschecker, J. P., and 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. doi:10.1073/pnas.97.22.11800.

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7701330 [Accessed October 14, 2014].

Rauschecker, J. P., Tian, B., Pons, T., and Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *J. Comp. Neurol.* 382, 89–103.

Recanzone, G. H. (2000). Spatial processing in the auditory cortex of the macaque monkey. *Proc Natl Acad Sci U S A* 97, 11829–11835. doi:10.1073/pnas.97.22.11829\r97/22/11829 [pii].

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci* 2, 1019–1025. doi:10.1038/14819.

Rizzolatti, G., and Arbib, M. a. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi:10.1016/S0166-2236(98)01260-0.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb. Cortex* 17 Suppl 1, i61–9. doi:10.1093/cercor/bhm099.

Romanski, L. M., and Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu. Rev. Neurosci.* 32, 315–46. doi:10.1146/annurev.neuro.051508.135431.

Romanski, L. M., Averbeck, B. B., and Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* 93, 734–47. doi:10.1152/jn.00675.2004.

Romanski, L. M., and Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–6. doi:10.1038/nn781.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–6. doi:10.1038/16056.

Sak, H., Senior, A., and Beaufays, F. (2014). Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *arXiv Prepr. arXiv1402.1128*.

Salminen, N. H., May, P. J. C., Alku, P., and Tiitinen, H. (2009). A population rate code of auditory space in the human cortex. *PLoS One* 4. doi:10.1371/journal.pone.0007600.

Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M., and Clarke, S. (2009). Hemispheric competence for auditory spatial representation. *Brain* 132, 1953–1966. doi:10.1093/brain/awp127.

Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (2005). Location coding by opponent neural populations in the auditory cortex. in *PLoS Biology*, 0520–0528. doi:10.1371/journal.pbio.0030078.

Stecker, G. C., and Middlebrooks, J. C. (2003). Distributed coding of sound locations in the auditory cortex. *Biol. Cybern.* 89, 341–349. doi:10.1007/s00422-003-0439-1.

Tian, B., Kuśmierek, P., and Rauschecker, J. P. (2013). Analogues of simple and complex cells in rhesus monkey auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7892–7. doi:10.1073/pnas.1221062110.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–3. doi:10.1126/science.1058911.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–4. doi:10.1126/science.1119983.

Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron* 56, 366–383. doi:10.1016/j.neuron.2007.10.012.

Warren, J. D., Zielinski, B. A., Green, G. G. R., Rauschecker, J. P., and Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron* 34, 139–148. doi:10.1016/S0896-6273(02)00637-2.

Werner-Reiss, U., and Groh, J. M. (2008). A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. *J. Neurosci.* 28, 3747–3758. doi:10.1523/JNEUROSCI.5044-07.2008.

Wernicke, C. C. (1970). The aphasic symptom-complex: A psychological study on an anatomical basis. *Arch. Neurol.* 22, 280–282. doi:10.1001/archneur.1970.00480210090013.

Wilke, M., and Lidzba, K. (2007). LI-tool: a new toolbox to assess lateralization in functional MR-data. *J. Neurosci. Methods* 163, 128–36. doi:10.1016/j.jneumeth.2007.01.026.

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–2. doi:10.1038/nn1263.

Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., and Recanzone, G. H. (2006). Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *J. Neurophysiol.* 96, 3323–3337. doi:10.1152/jn.00392.2006.

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex.* 48, 58–81. doi:10.1016/j.cortex.2011.03.004.

Yovel, G., and Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends Cogn. Sci.* 17, 263–71. doi:10.1016/j.tics.2013.04.004.

Zatorre, R. J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is "where" in the human auditory cortex? *Nat. Neurosci.* 5, 905–909. doi:10.1038/nn904.

6 LIST OF AUTHOR CONTRIBUTIONS

6.1 Kuśmierek P., Ortiz-Rios M., Rauschecker J.P. (2012). Sound-identity processing in early areas of the auditory ventral stream in the macaque. J. Neurophysiol. 107, 1123-41.

Paweł Kuśmierek co-designed the study and conducted recordings and most analyses, and wrote the manuscript. Michael Ortiz-Rios conducted analyses. Josef P. Rauschecker co-designed the study and participated in writing the manuscript.

6.2 Ortiz-Rios M., Kuśmierek P., DeWitt I., Archakov D., Azevedo F. A. C., Sams M., Jääskeläinen I. P., Keliris G. A., and Rauschecker J. P. (2015). Functional MRI of the vocalization-processing network in the macaque brain. *Front. Neurosci.* 9, 113.

Michael Ortiz-Rios co-designed the study, trained the animals, programmed stimulus presentation, acquired part of the data, conducted most analyses, and co-wrote the manuscript. Paweł Kuśmierek programmed the behavioural task and participated in writing the manuscript. Denis Archakov trained the animals and acquired part of the data. Iain DeWitt generated the scrambled stimuli and acquired part of the data. FACA contributed with data analyses and

participated in writing the manuscript. GAK, interpreted data and participated in writing the manuscript. MS, IPJ, and JPR co-designed the study and participated in writing the manuscript.

6.3 Ortiz-Rios M., Azevedo F.A.C., Kuśmierek P., Balla D.Z., Munk M.H., Keliris G.A., Rauschecker J.P., and Logothetis N.K. Widespread and opponent fMRI signals represent sound location in macaque auditory cortex. (f.e. *submitted manuscript*)

MOR designed the study, programmed stimulus delivery and event design, conducted all recordings and all data analyses and wrote the manuscript. FACA and GAK helped with event design, data acquisition for awake-fMRI experiments and revision of the manuscript. MHM helped with anesthetized experiments and revision of the manuscript. DB optimized the data acquisition for awake-fMRI experiments. PK provided some methodologies and participated in writing the manuscript. NKL and JPR co-designed this study and participated in writing the manuscript.

7 APPENDICES

7.1 Sound-identity processing in early areas of the auditory ventral stream in the macaque

Sound-identity processing in early areas of the auditory ventral stream in the macaque

Paweł Kuśmierek[1], Michael Ortiz[1,2], Josef P. Rauschecker[1]

Author contributions: PK co-designed the study, conducted recordings and most analyses, and wrote the manuscript. MO conducted analyses. JPR co-designed the study and participated in writing the manuscript.

Affiliation: [1]Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA, [2]Max Planck Intrenational School, Tübingen, Germany

Running head: Sound identity in early ventral stream

Corresponding author:

Paweł Kuśmierek

Georgetown University

NRB WP23

3970 Reservoir Rd NW

Washington, DC 20057

email: pk83@georgetown.edu

phone: +1 202 687 8851, fax +1 202 687 0617

Abstract

Auditory cortical processing is thought to be accomplished along two processing streams. The existence of a posterior/dorsal stream dealing, among others, with the processing of spatial aspects of sound has been corroborated by numerous studies in several species. An anterior/ventral stream for the processing of nonspatial sound qualities, including the identification of sounds such as species-specific vocalizations, has also received much support. Originally discovered in anterolateral belt cortex, most recent work on the anterior/ventral pathway has been performed on far anterior superior temporal (ST) areas and on ventrolateral prefrontal cortex (vlPFC). Regions of the anterior/ventral stream near its origin in early auditory areas have been less explored. In the present study, we examined three early auditory regions with different antero-posterior locations (caudal, middle, and rostral) in awake rhesus macaques. We analyzed how well classification based on sound-evoked activity patterns of neuronal populations replicates the original stimulus categories. Of the three regions, the rostral region (rR), which included core area R and medial belt area RM, yielded the greatest classification success. Starting from approximately 80 ms past stimulus onset, clustering based on the population response in rR became clearly more successful than clustering based on responses from any other region. Our study demonstrates that specialization for sound-identity processing can be found very early in the auditory ventral stream. Furthermore, the fact that this processing develops over time can shed light on underlying mechanisms. Finally, we show that population analysis is a more sensitive method for revealing functional specialization than conventional types of analysis.

Keywords

auditory cortex, rostral area, auditory object, cluster analysis

Introduction

The concept of two streams in auditory cortical processing, similar to the ones in visual cortex (e.g., Mishkin et al. 1983) has been proposed more than a decade ago (Rauschecker et al. 1997; Rauschecker 1998; Rauschecker and Tian, 2000). It was supported by contrasting patterns of anatomical connections in the macaque from anterior/ventral and posterior/dorsal belt regions of auditory cortex to segregated domains of lateral prefrontal cortex (Romanski et al. 1999) and by different physiological properties of these regions. In particular, the anterior lateral belt (area AL) exhibited enhanced selectivity for the identity of sounds (monkey vocalizations); the caudal lateral belt (CL) was particularly selective to sound location; the middle lateral belt area (ML) fell in-between and showed no specific preference for either of these properties (Tian et al. 2001).

Though refinements of the dual-pathway hypothesis have been proposed (e.g., Rauschecker and Scott 2009; Rauschecker 2011), its core concept has persisted and has been supported by numerous studies. Perhaps the most massive evidence for dual auditory processing streams comes from functional imaging in humans (e.g., Binder et al. 2000; Maeder et al. 2001; Alain et al. 2001; Arnott et al. 2004; Leaver and Rauschecker 2010; Chevillet et al. 2011). In non-human primates, Recanzone and colleagues have presented extensive data confirming the enhanced selectivity of neurons in caudal regions of macaque auditory cortex (especially area CL) to sound location (Recanzone et al. 2000a,b; Woods et al. 2006; Recanzone et al. 2010), thus providing further support for the existence of a caudal “where”-stream.

The concept of an anterior “what”-stream has been tested in various monkey studies as well. After the initial demonstration of increased selectivity to monkey calls in area AL by Tian et al. (2001), most of the evidence for stimulus-identity coding in the anterior pathway came from studies of far anterior regions of the superior temporal (ST) cortex and of ventrolateral prefrontal cortex (VLPFC) (Poremba et al. 2004; Romanski et al. 2005; Petkov et al. 2008; Cohen et al. 2009; Kikuchi et al. 2010). Fewer data are available from the earliest stages of the anterior stream, that is, from areas adjacent to primary auditory cortex (A1). Recanzone (2008) found no difference in monkey-call selectivity between the rostral core area (R) and A1, suggesting that feature selectivity may not emerge until the level of the belt. The same study failed to find specificity in belt area ML, consistent with Tian et al. (2001). However, Recanzone’s (2008) recordings did not extend to belt area AL, where Tian et al. (2001) found the earliest signs of selectivity to vocalizations, nor to other anterior belt regions, a fact noted both by the author himself (Recanzone 2008), as well as by subsequent commentators (Bizley and Walker 2009).

Indirect confirmation for the existence of an anterior “what”-stream in primates came from measures of temporal integration, which increase in areas anterior to A1 (Bendor and Wang 2008; Kuśmierek and Rauschecker 2009; Scott et al. 2011). Temporal integration of acoustic features is necessary for auditory stimulus identification (Rauschecker and Tian 2000). Combined with increasing spectral integration (Rauschecker et al. 1995) it provides the equivalent to the gradual increase in receptive field complexity along the ventral visual stream (e.g., Desimone et al. 1984; Connor et al. 2007).

As the anterior stream processes auditory structure in a hierarchical fashion (Chevillet et al. 2011), selectivity for sound stimuli is expected to develop gradually along the stream, which would make it more difficult to find indications of selectivity closer to A1. Still, the beginnings of selectivity constituting initial primitives of sound identification should be present even at relatively early processing stages. Coding may take place across larger populations of neurons, as has been found in posterior regions: while single neurons carry little information about sound location in auditory areas adjacent to A1 on the posterior side, analyzing neural *populations* allowed to extract more precise spatial information (Miller and Recanzone 2009). Population analysis has also been successful in studies of monkey-call selectivity in vlPFC (Romanski et al. 2005). Thus, we decided to investigate how stimulus identity is represented by population activity patterns in early areas of the auditory ventral stream.

Furthermore, many previous studies focused on selectivity of single neurons for stimuli *within* a stimulus class, e.g. monkey calls (e.g., Tian et al. 2001; Romanski et al. 2005; Recanzone 2008; Russ et al. 2008; Kuśmierek and Rauschecker 2009). In the present study, we examined how responses of neural populations can be used to discriminate *between* stimulus classes. Our expectation was that between-class categorization might require coarser discriminations than within-class categorization and would therefore be detectable more readily in early cortical areas.

Materials and Methods

This paper presents population analysis of data collected in two single/multiunit recording experiments from four male rhesus monkeys (Experiment 1: monkeys S and L; Experiment 2: monkeys B and N). Data from monkeys S and L have also been used in another study, but for a different purpose and analyzed in a different way (Kuśmierek and Rauschecker 2009). Experiments 1 and 2 were conceptually very similar, but differed in some details. For the sake of clarity, detailed information on methodological differences between Experiments 1 and 2 was moved to the last section of Materials and Methods. In earlier sections, we describe only those differences that are crucial to data interpretation.

Animals

Each animal was implanted with a plastic recording chamber (Crist Instruments, Hagerstown, MD, USA) over left auditory cortical areas. Implant locations were confirmed by 3T MRI with 1 mm^3 voxel size. Monkeys were water-restricted to provide adequate drive in a fluid-rewarded task. All experiments were conducted in accordance with NIH guidelines and approved by the Georgetown University Animal Care and Use Committee.

Stimuli and task

Monkeys were seated in a monkey chair (Crist Instruments) in a sound-attenuated chamber (IAC, Bronx, NY, USA) measuring 2.6 m x 2.6 m x 2.0 m (W x L x H).

The stimuli of interest included artificial sounds: pure tones (PT), 1/3-octave and 1-octave band-pass noise bursts (1/3-oct BPN, 1-oct BPN); and two classes of natural sounds: rhesus monkey calls (MC) and environmental sounds (ES). Duration of tones and noise bursts was 500 ms (Experiment 1) or 300 ms (Experiment 2). Due to limitation of the presentation system in Experiment 2, the range of PT and BPN frequencies was reduced compared to Experiment 1, and the number of MC and ES stimuli was 7 in each class instead of 10. (Throughout the paper, “BPN frequency” denotes BPN center frequency.) All MC used in Experiment 2 where previously used in Experiment 1, whereas only 5 out of 7 ES used in Experiment 2 were previously used in Experiment 1.

Data from each Experiment were analyzed separately. Then, in addition, responses to stimuli that were common to Experiment 1 and 2 were pooled and analyzed together. The respective analyses/results are labeled as “Experiment 1”, “Experiment 2”, and “Combined”. The stimulus sets used in the analyses were as follows:

Experiment 1: nine PT, nine 1/3-oct BPN, nine 1-oct BPN (frequencies of PT and BPN: 0.125-32 kHz), ten MC and ten ES.

Experiment 2: seven PT, seven 1/3-oct BPN, seven 1-oct BPN (frequencies of PT and BPN: 0.25-16 kHz), seven MC and seven ES.

Combined: seven PT, seven 1/3-oct BPN, seven 1-oct BPN (frequencies of PT and BPN: 0.25-16 kHz), seven MC and five ES.

Stimulus duration ranged from 151 to 2614 ms. Thus, to ensure that only stimulus-driven activity contributed to the results, all analyses covered the first 160 ms of neural responses (the approximate duration of the shortest stimulus).

Similarly, when acoustical properties of stimuli were examined, only the first 160 ms of each stimulus period were used. The stimulus presentation level was set to ~50 dB and ~30 dB above the macaque hearing threshold (Jackson et al. 1999) for Experiment 1 and 2, respectively.

The behavioral task was go/no-go auditory discrimination: a bar-release response to an infrequent (~10-15%) auditory target was rewarded by a small amount of juice, water, or a balanced electrolyte drink (Prang, BioServ, Frenchtown, NJ, USA). The purpose of the task was to keep the animals at an approximately constant level of attention. A block of all stimuli (including several repetitions of the behavioral target) was presented 10-13 times (Experiment 1) or 60 times (Experiment 2), in random order within each block presentation. Each trial started with a 300-400-ms pretrial period, during which the animal had to keep its hand on the bar.

Neural recordings

Single and multi-unit recordings were obtained by advancing 1-2 epoxylite- or glass-insulated 1-3 MΩ tungsten electrodes (FHC, Bowdoin, ME, USA or NAN Instruments, Nazareth Illit, Israel) into the auditory cortex by means of a micropositioner (Model 650, David Kopf, Tujunga, CA, USA, or FlexMT/EPS, Alpha Omega, Nazareth Illit, Israel). A stainless steel guide tube was used to puncture the dura. A 1 mm x 1 mm spacing grid (Crist Instruments) provided a repeatable spatial reference for electrode location. The electrode signal was amplified and filtered (Model 1800, A-M Systems, Sequim, WA, USA and PC1, TDT; or MCP Plus, Alpha Omega). In Experiment 1, neural activity was isolated with a window discriminator (SD1, TDT), and spike time stamps were recorded

with a custom-made program (“Fiordiligi”, Kuśmierek and Rauschecker 2009), which also presented stimuli and controlled the behavioral task. In Experiment 2, Power1401mk2 (CED) interface and Spike2 program (v. 6 or 7, CED) running custom-made scripts were used to record and isolate neural activity, present stimuli, and control behavior; in most cases, units were isolated post-hoc using principal component analysis. As the recording electrode was lowered, the surface of auditory cortex was determined from recording depth (in reference to MRI images) and from the presence of a “silent gap” corresponding to the lateral sulcus. In order to drive the neurons (whether identified by baseline activity, or silent when not stimulated), we used the same stimuli as those in the formal testing and/or natural sounds produced ad hoc (knocking, hissing, key jingling, clapping etc.). Only auditory-responsive units were tested further.

Cortical regions

Since the main purpose of the study was to investigate antero-posterior differences of acoustic stimulus representation in core and medial belt, recordings from core areas R and A1 were pooled with recordings from neighboring medial belt areas RM and MM, respectively. Total number of elements being an important factor in this type of study (Miller and Recanzone 2009), pooling enabled us to increase the size of analyzed neuronal populations. We have shown previously that response properties of MM and RM neurons are quite similar to responses of cells in A1 and R, respectively (Kuśmierek and Rauschecker 2009).

After pooling, three cortical regions were distinguished: a rostral region (rR) comprising the rostral core area R and rostro-medial area RM, a middle region (rM) consisting of the primary core area A1 and middle medial area MM, and a

caudal region (rC) which consisted of the caudo-medial area CM (

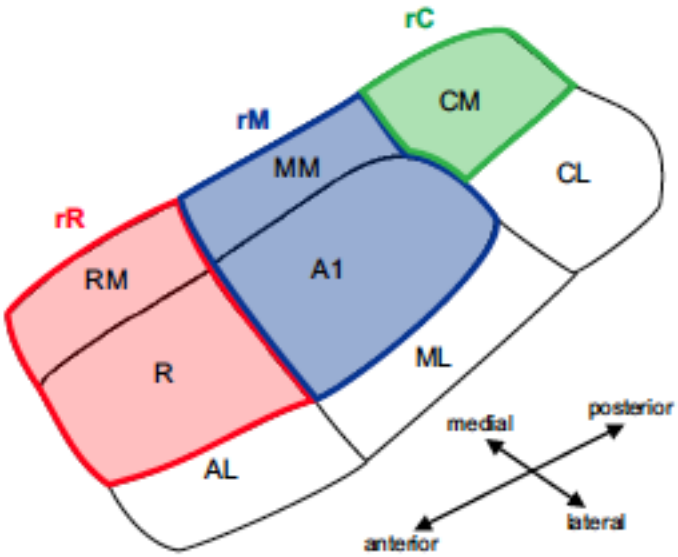


Figure **1**). Region rR data originated from Experiment 1 only, rC data from Experiment 2 only, and rM data from both Experiments 1 and 2. To account for the fact that Experiments 1 and 2 differed in several respects, rM datasets coming from Experiments 1 and 2 were analyzed separately and labeled as rM1 and rM2, respectively. This enabled us to distinguish differences between regions from differences between experiments.

Antero-posterior parcelation was based on best-frequency reversals, which were clear in both Experiment 1 (Kuśmierek and Rauschecker 2009) and 2. In Experiment 2, medio-lateral delimitation of area CM from CL was also required. In monkey N, CL was separated from CM based on longer latencies and higher selectivity to azimuth (Woods et al. 2006; Kuśmierek and Rauschecker in preparation). The resulting boundary ran approximately along the midline of the superior temporal plane, consistent with the placement of the CM/CL boundary

in the anatomical literature (Smiley et al. 2007). In monkey B, caudal recordings were performed only in the medial half of the superior temporal plane, and neither long latencies nor increased azimuth selectivity were found laterally in the recorded area. Thus, all recordings caudal to the best-frequency reversal at the caudal end of A1 in monkey B were considered to be from CM. The total number of neurons per cortical region was 159, 262, 95, and 69 for rR, rM1, rM2, and rC, respectively.

Neural data analysis

Spike time-stamps were exported to Matlab (Mathworks, Natick, MA, USA) and processed with custom-made Matlab scripts. Spike times were corrected for sound travel time from the loudspeaker to monkey ear, and, in Experiment 1, for spike discriminator delay.

The analysis of neural data was performed in several variants. First, as mentioned before, it was done separately for Experiment 1 or 2, or for both experiments Combined. Second, we analyzed all cortical regions together, or each of them separately to detect any between-region differences. Third, we performed computations on data from a single 160-ms temporal window starting at the stimulus onset, compared to a 160-ms window immediately preceding sound onset (pretrial), or we analyzed eight consecutive 20-ms windows within the stimulus starting at the stimulus onset and compared them to a 20-ms pretrial window. Fourth, analyses were performed for all stimuli (PT, BPN, MC, and ES), or for natural stimuli (MC and ES) only.

The first stage of analysis followed the method of Kiani et al. (2007). Specifically, for each unit and each stimulus, average firing rate within 20-ms or

160-ms temporal windows was calculated across all stimulus presentations. For each unit, values of firing rate in response to the entire stimulus set were treated as a vector that was normalized by subtracting the mean and dividing by the vector's Euclidean length. Correlation coefficients (*r*) between normalized population responses were calculated and visualized as similarity matrices. For natural stimuli MC and ES, representation of stimulus classes in the similarity matrices was quantified by comparing within-class correlation coefficients to between-class correlation coefficients (both between the given class and the other natural stimulus class, and between the given class and all artificial stimuli, that is, PT and BPN) with a t-test.

Next, the normalized responses were arranged into a units x stimuli matrix, and hierarchical clustering of stimuli based on a measure of neural distance (1-*r*) was calculated and visualized with dendrograms.

To quantify and compare the representation of stimulus classes in population responses, we assigned stimuli to *k* a-priori categories of stimuli. The choice of actual *k* values and of categories is described in the Results section. Next, we clustered the normalized firing rates into *k* clusters with the k-means procedure. The main measure obtained in this analysis was classification success, that is, the proportion of stimuli that were clustered into their a-priori classes: proportion of correct classifications (PCC).

Different numbers of neural units were available for different cortical regions. This could skew the results of clustering because the quality of stimulus representation by a neural population may depend on the population size (Miller and Recanzone 2009). To avoid this potential confound, we performed k-means

clustering on a subset of neurons from each region. The size of the subset was set to the number of units in the least numerous region of the analysis. Clustering was repeated 50 times with subsets drawn randomly from each region every time. The mean PCC (or mode PCC, see below) from these 50 repeats was taken as the representative value for a region.

The statistical significance of PCC values was quantified in two ways. First, to evaluate if quality of clustering was higher than the baseline, the mode PCC obtained in each temporal window during the stimulus was compared to the mode PCC derived from the pretrial with a one-way Fisher exact probability test. When cortical regions were analyzed separately, their separate mode PCCs were compared to one pretrial mode PCC from all regions combined.

Second, when regions were analyzed separately, we assessed if quality of clustering in a particular region deviated significantly from the “reference range”. The reference range was estimated by randomly re-assigning the neurons to regions and repeating the k-means analysis in an identical way as described above to obtain reference mean PCCs. The number of re-assignments was such that the number of reference PCCs was 400. For example, when 4 regions were analyzed, the analysis was run 100 times, each run producing four reference PCCs. The mean PCC of a region was considered significantly ($p \leq 0.05$) above or below the reference range if it was outside of the middle 95% of reference PCCs.

Again, this procedure could possibly be confounded by the unequal number of neurons per region. The reference range obtained by drawing neurons from the entire pool in a random fashion would be skewed towards values characteristic for the rM1 region (which contributed 45% neurons to the pool), and less

representative for the rC region (12%). Similarly, the reference range would be skewed towards values obtained from Experiment 1, which provided 72% of the analyzed units. Thus, the randomized pool of neurons was created by drawing the same number of neurons from each region. The number was equal to the mean number of neurons across the regions. Consequently, in each randomization a random subset of more numerous regions was used, whereas some neurons were re-drawn from less numerous regions. Still, the subset size used for k-means clustering of randomized data was the same as for the original data, that is, equal to the number of neurons in the least numerous region.

Analysis of sound stimuli

The purpose of the analysis of sound stimuli was to determine if classification of the stimuli based on responses of neural populations in the auditory cortex can be matched by classification based on acoustical properties of the stimuli. For this analysis, only the first 160-ms segment of each stimulus was used, to match the information used for the neural analysis. To avoid cut-off transient, 5-ms linear fade-out was applied to the segment. All stimuli used in any of the experiments were analyzed: 10 PT, 20 BPN, 10 MC and 12 ES. Three analysis approaches were used.

First, a log-frequency scale spectrogram was created by splitting each stimulus into 57 ⅛-octave frequency bands (center frequencies: 64 Hz - 41.3 kHz) with a 16384-point FIR filter and measuring RMS value (expressed in dB) in 33 consecutive non-overlapping 5-ms bins. Pearson correlation coefficient between spectrograms was used to measure the similarity between pairs of stimuli,

which was then visualized in the form of a similarity matrix. Between/within-class correlations of natural sounds were quantified, same as for the neural data. The similarity matrix was converted to a dissimilarity measure by subtracting from 1, and the resulting dissimilarity matrix was subjected to multidimensional scaling (MDS) with the number of dimensions set to three. As a result, each sound was assigned three parameter values derived from spectrogram dissimilarity.

Second, modulation spectrum analysis (Singh and Theunissen 2003) was performed for each sound with the STRFpak Matlab toolbox. We obtained a spectrogram of each sound by decomposing it into frequency bands using a bank of Gaussian filters (244 bands, filter width = 125 Hz). The filters were evenly spaced on the frequency axis (64-48000 Hz) and separated from each other by one standard deviation. The decomposition resulted in a set of narrow-band signals which were then cross-correlated with each other, including itself, to yield an autocorrelation matrix. This auto-correlation matrix was calculated for time delays of +/-150 ms. Two-dimensional Fourier transformation of this auto-correlation matrix was calculated to obtain the modulation spectrum (MS) of each sound. Just as for the spectrogram analysis, the Pearson correlation coefficient between MS was used to measure the similarity between all pairs of stimuli, displayed as similarity matrix, and quantified for natural stimuli. Then, MDS was used to calculate values of three parameters derived from MS dissimilarity.

Third, three direct acoustic measures were calculated for each stimulus using the program Praat (v. 5.1.04, Boersma and Weenik, University of Amsterdam, http://www.praat.org): center of gravity of spectrum (in logarithmic scale), mean

harmonicity (Boersma 1993), and standard deviation of intensity, with the purpose of estimating the frequency region with dominant energy, the ratio of periodic to aperiodic components, and the degree of amplitude modulation, respectively.

The last step of the analysis of sounds was an attempt to classify the stimuli based on calculated acoustic parameters in a similar way as for the neural responses. To this end, each of nine acoustic parameters (three from direct measurements, three from MDS based on spectrogram dissimilarity, three from MDS based on modulation spectrum dissimilarity) was converted to *Z*-scores and k-means clustering (*k*=4 and *k*=5 for all sounds, *k*=2 for natural sounds only) was performed:

- separately on each of three direct parameters (spectrum center of gravity, mean harmonicity, standard deviation of intensity),
- on all three parameters derived from spectrum dissimilarity (combined, i.e., used as three variables in a single clustering procedure), and
- on all three parameters derived from MS dissimilarity.

Furthermore, the clustering procedure was performed:

- on all three direct parameters combined,
- on all six parameters derived from spectrum and MS dissimilarity combined, and
- on all nine parameters combined (three direct measures, three parameters derived from spectrum dissimilarity, and three parameters derived from MS dissimilarity).

As for the neural data, classification quality was quantified as proportion of correct classifications (PCC), and the PCC values were compared to those obtained from the neural data in the 80-160 ms range.

Differences between Experiment 1 and 2 in more detail

In Experiment 1, a 19-mm diameter round recording chamber was used, while in Experiment 2 the chamber was oval and measured 19 x 38 mm.

In Experiment 1, auditory stimuli were played with an Audiophile 192 (M-Audio, Irwindale, CA, USA) sound card, PA4 attenuator (TDT, Alachua, FL, USA), SE 120 amplifier (Hafler, Tempe, AZ, USA) and Reveal-6 two-way studio “monitor” loudspeaker (Tannoy, Coatbridge, UK), located 1.7 m in front of the monkey. The stimuli were played at 96 kHz sampling frequency, 16-bit resolution.

In Experiment 2, auditory stimuli were played with a Power1401MkII laboratory interface (CED, Cambridge, UK), PA4 attenuator, SE 120 amplifier, and 400-312-10 3.5” one-way car speakers (CTS, Elkhart, IN, USA). Because spatial tuning data were also collected in Experiment 2 (not reported here), the speakers were arranged in a vertical arc-shaped array (Crist Instruments) which was rotated automatically around the monkey chair using a Unidex 100 controller (Aerotech, Pittsburgh, PA, USA) and 300SMB3-HM (Aerotech) stepper motor under control of the Spike2 v.6/7 software program (CED). In the first stage of the experiment, a subset of stimuli was presented at azimuths of 0, 45, 90, 135, 180, 215, 270 and 315 degrees and at elevation 0 to estimate spatial tuning in the horizontal plane. In the second stage, the full set of stimuli was presented at best azimuth and at five elevations of -60, -30, 0, 30 and 60 degrees. Data from Experiment 2 analyzed in this paper come from the second

stage only. The distance between the loudspeakers and the monkey's head was approximately 0.95 m. The stimuli were played at 48 kHz sampling frequency, 16-bit resolution.

The difference in loudspeaker size and quality as well as in sampling frequency (and consequently, bandwidth) resulted in a possibly lower playback quality in Experiment 2 compared to Experiment 1.

In addition to stimuli of interest, in Experiment 1, bursts of white noise (equal power per Hz) and pink noise (equal power per octave) were presented, and a short four-note melody was used as behavioral target. In Experiment 2, a white-noise burst was used as behavioral target. Although white-noise bursts were used in both experiments, they were excluded from analyses because of differing behavioral contingencies.

In Experiment 1, each stimulus block consisted of 49 stimuli and 8 repetitions of the target, whereas in Experiment 2, a block consisted of 35 stimuli + 4 repetitions of the target. The high number of block repeats in Experiment 2 (60) resulted from each block being played 12 times at each of 5 elevations.

In Experiment 1 the location of the speaker and monkey chair was adjusted to minimize the influence of the room on low-frequency response, this could not be done for Experiment 2 due to constraints of the spatial tuning study.

Results

As expected, similarity matrices revealed no correlations in activity of neuronal populations during pretrial, with all correlation coefficients (*r*) close to zero (Figure 2, upper row,

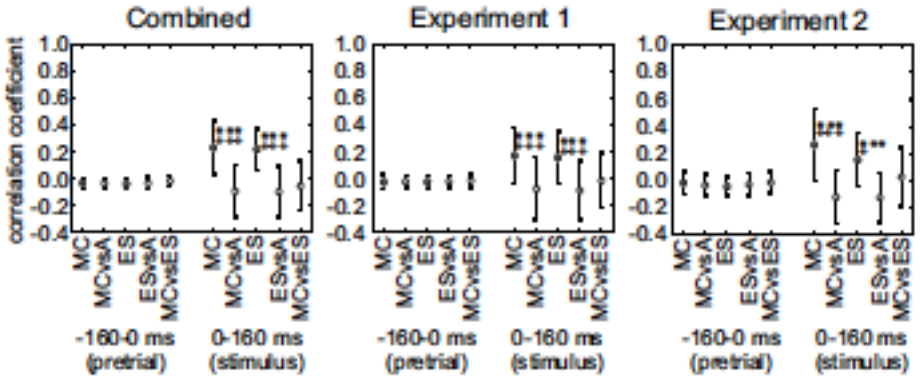


Figure **3**A-C). However, a clear correlation structure emerged during the first 160 ms of stimulus presentation (Figure 2, lower row). Oblique lines of high *r* values reflect similarity of population responses to PT, 1/3-oct BPN, and 1-oct BPN of the same frequency, that is, frequency tuning. Dark blue colors visible between the lines demonstrate that population responses to distant frequencies were anti-correlated, whether within one bandwidth or across bandwidths. Apparently, stimulus frequency was the main factor determining the population response to artificial sounds, whereas bandwidth played less of a role.

Responses to natural sounds were correlated within each natural stimulus class. This was particularly well visible for MC in Experiment 2 and in combined data, with the first 6 out of 7 MC all evoking clearly similar population responses. Response correlation within the ES class was less pronounced, but was still clearly noticeable.

These observations were confirmed quantitatively;

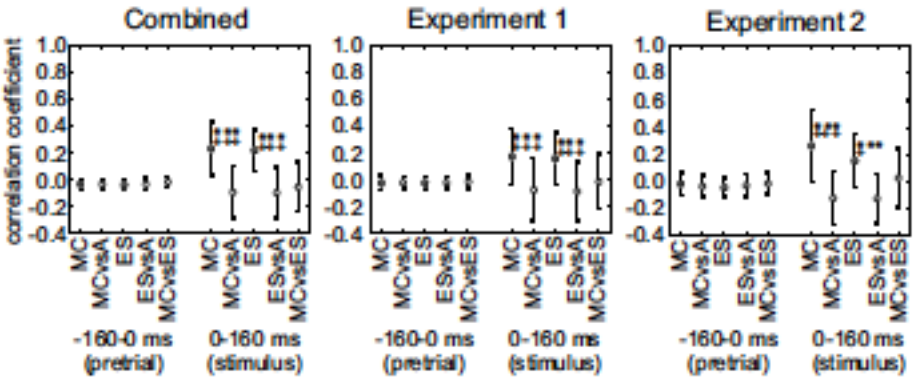


Figure **3** shows that within-class correlation coefficients of responses to each of natural stimulus classes were clearly and significantly higher than between-class *r* values. This was true both when response similarity within MC or ES class was compared to similarity of responses between this class and artificial stimuli, and when compared to similarity of responses between MC and ES classes.

Further confirmation of the findings based on similarity matrices was provided by hierarchical clustering (

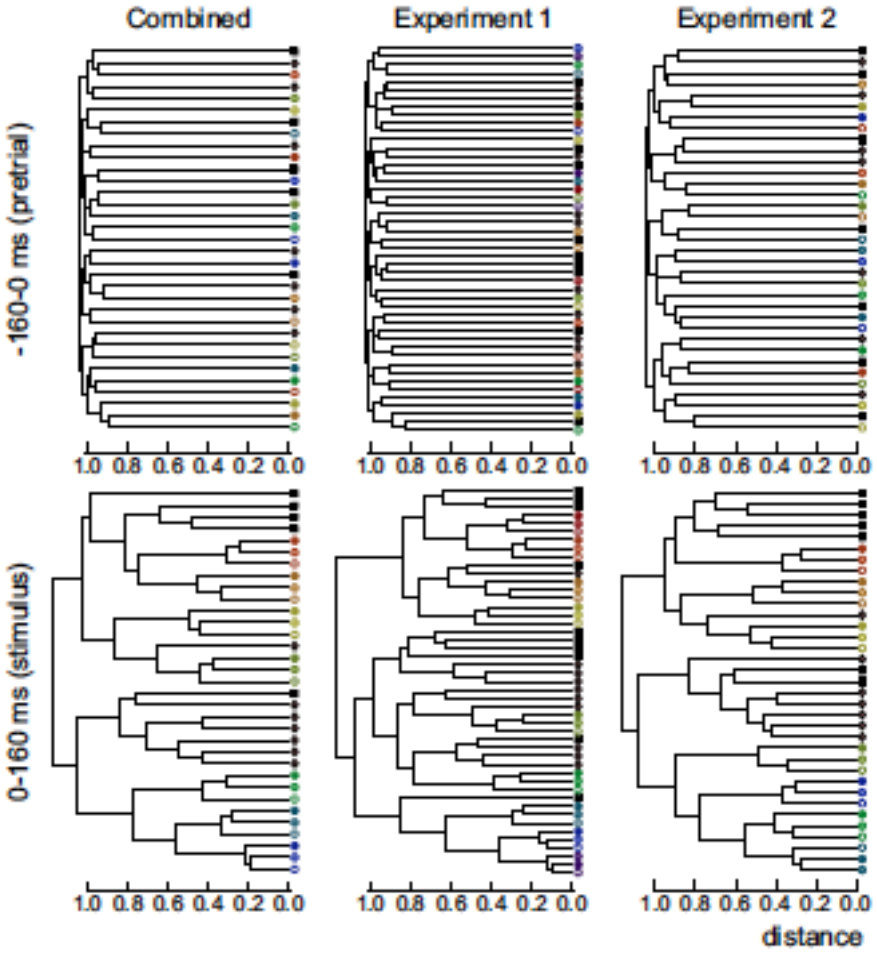


Figure **4**). Again, during the pretrial period, the stimulus structure was not reflected at all in the population responses (upper row), which is by itself trivial, but provides a control condition for the computational procedures.

During stimulus presentation, however, clustering of population responses replicated many features of the stimulus structure (lower row). Responses to artificial stimuli (PT and BPN) always clustered with PT and BPN of the same frequency, and usually fell close to PT and BPN of neighboring frequencies, indicating frequency tuning in the auditory cortex. Bandwidth was not a robust organizing principle of the population responses, as shown by the fact that responses to PT and BPN of the same frequency always clustered very closely together. However, the effect of bandwidth on the response was still detectable. Of all PT/BPN triplets of the same frequency, in only one case did the response to PT and 1-oct BPN (i.e., two *outlying* bandwidths) cluster together first, while

1/3-oct BPN (the *intermediate* bandwidth) joined at a larger distance (

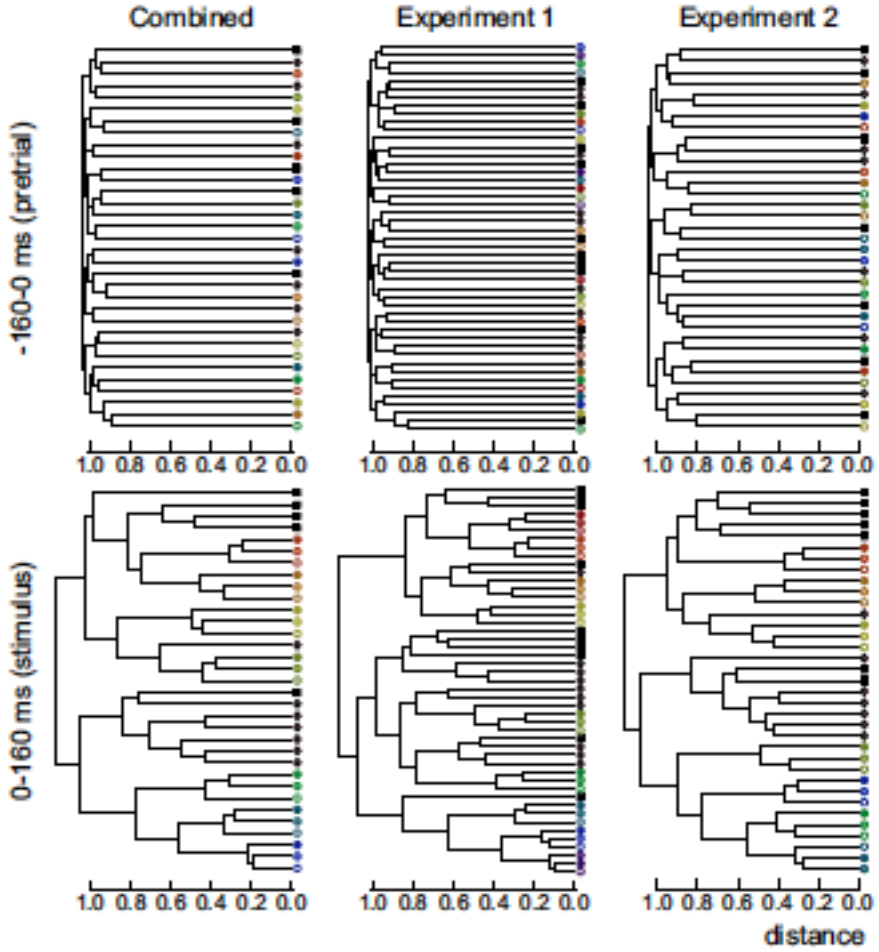


Figure **4**, Experiment 1, lower row, three lowermost branches). In all other cases, responses to *neighboring* bandwidths clustered together most closely (PT with 1/3-oct BPN or 1/3-oct BPN with 1-oct BPN), only later joined by the response to the remaining outlying bandwidth of the triplet. In quantitative terms, one out of 16 (6.25%) response triplets clustered inconsistently with a proportional effect of bandwidth on response clustering. The 95% confidence interval of this proportion is <0.01% to 30.31% and is below the chance level of 33.3%. (Only data from the Experiment 1 and Experiment 2 dendrograms were included in this calculation, because the Combined data are not independent from Experiment 1 and Experiment 2 data.)

In neither Experiment, nor in the Combined data, the responses to MC and ES (or to natural stimuli in general) were assigned to clusters that contained all responses to the class, and only responses to that class. However, in many cases they formed clear subgroups separated from responses to artificial sounds and, to some extent, to each other. In Experiment 1, possibly due to a larger number of ES and MC involved, the picture was not unequivocal; still, a certain degree of clustering of responses to ES and MC can be observed. In the data from Experiment 2, grouping is much clearer: all responses to MC except one were in a single cluster with two ES responses, with remaining responses to ES forming another big cluster. A similar picture emerged from the Combined data, with only one response to ES clustering with most MC responses. The remaining four ES responses grouped closely, also with responses to high-frequency artificial stimuli. This picture contrasts with one seen for pretrial responses, where all classes were mixed (

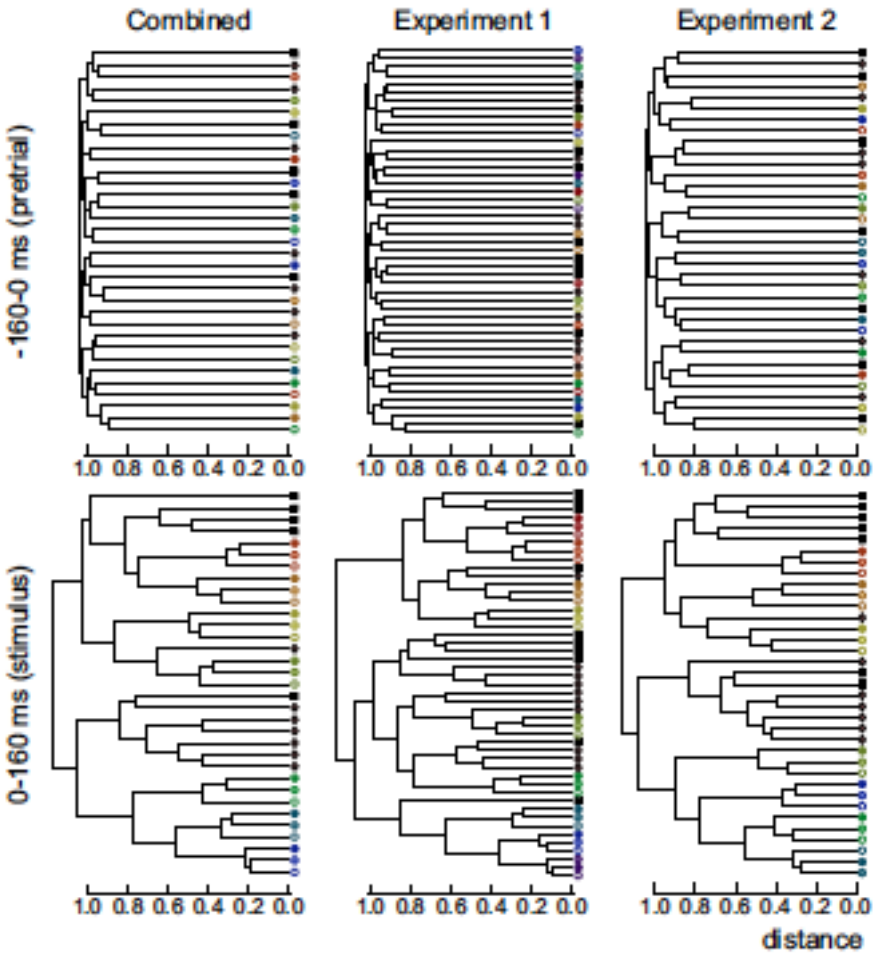


Figure **4**, upper row).

In summary, analyses of similarity matrices and of results of hierarchical clustering yielded several findings. Population responses to PT and BPN were determined mostly by the stimulus frequency. Bandwidth contributed less to the response. Responses to each of two classes of natural sounds formed distinguishable clusters and differed from responses to PT and BPN, and from responses to each other.

These findings were used to guide the choice of the cluster number (*k*) for the quantitative analysis of population responses to stimulus classes with k-means clustering. Not only does the procedure require deciding on an *a-priori* number of clusters (*k*) into which the data will be grouped; also, for the purpose of quantification of clustering quality with proportion of correct classifications (PCC), the stimuli must be pre-assigned to *k* classes. Stimulus assignments resulting from clustering are then compared to these original assignments.

For natural stimuli, the choice was simple, as they formed two obvious natural categories: MC and ES. Because analyses of similarity matrices and of hierarchical clustering showed that bandwidth of the artificial stimuli only weakly affected population responses in our data set, we decided to split PT and BPN into classes based on frequency only. The frequencies were evenly spaced, thus the decision had to be partly arbitrary. We split PT and BPN into two classes: low and high frequency, for a total k=4. Specifically, when analyzing data from Experiment 1 (9 PT/BPN frequencies), we placed four PT/BPN frequencies into the “low-frequency” class (125 Hz to 1 kHz) and five frequencies into the “high-frequency” class (2 kHz to 32 kHz), while the seven PT/BPN frequencies of Experiment 2/Combined were divided into ranges of 250

Hz to 1 kHz (3 members) and 2 kHz to 16 kHz (4 members). In addition, the analysis was replicated with PT/BPN frequencies split into low, middle and high range for a total *k*=5 (Experiment 1: 125-500 Hz, 1-4 kHz, and 8-32 kHz; Experiment 2/Combined: 250-500 Hz, 1-4 kHz, 8-16 kHz). Separate analyses on responses to natural sounds only (MC and ES) were performed with *k*=2.

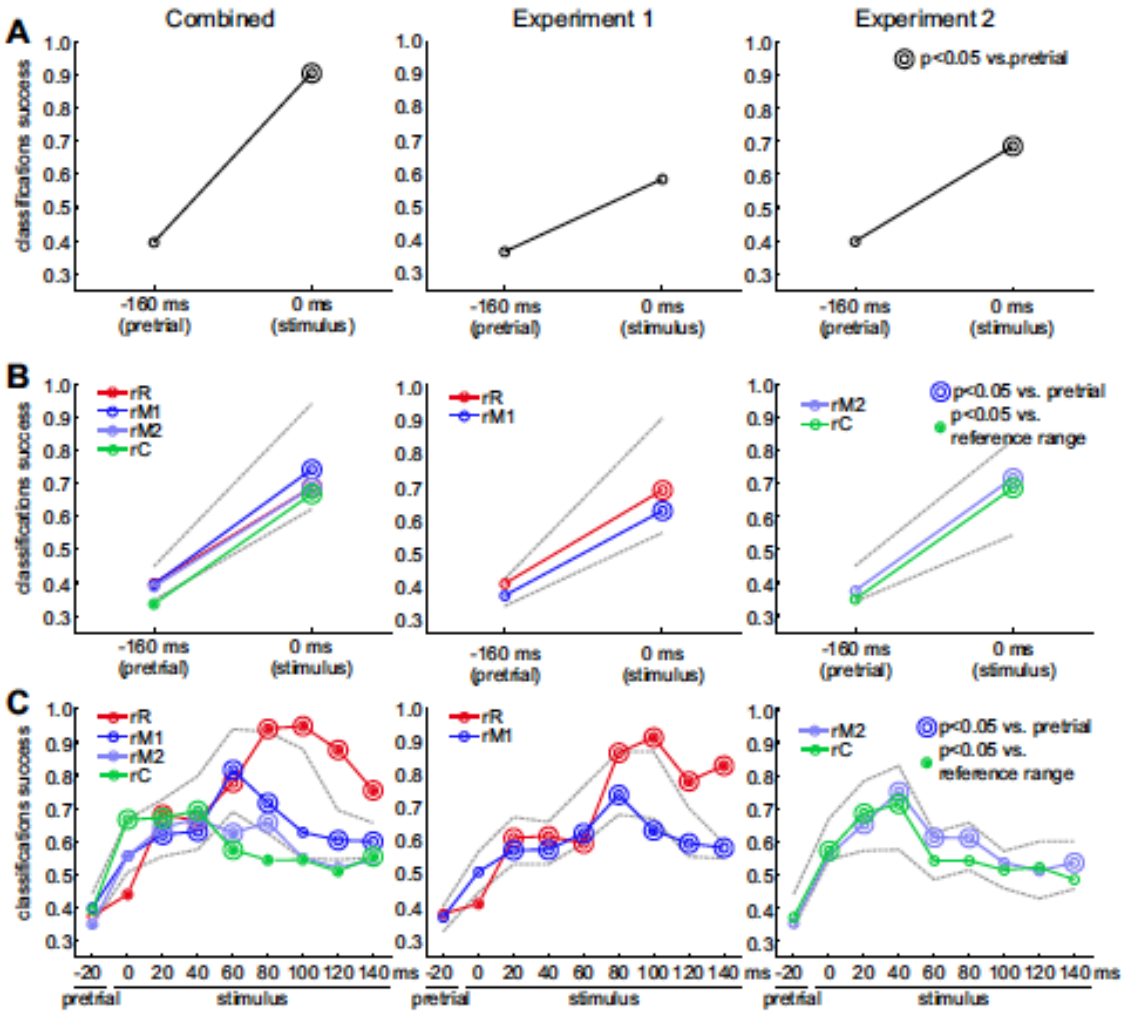


Figure 5A shows mean classification success (PCC) values from Experiment 1, Experiment 2, and Combined experiments, for *k*=4 (see also

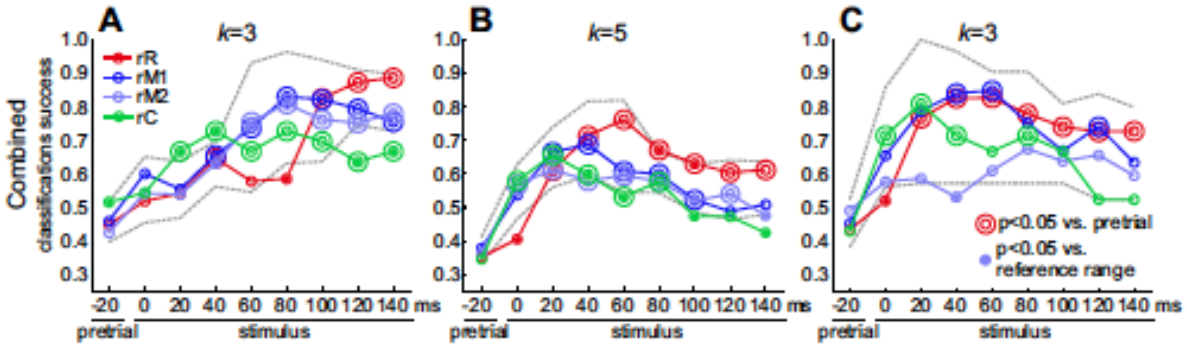


Figure **6**A for results obtained with *k*=5). Classification of responses recorded prior to stimulus presentation yielded mean PCC values from 0.35 to 0.40, considered chance value. Clustering of responses recorded during stimulus presentation replicated the original class structure with much higher accuracy of 0.57 to 0.90. Mode PCC for stimulus responses was significantly higher than mode PCC for pretrial responses for both *k* values in Combined data, as well as for *k*=5 in Experiment 1 and *k*=4 in Experiment 2 ($p<0.05$, one-way Fisher exact probability test). These results show that when a small number of discrete clusters are imposed on the data (as opposed to the basically continuous approach of hierarchical clustering) classification of population responses can still re-create original stimulus classes with reasonable accuracy.

The next step was to apply k-means clustering to population responses recorded from each cortical region separately.

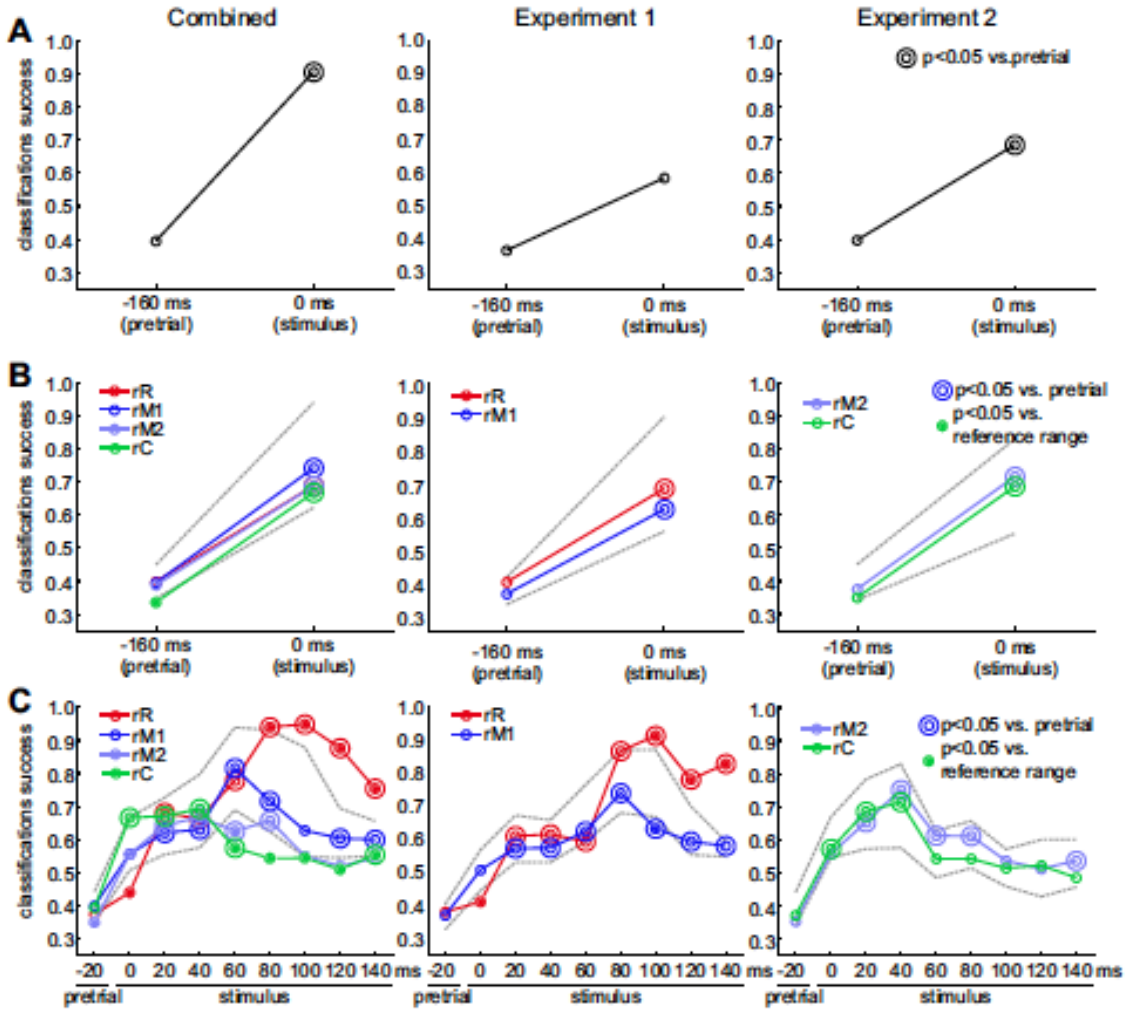


Figure 5B shows PCC values calculated for each region for pretrial (160 ms preceding stimulus onset) and the first 160 ms of stimulus (see also

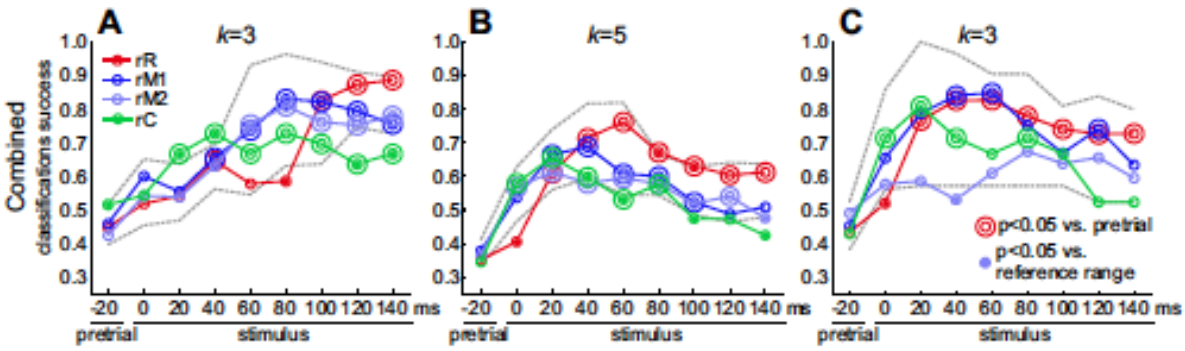


Figure **6**B for results obtained with *k*=5). In the Combined data at *k*=5, classification based on responses in regions rC and rM2 was poorer than in rM1 and rR, with rC and rM2 PCC being significantly below the reference range. This suggests that factors differing between Experiments 1 and 2 influenced the PCC values. It may be noteworthy that PCC obtained from the rC region was numerically lower in each case than PCC from rM2. Also, except for one measurement, classification success based on responses from rR numerically exceeded that from rM1. Still, these differences were small in the analysis that used 160-ms temporal windows, and none of them appeared to be significant.

A more compelling picture emerged from the next analysis, wherein we again calculated PCC for each region separately, but this time in eight 20-ms windows covering the first 160 ms of the stimulus period, with a single 20-ms window

preceding stimulus onset used to establish chance level (

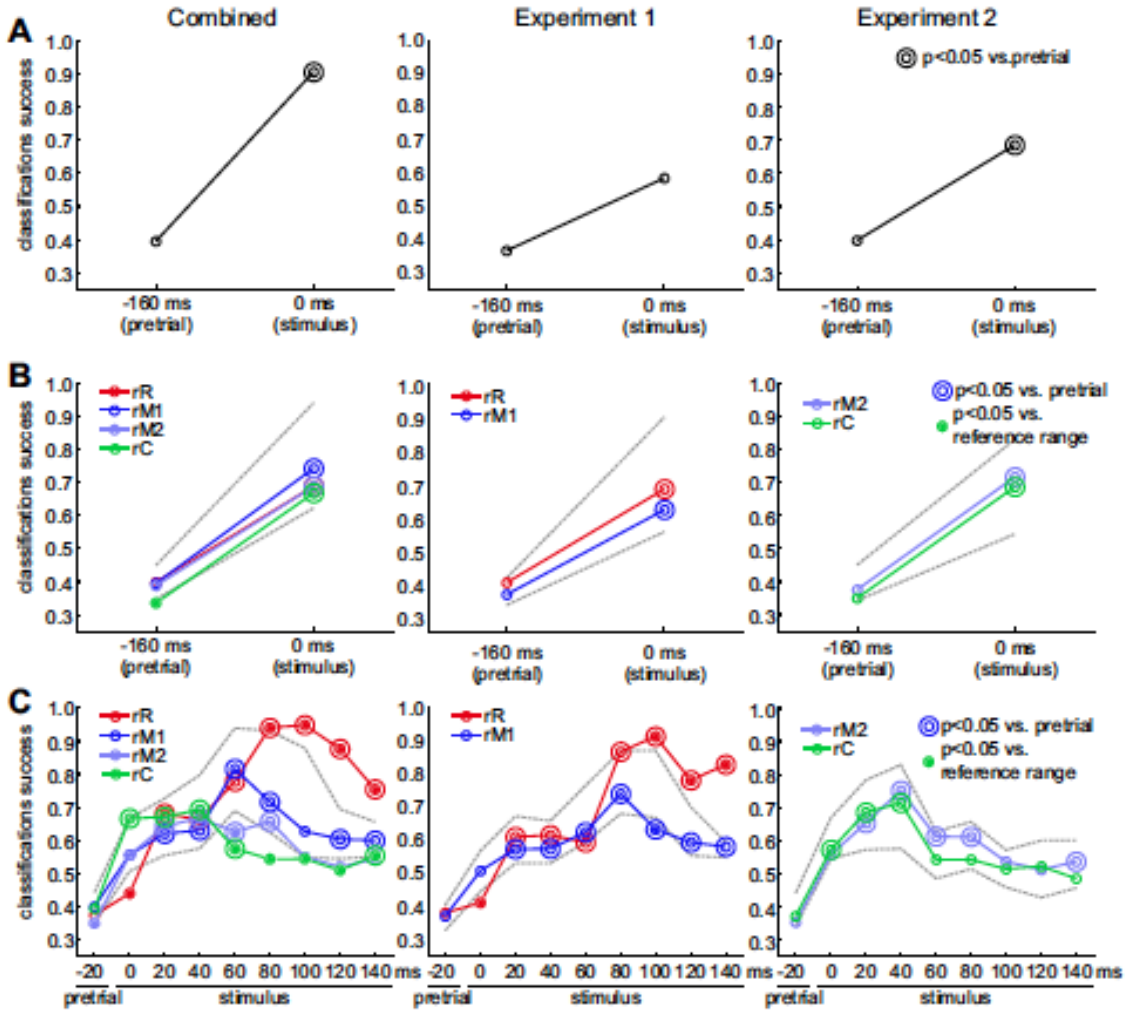


Figure 5C, see also

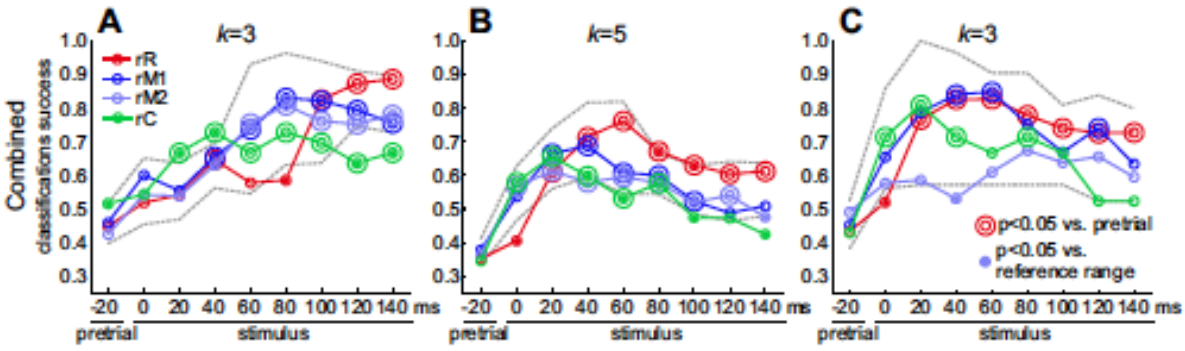


Figure **6**C for results obtained with *k*=5).

Classification based on responses from rC increased from chance level faster than classification based on other regions. This is particularly clear for *k*=4: in the first temporal window of the stimulus (0-20 ms), mode PCC from rC but not from other regions was significantly higher than pretrial PCC. For *k*=5, PCC from rM2 also reached significance in the first window, but it was still numerically lower than PCC from rC. Later in time, from approximately 60 ms past stimulus onset, classification based on responses from rC tended to be poorer than the one in rM2, but the difference was slight and there was much overlap.

The most interesting results, however, emerged from region rR. The development of classification success was notably slower in rR than in the other regions: in the first 20 ms of the stimulus period, PCC from rR remained significantly below the reference range, barely different from pretrial level. However, starting from approximately 60-80 ms past stimulus onset, clustering based on the population response from rR became clearly more successful in replicating the original classes than clustering based on response from any other region. For the four temporal windows starting at 80 ms after stimulus onset, the PCC ranges calculated from rR never overlapped with those from region rM1 within the same Experiment and *k* value. (Experiment 1 *k*=4: rR 0.783-0.915, rM1 0.582-0.742; Combined *k*=4: rR 0.755-0.949, rM1: 0.600-0.717; Experiment 1 *k*=5: rR 0.603-0.686, rM1 0.452-0.596; Combined *k*=5: rR 0.604-0.672, rM1 0.488-0.599). Within this temporal range, mode PCC calculated from rR was significantly above pretrial level in each window, which was not always the case for rM1. In many cases, mean PCC for rR was above the upper significance limit of the reference range, but this was never true for

rM1; actually, in one case (Experiment 1, *k*=4, at 100 ms) mean PCC for rM1 was below the lower significance limit.

Taken together, population responses in region rC provided information sufficient for stimulus classification early compared to region rM; later in time, however, the quality of classification based on rC responses was marginally worse than one based on rM responses. On the other hand, population responses in rR contributed very little to classification in the first 20 ms of the stimulus time. In the 20-60 ms period, the quality of classification based on rR responses was more or less similar to that based on responses from other regions; starting at approximately 60-80 ms, responses from rR allowed for much better classification of stimuli into original classes.

In summary, a clear pattern of differences emerged between response properties in the three regions. Responses in rC supported stimulus classification only with poor quality, but very quickly. Responses in rM supported marginally better classification, and they did so slightly later. Classification supported by responses in rR developed slower: only past 20 ms it matched classification based on rM or rC. Most importantly, however, starting at approximately 60-80 ms past stimulus onset the rR-based classification clearly surpassed classification based on any other region at any time point.

From the inspection of similarity matrices (Figure 2) and dendrograms (

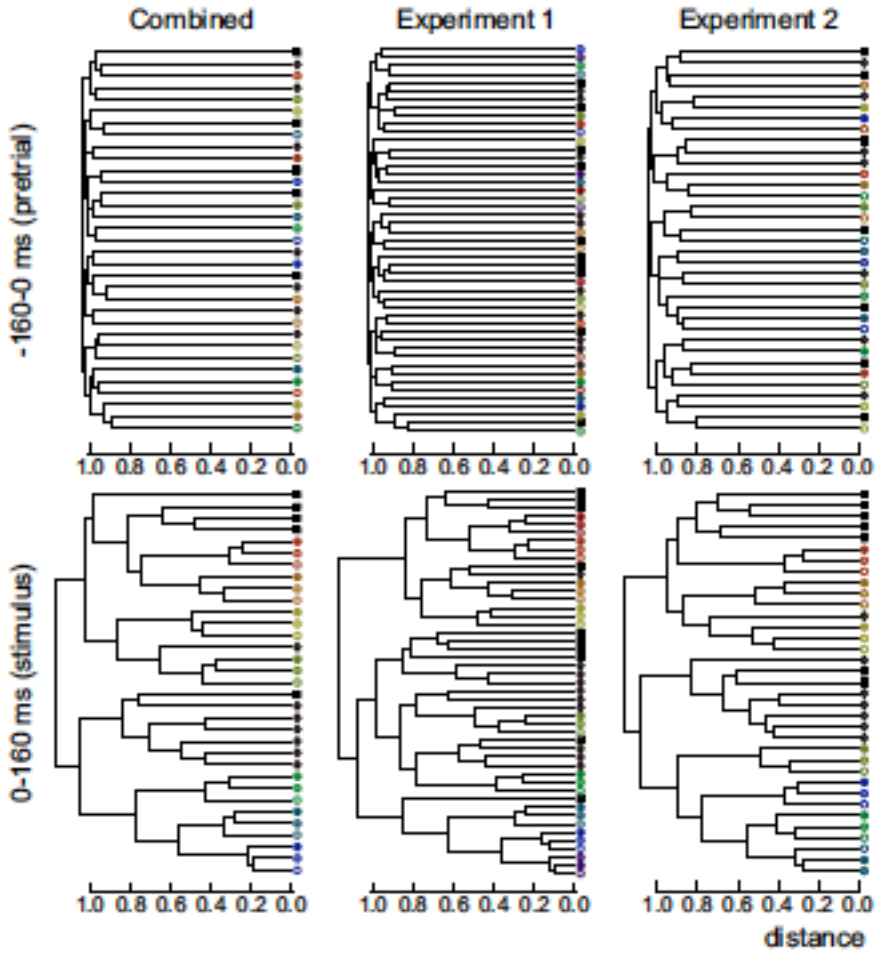


Figure **4**) we inferred that artificial stimulus frequency was a powerful factor that determined population responses. In the k-means analyses described above, the natural stimuli (ES and MC) were subjected to clustering together with the artificial stimuli (PT and BPN), the latter being classified based on frequency. Thus, the question remains whether the high classification success found in rR with k-means analyses arose only from very accurate classification of artificial stimuli, or region rR also excelled in classification of ES vs. MC stimuli. Many lines of evidence show that the latter was the case. First, the artificial stimuli constituted 57 or 64% (Experiment 1 and Combined, respectively) of stimuli, whereas more than 90% of stimuli were correctly classified based on rR responses at some temporal windows. Second, correlations of population responses within the MC and ES classes were very significantly higher than between these classes or between any of these classes and artificial sounds for rR in temporal windows at 80-160 ms, at which classification based on rR

responses was particularly successful (

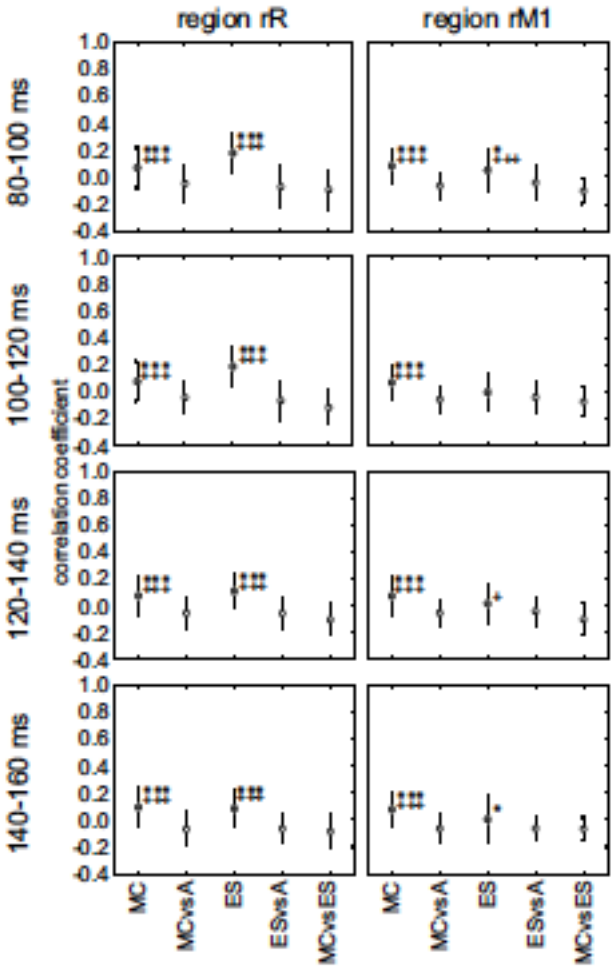


Figure **7**, left). Thus, there was a significant potential for MC vs. ES classification in rR responses. For comparison, correlations within the ES class did not exceed correlations between MC and ES or between ES and artificial

stimuli in rM1 responses as reliably as in rR responses (

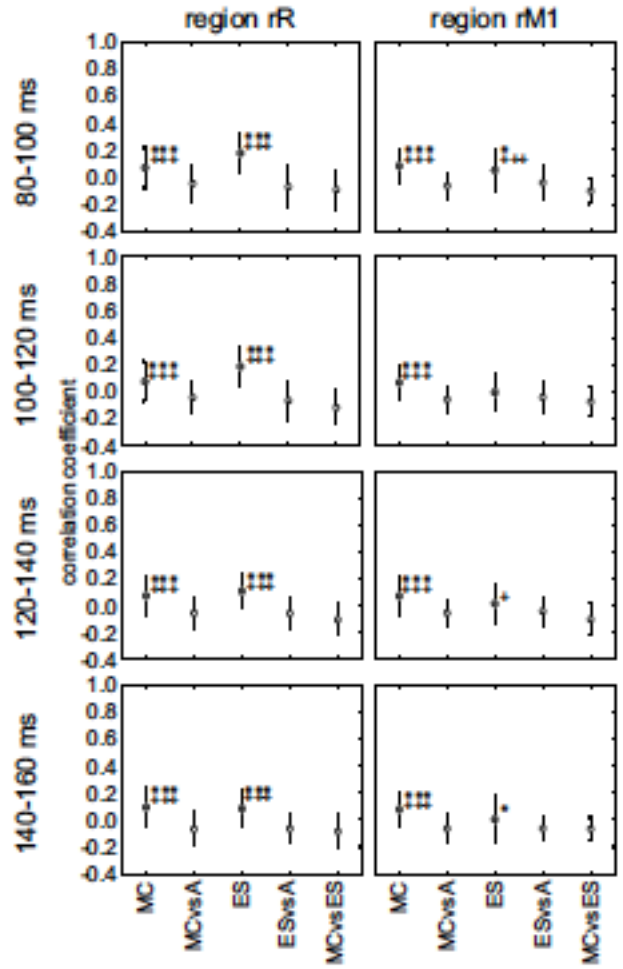
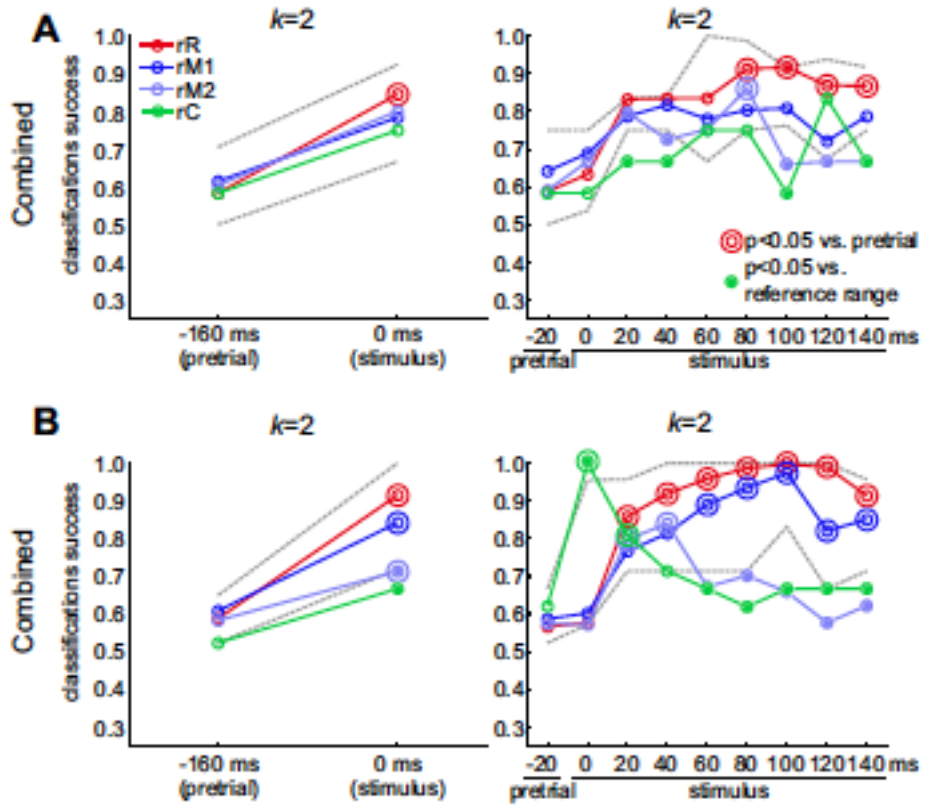


Figure **7**, right), indicating that the difference in general classification ability between rR and rM1 stemmed at least partially from different capability to classify natural sounds.



Finally,

Figure **8** shows PCC for all regions calculated for MC vs. ES classification only (*k*=2), for both 160-ms and 20-ms window sizes. In contrast to rM1 (as well as rM2 and rC), only classification based on rR responses was significantly better than classification based on pretrial data for the 0-160 ms window and for all 20-ms windows starting from 80 ms past stimulus onset. In that time period, classification success for responses from rR ranged 0.867-0.917, compared to 0.722-0.808 for responses from rM1.

The analyses of neural data described above provide evidence that population responses in early auditory cortex, specifically in region rR, carry sufficient information to allow for correct classification of stimuli that evoked these responses with an accuracy exceeding 90%. It would be interesting to know if this degree of classification accuracy can be supported by acoustic properties of

stimuli. Therefore, we measured a number of acoustic parameters of the stimuli and applied to them similar classification methods as we did for the neural data.

Correlations between log-frequency spectrograms and modulation spectra (MS) were visualized as similarity matrices (Figure 9A and C). A prominent feature of the similarity matrix calculated from spectrograms (Figure 9A) was the presence of oblique lines in the upper-left area, showing similarity of spectrograms calculated for PT/BPN of the same frequency, but different bandwidth. The picture resembled very closely the PT/BPN area seen in similarity matrices derived from neural data (Figure 2). Correlations within the group of natural stimuli were strong (lower right), but the separation of the MC and ES classes was not clear. Within-MC correlation coefficients were particularly high, but within-ES correlation coefficients appeared to be similar to correlation coefficients between ES and MC. These observations were confirmed quantitatively: mean within-MC correlation coefficient (*r*) values were significantly higher than *r* values between MC and PT/BPN, and than those between MC and ES. Within-ES *r* values, however, were significantly higher than *r* values between ES and PT/BPN, but did not differ from correlation coefficients between MC and ES (Figure 9B), in contrast to correlations derived from neural data (

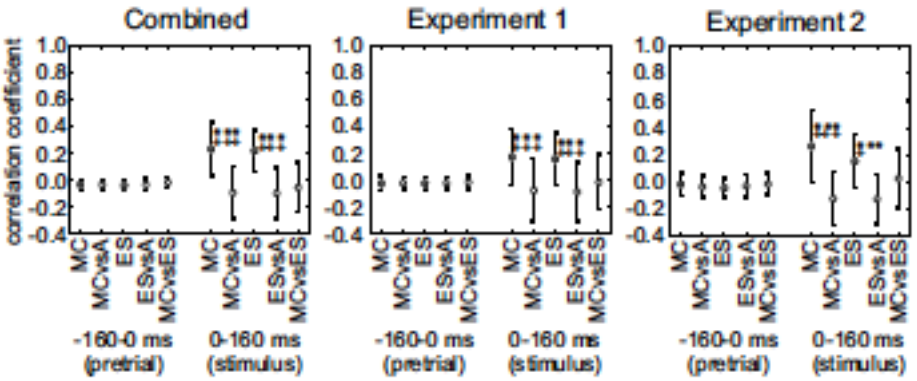


Figure **3**A).

The similarity matrix calculated from modulation spectra did not show the same characteristic patterns for the PT/BPN stimuli (Figure 9C), as did similarity

matrices calculated from spectrograms or neural data. The connection between PT/BPN of same frequency but different bandwidths was missed. On the other hand, pure tones seemed to be reasonably well separated from BPN.

At first glance, separation of the ES and MC classes was visible in the similarity matrix. Indeed, both within-MC and within-ES *r* values were significantly different from *r* values between ES and MC, as well as from *r* values between either natural sound class and artificial sounds (Figure 9D). However, within-MC *r* values were actually *lower* than those between MC and ES, showing that modulation spectra of various MC were on average more similar to modulation spectra of ES than to modulation spectra of other MC. This shows that differentiation of MC as a stimulus class cannot be supported by differences and similarities between MC (as measured with the acoustical parameters that we chose) and is in contrast to the findings from the neural data (Figure 2 and

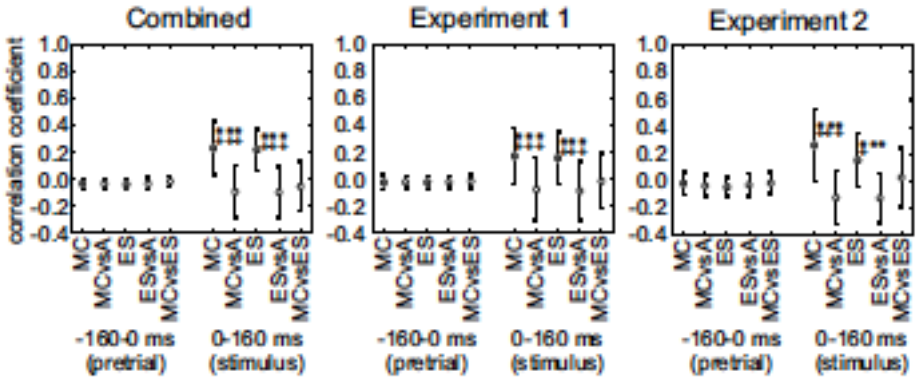


Figure **3**).

Finally, we attempted to classify the stimuli based on acoustic parameters using k-means clustering. As with neural data, we classified all stimuli into *k*=4 or *k*=5 clusters, and natural stimuli into *k*=2 clusters. The PCC values calculated based on acoustic parameters were compared to PCC values obtained from the neural

data in the 80-160 ms range (

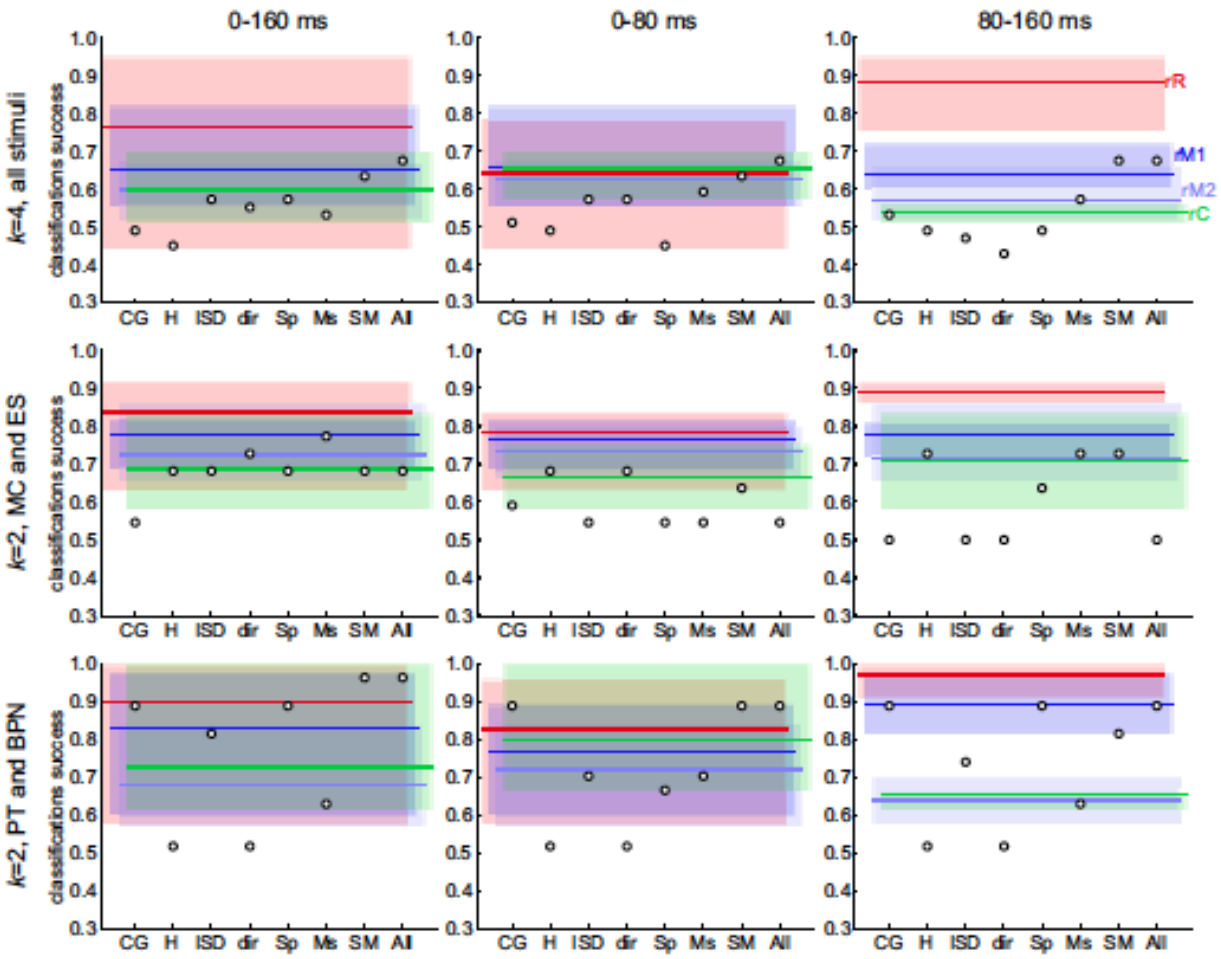


Figure **10**).

For *k*=4, PCC derived from any acoustic parameters, direct or derived from spectrograms or MS, or from any combination of acoustic parameters, remained below 0.6. The PCC at best reached values comparable to those derived from rC and rM2 recordings, while remaining below the range of values produced by analysis of rM1 population response, and far below those derived from rR responses. In two cases of spectrum center of gravity and mean harmonicity, acoustics-based PCC were below the range of PCC obtained from any of the regions.

For *k*=5, the acoustics-based PCC did not differ as much from neural based-PCC, being comparable to PCC obtained from responses from regions rC, rM1 and rM2. Still, they failed to match PCC derived from rR responses: only PCC from clustering based on parameters derived from dissimilarity of spectrograms exceeded 0.6 and reached rR-based PCC range, still remaining below its mean.

Finally, for clustering of natural sounds (MC and ES) only (

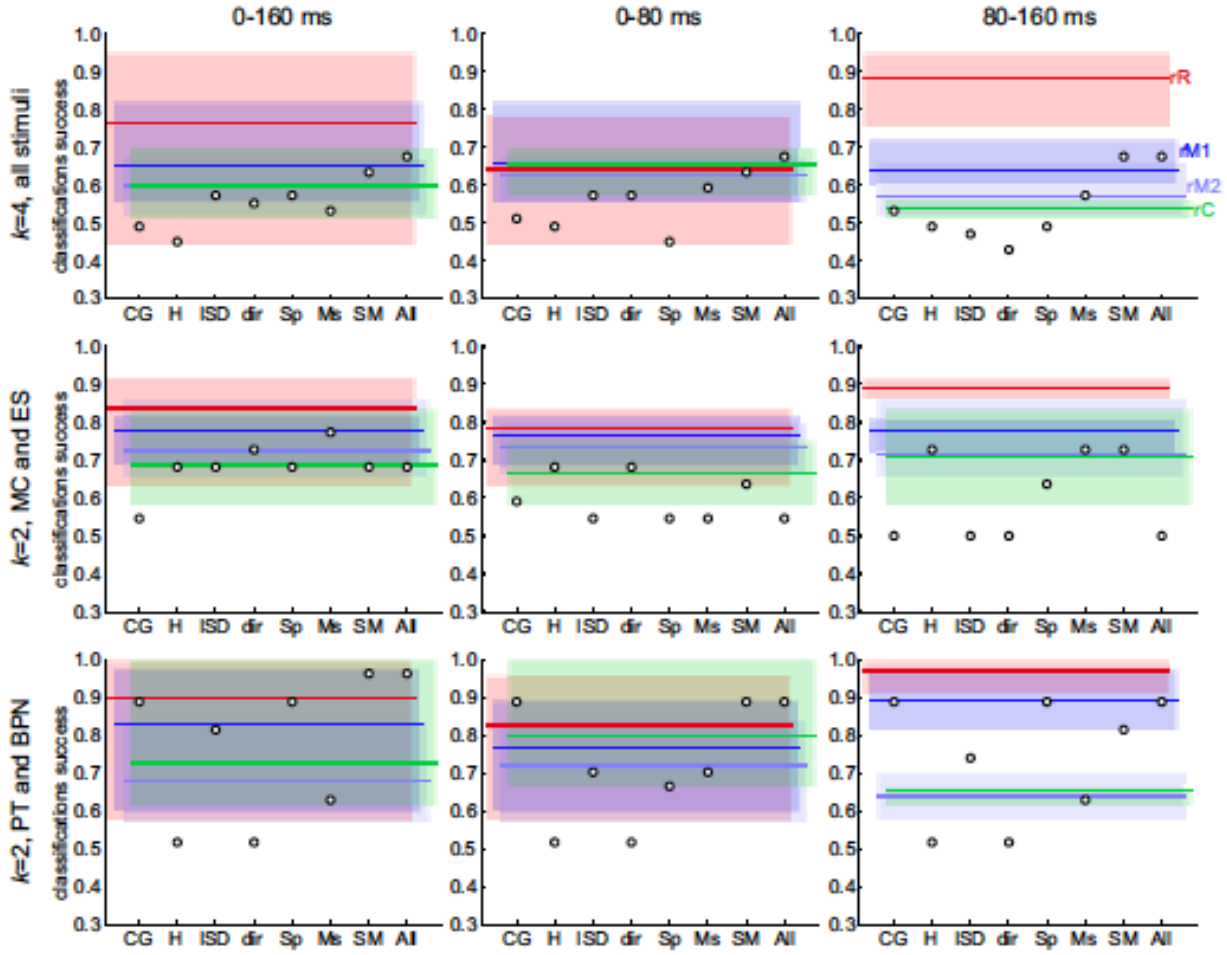


Figure **10**, *k*=2), some PCC (in this case, three) derived from acoustic parameters were below the range of all population activity-derived PCC. The remaining ones were spread across the neural PCC range from regions rC, rM1, and rM2. None reached the range of PCC obtained from the rR population response.

Taken together, the acoustic features that we analyzed could support classification success at best comparable to classification based on population responses collected in regions rC, rM1, and rM2. On the other hand, classification based on population responses recorded in rR within 80-160 ms past stimulus onset clearly surpassed not only classifications derived from the other regions' responses, but also those obtained from acoustic parameters, whether analyzed separately or in combinations.

Discussion

We applied neural population analysis to study the representation of stimulus *identity* in rhesus monkey auditory cortex. To our knowledge, such techniques have not been used for this purpose before, although population analysis has proved successful previously in studies of the representation of sound *location* in auditory cortex (Miller and Recanzone 2009; Recanzone et al. 2010) and of the representation of visual stimuli in visual cortex (e.g., Kiani et al. 2007; Kriegeskorte et al. 2008).

Methodological considerations

Kriegeskorte et al. (2008) emphasized that the methods they used assumed no or little structure in the data; only after classification was performed, correspondence between resulting clusters and natural categories was found. A similar approach was taken by Kiani et al. (2007). We initially followed this principle when we studied similarity matrices and dendrograms. In this way we confirmed the existence of a particular structure in the neural data and that this structure reflected our preconceived notions about stimulus categories. In the next step, however, we used k-means clustering, with *a-priori* structure explicitly sought in the data. This approach allowed us to achieve our main goal: quantifying how well stimulus structure is represented in different cortical areas and at different time points. It has to be pointed out that, should we have imposed an inappropriate structure on the clustering procedure, we would likely not have seen any difference between regions, and classification success scores would be low. What we saw instead were clear between-region

differences and high classification scores in rR (

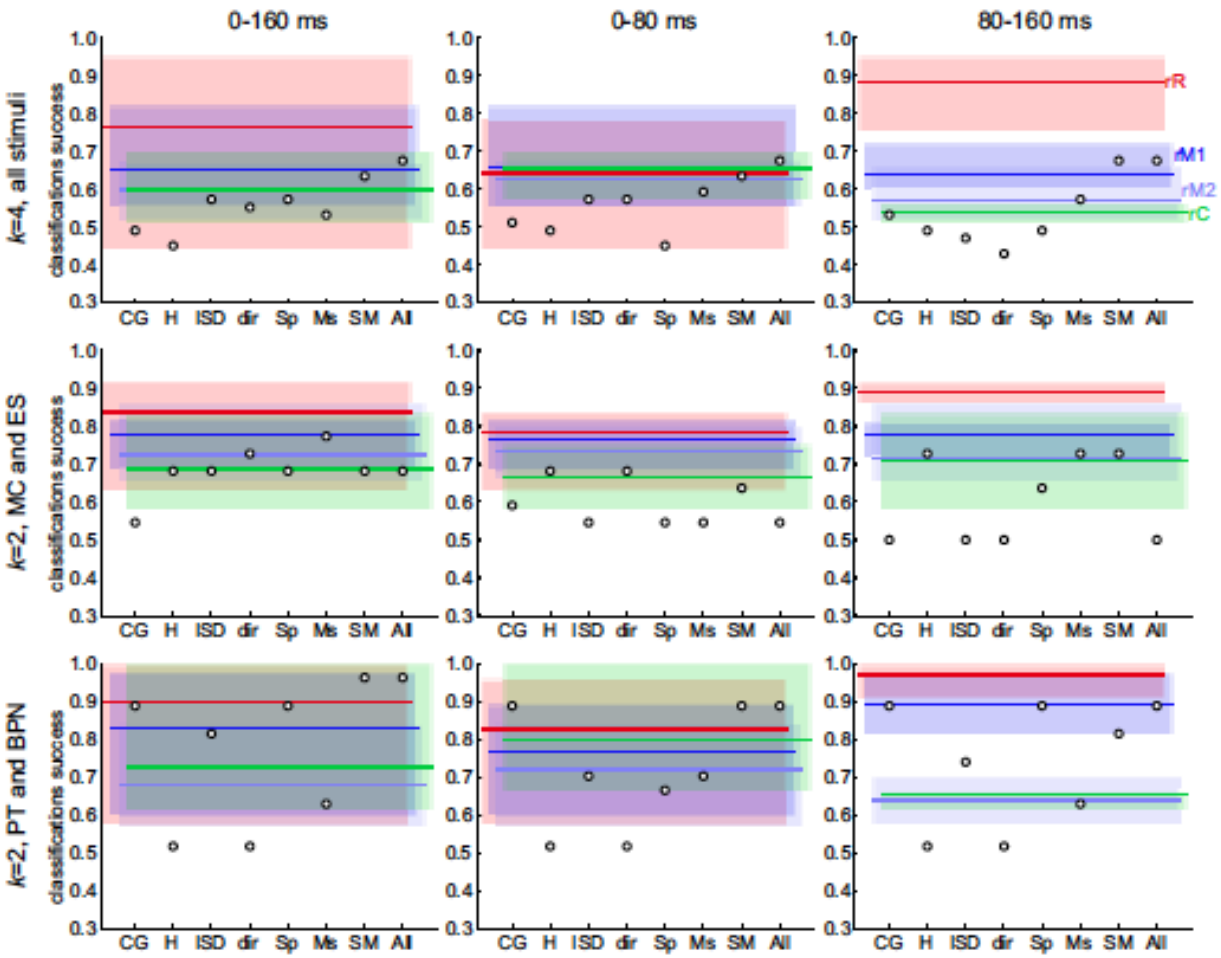


Figure **10**). Thus, the stimulus structure we imposed on k-means clustering appears to match a structure actually represented in the investigated cortical area, and, more importantly, a structure whose representation in anterior areas (region rR) exceeds that in more posterior areas (regions rM and rC). While we cannot assure that the structure was the most appropriate, our results, together with the choice of categories which arguably reflects natural categorization (i.e., vocalizations, environmental sounds, low frequencies, high frequencies), indicate that the structure imposed on k-means clustering was meaningful in terms of cortical sound processing.

In addition to employing k-means clustering to quantify classification quality, another novel component was to perform classification in successive short temporal windows, in addition to analysis over the entire stimulus duration (in our case, the duration of the shortest stimulus). This has proved very successful: when firing rates were averaged over the first 160 ms of the stimulus, virtually no differences in classification success between regions were detected (

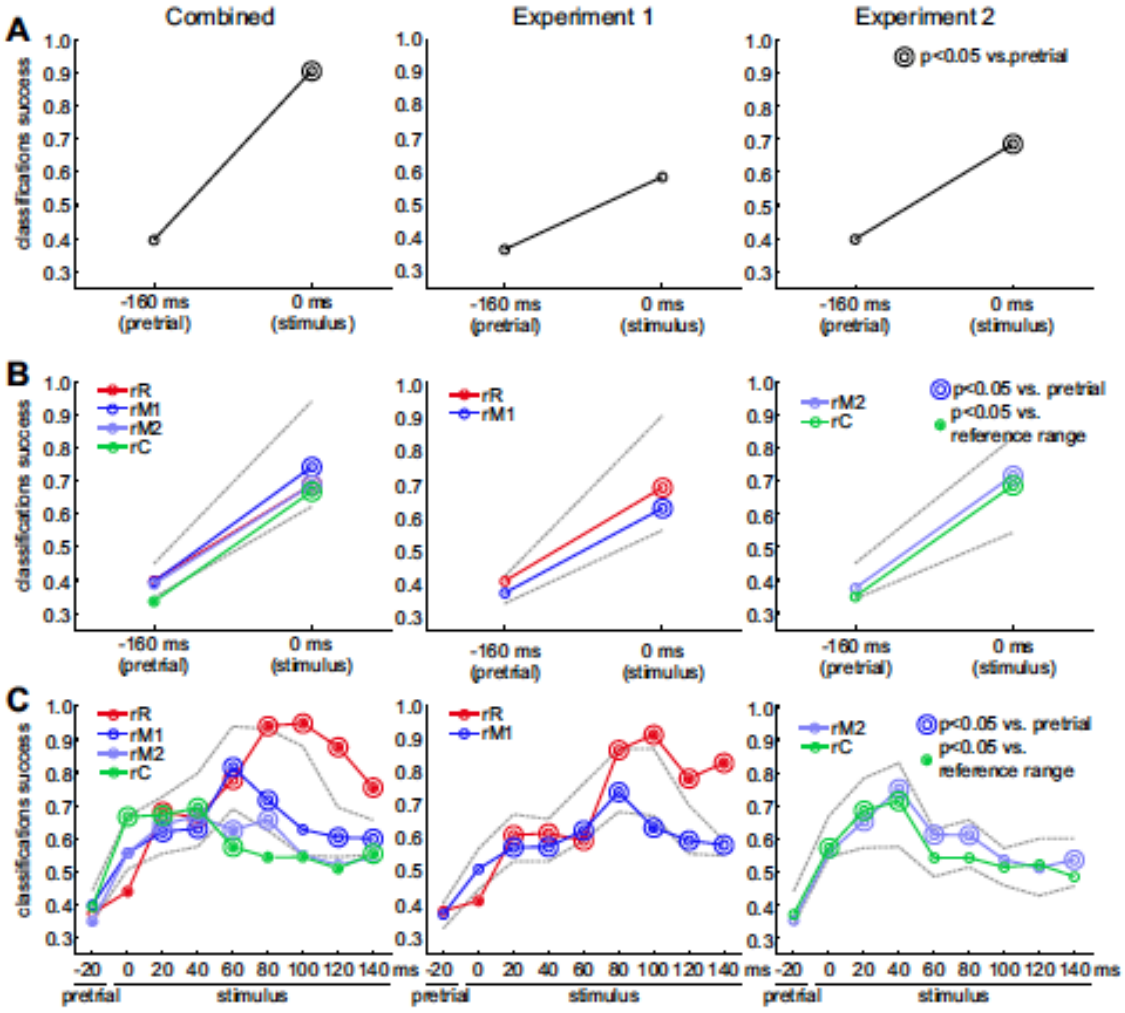


Figure 5B); however, analysis in 20-ms windows not only revealed a clear divergence between region rR and the other regions (

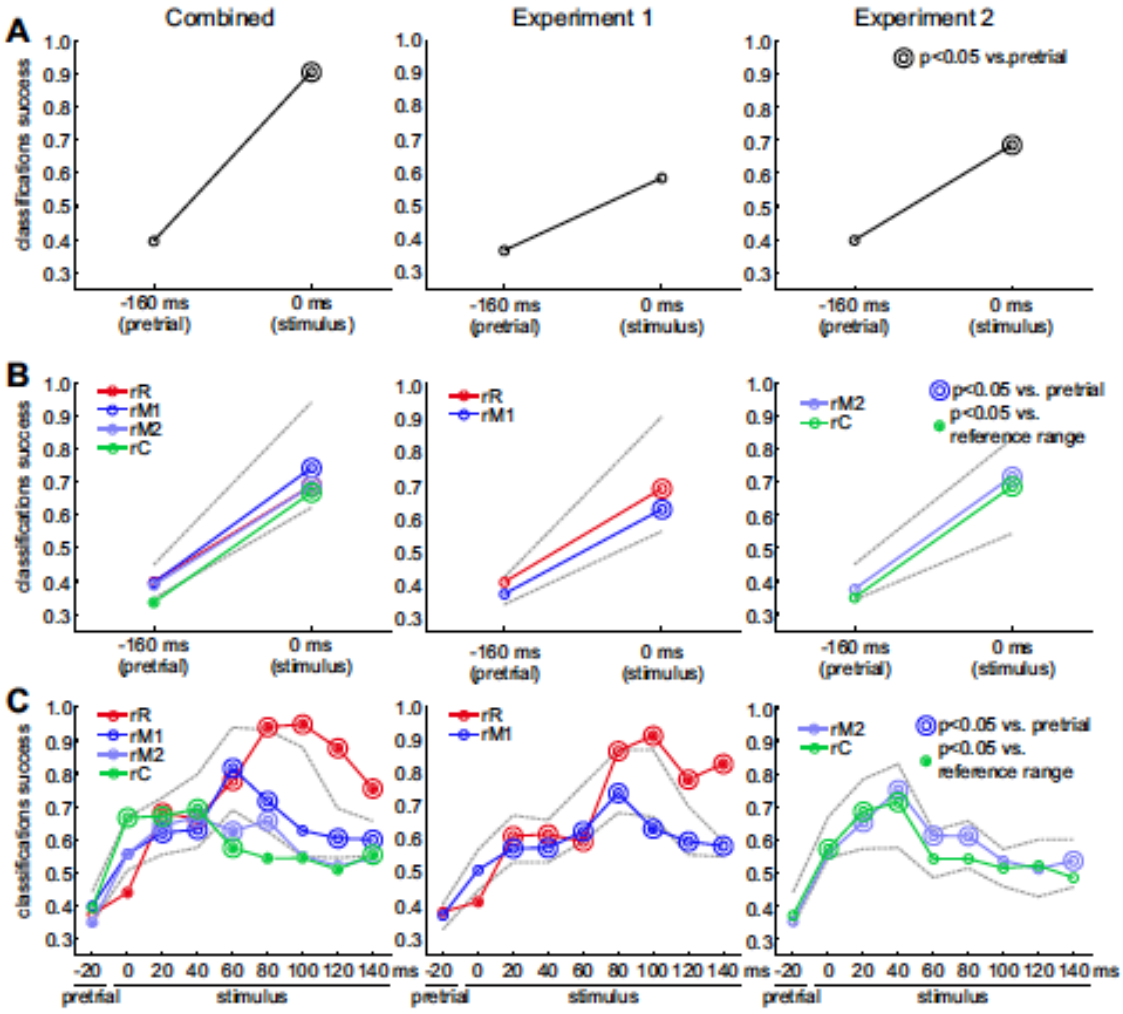


Figure 5C), but allowed us to trace the temporal dynamics of classification ability and discuss its origins (see further below).

A potential confound comes from the fact that the neural data came from two different experiments, with different stimulus presentation techniques, neural recording techniques, and slightly different stimulus sets; e.g. data from regions rC and rR came from different experiments. However, our key analyses were replicated within-experiment, that is, data from region rC were compared to those from region rM2 (region rM recorded in the same experiment as rC), and

data from rR to those from rM1 (region rM recorded in the same experiment as rR,

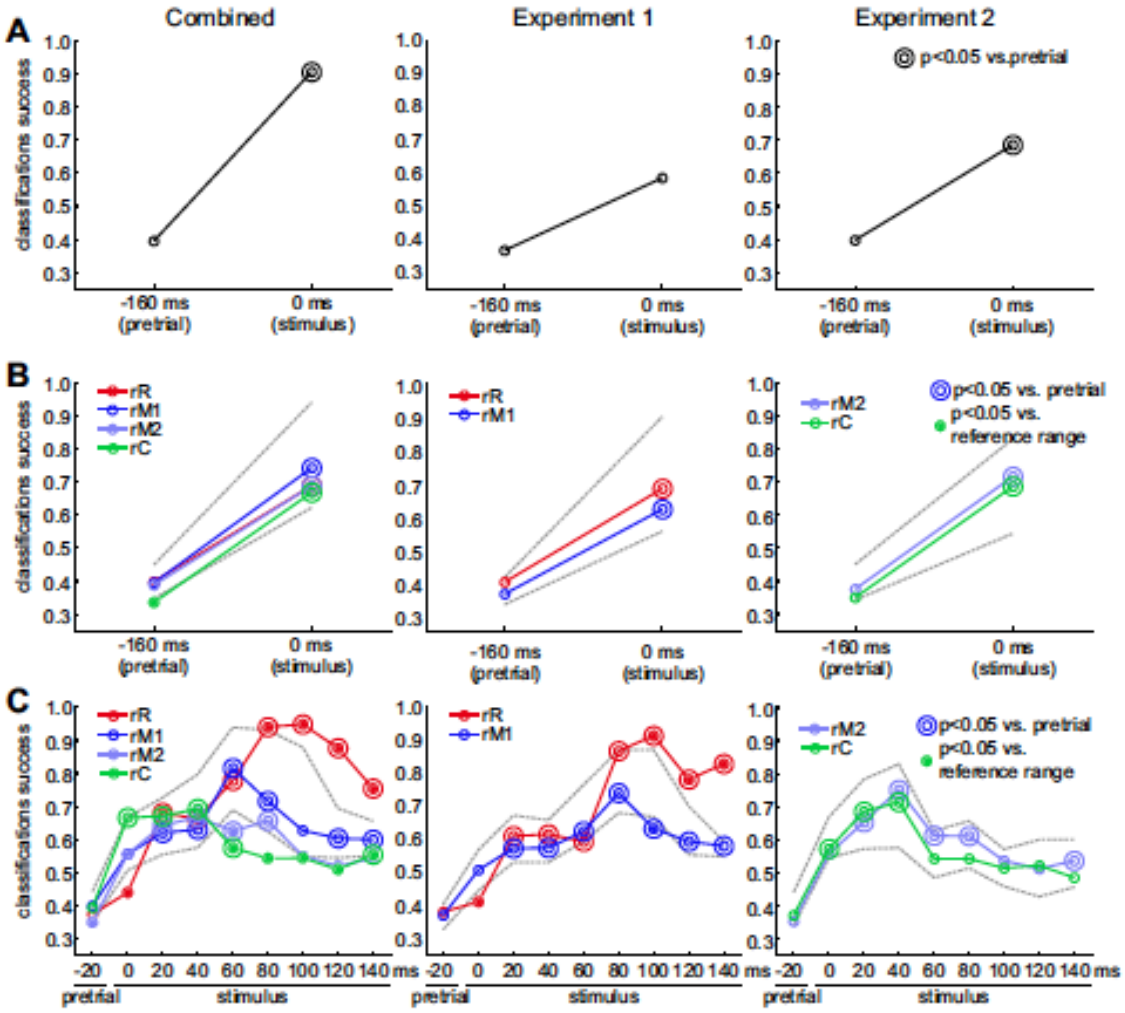
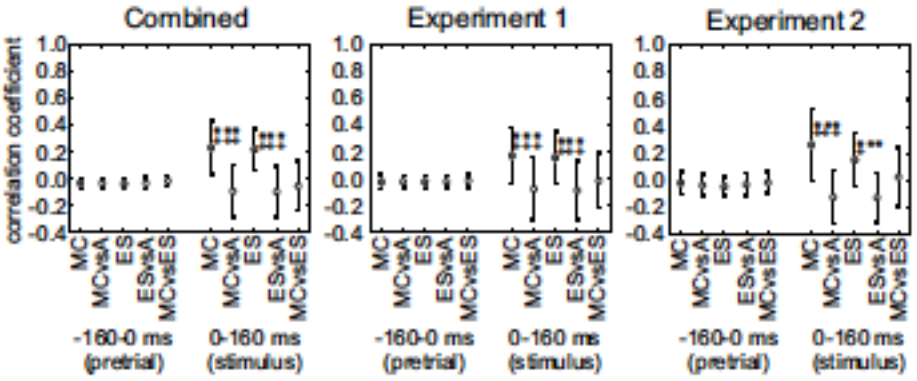


Figure 5C). In this way not only the results from the combination of the experiments were validated, but thanks to “anchoring” both experiments in region rM, we gained insight into the effects of experimental conditions under which neural data were gathered on stimulus classification based on these data. We did not systematically explore the effects of all differences between the experiments on classification, but we can speculate that reduced stimulus quality due to using small loudspeakers and/or lower sampling frequency (consequently, lower stimulus bandwidth) and/or lower stimulus intensity in Experiment 2 resulted in neural data that provided support for somewhat less

accurate classification than data from Experiment 1. This effect was not offset by supposedly more reliable estimation of firing rates in Experiment 2, wherein each stimulus was played 60 times, compared to 10-13 times in Experiment 1. In conclusion, quality of stimulus presentation should be carefully attended to when stimulus classification is investigated.

Band-pass noise

Bandwidth of band-pass noise (versus pure tone) has been shown to influence neural responses in the auditory cortex (e.g., Rauschecker et al. 1995; Rauschecker and Tian 2004; Petkov et al. 2006; Kuśmierek and Rauschecker 2009; Kajikawa et al. 2011); also discrimination of a tone from a ⅓-octave or 1-octave noise does not seem to be difficult at least for human listeners (although we are not aware of any formal tests in monkeys, see however below). It is therefore surprising that only a small effect of bandwidth on population response was found in the present study (see



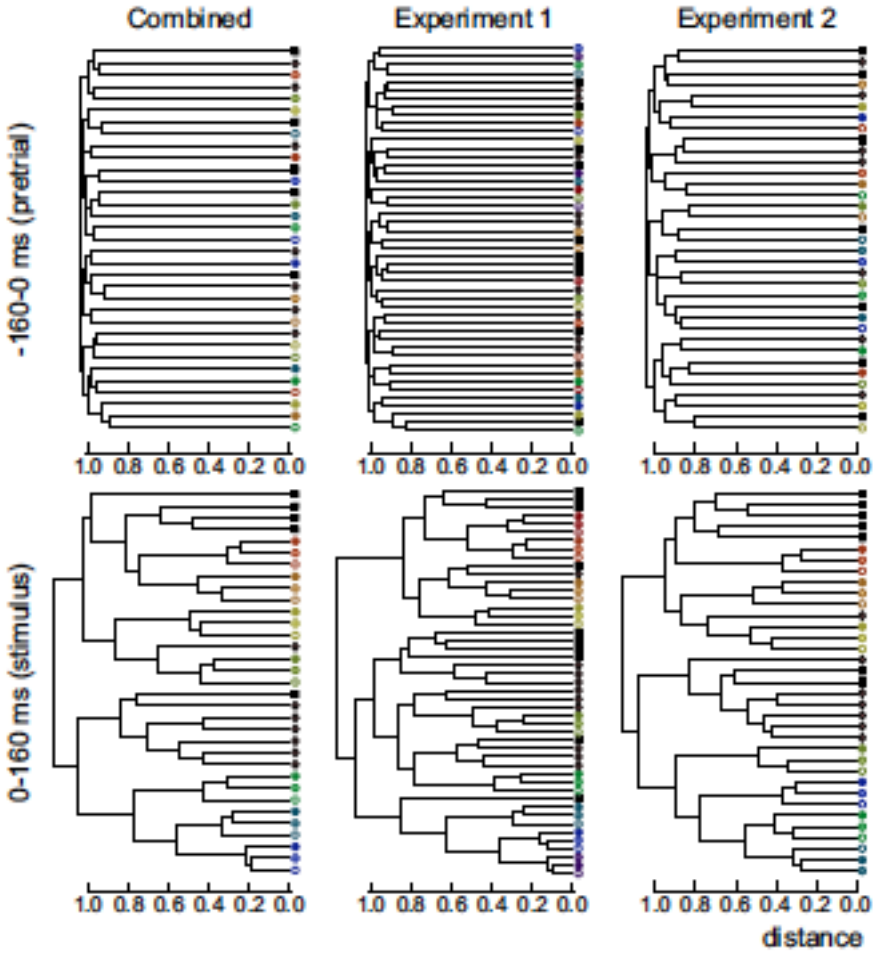


Figure **3**,

Figure **4**, and accompanying description in Results). It should be noted that the most robust effects of bandwidth have been found in the lateral belt (Rauschecker et al. 1995; Rauschecker and Tian 2004; Petkov et al. 2006), which was not included in the present study. Our recordings were derived from core and medial belt, and band-pass noise preference in the medial belt over core, although demonstrable, appears to be less pronounced than in the lateral belt (Petkov et al. 2006; Kuśmierek and Rauschecker 2009; Kajikawa et al. 2011). Thus, the picture might be different if the present study was replicated in the lateral belt or in cortical areas further anterior or lateral. Furthermore, behavioral data from Experiment 1 offer an insight into perception of bandwidth and frequency in rhesus monkeys. In this experiment, the behavioral target was a short “melody” consisting of a rapid succession of four 125-ms pure tones; the first tone’s frequency was 523 Hz, less than a semitone from frequency of a non-target 500-Hz stimulus. Consequently, the monkeys often responded to the 500-Hz tone. In addition, they quite often responded to ⅓-octave and 1-octave noise bursts centered at 500 Hz, and much more rarely to tones or noises at

neighboring frequencies of 250 or 1000 Hz, let alone more distant frequencies (

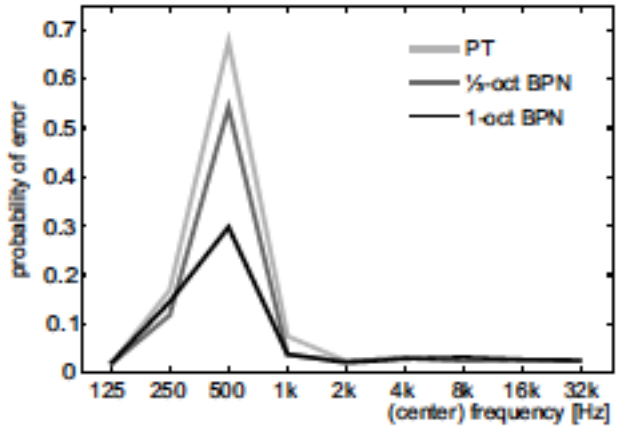


Figure **11**). Apparently, the perceptual difference between a pure tone and a band-pass noise burst at the same frequency was smaller than a step to a tone (or band-pass noise) one octave apart. The small effect of bandwidth on population response may be a correlate of this behavioral finding.

Stimulus classification in the 0-20 ms window

Within the first 20 ms after stimulus onset, the classification capability measured by PCC developed differently in the cortical regions (

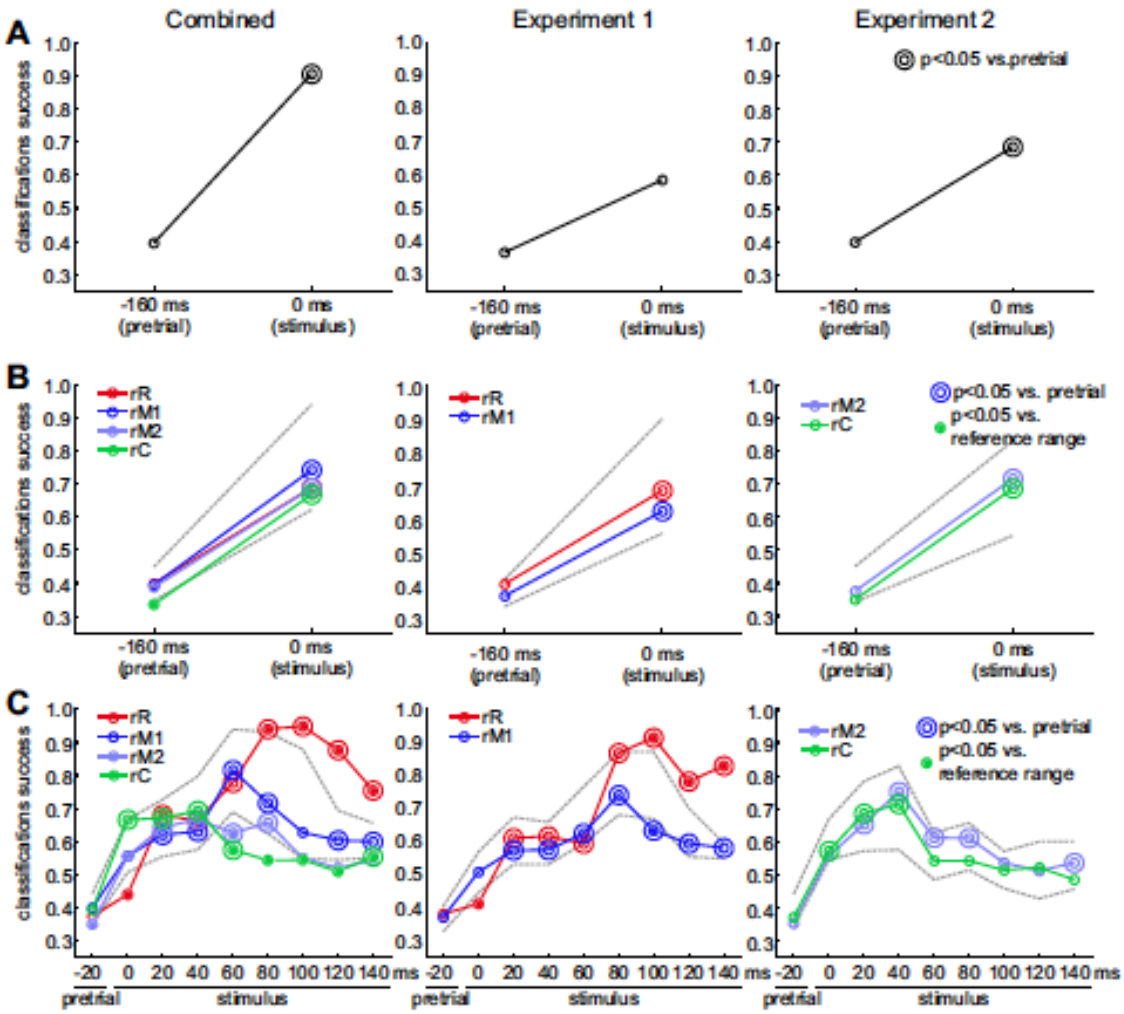


Figure 5C). Data from region rC supported high classification accuracy, comparable to accuracy in any subsequent time window. On the other hand, responses from rR barely differed from chance level in the 0-20 ms window, and the responses from region rM fell in-between. Latencies in area CM are shorter than latencies in area A1 in rhesus monkeys (Scott et al. 2011; also Kuśmierek and Rauschecker, in preparation); the same was found by Kajikawa et al. (2005) in the marmoset. Latencies in R and RM were shown to be longer than in A1 and MM in macaques (Recanzone et al. 2000a; Kuśmierek and Rauschecker 2009; Scott et al. 2011) as well as in marmosets (Bendor and

Wang 2008). The differences between classification capabilities found in the 0-20 ms window likely reflect this posterior-to-anterior progression of neural latencies in the core and medial belt. In the first 20 ms of the stimulus, fewer region-rM neurons than region-rC neurons began firing, and consequently, contributed to classification, and only few rR neurons were active at that time.

Object categories or combinations of features?

Finding category specificity in neural responses always raises an important question: Does this effect correspond to genuine object categorization, or simply to low-level stimulus features that were unequally distributed across different stimulus classes? A direct answer can be given if sharp categorical boundaries are found in response to gradual continua of low-level features: invariance of responses to large changes of features irrelevant to object identity, and specificity of responses when small changes are introduced to highly relevant features. This approach requires more knowledge about feature relevance and on parameters of object invariance than is available for auditory perception in macaques. Therefore, we have chosen a simpler alternative approach (Kiani et al. 2007), that is, we attempted to classify the stimuli based on low-level acoustic features and compared the results to results of classification based on neural data. An important caveat is that a failure of feature-based classification to match neural data-based classification does not prove that the cortical region in question operates on representation of sound objects beyond simple combinations of features. Another explanation might be that crucial low-level features were not entered into the analysis. We have chosen three direct stimulus features that covered a wide range of qualities:

spectrum center of gravity approximated which cochlear channels were activated by the sound; harmonicity measured whether a stimulus was more periodic or more noisy; and standard deviation of harmonicity revealed amplitude modulation structure. In addition, parameters were derived from dissimilarity of simple frequency-time representation (log-frequency spectrogram), and from dissimilarity of spectrotemporal modulation spectra (Singh and Theunissen 2003; Cohen et al. 2007).

Inspection of

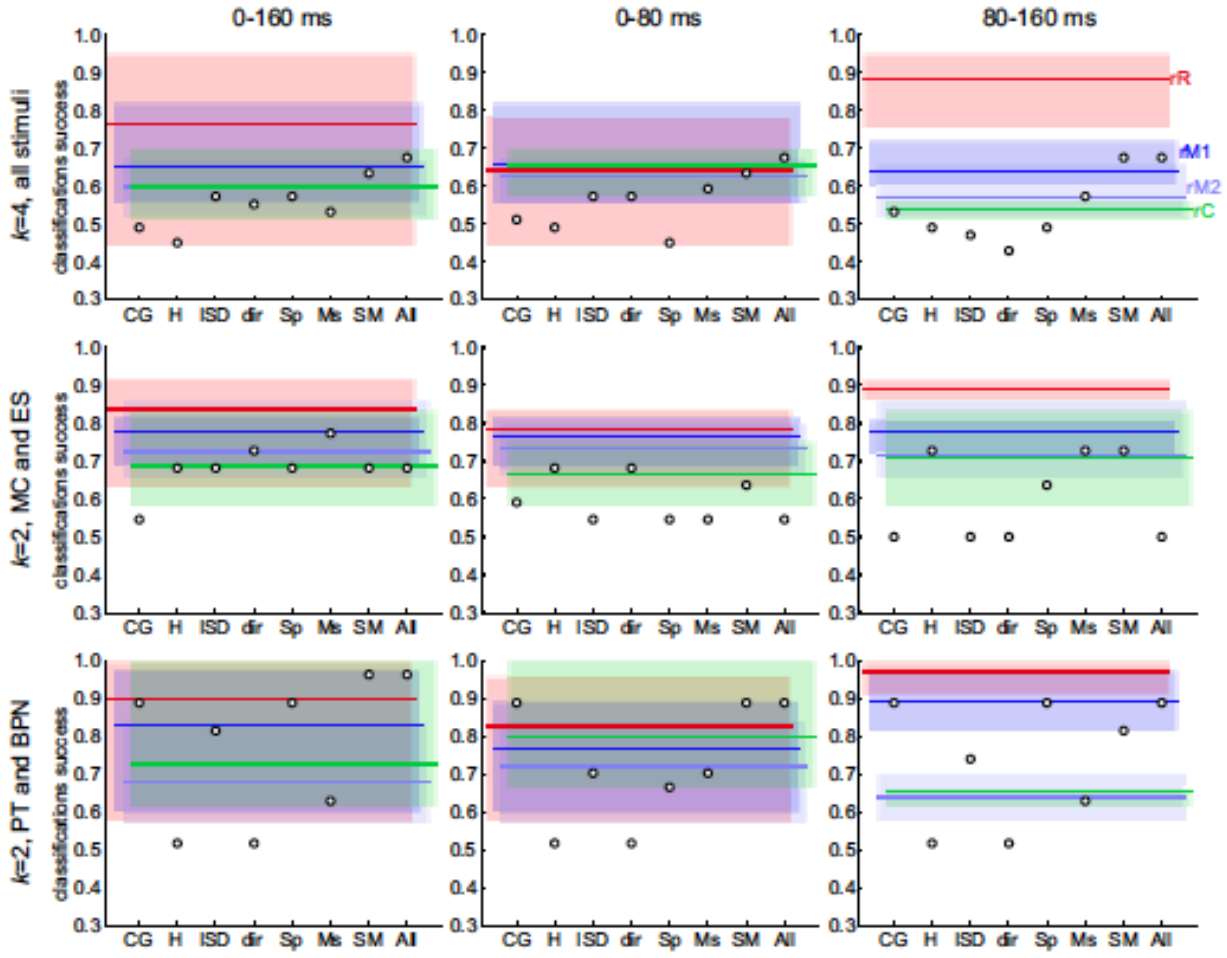


Figure **10** suggests that representation of stimuli in regions rM and rC may be driven by low-level features, as the performance of stimulus classification based on low-level features overlapped with classification performance based on neural responses from these regions. Conversely, population responses in rR

(in the 80-160 ms time window) supported much higher classification performance than achieved by way of stimulus features. Although more research is needed before definite conclusions can be drawn, this result may suggest that already in region rR the stimulus representation begins to shift from feature-based towards object-based.

Processing streams in the auditory cortex

The function of the dorsal stream of auditory cortical processing compared to the ventral stream remains less sharply defined. Although processing of space (i.e., sound source location) has been emphasized (e.g., Romanski et al. 1999; Rauschecker and Tian 2000), more recent proposals include sensorimotor integration with the purpose of learning and control of auditory production (Rauschecker and Scott 2009; Rauschecker 2011). Specifically, parietal regions are fed by posterior areas of auditory cortex with a fast, temporally precise, but relatively rough “primal sketch” of ongoing auditory information (Rauschecker 2011). Properties of area CM, as shown in the present study, fit this description well: neural data collected from CM support stimulus classification at an earlier

time point than data from any other region (

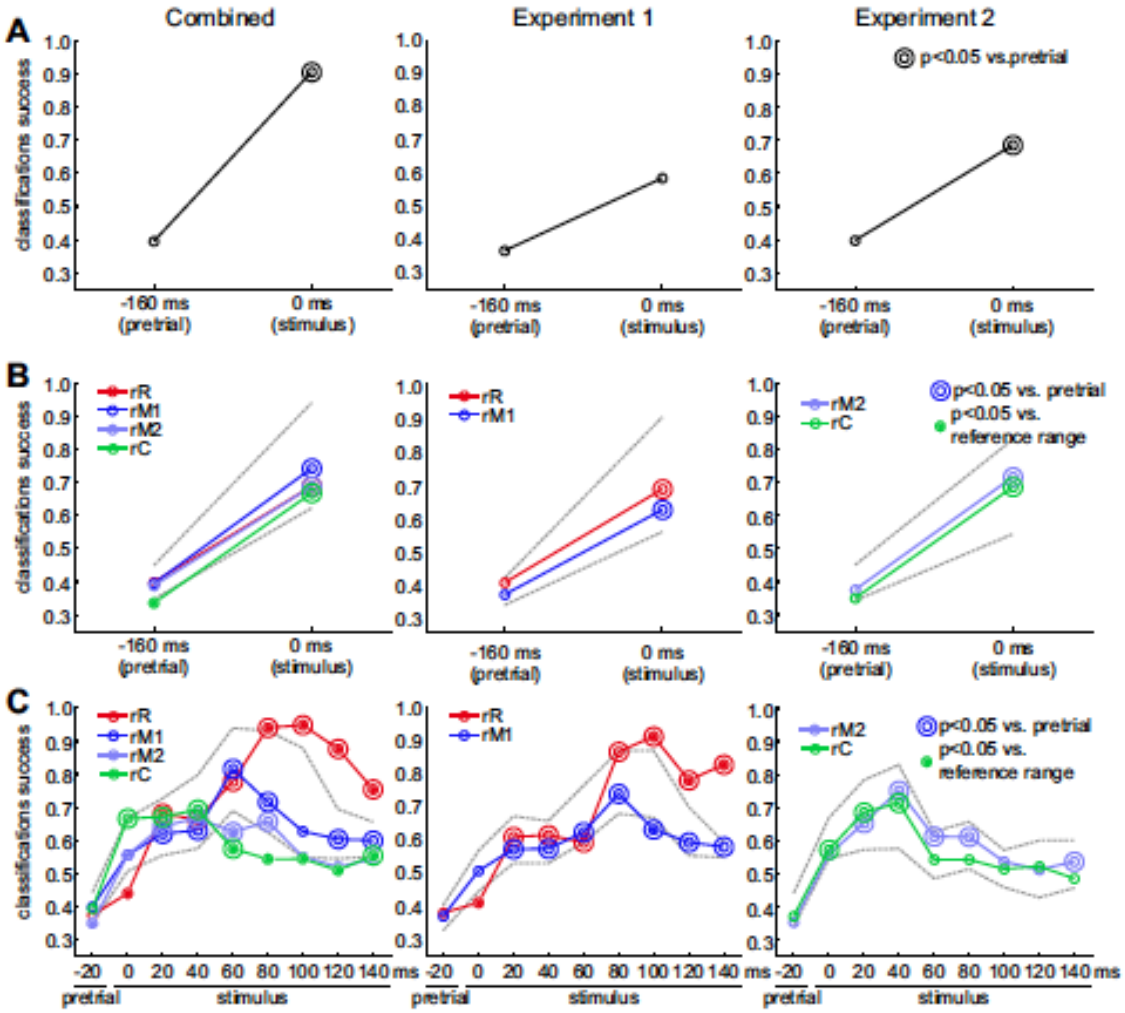


Figure 5C), but with relatively low classification quality (

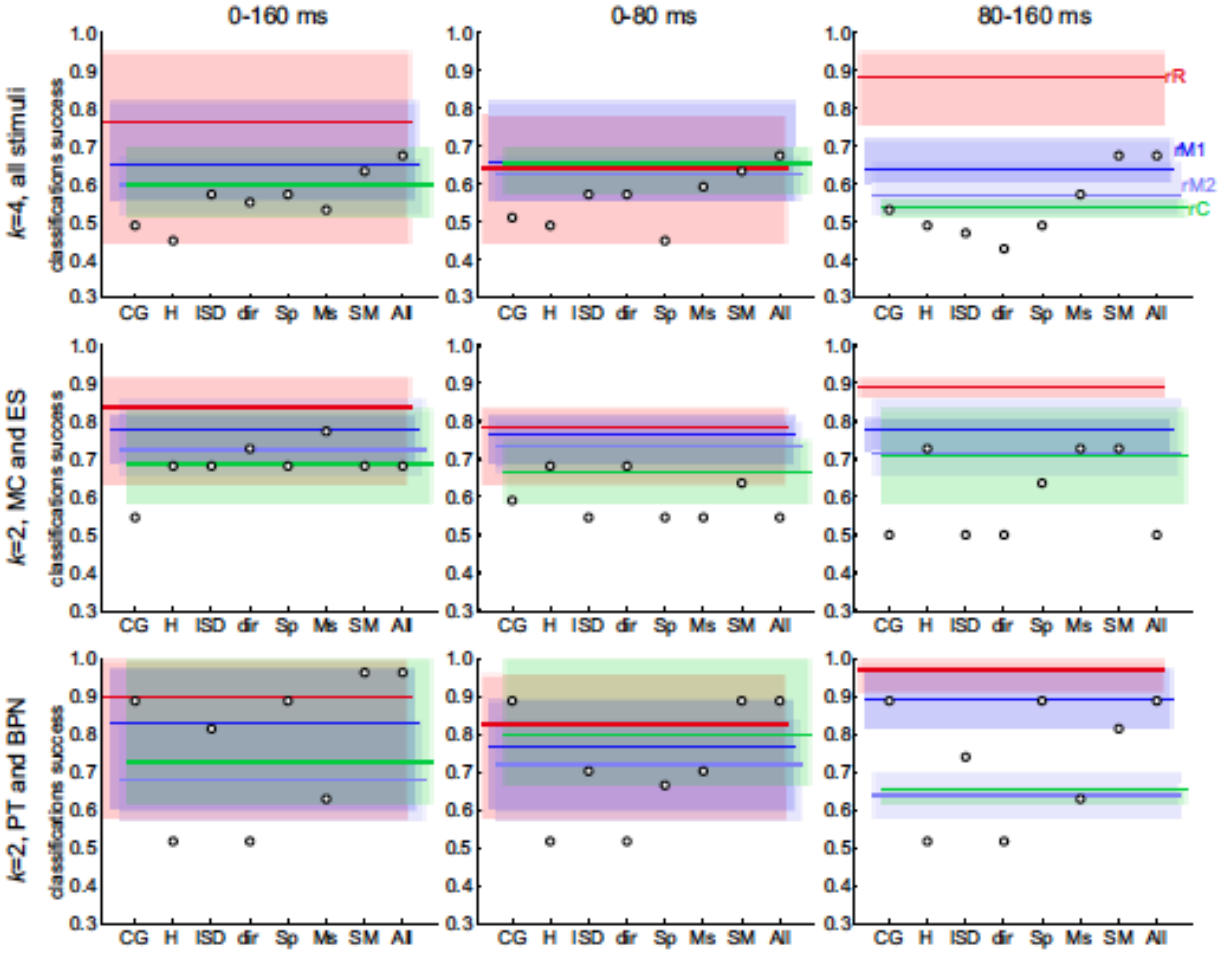


Figure **10**). While a common denominator can be found for both functions postulated for the dorsal stream (i.e., spatial processing and audio-motor integration, Rauschecker 2011), emerging properties of areas CM and CL hint at a possible division of labor. The spatial sensitivity appears to be more pronounced in CL (Tian et al. 2001; Miller and Recanzone 2009; also Kuśmierek and Rauschecker, in preparation), while CM seems to respond faster and with a higher temporal precision (Kajikawa et al. 2005; Scott et al. 2011; the present study; also Kuśmierek and Rauschecker, in preparation).

The results of the present study show that specialization for stimulus identification, a feature postulated for the anterior stream, can be found as early

as in areas R/RM. The finding adds to the earlier result of Tian et al. (2001) who found increased selectivity for discrimination between monkey calls in area AL, a lateral belt area adjacent to area R. Strongly specific responses are expected in more anterior regions of the superior temporal lobe, or even in ventrolateral prefrontal cortex (e.g., Poremba et al. 2004; Romanski et al. 2005; Petkov et al. 2008; Kikuchi et al. 2010), but the findings of Tian et al. (2001) and of the present study demonstrate that primitives of "what" specialization are already present much closer to the primary areas.

How can these findings be reconciled with results from Recanzone (2008), who described no increased stimulus selectivity in area R compared to more posterior areas? One explanation might be that high selectivity is limited to belt areas, which were studied by Tian et al. (2001) and in the present study (lateral belt area AL and medial belt area RM being a part of our region rR, respectively), while Recanzone (2008) studied the core area R, but not adjacent belt areas AL or RM. Still, this explanation would imply that enhanced stimulus selectivity in RM is strong enough to be detectable even when RM neurons are, like in the present study, pooled with R neurons, which supposedly show no enhanced selectivity. Given that RM units constituted only about ⅓ of our region rR, and that response properties of RM cells are in general similar to R units (Kuśmierek and Rauschecker 2009), this explanation seems unlikely. The crucial argument, however, is that the difference in classification success between the rR and rM1 regions late after stimulus onset is still visible even

when the analysis is limited only to core components, that is, areas R and A1 (

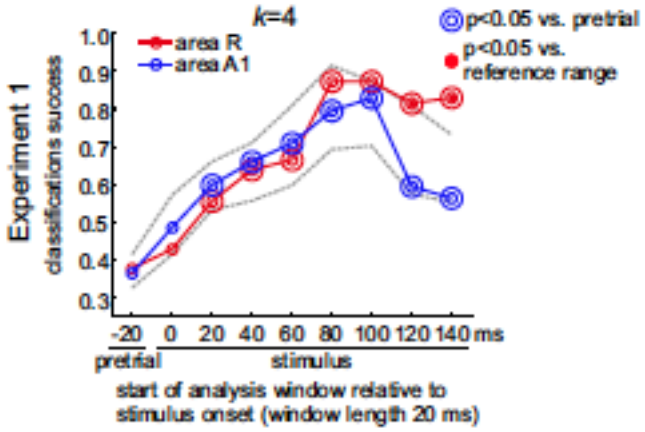


Figure 12).

Another possible explanation is that Recanzone (2008) and Tian et al. (2001) studied stimulus selectivity within the same class and in individual neural units, whereas we looked at between-stimulus-class differences in responses evoked in larger neural populations. Our choice of methods might have provided adequate sensitivity to detect selectivity in R/RM, whereas classical methods used by Recanzone (2008) and Tian et al. (2001) were sufficient to detect only more pronounced selectivity, such as the one present in AL.

Finally, one should not overlook that Recanzone (2008) did, in fact, show enhanced selectivity in area R using the linear discrimination method, although the effect was limited to reversed vocalizations and to a subset of linear discriminator bin widths. The effect was absent at the shortest bin widths (~2-10 ms), at which the discriminator performance reached the highest values. It is possible, however, that these high values at short bin widths were a byproduct of a particular implementation of the algorithm, and that results obtained at

longer bin widths actually reflected genuine properties of the investigated areas (see Supplementary Information).

An intriguing finding in the present study was that classification based on region rR data initially was only as good as classification based on data from other regions. But about 60-80 ms after stimulus onset, rR-based PCC began to clearly exceed those derived from other regions. This delay might have been due to time needed by rR neurons to integrate incoming information and perform computations needed to discriminate stimuli at a high accuracy level, beyond what was possible based on simple analysis of acoustic features. We have shown previously that (for temporally structured stimuli) the best linear discriminator bin width, which can be interpreted as a measure of temporal integration scale, was on the order of 40-50 ms in area R, and 20-30 ms in A1 (Kuśmierek and Rauschecker 2009). From synchronization cut-off frequencies, Scott et al. (2011) estimated temporal integration windows to be approx. 100 ms in R and 20-30 ms in A1.

Instead of (or in addition to) the computation in local circuits, arrival of feedback or "top-down" information from areas further downstream could have contributed to the late (60-80 ms past stimulus onset) surge in classification accuracy. Based on latency estimates from Kikuchi et al. (2010), candidates for the source of such feedback information include area RT, reported to respond with mean latencies of 40-70 ms, while area RTp would have to be excluded (70-110 ms latencies). Romanski and Hwang (submitted), on the other hand, found mean latencies exceeding 100 ms in the ventrolateral prefrontal cortex, but a subset of cells responded with latencies as fast as 50 ms; thus, these neurons could theoretically be the source of increased classification accuracy

found in region rR in the present study. It has to be noted as well that relying on latencies measured across different studies is somewhat risky. For example, Kikuchi et al. (2010) reported mean latencies in area A1 to be on the order of 40-60 ms, far longer than found by us (median 17-20 ms, Kuśmierek and Rauschecker 2009) or by others (Scott et al. (2011): median 20 ms, Recanzone et al. (2000a): mean 32.4 ms).

Methodological notes on linear discriminator analysis

The results of linear discriminator analysis reported by Recanzone (2008), as well as by Russ et al. (2008), and presented again by Recanzone (2011), raise some questions. In both studies, linear discriminator performance was found to reach very high values (80-90%) at a very short bin (window) width of 2 ms.

This finding implies that almost all neurons produced highly replicable and clearly different spiking patterns to all stimuli and that they were driven by almost all these stimuli. A cell that does not respond to a subset of stimuli would not produce spiking patterns supporting such high discriminator performance because these non-driving stimuli would not be distinguished by the procedure. Raster plots presented by Recanzone (2008) and Russ et al. (2008) do not seem to show firing that is highly replicable within each stimulus, differently patterned across stimuli, and vigorous for all stimuli.

Furthermore, the firing patterns should be replicable with 2-ms accuracy. Again, the raster plots in both papers do not seem to show such accuracy. Moreover, the replicability of spiking patterns could be based on one of two mechanisms: a precise abstract temporal code, or precise following of spectrotemporal features

in the ongoing stimulus. We have shown that spiking patterns in the auditory cortex seem to follow features of the ongoing stimulus (Kuśmierek and Rauschecker 2009), suggesting that the second mechanism may play a role in the generation of highly replicable spike patterns. Furthermore, if neurons can replicate spiking patterns on a 2-ms temporal scale, they should also be capable of locking to stimulus modulations on the order of 500 Hz. (One can conceive of a neuron capable of firing to some stimulus features with very high temporal accuracy but only with relatively long inter-spike intervals, and hence incapable of locking to modulations at 1/accuracy Hz. However, such a neuron would not necessarily be able to produce *replicable* spike patterns to repetitions of a stimulus, as in some cases the first potentially spike-evoking feature of the stimulus could occur during the inter-spike interval resulting in a different spike pattern.) Such capability has not previously been demonstrated in the auditory cortex, let alone in a majority of neurons (Malone et al. 2007; Bendor and Wang 2008; Oshurkova et al. 2008; Scott et al. 2011). Finally, when various other authors utilized linear discriminator or similar methods employing the concept of bin/window, the best window has never been shown to be as short as 2 ms. In our hands, linear discriminator performance peaked at around 10-50 ms bin size (Kuśmierek and Rauschecker 2009; Kuśmierek and Rauschecker in preparation, see also below). Malone et al.'s (2007) discriminators performed best at 5-10 ms, while Schnupp et al. (2006) showed that the mutual information for their "highly informative" units in ferret A1 reached the maximum at 10-20 ms and declined above 40 ms and below 5 ms. In the prefrontal cortex, Averbeck and Romanski (2006) obtained the best classification performance using a hidden Markov model at 60-ms bin size. Interestingly, Russ et al. (2008) also

used another measure than the linear discriminator, i.e., mutual information, and this measure's performance appeared to peak around 20 ms.

Given all of the above, it is problematic to consider 80-90% performance of a linear discriminator at 2-ms bin width in prefrontal cortex (Russ et al. 2008) or in certain areas of auditory cortex (Recanzone 2008) a reliable result. It is difficult to determine the cause of such an outcome without knowing the exact implementation of the linear discriminator algorithm used in these two studies. What may be significant, however, is that we have identified a step in the algorithm described by Russ et al. (2008) and Recanzone (2008) in their Methods sections which, if slightly modified from the published form, produces results resembling theirs. The modification is, in our opinion, of a kind that one could easily make by accident while streamlining the code and optimizing performance.

Recanzone (2008) described the first steps of the algorithm as follows: "For each neuron, the first step was to select one spike train from one trial, referred to as the test trial. A stimulus PSTH was then constructed using all 12 trials for the other 7 stimuli and the remaining 11 trials for that particular stimulus." We have found that if this procedure was exactly followed ("original discriminator"), the performance of the linear discriminator based on our data peaked at a 10-50 ms window size, with the best average performance dependent on stimulus type (likely, on the temporal structure available in the stimulus, Kuśmierek and

Rauschecker 2009), but remaining below 65% (

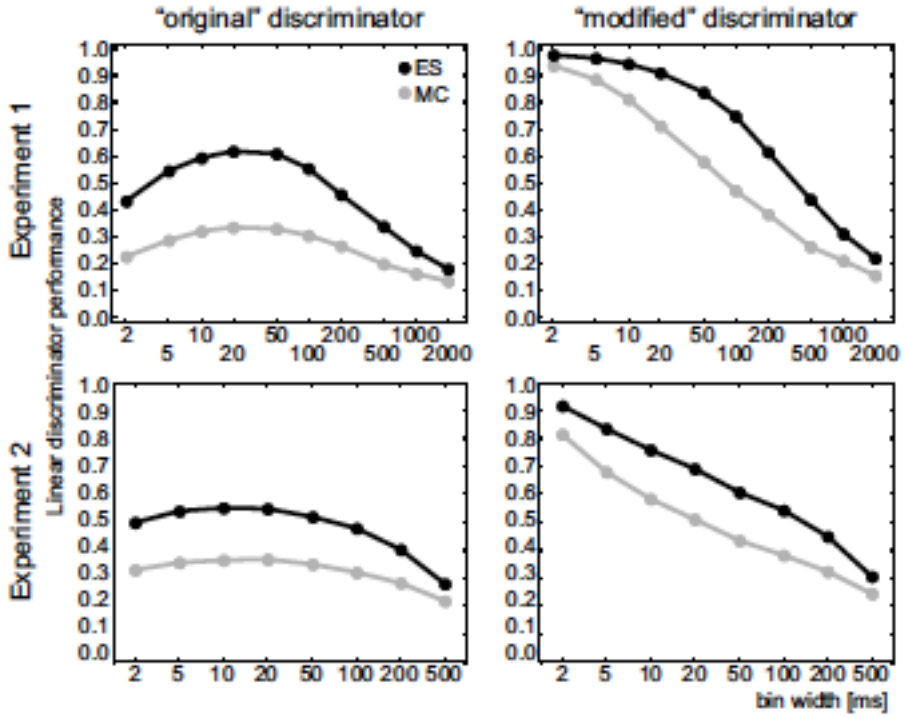


Figure **13**, Kuśmierek and Rauschecker 2009). However, if the "same" stimulus PSTH contained not only the *remaining* trials, but *all* trials for that particular stimulus, including the test trial, the performance of this "modified" linear discriminator peaked at >80% accuracy at the shortest bin width of 2 ms (

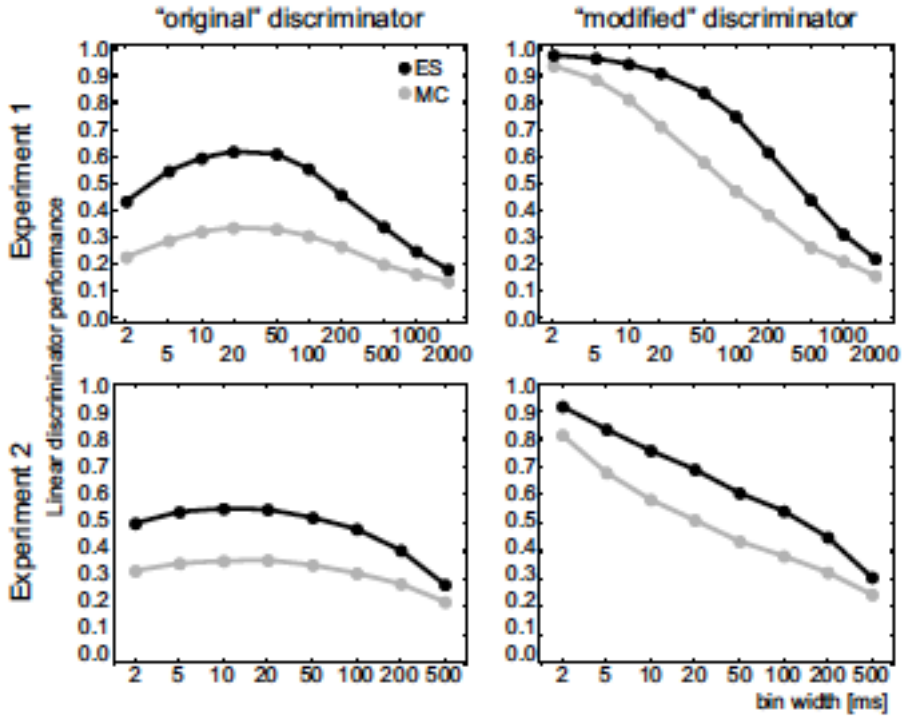


Figure **13**). The effect was similar whether data from Experiment 1 or 2 were used, thus it was little influenced by differences in stimuli, presentation manner, spike acquisition system, or cortical areas covered. Moreover, when the discriminator was applied to randomly generated time-stamps, the modification described above changed the result from expected chance performance at all windows to one that reached almost 100% performance at narrow bins (

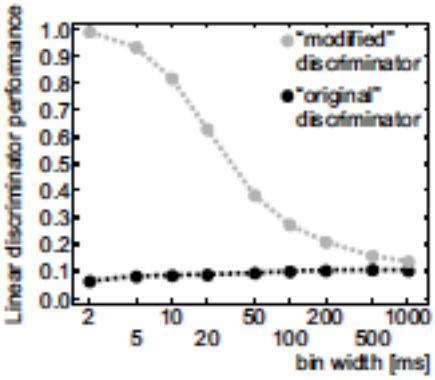


Figure **14**). The Matlab code which we used to perform both the “original” and “modified” linear discriminator analyses, as well as one used to generate and analyze the random time-stamps, are provided as Supplementary Material.

This analysis shows how a departure from the original, published linear discriminator algorithm resulted in altering the linear discriminator performance. It is possible that the unique results shown by Russ et al. (2008) and Recanzone (2008) might have stemmed from a similar departure from the published algorithm, or from another factor of comparable consequences. In the context of the present paper, we propose that the results of Recanzone (2008) should be reinterpreted as being consistent with increased stimulus selectivity in area R compared to other areas studied.

Summary

We have shown that specialization for sound-identity processing in the ventral stream can be found at its earliest stages, at the level of areas R and RM. The processing appears to develop in two stages: within 20-60 ms after stimulus onset, stimulus clustering quality based on R+RM responses cannot be distinguished from responses of more posterior early areas (A1, MM, CM), as well as clustering based on acoustic measures. Later on, it surpasses both.

A methodological achievement of the study is the demonstration of how the use of population responses analyzed in short temporal windows yields substantially more information than is available with more conventional methods.

Acknowledgements

We would like to express our gratitude to Dr. Hans Engler and Dr. Mark Chevillet for data analysis suggestions, to Mr. Michael Lawson and Ms. Carrie Silver for assistance with animal training and care, and to Dr. John VanMeter for help with MRI scanning.

Grant Support

This work was supported by grants to J. P. Rauschecker from National Institute of Neurological Disorders and Stroke (R01-NS-052494) and from National Science Foundation (OISE-0730255).

Disclosures

The authors declare no conflict of interest

References

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. "What" and "where" in the human auditory system. *Proc Natl Acad Sci USA* 98: 12301-12306, 2001.

Arnott SR, Binns MA, Grady CL, Alain C. Assessing the auditory dual pathway model in humans. *Neuroimage* 22: 401-408, 2004.

Averbeck BB, Romanski LM. Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. *J Neurosci* 26: 11023–11033, 2006.

Bendor D, Wang X. Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *J Neurophysiol* 100: 888–906, 2008.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10: 512-528, 2000.

Bizley JK, Walker KM. Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. *J Neurosci* 29: 3011-3013, 2009.

Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proc* 17: 97-110, 1993.

Chevillet M, Riesenhuber M, Rauschecker JP. Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *J Neurosci* 31: 9345-9352, 2011.

Cohen YE, Russ BE, Davis SJ, Baker AE, Ackelson AL, Nitecki R. A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proc Natl Acad Sci USA* 106: 20045-2005, 2009.

Cohen YE, Theunissen F, Russ BE, Gill P. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol* 97: 1470-1484, 2007.

Connor CE, Brincat SL, Pasupathy A. Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17:140-147, 2007.

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051-2062, 1984.

Jackson LL, Heffner RS, Heffner HE. Free-field audiogram of the Japanese macaque (*Macaca fuscata*). *J Acoust Soc Am* 106: 3017-3023, 1999.

Kajikawa Y, Camalier CR, de la Mothe LA, D'Angelo WR, Sterbing-D'Angelo SJ, Hackett TA. Auditory cortical tuning to band-pass noise in primate A1 and CM: A comparison to pure tones. *Neurosci Res* 70: 401-407, 2011.

Kajikawa Y, de la Mothe L, Blumell S, Hackett TA. A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. *J Neurophysiol* 93: 22-34, 2005.

Kiani R, Esteky H, Mirpour K, Tanaka K. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97: 4296-4309, 2007.

Kikuchi Y, Horwitz B, Mishkin M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci* 30: 13021-13030, 2010.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126-1141, 2008.

Kuśmierek P, Rauschecker JP. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606-1622, 2009.

Leaver A, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30: 7604-7612, 2010.

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S. Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14: 802-816, 2001.

Malone BJ, Scott BH, Semple MN. Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* 98: 1451-1474, 2007.

Miller LM, Recanzone GH. Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc Natl Acad Sci USA* 106: 5931-5935, 2009.

Mishkin M, Ungerleider LG, Macko KA. Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 6: 414-417, 1983.

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

Petkov CI, Kayser C, Augath M, Logothetis NK. Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4: 1213-1226, 2006.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nat Neurosci* 11: 367-374, 2008.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427: 448-451, 2004.

Rauschecker JP. Cortical processing of complex sounds. *Curr Opin Neurobiol* 8: 516-521, 1998.

Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res* 271: 16-25, 2011.

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12: 718-724, 2009.

Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 97: 11800–11806, 2000.

Rauschecker JP, Tian B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578-2589, 2004.

Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114, 1995.

Rauschecker JP, Tian B, Pons T, Mishkin M. Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* 382: 89-103, 1997.

Recanzone GH. Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *J Neurosci* 28: 13184-13193, 2008.

Recanzone GH. Perception of auditory signals. *Ann N Y Acad Sci* 1224: 96-108, 2011.

Recanzone GH, Engle JR, Juarez-Salinas DL. Spatial and temporal processing of single auditory cortical neurons and populations of neurons in the macaque monkey. *Hear Res* 271: 115-122, 2010.

Recanzone GH, Guard DC, Phan ML. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkeys. *J Neurophysiol* 83: 2315-2331, 2000a.

Recanzone GH, Guard DC, Phan ML, Su TK. Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol* 83: 2723-2739, 2000b.

Romanski LM, Averbeck BB. The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci* 32: 315-346, 2009.

Romanski LM, Averbeck BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol* 93: 734-747, 2005.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2: 1131-1136, 1999.

Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol* 99: 87-95, 2008.

Scott BH, Malone BJ, Semple MN. Transformation of temporal processing across auditory cortex of awake macaques. *J Neurophysiol* 105: 712-730, 2011.

Schnupp JWH, Hall TM, Kokelar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26: 4785–4795, 2006.

Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114: 3394-3411, 2003.

Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502: 894–923, 2007.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290-293, 2001.

Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH. Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory

cortex of the alert macaque monkey. *J Neurophysiol* 96: 3323-3337, 2006.

Figure Legends

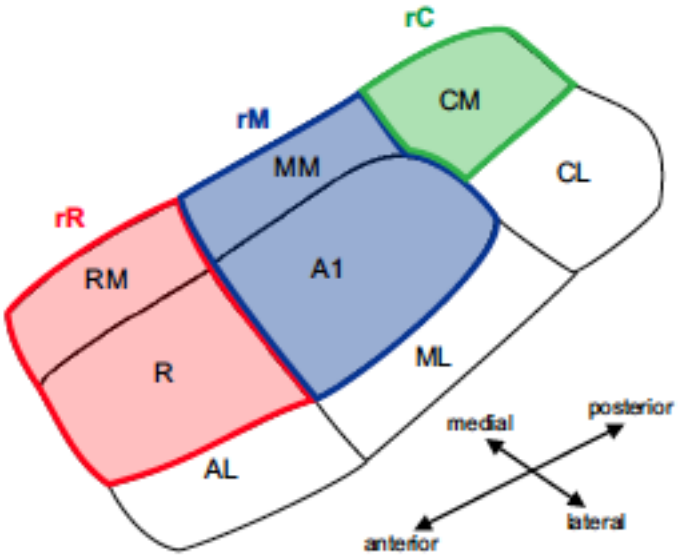


Figure 1. Schematic representation of anatomical areas of rhesus monkey auditory cortex. Medial belt areas: RM, rostromedial area; MM middle medial area; CM, caudomedial area. Core areas: R, rostral area; A1 primary auditory area. Lateral belt areas: AL, anterolateral area; ML, middle lateral area; CL, caudolateral area. Overlaid in color are regions rR, rM and rC, distinguished for the purpose of the present paper.

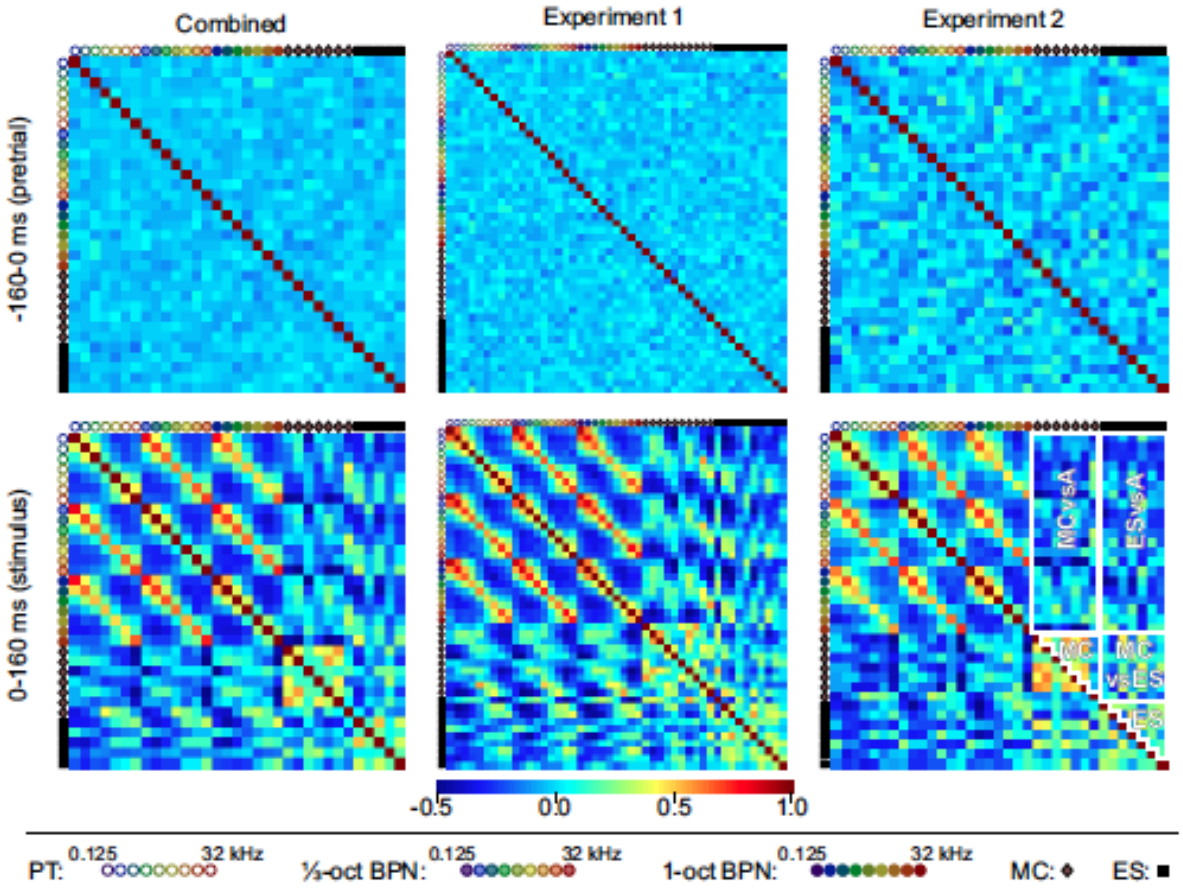


Figure 2. Similarity matrices showing correlations between population responses to PT, BPN, MC and ES stimuli. Left column: data combined from Experiments 1 and 2; middle column: Experiment 1 only; right column: Experiment 2 only. Top row: pretrial data from 160 ms preceding stimulus onset; bottom row: data from the first 160 ms of stimulus. Stimuli are coded by symbols along matrices' edges, explanation of symbols at the bottom of the figure. Correlation coefficient (*r*) value color-coded in the -0.5..1 range.

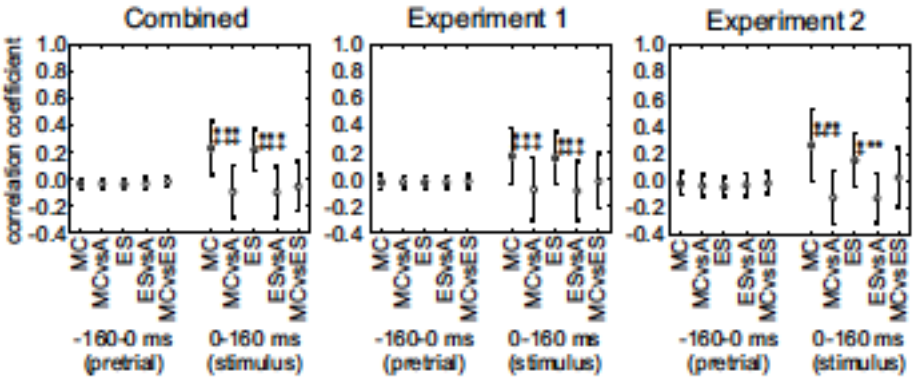


Figure 3. Quantification of correlations between population responses to monkey calls (MC) and environmental sounds (ES), for all cortical regions combined. **A**-**C**. Mean within-class correlation coefficients (+/- SD) within a stimulus class are shown with filled symbols separately for MC and ES. Empty symbols show between-class correlations, MC/ESvsA: between MC or ES and artificial sounds (i.e., PT and BPN, marked as A); MCvsES: between MC and ES. Asterisks denote significant differences between within-class correlations and the correlation of the same class with artificial sounds, e.g. between MC and MCvsA. Plus signs denote significant differences between within-class correlations and the MCvsES correlation. One symbol: $p<0.05$, two symbols: $p<0.01$, three symbols: $p<0.001$, t-test. Left side of each panel shows “pretrial” data, that is, data from 160 ms preceding the stimulus onset. Right side shows “stimulus” data, i.e., data from the first 160 ms after stimulus onset. **D**. Illustration of the relationship between a similarity matrix and values used in **A**-**C**. Stimuli coded by symbols, see Figure 2.

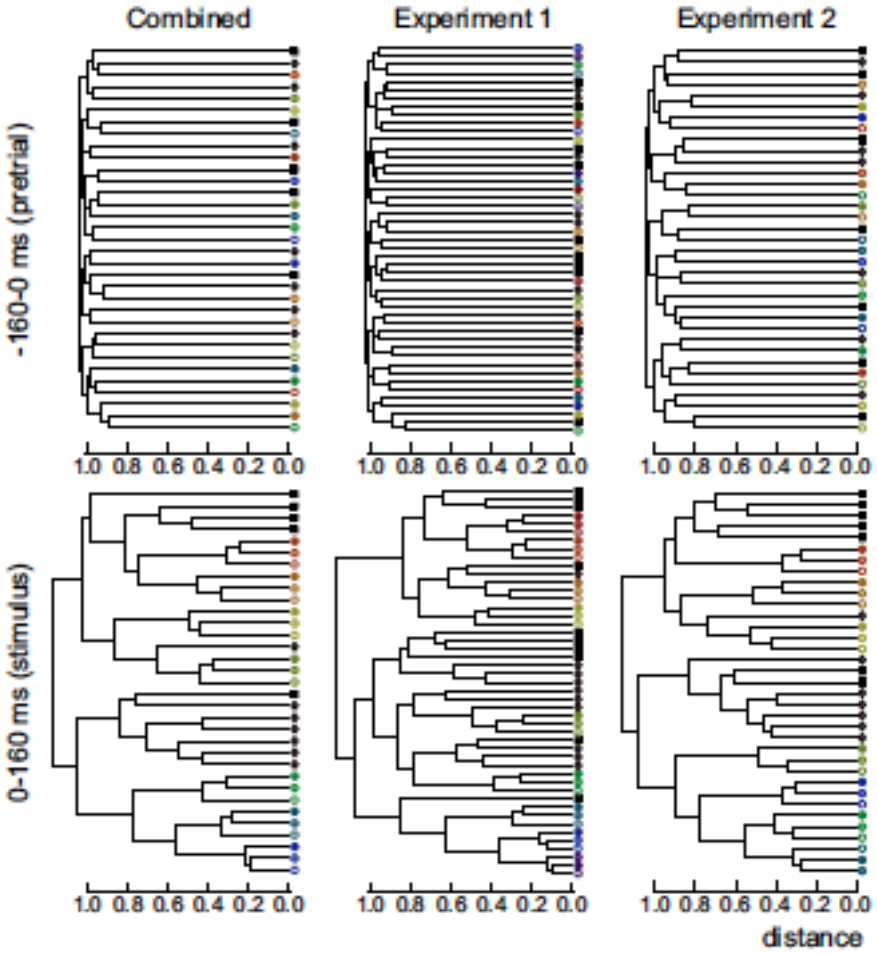


Figure 4. Dendrograms obtained from hierarchical clustering of the population responses to auditory stimuli. Stimuli coded by symbols, see Figure 2. Top row: pretrial data from 160 ms preceding stimulus onset; bottom row: stimulus data from the first 160 ms of the stimulus. Left column: data combined from Experiment 1 and 2; middle column: Experiment 1 only; right column: Experiment 2 only.

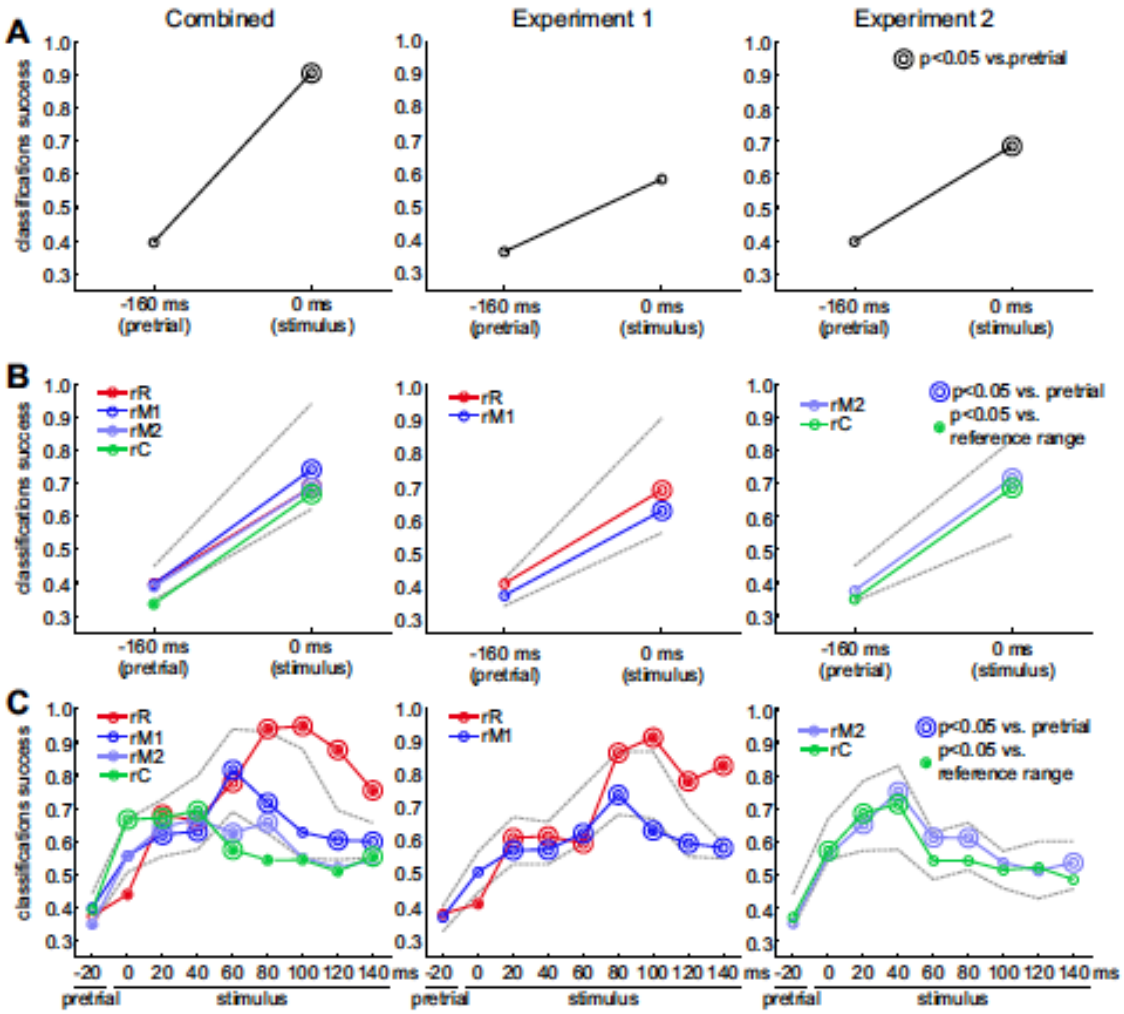


Figure 5. Classification of all sound stimuli based on neural population response. Mean classification success (proportion of correct classifications) with k-means clustering, *k*=4. **A**. Data from brain regions pooled: rR and rM1 for Experiment 1, rM1 and rC for Experiment 2, and all four regions for Combined. Within each panel, pretrial data from 160 ms preceding the stimulus onset are shown on the left, stimulus data from the first 160 ms of stimulus are shown on the right. **B**. Data from each brain region (rR, rM1, rM2, rC) plotted separately. Again, "pretrial" data from 160 ms before the onset on the left, "stimulus" data from the first 160 ms of stimulus on the right. **C**. Data from each brain region (rR, rM1, rM2, rC) plotted separately. Within each panel, "pretrial" data from 20 ms preceding the stimulus onset are shown on the left, followed by "stimulus" data from eight consecutive 20-ms windows covering 0-160 ms of stimulus. Double circles denote classification success that was significantly higher ($p \leq 0.05$) than pretrial classification success. (Note that this statistical test is

based on mode classification success whereas mean classification success is plotted in the figure). Filled circles represent stimulus classification success outside reference range (p≤0.05, see Methods). Grey broken lines show boundaries of the reference range for p≤0.05. For plots showing classification with *k*=5, see

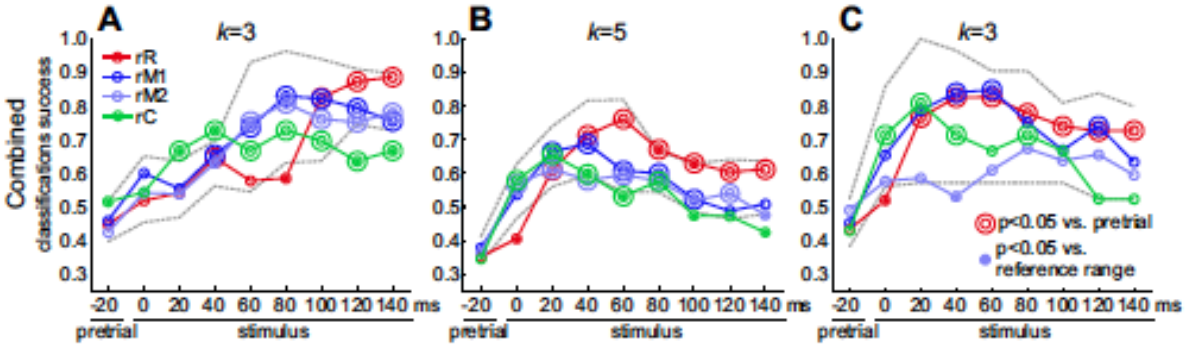


Figure **6**. The values on the abscissa denote onset of analysis window.

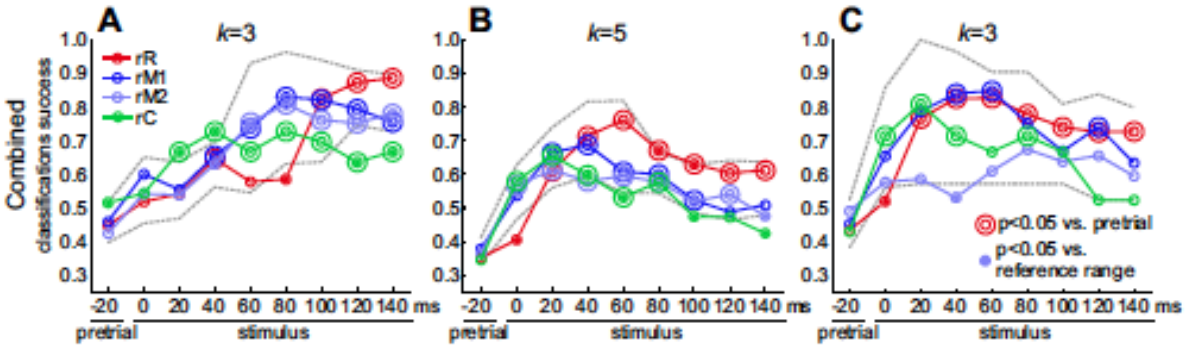


Figure 6. Classification of all sound stimuli based on neural population response. Mean classification success (proportion of correct classifications) with k-means clustering, *k*=5. **A**. Data from brain regions pooled: rR and rM1 for Experiment 1, rM1 and rC for Experiment 2, and all four regions for Combined. Within each panel, pretrial data from 160 ms preceding the stimulus onset are shown on the left, stimulus data from the first 160 ms of stimulus are shown on the right. **B**. Data from each brain region (rR, rM1, rM2, rC) plotted separately. **C**. Data from each brain region (rR, rM1, rM2, rC) plotted separately. Within each panel, “pretrial” data from 20 ms preceding the stimulus onset are shown

on the left, followed by “stimulus” data from eight consecutive 20-ms windows covering 0-160 ms of stimulus. For detailed legend and respective data obtained with *k*=4, see

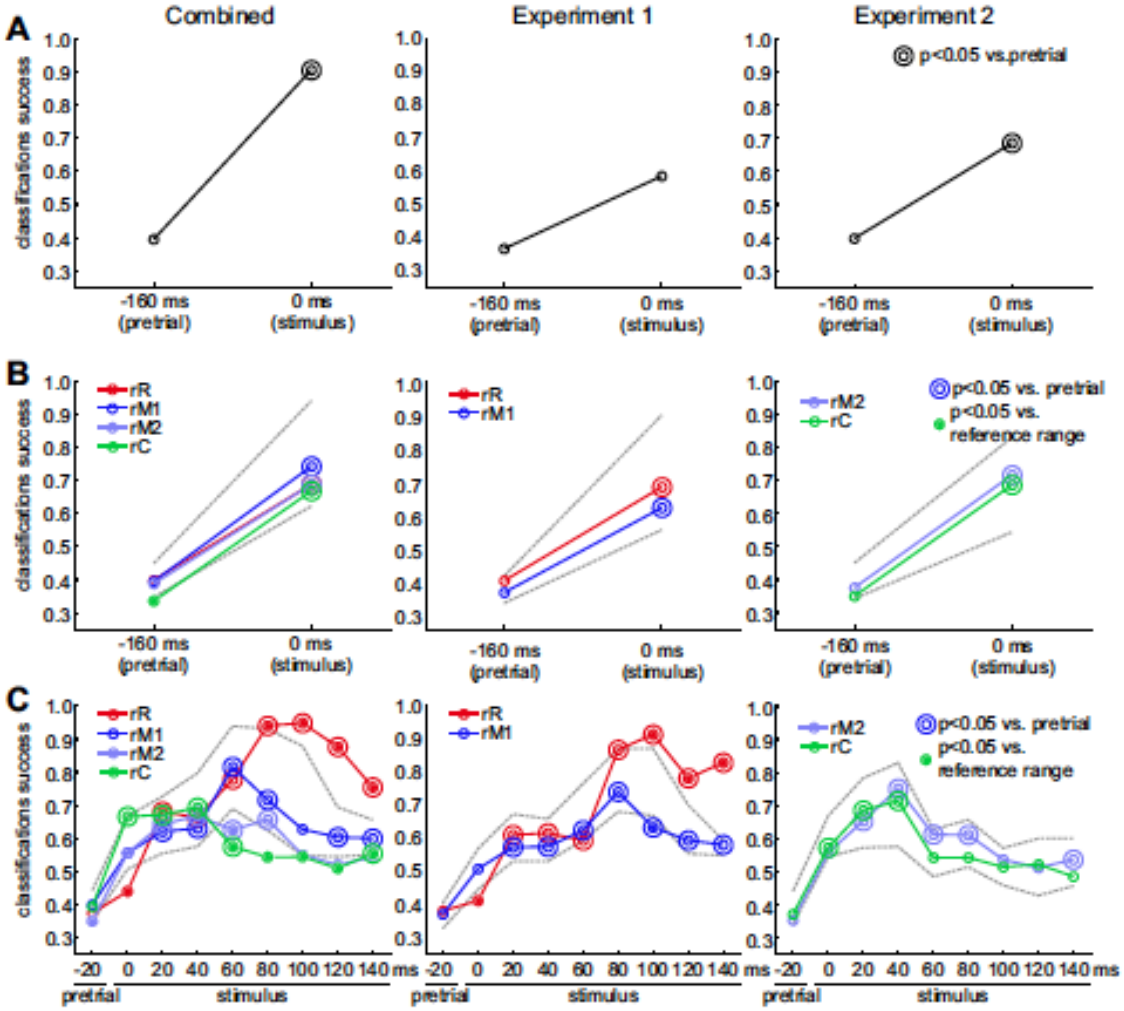


Figure **5**.

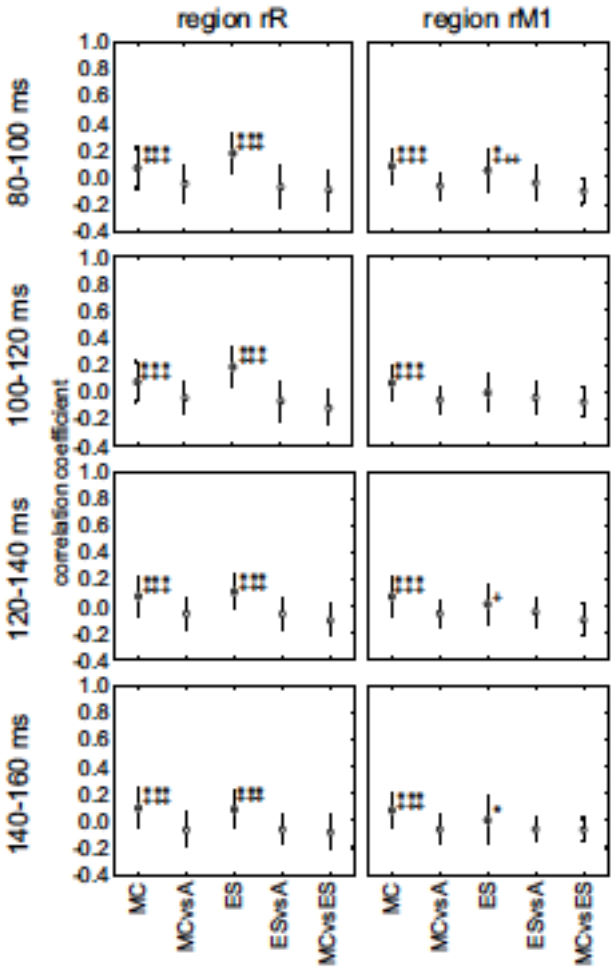
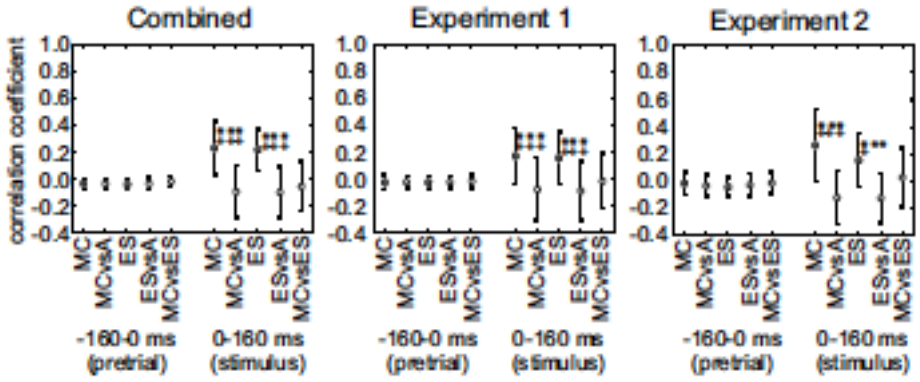


Figure 7. Quantification of correlations between population responses to monkey calls (MC) and environmental sounds (ES) for region rR (left) and rM1 (right), in four temporal windows from 80 to 160 ms after stimulus onset (rows). Data from Combined experiments. Conventions and explanations of symbols:



see

Figure 3A-C.

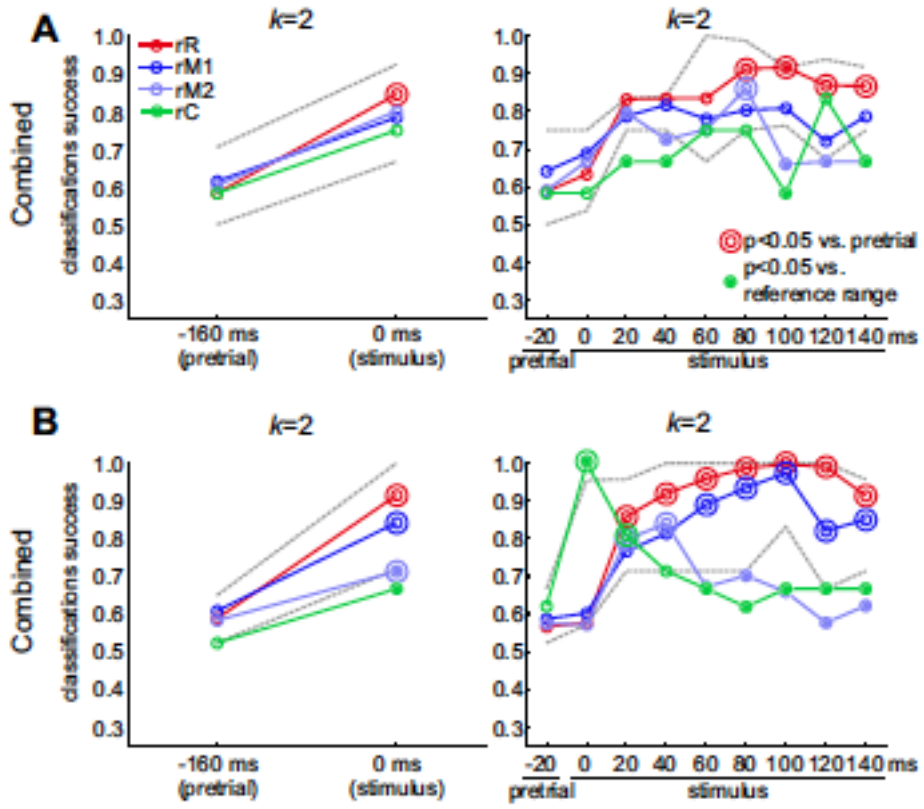


Figure 8. Classification of sound stimuli based on neural population responses to natural stimuli. Mean classification success (proportion of correct classifications) with k-means clustering, for MC and ES stimuli only (*k*=2). Data from each brain region (rR, rM1, rM2, rC) presented separately. Left panel: pretrial data from 160 ms preceding the stimulus onset on the left, stimulus data from the first 160 ms of stimulus on the right. Right panel: pretrial data from 20 ms preceding the stimulus onset are shown on the left, followed by stimulus data from eight consecutive 20-ms windows covering 0-160 ms of stimulus. For

detailed legend, see

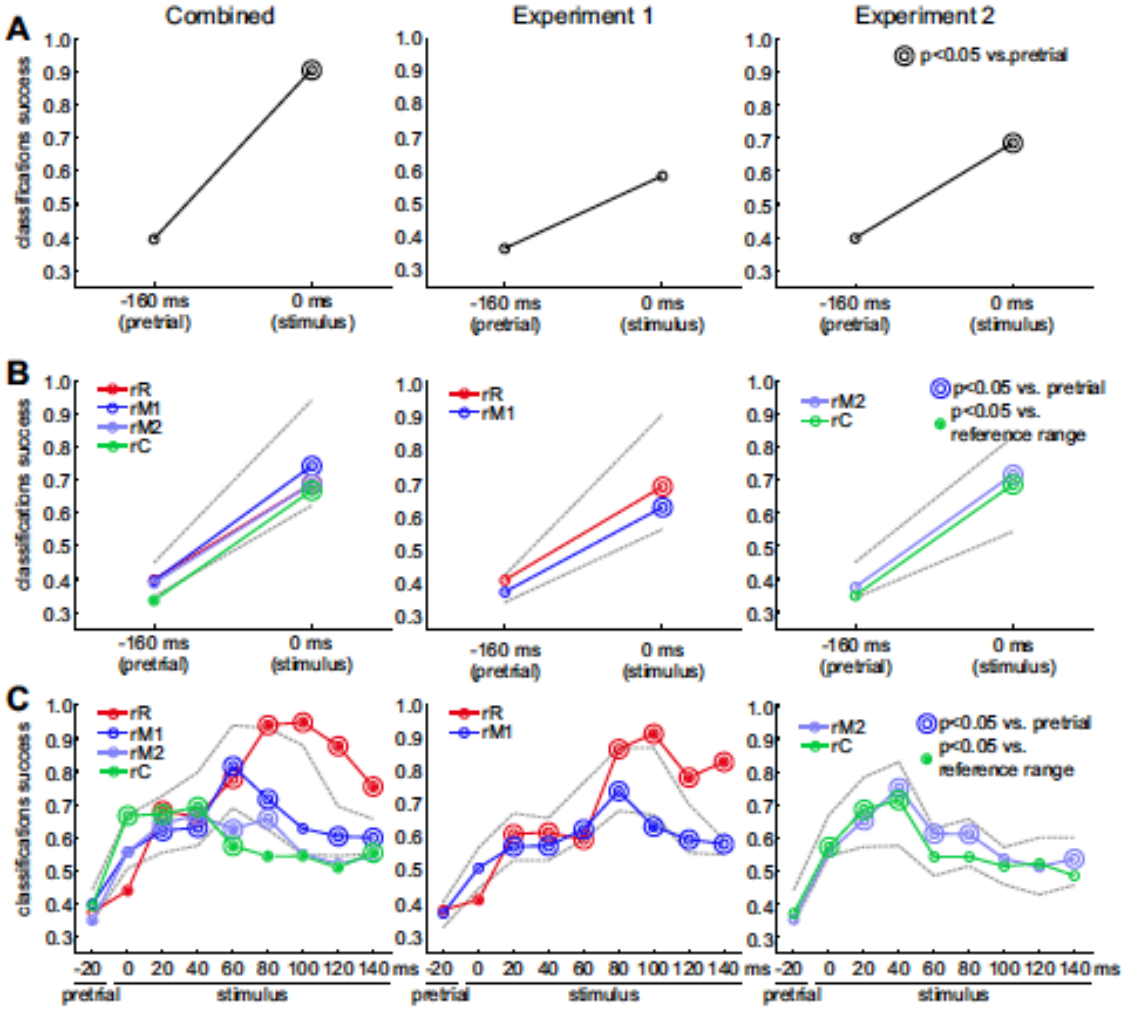


Figure **5**.

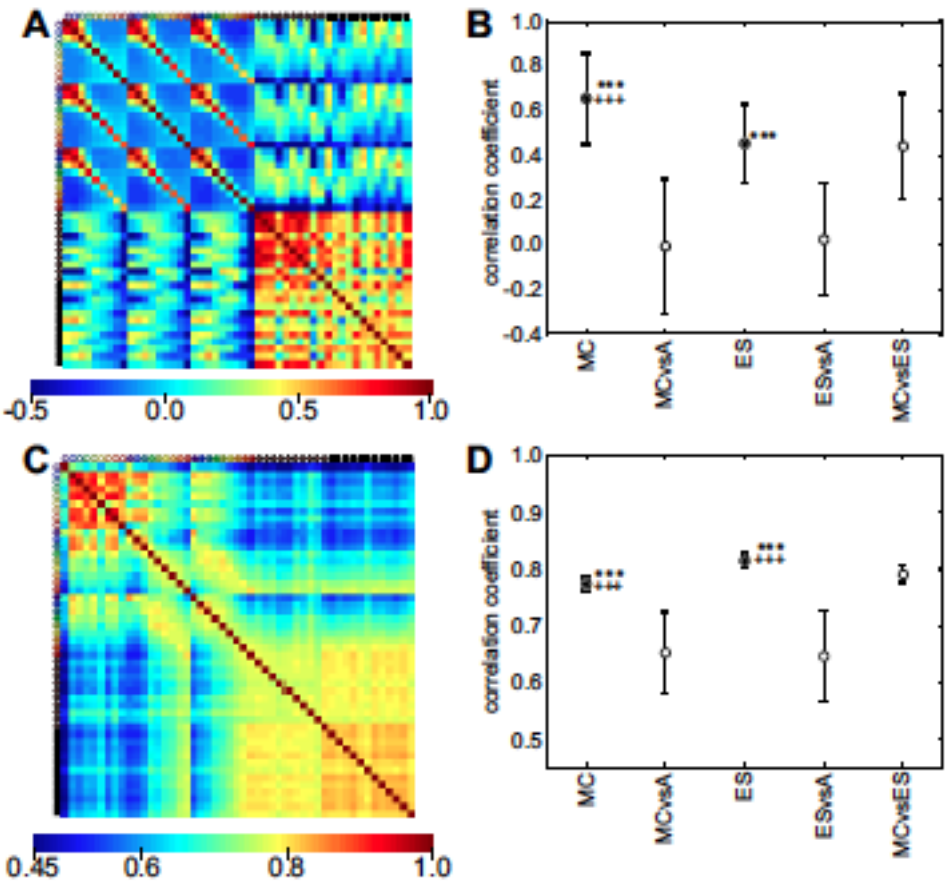


Figure 9. **A.** Similarity matrix showing correlations between log-frequency spectrograms of all stimuli. **B**. Quantification of correlations between log-frequency spectrograms of monkey calls (MC) and environmental sounds (ES). **C**. Similarity matrix showing correlations between modulation spectra of all stimuli. **D**. Quantification of correlations between modulation spectra of monkey calls (MC) and environmental sounds (ES). Note that while in **D** the mean correlation coefficient within the MC class (MC) differs significantly from the mean correlation coefficient between MC and ES classes (MCvsES), the value

for MC is lower than for MCvsES. Conventions same as in Figure 2 and

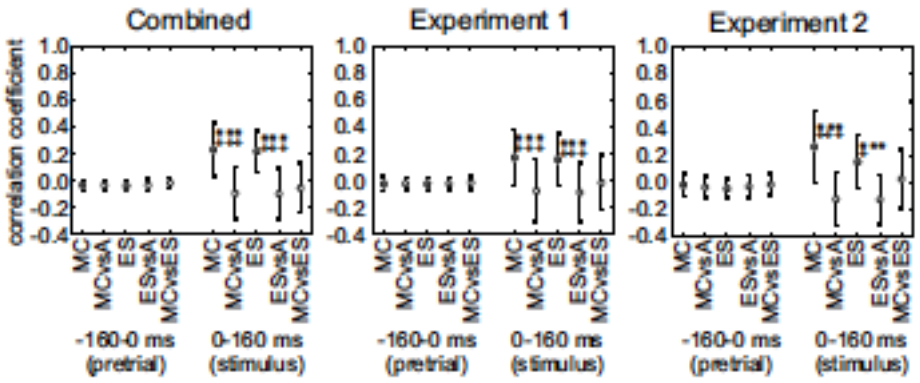


Figure 3 except for scale in **C** and **D**.

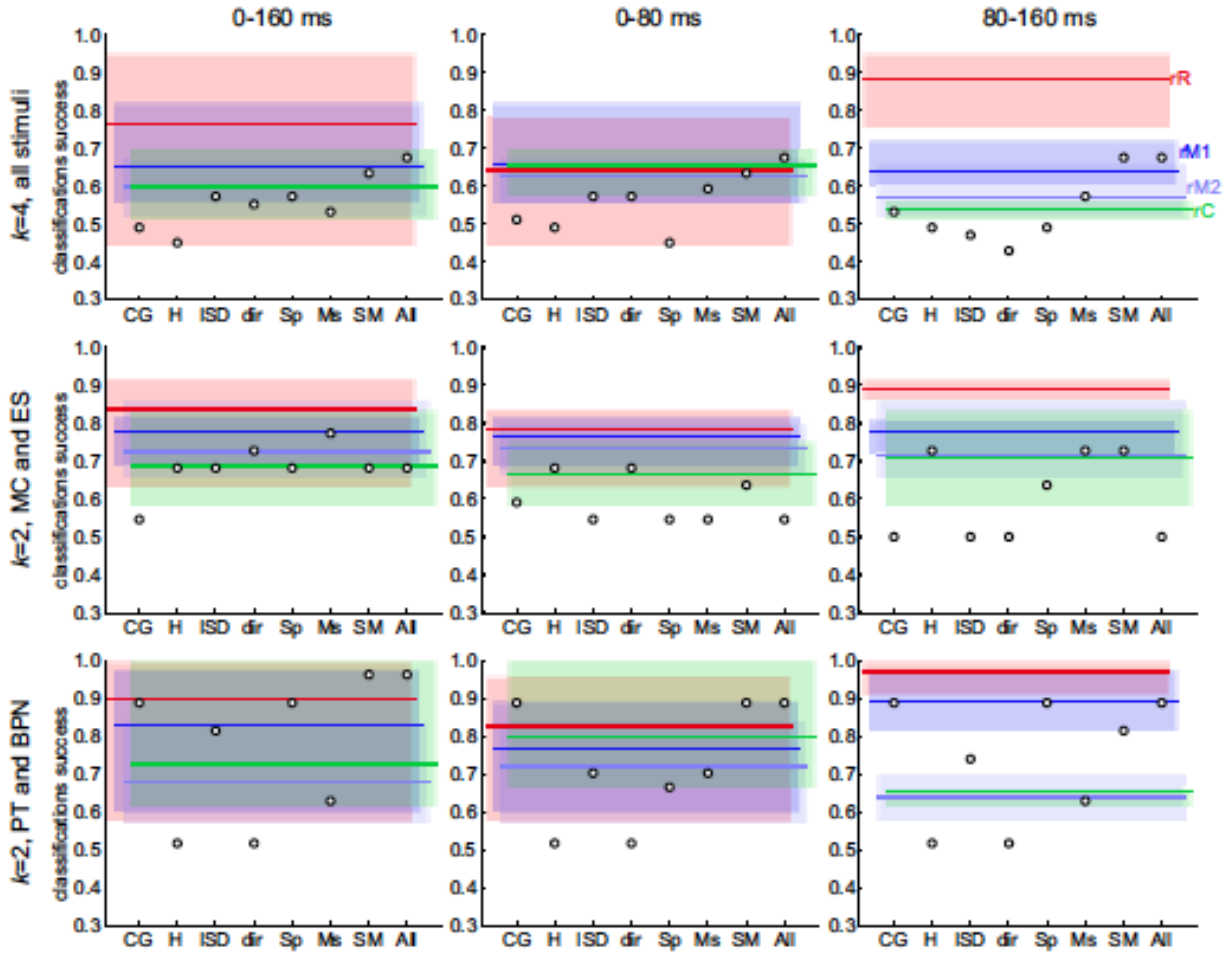


Figure 10. Classification of sound stimuli based on acoustical parameters (black circles), compared to classification based on neural responses (color lines and shading). Classification success (proportion of correct classifications) with k-

means clustering. Left: all stimuli, *k*=4 clusters; middle: all stimuli, *k*=5 clusters, right: natural (MC and ES) stimuli, *k*=2 clusters. Abbreviations represent acoustic parameters used as the basis for clustering: CG, spectrum center of gravity; H, mean harmonicity; ISD, standard deviation of intensity; dir, all direct parameters (CG, H, ISD); Sp, three parameters derived from dissimilarity of log-frequency spectrograms; Ms, three parameters derived from dissimilarity of modulation spectra; SM, all six parameters derived from dissimilarity of spectrograms and modulation spectra; All, all nine parameters. For reference, respective ranges (shaded areas) and means (lines) of mean classification success values based on neural data from the 80-160 ms range are shown in color. Red, region rR; darker blue, region rM1; paler blue, region rM2; green, region rC. Neural data from the Combined experiment, see

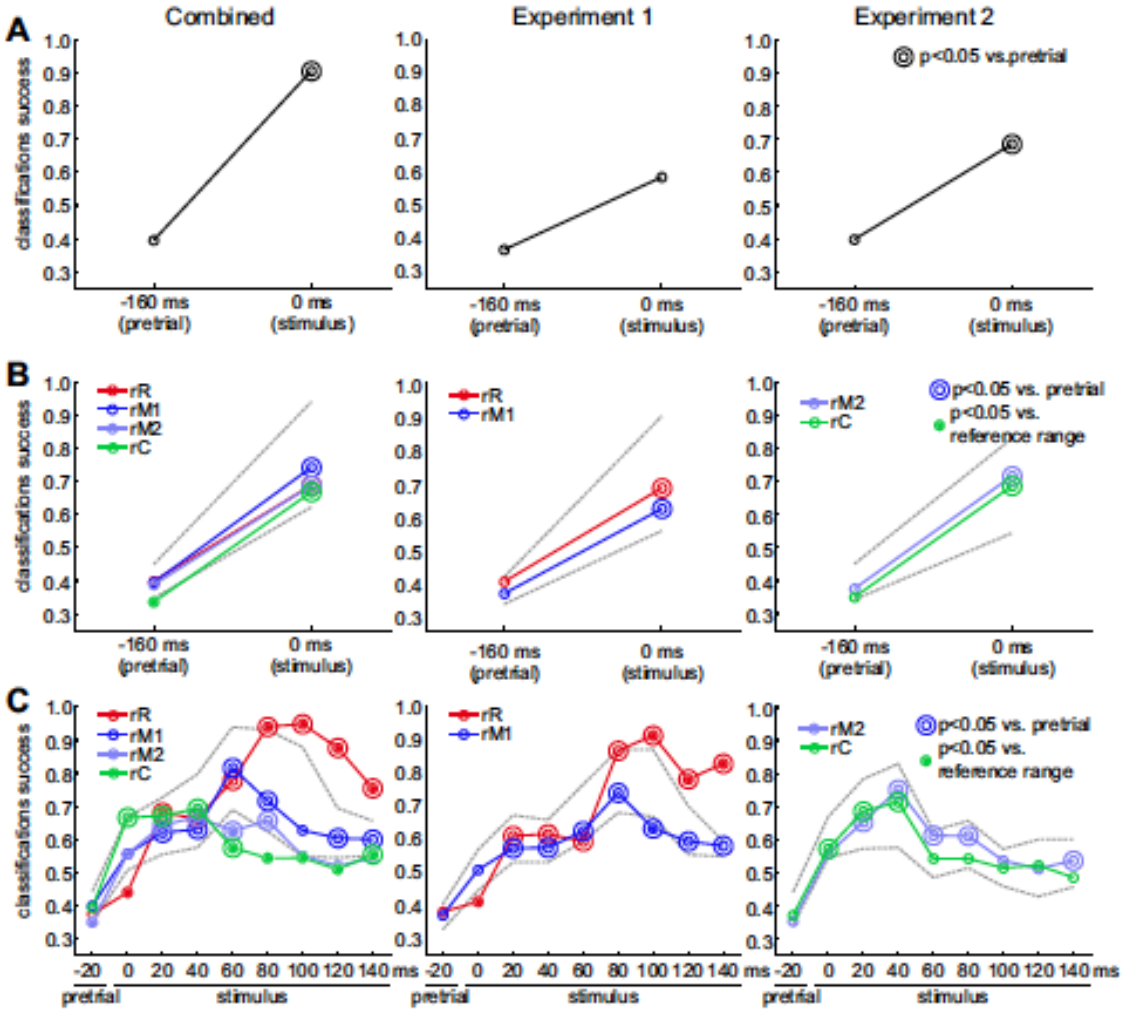


Figure **5**,

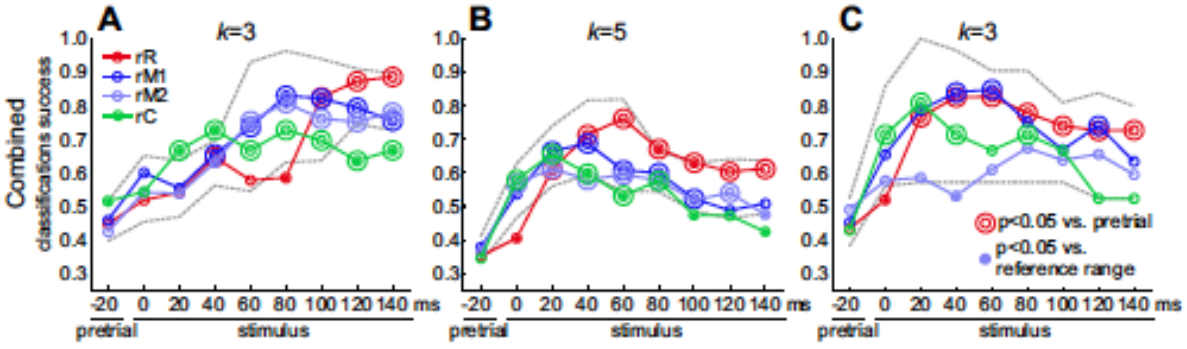


Figure 6, and

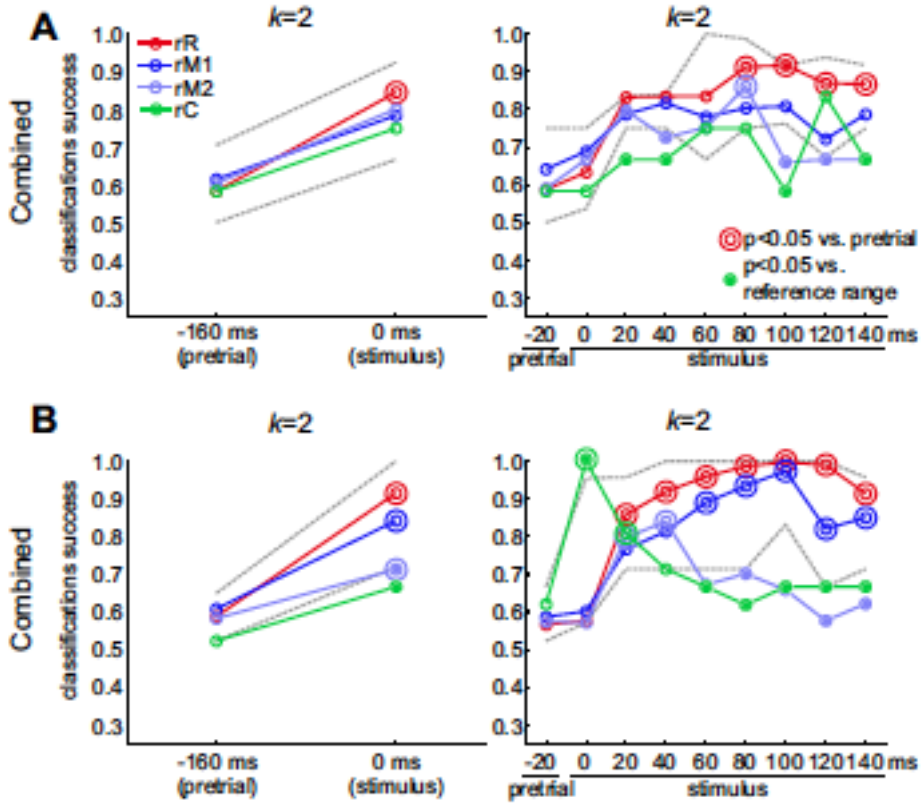


Figure 8 for source data. The shaded areas are shifted along the abscissa relative to each other to improve readability.

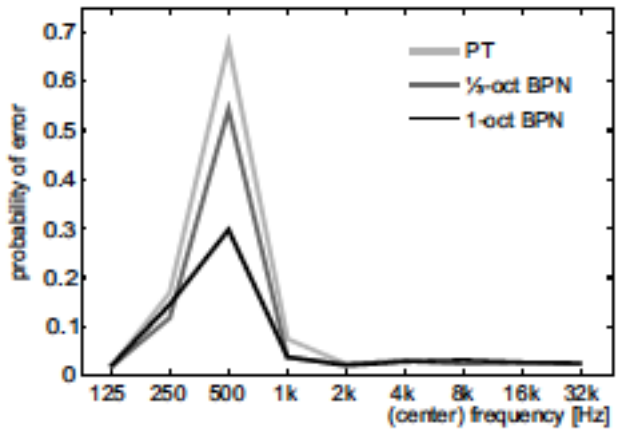


Figure 11. Effect of frequency and bandwidth on probability of false alarm error to pure tones (PT) or band-pass noise bursts (BPN) in monkeys L and S (Experiment 1). The rewarded stimulus began with a 523 Hz tone. Bandwidth errors are more likely than frequency errors.

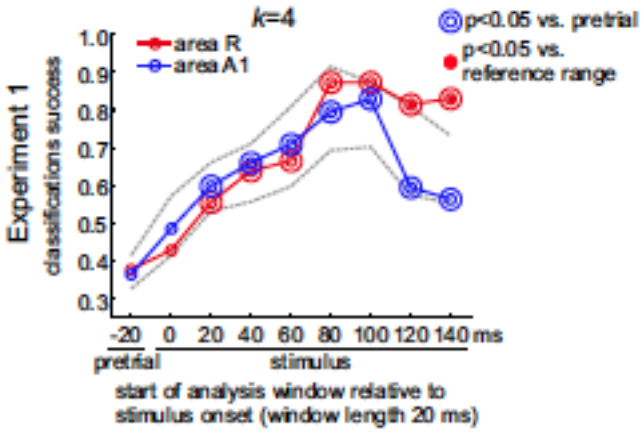


Figure 12. Classification of all sound stimuli based on neural population responses for core areas R (subset of region rR) and A1 (subset of region rM1)

into four clusters (k=4) as a function of time window in Experiment 1. Cf.

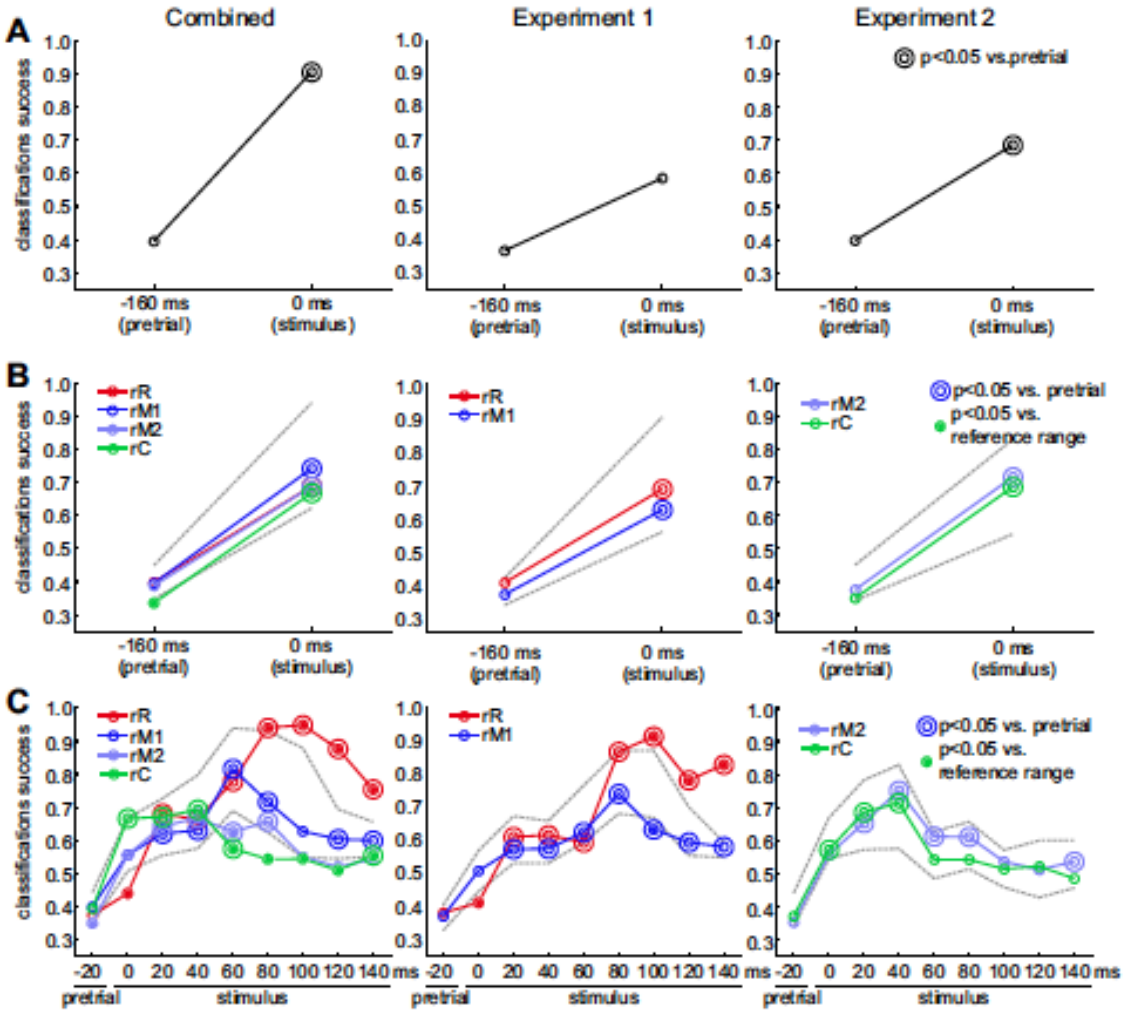


Figure 5, bottom center panel.

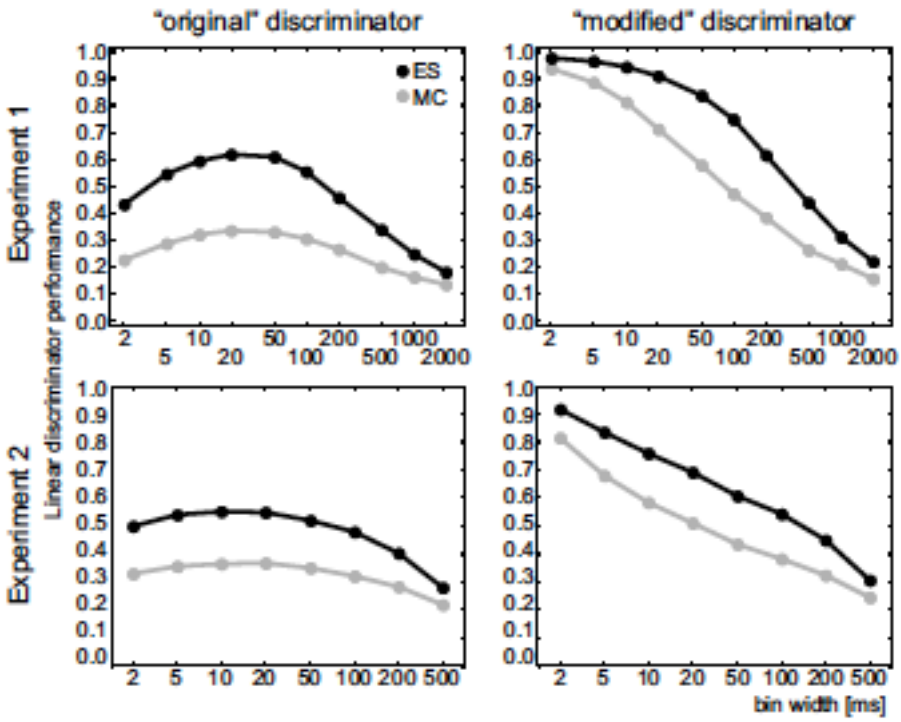


Figure 13. Performance of the linear discriminator for environmental sounds (ES) and monkey calls (MC) plotted against analysis bin width. In Experiment 1 (regions rR and rM1 pooled), there were 10 stimuli in each class, hence the chance level was 0.1. In Experiment 2 (regions rC and rM2 pooled), there were 7 stimuli in each class, for a chance level of 0.143. Left: “original” discriminator according to descriptions in Recanzone (2008), Russ et al. (2008), and Kuśmierek and Rauschecker (2009). Right, “modified” discriminator, as described in the Discussion of the present paper.

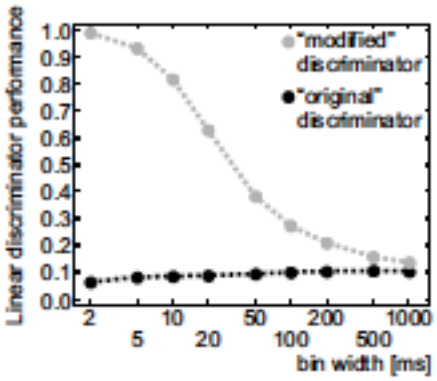


Figure 14. Performance of the “original” and “modified” linear discriminator for random time-stamps, plotted against analysis bin size. For 100 “neurons”, time-stamps were generated as 10-50 values in the 0-1000 ms range, for 10 “stimuli” and 12 “trials” per “stimulus”. Both the number of time-stamps per “trial” and the value of the time-stamps were randomly and uniformly distributed.

7.2 Functional MRI of the vocalization-processing networks in the macaque brain

Functional MRI of the vocalization-processing network in the macaque brain

Michael Ortiz-Rios[1, 2, 6*], Paweł Kuśmierek[1], Iain DeWitt[1], Denis Archakov[1, 3], Frederico A.C. Azevedo[2, 6], Mikko Sams[3], Iiro P. Jääskeläinen[3], Georgios A. Keliris[2, 4], Josef P. Rauschecker[1, 3, 5]

[1]Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA

[2] Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[3] Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Aalto, Finland

[4] Bernstein Centre for Computational Neuroscience, Tübingen, Germany

[5] Institute for Advanced Study and Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, München, Germany

[6]IMPRS for Cognitive and Systems Neuroscience, Tübingen, Germany

[7]Department of Biomedical Sciences, University of Antwerp, Wilrijk, Belgium

Running title: fMRI of the macaque vocalization-processing network

***Correspondence**: Michael Ortiz-Rios and Josef P. Rauschecker, Department of Neuroscience, Georgetown University Medical Center, NRB WP19, 3970 Reservoir Rd NW, Washington, DC 20057, USA

Email: rauschej@georgetown.edu, phone: +1 202 687 8842, fax +1 202 687 0617

Abstract

Using functional magnetic resonance imaging in awake behaving monkeys we investigated how species-specific vocalizations are represented in auditory and auditory-related regions of the macaque brain. We found clusters of active voxels along the ascending auditory pathway that responded to various types of complex sounds: inferior colliculus (IC), medial geniculate nucleus (MGN), auditory core, belt, and parabelt cortex, and other parts of the superior temporal gyrus (STG) and sulcus (STS). Regions sensitive to monkey calls were most prevalent in the anterior STG, but some clusters were also found in frontal and parietal cortex on the basis of comparisons between responses to calls and environmental sounds. Surprisingly, we found that spectrotemporal control sounds derived from the monkey calls ("scrambled calls") also activated the parietal and frontal regions. Taken together, our results demonstrate that species-specific vocalizations in rhesus monkeys activate preferentially the auditory ventral stream, and in particular areas of the antero-lateral belt and parabelt.

Keywords: Auditory cortex, monkey, species-specific calls, spectrotemporal features, higher-level representations

Introduction

The concept of two streams in auditory cortical processing, analogous to that in visual cortex (Mishkin et al., 1983), was proposed more than a decade ago (Rauschecker, 1998a; Rauschecker and Tian, 2000). The concept was supported by contrasting patterns of anatomical connections in the macaque from anterior/ventral and posterior/dorsal belt regions of auditory cortex to segregated domains of lateral prefrontal cortex (Romanski et al., 1999) and by different physiological properties of these belt regions. In particular, the anterior lateral belt (area AL) in the macaque exhibited enhanced selectivity for the identity of sounds (monkey vocalizations), whereas the caudal lateral belt (area CL) was particularly selective to sound location (Tian et al., 2001; see also Kuśmierek and Rauschecker, 2014). Evidence for segregated streams of auditory cortical processing has also been provided in human studies (Maeder et al., 2001; Arnott et al., 2004; Ahveninen et al., 2006).

Use of species-specific vocalizations for auditory stimulation in the macaque is of particular interest in the context of the ongoing debate about the evolution of speech and language (Rauschecker, 2012; Bornkessel-Schlesewsky et al., 2015). Comparative approaches have focused on identifying the common neural networks involved in the processing of speech in humans and of vocalizations in non-human primates (Gil-da-Costa, 2004; Petrides and Pandya, 2009; Joly et al., 2012b; Frey et al., 2008, 2014). Monkey calls convey semantic information about objects and events in the environment as well as about affective states of individuals, similar to information contained in human communication sounds and speech (Cheney and Seyfarth, 1990; Ghazanfar and Hauser, 1999; Yovel and Belin, 2013). An open question regarding the vocalization-processing network in the macaque brain is whether it also carries information about the

motor actions necessary to produce the vocalizations, as has been shown in humans listening to speech and music (Wilson et al., 2004; Leaver et al., 2009).

Several studies have examined the representation of complex sounds, including vocalizations, in the macaque brain using neuroimaging techniques (Poremba et al., 2003; Petkov et al., 2008; Joly et al., 2012b). In particular, the first fMRI study by Petkov et al. (2008) found activation specific to monkey vocalizations in the anterior STG region. One of the aims in later studies has been to characterize the physiological properties of the anterior superior temporal (aSTG) region that shows sensitivity to higher-level spectrotemporal features in vocalizations (Russ et al., 2008; Kikuchi et al., 2010, 2014; Perrodin et al., 2011; Fukushima et al., 2014). A recent comparative study by Joly et al. (2012b) replicated and extended these results by analyzing fMRI images of the entire brain and found an involvement of orbitofrontal cortex in the processing of monkey vocalizations. Given that the ventral pathway continues into orbitofrontal and ventrolateral prefrontal cortex (vlPFC) (Barbas, 1993; Romanski et al., 1999; Cohen et al., 2007; Petkov et al., 2015), this finding is of particular interest.

In humans, the ventral auditory pathway is thought to be particularly involved in the recognition and identification of vocalizations as well as speech (Binder et al., 2000; DeWitt and Rauschecker, 2012). By contrast, the dorsal pathway is involved primarily in processing sound source location and motion in both humans and animals (Maeder et al., 2001; Tian et al., 2001; Arnott et al., 2004). However, a recent proposal, derived from both human and nonhuman primate studies, suggests that the dorsal stream may also play a role in sensorimotor integration and control of complex sounds, including speech (Rauschecker and Scott, 2009; Rauschecker, 2011). Thus, activation of frontal and parietal regions might also be expected when monkeys are presented with conspecific vocalization sounds.

Here we identified which brain regions of the macaque monkey are sensitive to conspecific vocalizations using whole-brain functional magnetic resonance imaging (fMRI). We found the most distinct activation in the anterior STG and along the auditory ventral stream, but some clusters of activation were also found in prefrontal, premotor and parietal cortex when comparing monkey vocalizations to environmental sounds. These findings are discussed in terms of their functional significance.

Material and Methods

Subjects

Two male rhesus monkeys *(Macaca mulatta*) weighing 10-12 kg participated in our awake-fMRI experiments. Each animal was implanted with an MRI-compatible headpost (Applied Prototype) secured to the skull with ceramic screws (Thomas Recording), plastic strips and bone cement (Osteobond, Zimmer). All surgical procedures were performed under general anesthesia with isoflurane (1-2%) following pre-anesthetic medication with ketamine (13 mg/kg) and midazolam (0.12 mg/kg). The experiments were approved by the Georgetown University Animal Care and Use Committee and conducted in accordance with standard NIH guidelines.

Behavioral Training

To ensure the monkeys attended to each stimulus for which a brain volume was acquired, we adapted a go/no-go auditory discrimination task (Kuśmierek and Rauschecker, 2009; Kikuchi et al., 2010) for sparse-sampling functional MRI.

First, each monkey was trained to lie in sphinx position in an MRI-compatible primate chair (Applied Prototype) placed inside a double-walled acoustic chamber simulating the scanner environment. Inside the chamber, the animals were trained to be accustomed to wearing headphone equipment and hearing (simulated) scanner noise, presented by a loudspeaker. Eye movements were monitored using an infrared eye-tracking system (ISCAN). Analog output of the tracker was sampled with an analogue-to-digital conversion device (National Instruments). A PC running Presentation software (Neurobehavioral Systems) was used to present visual and auditory stimuli, control the reward system, and trigger imaging data acquisition (see below).

After the animal completed the fixation training, a go/no-go auditory discrimination task was introduced, in which the monkeys could initiate a trial by holding fixation on a central red spot while a block of auditory stimuli would be simultaneously presented. After the first 6 seconds of auditory stimulation, a trigger was sent to the scanner, starting the acquisition of an image volume (**Figure 1B)**. Following acquisition and a random delay, the target sound (white noise) was presented, cueing a saccade to the left or to the right side as signaled at the beginning of each experimental session (**Figure 1A)**. To provide feedback, after the response window, a yellow spot was shown indicating the correct target location. Finally, contingent on performance, the animal received a juice reward. An inter-trial interval of at least 2 s was enforced before the next trial could be initiated by fixation. Every sound presentation trial was followed by a "silence" trial, allowing for measurement of baseline blood oxygen level dependent (BOLD) signal. Monkey 1 (M1) performed the task correctly for over 90% of the trials. Monkey 2 (M2) was not able to perform the saccadic go/no-go discrimination task with high accuracy and was therefore scanned while passively listening to the acoustic stimuli. To ensure stable attention, M2 was rewarded for successfully holding fixation throughout the trial.

Auditory stimuli

Three sound categories were used in the experiments: environmental sounds (Env), monkey vocalizations or calls (MC), and scrambled monkey calls (SMC). Spectrograms of example clips from each of these three categories are illustrated in **Figure 1C.** Environmental sounds were obtained from multiple online sources and from recordings made in our laboratory facilities (Kuśmierek and Rauschecker, 2009). They included the sounds of vehicles, cages, water, food containers, clocks, cameras, applause, coins, footsteps, chewing, heartbeats, horns, and telephones (n = 56). The mean duration of the Env stimuli was 1.14 s (range: 0.96 – 2.6 s). Monkey calls were obtained from recordings made outside our colony (M. Hauser and/or Laboratory of Neuropsychology [LN] library). Monkey vocalizations (n = 63) consisted of grunts, barks, warbles, coos and screams, as used in prior studies (Rauschecker et al., 1995; Tian et al., 2001, Kuśmierek et al., 2012). The mean duration of the vocalization stimuli was 0.67 s (range: 0.13 - 2.34 s). SMC were generated by randomly rearranging 200 ms by 1-octave tiles of the constant-Q spectrogram (Brown, 1991) for each monkey call and reconstructing a time-domain waveform with an inverse transform (Schörkhuber & Klapuri, 2010). Transposition along the time axis was not constrained while transposition along the frequency axis was restricted to displacement by a single octave. For each trial, a random selection of stimuli from one class (MC, Env, or SMC) was arranged sequentially into a smooth auditory clip that lasted for the duration of the trial (8 s).

Sounds were presented through modified electrostatic in-ear headphones (SRS-005S + SRM-252S, STAX), mounted on ear-mold impressions of each animal's pinna (Sarkey Eden Prairie) and covered with a custom-made earmuff system for sound attenuation.

To match loudness, the stimuli were played through the sound presentation system and re-recorded with a probe microphone (Brüel and Kjær, type 4182 SPL meter) inserted in the ear-mold of an anesthetized monkey. The recordings were then filtered with an inverted macaque audiogram (Jackson et al. 1999) to simulate the effect of different ear sensitivity at different frequencies, analogous to the dB(A) scale for humans. The stimuli were finally equalized so that they produced equal maximum root mean square (RMS) amplitude (using a 200-ms sliding window) in filtered recordings (Kuśmierek and Rauschecker, 2009). During experiments, all stimuli were amplified (Yamaha AX-496) and delivered at a calibrated RMS amplitude of ~80 dB SPL.

Analyses of sound categories

A modulation spectrum analysis (Singh and Theunissen, 2003) was performed for each sound with the STRFpak Matlab toolbox (http://strfpak.berkeley.edu). We obtained a spectrogram of each sound by decomposing it into frequency bands using a bank of Gaussian filters (244 bands, filter width = 125 Hz). The filters were evenly spaced on the frequency axis (64-48000 Hz) and separated from each other by one standard deviation. The decomposition resulted in a set of narrow-band signals, which were then cross-correlated with each other and themselves to yield a cross-correlation matrix. This matrix was calculated for time delays of +/-150 ms, and the two-dimensional Fourier transform of this matrix was calculated to obtain the modulation spectrum of each sound (**Figure 1D**).

Data acquisition

Images were acquired with a horizontal MAGNETOM Trio 3-T scanner (Siemens) with a 60-cm bore diameter. A 12-cm custom-made saddle shape radiofrequency coil (Windmiller Kolster Scientific) covered the entire brain and was optimized for imaging

the temporal lobe. The time series consisted of gradient-echo echo-planar (GE-EPI) whole-brain images obtained in a sparse acquisition design. Sparse sampling allows single volumes to be recorded coincidentally with the predicted peak of the evoked hemodynamic response (Hall et al., 1999). This helps to avoid contamination of the measured stimulus-specific BOLD response by the scanner-noise-evoked BOLD response. Further, by triggering acquisition 6 s after stimulus onset, the auditory stimulus was presented without acoustic interference from gradient-switching noise, typical of a continuous fMRI design. For the functional data, individual volumes with 25 ordinal slices were acquired with an interleaved single-shot GE-EPI sequence (TE = 34 ms, TA = 2.18 s, flip angle = 90°, field of view (FOV) = 100 x 100 mm^2, matrix size = 66 x 66 voxels, slice thickness = 1.9 mm, voxel size = 1.5 x 1.5 x 1.9 mm^3). On each experiment day, a low-resolution FLASH anatomical scan was acquired with the same geometry as the functional images (TE = 14 ms, TR= 3 s, TA = 2.18 s, FOV = 100 x 100 mm^2, matrix = 512 x 512 voxels, slice thickness = 1.9 mm, number of averages = 2, flip angle = 150°). For overlaying our functional images, we created a high-resolution anatomical template (0.5 x 0.5 x 0.5 mm^3 isotropic voxels) by averaging five high-resolution anatomical scans acquired under general anesthesia with an MP-RAGE sequence (TE = 3.0 ms, TR = 2.5 s, flip angle = 8°, FOV = 116 x 96 x 128 mm^3; matrix = 232 x 192 x 256 voxels).

Data analysis

For M1, 9 EPI runs (180 time points each) were acquired over 6 sessions. For M2, 7 runs were acquired over 4 sessions. All data analyses were performed using AFNI (Cox, 1996) (http://afni.nimh.nih.gov/afni), FreeSurfer (Dale et al., 1999; Fischl et al., 1999) (http://surfer.nmr.mgh.harvard.edu/*),* SUMA (http://afni.nimh.nih.gov/*)* and custom

code written in Matlab (MathWorks). Preprocessing involved slice timing correction, motion correction (relative to the run-specific mean GE-EPI), spatial smoothing with a 3.0 mm full width at half-maximum Gaussian kernel, and normalization of the time series at each voxel by its mean. All volumes that had motion values with shifts > 0.5 mm and/or rotations > 0.5 deg were excluded from further analyses. Lastly, we performed linear least-squares detrending to remove nonspecific variations (i.e. scanner drift). Following preprocessing, data were submitted to generalized linear model analyses. The model included three stimulus-specific regressors and six estimated motion regressors of no interest. For each stimulus category (Env, MC, SMC) we estimated a regressor by convolving a one-parameter gamma distribution estimate of the hemodynamic response function with the square-wave stimulus function. We performed t-tests contrasting all sounds vs. baseline ("silence" trials), MC vs. Env and MC vs. SMC. Finally we coregistered and normalized our functional data to the population-average MRI-based template for rhesus monkeys 112RM-SL (McLaren et al., 2009) and then displayed the results on a semi-inflated cortical surface of the template extracted with Freesurfer and displayed with SUMA to facilitate visualization and identification of cortical activations. The anatomical boundaries described here are based on the macaque brain atlas of Saleem and Logothetis (2012).

To quantify the lateralization of the BOLD response across hemispheres we measured a lateralization index ($LI = (R_h-L_h)/(R_h+L_h)$), where R_h and L_h are the mean responses in the right and left hemisphere, respectively. The LI curve analyses ensure that the lateralization effect is not caused by small numbers of highly activated voxels across hemispheres. The LI curves were based on the t-values obtained from each contrast condition and were calculated using the LI-toolbox (Wilke and Lidzba, 2007) with the

following options: ±5 mm mid-sagittal exclusive mask, clustering with a minimum of 5 voxels and default bootstrapping parameters (min/max sample size 5/10000 and bootstrapping set to 25% of data). The bootstrapping method calculates 10,000 times LIs using different thresholds ranging from 0 until the maximum t-value for a specific contrast condition. For each threshold a cut-off mean value is obtained from which a weighted mean (LI-wm) index value can then be calculated (Wilke and Lidzba, 2007). This yields a single value between − 1 and 1 indicating right- or left-sided hemisphere dominance.

Results

Our first goal was to identify brain regions involved in the processing of conspecific vocalizations by the macaque brain. To this end, we collected functional MR images of two monkeys in a horizontal 3-T scanner while stimuli from three different sound categories were presented to the animals. Complex sounds are characterized by having a wide range of spectrotemporal features. While environmental sounds typically contain sharp temporal onsets, monkey vocalizations contain greater modulations in the spectral domain because of the harmonics contained in these sounds. Environmental sounds also carry abstract information about the identity of objects, so a comparison between BOLD responses to monkey vocalizations and environmental sounds is useful in determining brain structures involved in higher-level processing. However, specific spectrotemporal differences exist between these two types of sounds. This can be seen, for instance, in the spectral modulation of monkey vocalizations at approximately 1.5 - 2 cycles/kHz, which is not present for other sound categories (**Figure 1D**). Thus, scrambled versions of monkey calls (SMC) were used to further control for the local spectrotemporal

features in the vocalizations (see **Figure 1C** and **Material and Methods**). Comparison of average modulation spectra between categories showed that SMC were acoustically better matched to MC than Env (correlation coefficient between the modulation spectra: SMC vs. MC: 0.92, Env vs. MC: 0.86; **Figure 1D**).

Overall, sound stimulation elicited significant BOLD responses compared to silent trials irrespective of auditory stimulus category (q [FDR] < 0.05, $p < 10^{-3}$, one-tailed t-test, t range: 2.3 - 10, cluster size > 10 voxels) in a broad network of brain regions, including subcortical auditory pathways, classical auditory areas of the superior temporal gyrus (STG), but also regions in parietal and prefrontal cortices (**Figure 2**). The clusters in **Figure 2A** highlight the main activation sites on the cortical surface of monkey M1. **Figure 2B** shows selected coronal slices for both animals (M1 and M2) showing activation in the ascending auditory pathway. These regions include the cochlear nucleus (CN), the inferior colliculus (IC), the medial geniculate nucleus (MGN), the primary auditory cortex (A1), and areas in the anterior superior temporal cortex, including the rostral (R) and anterolateral (AL) areas, the rostrotemporolateral area (RTL) and the rostrotemporal pole (RTp) region.

Activation clusters (averaged across animals and hemispheres) taken from a normalized number of voxels (i.e. equal number of left and right voxels) were found in: IC [N = 84 voxels, peak coordinate = (4, -1, 12)]; A1 [N = 198 voxels, peak coordinate = (22, 6, 24)]; R/AL [N = 131 voxels, peak coordinate = (24, 17, 12)]; and RTL/RTp [N = 165 voxels, peak coordinate = (23, 22, 8)].

For both animals we observed a larger amplitude and spatial extent of the BOLD response in the right hemisphere as compared to the left hemisphere (**Figure 2B**). Activation (percent signal change) in selected clusters for each hemisphere is shown in **Figure 2C**. We compared the activation between the two hemispheres by calculating a

laterality index (LI), with a positive index indicating a left-hemisphere bias and a negative index indicating a right-hemisphere bias. Given the fact that LIs show a threshold dependency (Nagata et al., 2001), we measured LI curves to provide a more comprehensive estimate over a whole range of thresholds (Wilke and Lidzba, 2007). Using this adaptive thresholding approach we found a right-hemisphere bias in the LI curves for general auditory activation (all sounds versus baseline) in both monkeys (M1, weighted mean = -0.33; M2, weighted mean = -0.66). For higher thresholds, the activation was clustered in primary auditory cortex (A1) of the right hemisphere in each animal.

Vocalizations are complex naturalistic stimuli that contain behaviorally relevant information. In order to investigate if the auditory system contained representations that are sensitive to this sound category versus other types of behaviorally relevant complex sounds, we contrasted monkey calls against environmental sounds (see **Material and Methods**). Environmental sounds also carry abstract information about object identity in their spectrotemporal patterns. We, therefore, also looked for areas showing elevated response to these sounds relative to monkey vocalizations. When correcting for multiple comparisons (q [FDR] < 0.05), no differences were observed for the contrast of MC vs. Env. However, at uncorrected thresholds, we found significantly higher activations by MC as compared to Env in both monkeys across regions in temporal, parietal and prefrontal cortices (M1, $p < 10^{-3}$ uncorrected, t-value range: -4.2 to 6.1, cluster size > 5 voxels; M2, $p < 10^{-2}$ uncorrected, t-value range: -3.6 to 5.9, cluster size > 5 voxels) (**Figure 3A**). Specifically, activations sensitive to MC were found in the anterior STG region, including areas AL and RTp of the rostral belt/parabelt, and further along the auditory ventral stream in ventrolateral prefrontal cortex (vlPFC). In addition, we observed activation patches in the inferior parietal lobule (areas PF/PFG)

of the right parietal cortex, and bilaterally inside the inferior branch of the arcuate sulcus, possibly corresponding to Brodmann's area (BA) 44, and posterior to the arcuate sulcus, in a region that is part of ventral premotor cortex (PMv). In addition, we found regions sensitive to environmental sounds (blue) along the superior temporal sulcus (STS) and inferotemporal (IT) cortex. To investigate hemispheric lateralization in the processing of vocalizations, we measured LI curves for this contrast (Mc > Env), finding a slight right hemispheric bias in monkey M1 (weighted mean = -0.19) and a moderate right-hemisphere bias in monkey M2 (weighted mean = -0.42).

In order to determine whether spectrotemporal features alone could have driven the activation in these areas, we further contrasted monkey calls (MC) with scrambled monkey calls (SMC). The results showed similar patterns of MC activation in both monkeys in the RTL region of the aSTG (M1; $p < 10^{-3}$ uncorrected, t-value range > -4.8 – 7.5, cluster size > 5 voxels and for M2, $p < 10^{-2}$ uncorrected, t-value range > -4.3 – 6.1, cluster size > 5 voxels) in both monkeys specifically in the RTL region of the aSTG (**Figure 3B**). In monkey M2, a second region, the middle medial belt (MB), was also more strongly activated by monkey vocalizations than by their scrambled counterparts. The weighted-mean lateralization index (LI) for this contrast (MC > SMc) also showed higher values towards the right hemisphere (M1: weighted mean = -0.34; M2: weighted mean = -0.44). A summary is shown in **Table 1**.

Some differences in the patterns of activity were observed across the two animals. These differences might be explained either by variability across subjects or by differences in attentional state: M1 was significantly engaged in completing the task (> 90% success), whereas M2 was scanned passively while holding fixation. To compensate for this variability, we calculated the minimum *t*-statistic ($p < 0.01$ uncorrected) across contrasts in each monkey (a conjunction test) and across monkeys in each contrast (**Figure 4**). Conjunction across contrasts (MC > Env and MC > SMC)

and monkeys (M1 and M2) found a single area in the right hemisphere to be specifically involved across both conjunction analyses, area RTL/RTp (peak coordinate: 24, 17, 12).

Discussion

Species-specific vocalizations in non-human primates ("monkey calls") convey important information about affective/emotional states as well as the recognition of objects and individuals (Ghazanfar and Hauser, 1999). We used whole-brain functional magnetic resonance imaging (fMRI) in awake behaving monkeys to examine auditory responses to stimuli from three different sound categories: a) multiple types of conspecific monkey calls, b) environmental sounds, and c) scrambled versions of the same monkey calls largely preserving their local spectrotemporal features.

For all three sound categories combined we found robust BOLD responses along various regions in the ascending auditory pathways (CN, IC, MGB and A1, **Figure 2A**,**B**). These results, using a 3-T scanner without contrast agent, corroborate previous fMRI findings obtained on a 1.5-T magnet with the contrast agent MION, showing activation by complex sounds along the auditory pathway (Joly et al., 2012a). The results further attest to the fact that complex sounds are highly effective for mapping subcortical and cortical auditory structures (Rauschecker et al., 1995; Rauschecker, 1998b; Poremba et al., 2003). Furthermore, our results confirm the general trend of a slight right-hemisphere bias (**Table 1**) in the processing of complex sounds in the macaque auditory cortex, as measured with fMRI (Joly et al., 2012a; Petkov et al., 2008). Similar results have been found in humans for non-speech voice sounds (Belin et al., 2000).

When we compared activations produced by monkey vocalizations versus the other two sound categories using a conjunction analysis, we found consistent activations in regions along the anterior STG, in particular in areas AL, RTL and RTp, in both animals (**Figure 4**). Our results extend previous findings of increased sensitivity to monkey vocalizations in anterior STG regions (Poremba et al., 2003; Petkov et al., 2008; Kikuchi et al., 2010; Joly et al., 2012a,b; Fukushima et al., 2014) by using control stimuli (SMC) that retained the low-level acoustic information of macaque vocalizations and whose acoustic structure was better matched to the vocalizations than the acoustic structure of other complex sounds (**Figure 1D**). Single-unit studies of the R/AL region have also found increased selectivity either to monkey calls, or to sound categories including vocalizations (Tian et al., 2001; Kuśmierek et al., 2012), consistent with the present results (**Figure 3** and **4**).

Thus, the cortical representation of vocalizations involves an auditory ventral pathway, consisting of a chain of interconnected regions in anterior STG and vlPFC that extract abstract information for the recognition and categorization of vocalizations (Rauschecker, 2012). The rostral belt, parabelt and aSTG send afferent projections into ventrolateral, polar, orbital and medial regions of the prefrontal cortex (PFC) (Jones and Powell, 1970; Hackett et al., 1999; Romanski et al., 1999; Kaas and Hackett, 2000; Cavada et al., 2000; Hackett, 2011; Yeterian et al., 2012), and together these regions form the ventral cortical stream in audition. Vocalization-sensitive neurons are found along with face-sensitive neurons in the vlPFC (Romanski et al., 2005), allowing these regions to integrate vocalizations with the corresponding facial gestures (Romanski and Goldman-Rakic, 2002; Cohen et al., 2007; Diehl and Romanski, 2014). The PFC is involved in higher-level integrative processes for the cognitive control of vocalizations as well as in the interpretation of semantic content in vocalizations (Romanski and Averbeck, 2009). The activation patterns observed in PFC (**Figure 3A**) could represent

categorical or affective information reflected in the vocalizations. Further imaging studies and multivariate analyses comparing multiple vocalization types might elucidate the differential contribution of each subregion of the PFC.

Our stimuli also activated higher-level visual areas, such as the middle temporal (MT) and inferior temporal areas (IT). These areas are known to be involved in the processing of visual motion (Maunsell and Van Essen, 1983) and in object perception (including faces), respectively (Tsao et al., 2006; Ku et al., 2011). Their activation by purely auditory stimuli raises interesting questions regarding their possible role in the multisensory processing of dynamic audio-visual stimuli, such as facial expressions that naturally occur in conjunction with vocalizations and/or motion of the face (Furl et al., 2012; Polosecki et al., 2013; Perrodin et al., 2014). However, to answer these questions more definitively, further imaging experiments utilizing dynamic audio-visual stimuli would be necessary. Such studies could enlighten us on how auditory information combines with visual information in both the ventral and dorsal pathways building multimodal representations from dynamic facial expressions combined with vocalizations (Ghazanfar and Logothetis, 2003).

When we contrasted monkey calls to environmental sounds, we also found differential activation in regions PF/PFG (area 7b) (Pandya and Seltzer, 1982; Rozzi et al., 2006) of the inferior parietal lobule (IPL), in addition to the well-known regions in the STG sensitive to monkey vocalizations. Parietal regions inside the intraparietal sulcus (IPS) have been known to receive auditory projections (Lewis and Van Essen, 2000) and to contain neurons that respond to auditory and multimodal stimuli (Stricanne et al., 1996; Bushara et al, 1999; Grunewald et al., 1999; Cohen and Andersen, 2000; Cohen, 2009),

but the role of these regions has traditionally been assumed to lie in spatial processing and control of eye movements.

Similarly, we found an engagement of the ventral premotor cortex (PMv) in the processing of monkey vocalizations (**Figure 3A**). This region has previously been thought to be involved in the processing of the location (but not quality) of nearby sounds (Graziano et al., 1999). Surprisingly, when we compared the effects of vocalizations (MC) against vocalizations that were scrambled in both the spectral and temporal domains (SMC), we did not observe greater activation in parietal or prefrontal areas for MC, suggesting that the scrambled versions of the MC evoked the same amount of activity in these regions. Similar results where obtained by Joly et al. (2012b) with temporally scrambled vocalizations activating large regions of premotor and parietal cortices. Ventral premotor cortex (PMv) has also been implicated in the initiation of vocalizations in the macaque monkey (Hage and Nieder, 2013). It appears possible, therefore, that the same neurons are the source of an efference copy signal (Kauramäki et al., 2010), which is responsible for the suppression of auditory cortex during self-initiated vocalizations (Eliades and Wang, 2003). More generally, they could be part of an audio-motor network connecting perception and production of sounds (Rauschecker and Scott, 2009; Rauschecker, 2011).

Acknowledgments

Special thanks to Josie Cui for animal care and assistance with the experiments, and John VanMeter for fMRI data optimization. This work was supported by grants from the National Institutes of Health (R01-DC03489, R01-NS052494 and R56-NS052494 to

J.P.R.), a PIRE grant from the National Science Foundation (OISE-0730255 to J.P.R.), and a FiDiPro award from the Academy of Finland (J.P.R.).

Author contributions

MOR co-designed the study, trained the animals, programmed stimulus presentation, acquired part of the data, conducted most analyses, and co-wrote the manuscript. PK programmed the behavioral task and participated in writing the manuscript. DA trained the animals and acquired part of the data. ID generated the scrambled stimuli and acquired part of the data. FACA contributed with data analyses and participated in writing the manuscript. GAK, interpreted data and participated in writing the manuscript. MS, IPJ, and JPR co-designed the study and participated in writing the manuscript.

References

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B. G., et al. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14608–13. doi:10.1073/pnas.0510480103.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage* 22, 401–8. doi:10.1016/j.neuroimage.2004.01.014.

Barbas, H. (1993) Organization of cortical afferent input to orbitofrontal areas in the rhesus monkey. *Neuroscience* 56, 841-864.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*. 403, 309-312. doi:10.1038/35002078

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S.L., and Rauschecker, J.P. (2015) Neurobiological roots of language in primate audition: common computational properties. *Trends in Cognitive Sciences* 19, 142-50. pii: S1364-6613(14)00275-7. doi: 10.1016/j.tics.2014.12.008.

Brown, J. C. (1991). Calculation of a constant Q spectral transform. *J. Acoust. Soc. Am*. 89, 425. doi:10.1121/1.400476.

Bushara, K.O., Weeks R.A., Ishii K., Catalan M.-J., Rauschecker J.P. and Hallett M. (1999) Evidence for modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience* 2, 759-766.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., and Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb. Cortex* 10, 220–42.

Cheney, D.L. and Seyfarth, R.M. 1990. *How monkeys see the world*. Chicago, University of Chicago Press.

Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hear. Res*. 258, 100–5. doi:10.1016/j.heares.2009.01.011.

Cohen, Y. E., and Andersen, R. A. (2000). Reaches to Sounds Encoded in an Eye-Centered Reference Frame. *Neuron* 27, 647–652. doi:10.1016/S0896-6273(00)00073-8.

Cohen, Y. E., Theunissen, F., Russ, B. E., and Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol*. 97, 1470–84. doi:10.1152/jn.00769.2006.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res*. 29, 162–73.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–94. doi:10.1006/nimg.1998.0395.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U. S. A.* 109, E505–14. doi:10.1073/pnas.1113427109.

Diehl, M. M., and Romanski, L. M. (2014). Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J. Neurosci.* 34, 11233–43. doi:10.1523/JNEUROSCI.5168-13.2014

Eliades, S. J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89, 2194–207. doi:10.1152/jn.00627.2002.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi:10.1006/nimg.1998.0396.

Frey, S. Campbell, J.S., Pike, G.B., Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. J Neurosci. 5, 11435-44.

Frey, S., Mackey, S., Petrides, M. (2014). Cortico-cortical connections of areas 44 and 45B in the macaque monkey. Brain Lang. 131, 36-55.

Fukushima, M., Saunders, R. C., Leopold, D. A., Mishkin, M., and Averbeck, B. B. (2014). Differential coding of conspecific vocalizations in the ventral auditory cortical stream. *J. Neurosci.* 34, 4665–76. doi:10.1523/JNEUROSCI.3969-13.2014.

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B. B., and Ungerleider, L. G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J. Neurosci.* 32, 15952–62. doi:10.1523/JNEUROSCI.1992-12.2012.

Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M.D., Carson, R.E., Herscovitch, P., Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc. Natl. Acad. Sci. U. S. A*, 101, 17516-21.doi: 10.1073/pnas.0408077101.

Ghazanfar, A. A., and Logothetis, N. K. (2003). Neuroperception: facial expressions linked to monkey calls. *Nature* 423, 937–8. doi:10.1038/423937a.

Ghazanfar, A., and Hauser, M. (1999). The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cogn. Sci.* 3, 377–384.

Graziano, M. S., Reiss, L. A., and Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature* 397, 428–30. doi:10.1038/17115.

Grunewald, A., Linden, J. F., and Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. *J. Neurophysiol.* 82, 330–42.

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–46. doi:10.1016/j.heares.2010.01.011.

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

Hage, S. R., and Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* 4, 2409. doi:10.1038/ncomms3409.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., and Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–23.

Jackson L.L., Heffner R.S., Heffner H.E. (1999). Free-field audiogram of the Japanese macaque (*Macaca fuscata*). *J Acoust Soc Am* 106, 3017–3023.doi:10.1121/1.428121

Joly, O., Ramus, F., Pressnitzer, D., Vanduffel, W., and Orban, G. A. (2012a). Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cereb. Cortex* 22, 838–53. doi:10.1093/cercor/bhr150.

Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban G.A. (2012b). Processing of vocalizations in humans and monkeys: a comparative fMRI study. *Neuroimage*. 62, 1376-89. doi:10.1016/j.neuroimage.2012.05.070.

Jones, E. G., and Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820.

Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11793–9. doi:10.1073/pnas.97.22.11793.

Kauramäki, J., Jääskeläinen, I.P., Hari, R., Möttönen, R., Rauschecker, J.P., and Sams, M. (2010) Transient adaptation of auditory cortex organization by lipreading and own speech production. *J. Neurosci*. 30(4), 1314 –1321.doi:10.1523/JNEUROSCI.1950-09.2010

Kikuchi, Y., Horwitz, B., and Mishkin, M. (2010). Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J. Neurosci*. 30, 13021–30. doi:10.1523/JNEUROSCI.2267-10.2010.

Kikuchi, Y., Horwitz, B., Mishkin, M., and Rauschecker, J. P. (2014). Processing of harmonics in the lateral belt of macaque auditory cortex. *Front. Neurosci*. 8, 204. doi:10.3389/fnins.2014.00204.

Ku, S.-P., Tolias, A. S., Logothetis, N. K., and Goense, J. (2011). fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70, 352–62. doi:10.1016/j.neuron.2011.02.048.

Kuśmierek, P., and Rauschecker, J. P. (2009). Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J. Neurophysiol*. 102, 1606–22. doi:10.1152/jn.00167.2009.

Kuśmierek, P., Ortiz, M., and Rauschecker, J. P. (2012). Sound-identity processing in early areas of the auditory ventral stream in the macaque. *J. Neurophysiol*. 107, 1123–41. doi:10.1152/jn.00793.2011.

Kuśmierek, P., and Rauschecker, J. P. (2014). Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol*. 111, 1671–85. doi:10.1152/jn.00436.2013.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., and Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *J. Neurosci*. 29, 2477–85. doi:10.1523/JNEUROSCI.4921-08.2009.

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–37. DOI: 10.1002/1096-9861(20001204)428:1<112::AID-CNE8>3.0.CO;2-9.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–16. doi:10.1006/nimg.2001.0888.

Maunsell, J. H., and Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* 49, 1127–1147.

McLaren, D. G., Kosmatka, K. J., Oakes, T. R., Kroenke, C. D., Kohama, S. G., Matochik, J. A., Ingram, D. K., and Johnson, S. C. (2009). A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage* 45, 52–9. doi:10.1016/j.neuroimage.2008.10.058.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417. doi:10.1016/0166-2236(83)90190-X.

Nagata, S. I., Uchimura, K., Hirakawa, W., and Kuratsu, J. I. (2001). Method for quantitatively evaluating the lateralization of linguistic function using functional MR imaging. *AJNR. Am. J. Neuroradiol.* 22, 985–91.

Pandya, D.N., and Seltzer, B. (1982) Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey. *J. Comp. Neurol.* 204(2), 196-210.

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2014). Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J. Neurosci.* 34, 2524–37. doi:10.1523/JNEUROSCI.2805-13.2014.

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Curr. Biol.* 21, 1408–15. doi:10.1016/j.cub.2011.07.028.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–74. doi:10.1038/nn2043.

Petkov, C.I., Kikuchi, Y., Milne, A., Mishkin, M., Rauschecker, J.P., and Logothetis, N.K. (2015) Different forms of effective connectivity in primate frontotemporal pathways. *Nature Communications* 6, 6000. doi: 10.1038/ncomms7000.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol*. 7, e1000170. doi:10.1371/journal.pbio.1000170.

Polosecki, P., Moeller, S., Schweers, N., Romanski, L. M., Tsao, D. Y., and Freiwald, W. A. (2013). Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J. Neurosci*. 33, 11768–73. doi:10.1523/JNEUROSCI.5402-11.2013.

Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., and Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science* 299, 568–72. doi:10.1126/science.1078900.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res*. 271, 16–25. doi:10.1016/j.heares.2010.09.001.

Rauschecker, J. P. (1998a) Parallel processing in the auditory cortex of primates. *Audiol. Neurootol*. 3, 86–103.

Rauschecker, J.P. (1998b) Cortical processing of complex sounds. *Current Opinion in Neurobiology* 8, 516-521

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci*. 4, 7. doi:10.3389/fnevo.2012.00007.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci*. 12, 718–24. doi:10.1038/nn.2331.

Rauschecker, J. P., and 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. doi:10.1073/pnas.97.22.11800.

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

Romanski, L. M., and Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu. Rev. Neurosci.* 32, 315–46. doi:10.1146/annurev.neuro.051508.135431.

Romanski, L. M., Averbeck, B. B., and Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* 93, 734–47. doi:10.1152/jn.00675.2004.

Romanski, L. M., and Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–6. doi:10.1038/nn781.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–6. doi:10.1038/16056.

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., and Luppino, G. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb. Cortex* 16, 1389–417. doi:10.1093/cercor/bhj076.

Russ, B. E., Ackelson, A. L., Baker, A. E., and Cohen, Y. E. (2008). Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J. Neurophysiol.* 99, 87–95. doi:10.1152/jn.01069.2007.

Saleem KS, Logothetis NK (2012). A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates. 2nd edition with Horizontal, Coronal and Sagittal series; San Diego: Elsevier/Academic press.

Schörkhuber, C., and Klapuri, A. (2010). Constant-Q transform toolbox for music processing. 7th Sound and Music Computing Conference, Barcelona, Spain. (http://www.iem.at/~schoerkhuber/cqt2010/)

Singh, N. C., and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* 114, 3394–411.

Stricanne, B., Andersen, R. A., and Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol.* 76, 2071–6.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–3. doi:10.1126/science.1058911.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–4. doi:10.1126/science.1119983.

Wilke, M., and Lidzba, K. (2007). LI-tool: a new toolbox to assess lateralization in functional MR-data. *J. Neurosci. Methods* 163, 128–36. doi:10.1016/j.jneumeth.2007.01.026.

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–2. doi:10.1038/nn1263.

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex*. 48, 58–81. doi:10.1016/j.cortex.2011.03.004.

Yovel, G., and Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends Cogn. Sci.* 17, 263–71. doi:10.1016/j.tics.2013.04.004.

Figure Legends:

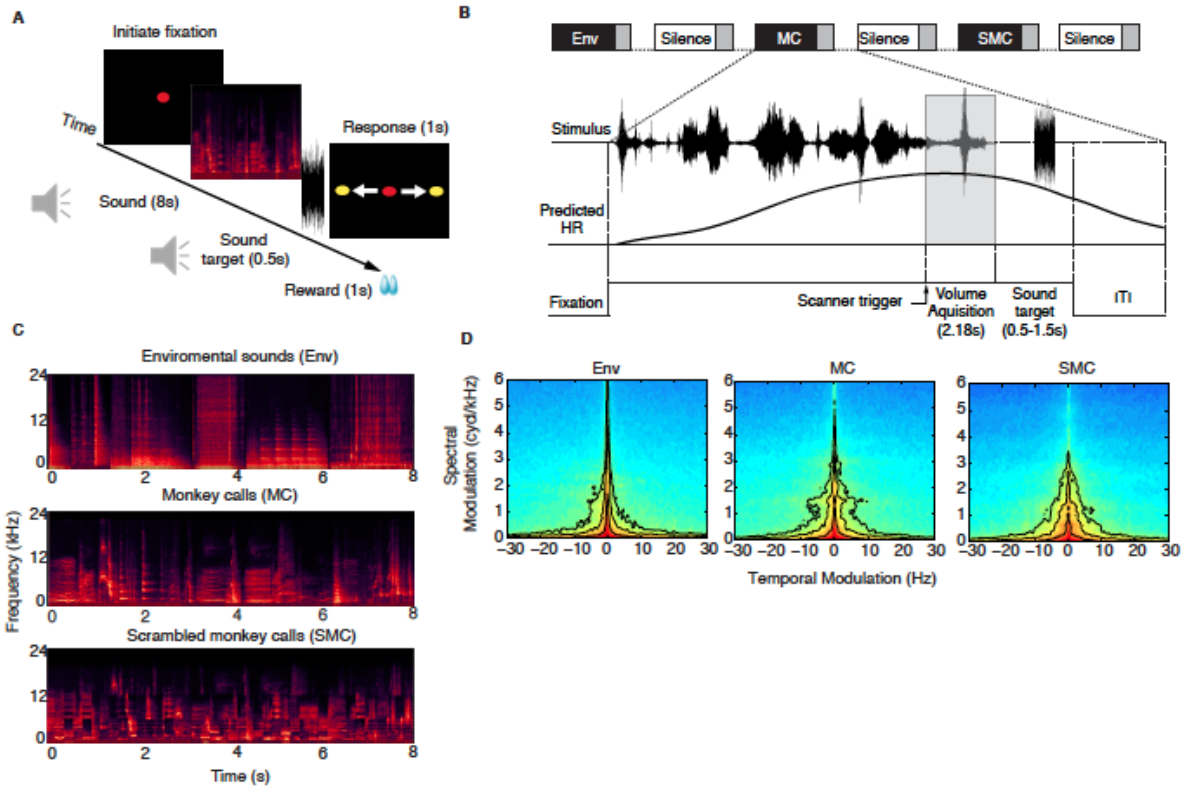


Figure 15. Behavioral paradigm and example stimuli for each sound category. (**A**) The monkey had to keep fixation on a central spot while stimuli were presented from one of three sound categories: environmental sounds (Env), monkey calls (MC) and scrambled monkey calls (SMC). Next, a target sound (white noise, 500 ms) was played after a random delay of 0.5 – 1.5 s at the end of each stimulus period, and the animals were required to make a saccade to an imaginary cue position (yellow cue). The imaginary target was chosen to be either on the left or the right side of the screen, and the animal was instructed at the beginning of each session where the target was going to appear (sees **Methods**). (**B**) All conditions were presented in a sparse-sampling design to avoid interference between the hemodynamic response (HR) generated by the scanner noise and by the stimuli. The inter-trial interval (ITI) lasted for 2 s, and the monkey was then allowed to start a new trial by initiating fixation once again. (**C**) Eight-second series of spectrograms from the three sound categories presented. (**D**)

Average modulation spectra for each stimulus category. Pearson correlations between average modulation spectra were: MC vs. SMC = 0.92, MC vs. Env = 0.86.

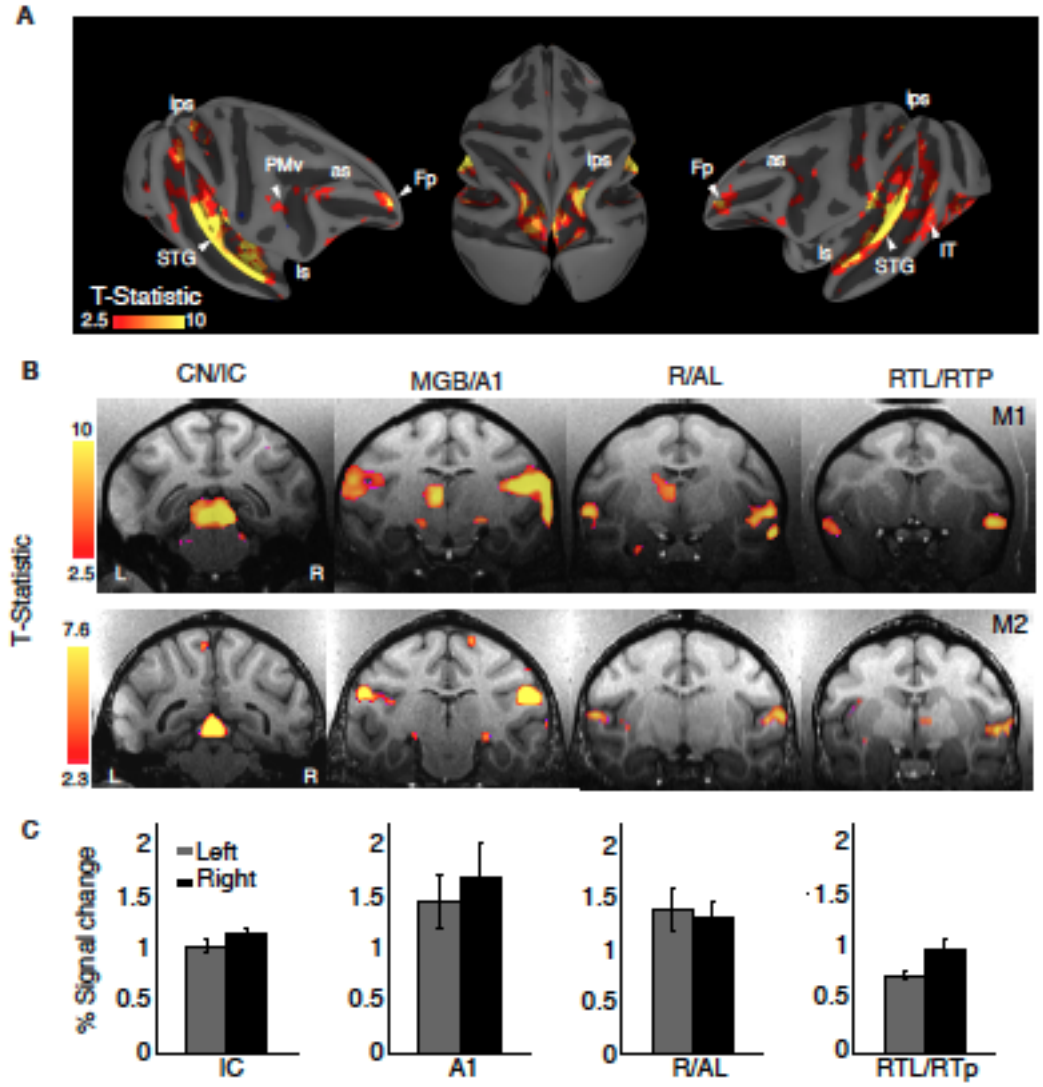


Figure 16. Mapping auditory and auditory-related regions with complex sounds. (**A**) Representative cortical responses from monkey (M1) for all sound conditions combined (q FDR < 0.05, $p < 10^{-2}$; cluster size > 10 voxels). The projection onto the semi-inflated surface preserves sulcal and gyral landmarks while allowing visualization inside the intraparietal sulcus (ips) and lateral sulcus (ls). Activation was observed along the auditory ventral stream in the superior temporal gyrus (STG), the superior temporal sulcus (STS), ventral intraparietal area (VIP) and the frontal pole (Fp). Activated dorsal-

stream regions included the ips and ventral premotor cortex (PMv). Some active clusters were also observed in the middle temporal area (MT) and the inferior temporal cortex (IT). (**B**) Activation was robust across regions in the ascending auditory pathway of the two monkeys: cochlear nuclei (CN), inferior colliculus (IC), medial geniculate nucleus (MGN), primary auditory cortex (A1), rostral area (R), anterolateral area (AL), lateral rostrotemporal area (RTL) and the rostrotemporal pole region (RTp). (**C**) The average BOLD response for the main auditory activation showed a right-hemisphere bias in both animals (M1, weighted mean = -0.33, M2, weighted mean = -0.66).

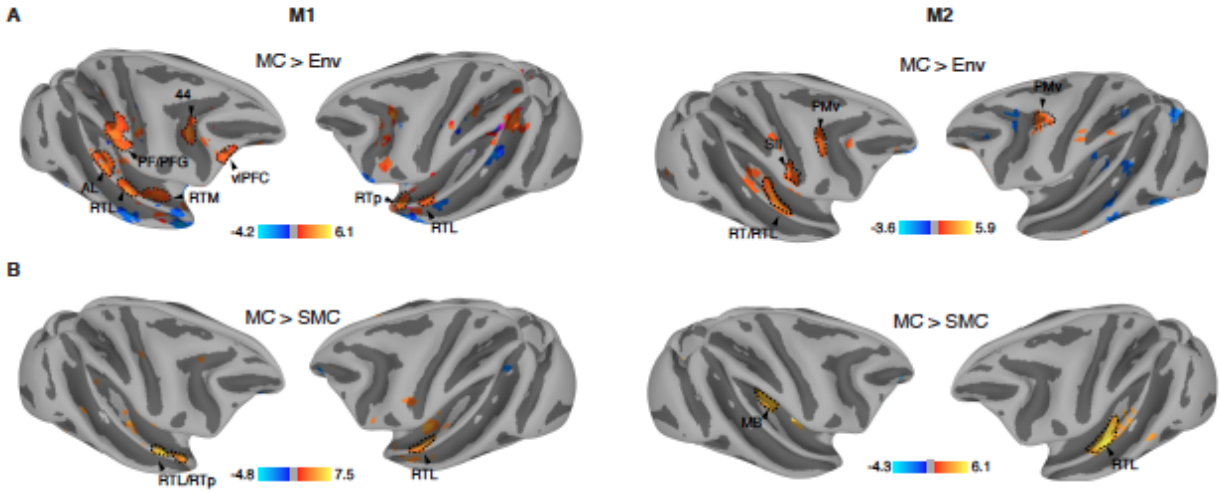


Figure 3. Regions specifically activated by monkey vocalizations. (**A**) Vocalization-sensitive regions obtained from comparison between the effects of monkey calls and environmental sounds. All activation maps were displayed on a semi-flattened surface of the macaque monkey template. Active regions were found in the anterolateral area (AL), lateral rostrotemporal area (RTL), rostrotemporal pole (RTp), secondary somatosensory (SII) cortex, ventral premotor cortex (PMv), ventrolateral prefrontal cortex (vlPFC), and inferior parietal areas (PF and PFG). (**B**) Regions significantly more activated by monkey vocalizations than by scrambled monkey vocalizations include areas in the anterior STG, RTL/RTp. Red/orange: significantly higher

activation by MC than by control sounds (SMC or Env); blue: significantly higher activation by SMC or Env than by MC.

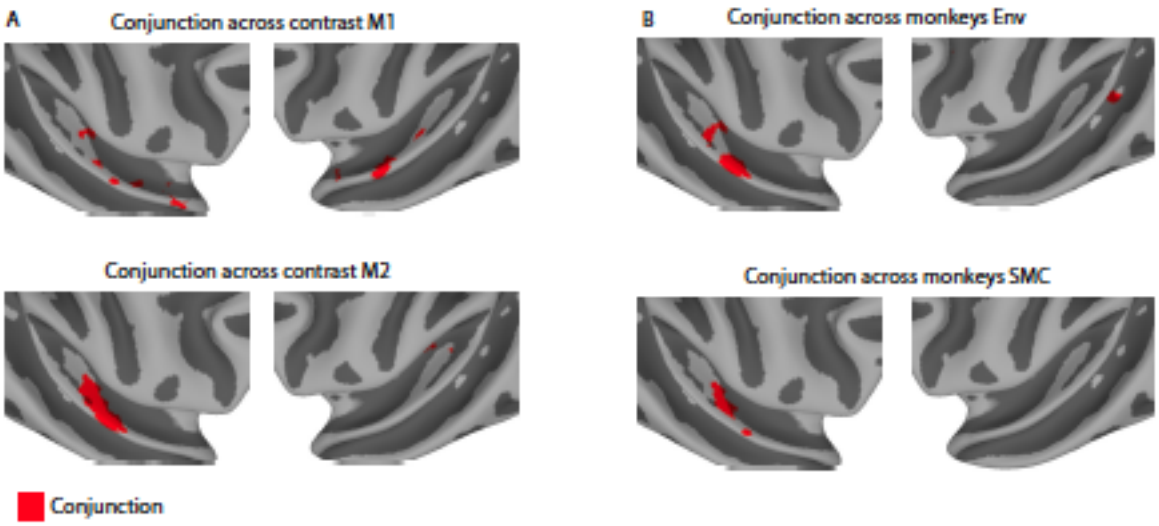


Figure 4. Conjunction results across contrast conditions and across monkeys. (**A**) Conjunction across contrast (MC > Env and MC > SMC) for monkey M1 (n voxels = 235, top panel) and for monkey M2 (n voxels = 89, lower panel). (**B**) Conjunction across monkeys (M1 and M2) for contrast MC > Env (n voxels = 248, top panel) and for contrast MC > SMC (n voxels = 58, lower panel). Each individual contrast map was thresholded at $p < 0.01$ (uncorrected). Red region indicates conjunction voxels that were differentially activated for both contrasts in each monkey or in both monkeys for each contrast.

Table 1. LI-weighted-mean values for the overall sound activation and for each contrast condition. Mean lateralization index values (LI-wm) are shown that were obtained from LI curves measured as a function of the statistical threshold (t-value) for the overall auditory activation (all sounds vs. baseline), for the contrast between monkey calls and environmental sounds (MC > Env) and for the contrast between

monkey calls and scrambled monkey calls (MC > SMC). A positive index indicates a left-hemisphere bias, while a negative index indicates a right-hemisphere bias. LI-wm values are shown separately for monkeys M1 and M2.

	All > baseline	**MC > Env**	**MC > SMC**
M1	-0.33	-0.19	-0.34
M2	-0.66	-0.42	-0.44

7.3 Widespread and opponent fMRI signals represent sound location in macaque auditory cortex

Widespread and opponent fMRI signals represent sound location in macaque auditory cortex

Michael Ortiz-Rios[1,2*], Frederico A.C. Azevedo[1,2], Dávid Z. Balla[1], Matthias H. Munk[1,3], Georgios A. Keliris[1, 4], Josef P. Rauschecker[5,6] and Nikos K. Logothetis[1,7]

Affiliations: [1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany [2]International Max Planck Research School (IMPRS), Tübingen, Germany, [3]Department of Systems Neurophysiology, Fachbereich Biologie, Technische Universität Darmstadt, Germany, [4]Bio-Imaging Lab, Department of Biomedical Sciences, University of Antwerp, Wilrijk, Belgium, [5]Institute for Advanced Study, Technische Universität München, Germany, [6]Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA, [7]Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, UK.

***Corresponding author:** Michael Ortiz-Rios

Department of Physiology of Cognitive Processes

Spemannstraße 38

72076 Tübingen

Germany

E-Mail : michael.ortiz@tuebingen.mpg.de

Phone: +49 7071 601 675

Fax: +49 7071 601 652

Important words used throughout this text: BOLD-fMRI, acoustic space; auditory cortex; hemifield tuning; ITD; posterior superior temporal cortex.

Abstract (150 words):

In primates, posterior auditory areas are thought to be part of a dorsal auditory pathway specialized in processing spatial information. While a number of human studies support this notion, neuroimaging evidence in macaques is missing. Here we provide new evidence based on functional magnetic resonance imaging (fMRI) of the macaque monkey indicating that auditory space is represented based on a distributed hemifield code rather than on a local spatial topography. Hemifield tuning emerges from an opponent pattern of positive and negative fMRI signals across the cerebral hemispheres that dependents on the suppression carried by interaural delay cues. Importantly, this pattern in right posterior region generates enough spatial information to segregate space across hemifields. Taken together, our results suggest that primate auditory cortex represents space based on a distributed hemifield code and that posterior specialization for space arises from this form of coding.

Introduction:

The ability to localize sounds is essential for an animal's survival. In primates, auditory spatial information is thought to be processed along a dorsal auditory pathway[1], a hierarchical system with reciprocal connections that includes areas CL/CM of the posterior superior temporal (pST) region[2,3]. A central question regarding the cortical mechanisms for sound-source localization in primates is whether the spatial representation of sounds is localized to distinct areas of the pST region or distributed throughout the superior temporal cortex.

Most of the original results that suggested regional specificity in posterior regions of auditory cortex (AC) for sound source localization were based on single-unit recordings in macaque monkeys[3], followed by neuroimaging studies in humans[4,5]. In the macaque, neurons in area Cl were found to be sharply tuned[3] and more selective to spatial position than other fields of AC[3,6,7]. However, more recent evidence in both monkeys[8] and humans[9,10] suggest that auditory space is represented based on a hemifield code by neurons broadly tuned to either side of space.

These new evidence is consistent with a competing perspective which suggests that space is coded by an opponent-channel mechanism [11,12] as similarly found in the brainstem [13,14]. Strong evidence supporting this notion comes from previous lesion studies in cats[15] and monkeys[16] demonstrating that unilateral or reversible lesions[17] of either left or right AC resulted in severe localization deficits for sound sources contralateral to the lesion. Comparably, single-unit studies in cats[12] and optical imaging experiments[18] in ferrets have provided data predominantly supporting a distributed mechanism by showing neurons responding broadly with a maximal response to sound sources in contralateral space (e.g. near the opposite ear).

While several animal studies have addressed the issue of space coding in AC utilizing optical imaging[18] and single-unit methods[19], these techniques are not ideal for studying long-range functional interactions among the distant hemispheres, particularly in primates, where auditory areas reside deep inside sulci. Functional magnetic resonance imaging (fMRI) based on blood oxygen level depend (BOLD) signals provides a complementary method to overcome this problem.

To visualize the functional representation of space across auditory cortical fields (CFs) of each hemisphere we first mapped the frequency organization in AC and identified mirror reversal tonotopic maps[20,21]. Secondly, we presented spatial stimuli from an individualized virtual azimuth space (**Fig. 1g**) and found a lack of topographical organization. Instead, we found a distributed representation of azimuth with an opponent pattern of positive and negative BOLD responses (PBRs and NBRs respectively) across hemispheres with a predominant NBR in pST of the right hemisphere. We explored these results further by presenting the same original spatial sounds but without ITD cues and found that such manipulation resulted in a loss of hemifield tuning. Finally, we show by using multivariate pattern dissimilarity[22] of the BOLD responses to each azimuth sector that right pST can segregate space similarly to a hemifield code.

Results:

Our first aim was to identify auditory CFs based on the frequency organization of AC and then map the spatial dimension using phase-encoding analytical methods[23,24].

Functional experiments were conducted under general anesthesia[25] in two male rhesus monkeys (*Macaca mulatta*) using a 7-Tesla MRI scanner (Biospec 7/60v, Bruker),

while awake-monkey experiments were conducted in two female monkeys using a 4.7-Tesla MRI scanner (Biospec 47/40v, Bruker). All images were acquired in-plane resolution of 0.75 x 0.75 with a slice thickness of 2 mm aligned parallel to the superior temporal gyrus (STG) (**Fig. 1a**). Sparse sampling acquisition was used to avoid scanner noise interference with the stimulus induced BOLD responses (**Fig. 1b**). Auditory stimulation elicited significant BOLD responses (q FDR < 0.05, $p < 10^{-7}$, cluster size > 10 voxels) along the auditory pathway; inferior colliculli (IC), medial geniculate body and AC in both anesthetized and awake monkeys (**Supplementary Fig. 1**).

After optimizing the sparse acquisition design, we conducted tonotopic experiments to identify individual CFs utilizing phase-mapping analytical techniques (see **Online Methods**). The method consisted of presenting tones and narrow band noise (**Fig. 1b**) in blocks of one-octave steps in ascending frequency order and repeating the stimulus cycle twelve times (**Supplementary Fig. 2a**).

The resulting average BOLD response to each frequency range was narrow and gradually shifted from low frequency A1 to anterior and posterior regions of AC with a distinct wave pattern of PBRs and NBRs (**Fig. 1c**). The PBR/NBR pattern reversed drastically around 2 kHz indicating a shift towards high frequency regions. Voxels with significant modulation (coherence > 0.3) to the stimulus rate (0.01 Hz, 12 cycles/ 1200 s) were mapped by their scaled phase values to the frequency range of the presented stimuli (0.1 - 16 kHz) (**Fig. 1d**). For visualization of the overall maps inside the lateral sulcus we superimposed all results into a rendered surface of the STG of each animal. Subsequently, we defined four CFs (posterior, primary, rostral and anterior) based on the mirror reversal boundaries with the same population response mediolaterally (**Fig. 1f** and **Supplementary Fig. 2b-d**).

We then aimed at mapping the spatial domain utilizing the same analytical methods but at a stimulation rate of 0.008 Hz (12 cycles/ 1800 s). Prior to experiments we obtained individualized 3D-head-related transfer function (HRTF) sounds (broad-band noise bursts, 0.2-16 kHz, 80 dB SPL, 100 ms in duration) from binaural recordings of each individual monkey (see **Online Methods**). The recorded stimuli contained all individual spatial cues (ITDs, interaural level differences (ILDs) and spectral cues, **Supplementary Fig. 3a**, **b).** In addition, the virtual noise bursts changed in azimuth direction (leftward, rightward and in distance) within a 30° sector over time (**Fig. 1g**). Such innovative design allowed us to keep space partitioned while we estimated the response to dynamic spatial sounds within a constrained spatial sector and thus avoided repetition suppression in the BOLD response. A total of 12 sectors around a virtual plane surrounding the head of the monkey were used to image the BOLD signal at any given point in time.

As opposite to the frequency domain, the mean BOLD responses to spatial sounds were broad and shifted between two phases across the cerebral hemispheres (**Fig. 1h**). Overall, the maps showed no topographic organization, but instead two color codes on each hemisphere reflected the broad peak response across sounds in the contralateral hemifield (**Fig. 1i** and **Supplementary Fig. 3c**). We investigated these results further by plotting region-of-interest (ROI) in significant coherent voxels crossing parallel and orthogonal to the primary field and confirmed a lack of topography in the flat phase-peak response across cortical space spanning 10 mm (**Supplementary Fig. 3d**).

In summary, our mapping experiments corroborated previous electrophysiological[26] and imaging studies of macaque AC showing mirror-reversal tonotopic maps[21] and provided new evidence indicating that the functional representation of azimuth as measured by fMRI at the millimeter scale is not organized based on spatial topography.

Positive and negative BOLD responses across auditory fields

How is azimuth space represented on each auditory CFs? We investigated our data further by analyzing each time series with a general linear model (GLM) of the BOLD signal. We tested the significance of the model from the measured BOLD responses to each condition (n = 12) as compared to the baseline/silence periods (q FDR < 0.05, $p < 10^{-6}$, cluster size > 10 voxels). Surprisingly, we found distinct patterns of PBRs and NBRs within each hemisphere that changed dynamically as a function of each spatial sector (**Fig. 2a** and **Supplementary Fig. 4**). Spatial tuning curves calculated from the spatial spread of PBRs and NBRs in each AC showed that the overall tuning was in opposite polarity between signals, with PBRs oriented approximately at ±120° and NBRs at ±60° between hemispheres. Similarly, hemispheric differences in PBRs and NBRs showed opposite polarity between signals, with NBRs showing a peak around frontal right sectors (**Fig. 2d**). The peak for PBRs was observed for sectors near the contralateral ear (e.g. ±90°/120°), with a cluster size extending the overall spatial volume of AC. However, the ipsilateral PBRs were greatly reduced in size and were accompanied by a NBR pattern in anterior and posterior regions of AC. In the primary field, the responses exhibited a concentric pattern (e.g. at +30°- 60°) with positive voxels expanding the overall anterior-posterior frequency axis and negative voxels mostly lying anteriorly and posteriorly (**Fig. 2b**).

The average BOLD signal in CFs of each monkey (mean and ±SEM, including both PBRs and NBRs) showed a marked shift in amplitude around the midline, which further indicated hemifield tuning across all CFs (**Fig. 2c**). Similarly, spatial tuning curves obtained from PBRs showed highly significant deviations from circular uniformity (Rayleigh test, $p < 2.3^{-20}$) with angular means oriented in opposite polarity (~ ±120°)

between hemispheres but not within CFs of the same hemisphere (**Fig. 3**). Vector length showed that more than half of the total numbers of voxels were active to contralateral sectors (~ ±120°) for all CFs. The overall tuning in central fields (primary and rostral) was slightly broader than in anterior and posterior fields of the same hemisphere based on the standard deviations (see **Supplementary table 1** for details).

In summary, our functional analyses showed that azimuth space as measured by fMRI is represented by opponent hemifield responses of positive and negative BOLD across hemispheres. The dynamic change between PBRs and NBRs strongly supports the idea of an opponent-channel mechanism[11,12] in the representation of space in the macaque monkey.

Contralateral bias measured with BOLD response contrast

While animal studies have shown a clear contralateral bias in the firing rate of cortical neurons[6–8,12], neuroimaging studies in humans have obtained mixed results in respect to the degree of contralaterality[8,27,28] in AC to spatial sounds. Whether these discrepancies are due to species differences in neural coding or to methodological differences (e.g. sound stimulation, single-unit and/or fMRI) between studies in animals and humans remain a matter of debate[8].

Here, we provide evidence showing a contralateral bias in the fMRI BOLD contrast between equidistance spatial sectors (**Fig. 4a**). The differential activation maps (q FDR < 0.05, $p < 10^{-3}$, cluster size > 10 voxels) showed whether the responses were greater for the left (blue to cyan) or the right hemifield (red to yellow). The strength of the BOLD response showed a robust contralateral bias for spatial sectors near the lateral axis (e.g. ~ ±90° - 120°). The contrast in frontal sectors (± 0 - 30°) showed greater

differential response only in the right hemisphere, while contrast for backward sectors (± 150 - 180°) showed almost no differential activation at equal threshold values (q FDR < 0.05).

We quantified these results further by calculating a laterality index (LI = L − R/ |L + R|) between correspondent CFs in the opposite hemisphere, including AC as a whole (all fields included). Since laterality indexes (LI) in fMRI typically show a threshold dependency[29], we measured LI curves by bootstrapping LI values as a function of the t-value threshold and then calculated a mean weighted laterality index (LIwm) (see **Online Methods** for details). The LIwm ranges between −1 and +1 with a positive index indicating left-hemisphere bias and a negative index indicating a right-hemisphere bias. The resulting indexes for all CFs showed a very strong right hemisphere bias (LIwm < −0.5) for sectors in the left hemifield and a left hemisphere bias (LIwm > 0.5) for sectors in the right hemifield, except for the right backward sector (+150 - 180°), where indexes were more variable across CFs and monkeys (**Fig. 4b**). The changes around the midline were drastic as observed in the slope line in frontal sectors (±0 - 30°) as opposed to the slope lines for sectors within each hemifield.

Similar contralateral representations were found in AC of the awake-monkey (**Fig. 5**). Additionally however, we obtained reliable BOLD signals from IC of the awake-monkey. Similarly to the anesthetized monkey, the average time courses showed an overall suppression effect to sound sources in the ipsilateral side (**Fig. 5a**). For contrast between spatial sectors of the awake-monkey we collapsed across all left and all right hemifield sectors (q FDR < 0.05, $p < 10^{-3}$, cluster size > 10) and confirmed a robust contralateral bias (**Fig. 5b**). The LIwm values in the awake-monkey resembled those in the anesthetized animal showing contralateral biases (**Fig. 5c**).

Overall, our analyses showed a robust contralateral bias in both anesthetized and awake monkeys as measured with fMRI. Moreover, our finding in IC further supports the neurophysiological evidence showing hemifield coding below the cortical level[30] and attest the feasibility of fMRI as a tool for measuring spatial coding in cortical and subcortical structures of primates.

Removing ITD cues from spatial sounds

Previous, work has suggested that NBRs are related to decreases in neuronal activity[31]. Given the strong inhibitory roles involved in ITD coding at subcortical levels[32,33] we explored the effects of removing ITD cues from the original recorded sounds in the PBR/NBR pattern in AC.

First, we replicated our previous findings (**Fig. 6a**) showing PBRs and NBRs in AC for spatial sounds carrying all spatial cues in frontal azimuth (± 0 - 60°) of the anesthetized monkey (All-cues condition, q FDR <0.01, $p < 10^{-6}$, cluster size > 10 voxels, t-value range −6.4 to 8.7). Secondly, we measured the BOLD responses to the same sounds, but without ITD cues, sounds mainly carrying ILD and spectral cues (NO-ITD condition, q FDR < 0.01, $p < 10^{-3}$, cluster size > 10 voxels, t-value range −6 to 18.8). The BOLD response to leftward sounds (e.g. NO-ITD at −60°) showed greater activation in the right hemisphere as compared to the left hemisphere (**Fig. 6a**). However, the BOLD response to rightward sounds (e.g. NO-ITD at +60°) showed a bilateral activation in both left and right AC.

Moreover, while spatial tuning curves for the All-cues condition showed contralateral tuning (~ ±30°), the ITD-control condition showed a lack of contralateral tuning (**Fig. 6b**). Laterality indexes showed comparable values between −60° and −30°, however the

LIwm values near the midline (−15° to +15°) shifted drastically towards zero and just increased slightly (LIwm < 0.5) for more rightwards sounds as compared to the All-cue condition. Thus, removal of ITD cues facilitated the responses of the right hemisphere and consequently the difference in BOLD activity between both hemispheres plateaus near the midline (**Fig. 6c**).

Taken together, our results suggest that the suppression effects (either in the form of smaller positive clusters and/or negative BOLD responses) might be due to inhibitory inter-hemispheric processes brought by ITD cues. In addition, our results indicate that the lack of suppression caused by the removal of ITD cues particularly affected the right hemisphere response necessary for contralateral tuning. Overall, we provide new evidence for the role of ITD mechanisms in the representation of azimuth space in the right hemisphere of the macaque monkey.

Relating cortical representations with hemifield model

Previous work has strongly suggested that the form of coding in AC for azimuth space follows a hemifield rate code [8,9,11]. The neural population code could be investigated by representational similarity analyses (RSA) [22]. Here, RSA measures dissimilarity between the responses patterns to each spatial sector.

The beta coefficients (β) obtained from the fitted GLM to each stimulus condition (n =12) were subjected to pairwise Pearson's correlation (R) and the distance ($1 - R$) to each spatial sector was ordered into a 12 x 12 representational dissimilarity matrix (RDM). The RDM characterizes the BOLD response patterns to each spatial sector and captures distinctions within and between hemifield responses (**Fig. 7a**).

This analysis was repeated for each CF, AC and hemifield model, providing a total of 11 RDMs (**Supplementary Fig. 6**). Visual inspection of each RDM from the left

hemisphere revealed a small dissimilarity (blue) distance between spatial sectors in the right hemifield, while RDMs from the right hemisphere showed a small dissimilarity distance between spatial sectors in the left hemifield. These results largely confirm our previous results showing contralateral tuning (**Fig. 4**, **5**). More importantly however, these analyses revealed that while most regions showed variability within ipsilateral sectors, the pST region of the right hemisphere showed small dissimilarity within hemifields and a graded dissimilarity distance (red) between hemifields (**Fig. 7b**), indicating that right pST carried spatial information in the NBRs to ipsilateral sound sources. When subjecting RDMs to clustering analyses and multidimensional scaling (MDS), we were able to visualize how the right pST region segregated the response patterns in an ordered distance (**Fig. 7c, d**). Interestingly, the elicited response patterns to frontal sectors (±30°) were very dissimilar, generating a larger distance between them that further indicated drastic changes of responses around the midline.

We also examine the dissimilarity between spatial representations obtained from individual CFs (including AC) and hemifield model by computing Spearman's rank-order correlations between RDMs (1 − Spearman's *R*). This analysis resulted in a second-order RDM (see **Online Methods** and **Supplementary Fig. 6**). The second-order RDM when subjected to MDS showed how the pST clustered at a closer distance to a hemifield code (**Fig. 7e**). Using a sign-rank permutation test (FDR $p < 0.01$, 95% confidence intervals by bootstrap) we determined that the right pST RDM was significantly more similar to a hemifield model RDM than any other cortical RDM (**Fig. 7f**). Such great distinctions found for the pST of the right hemisphere suggested a distinct cortical region in macaque sensitive to auditory space and motion.

Discussion:

Using fMRI and multivariate analytical methods, we mapped auditory CFs and then measured the BOLD responses to spatial auditory stimuli. We demonstrated that the functional representation of azimuth in the macaque as measured by the BOLD signal is not topographically organized but instead distributed throughout AC. We further show that an opponent pattern of positive and negative BOLD responses across the cerebral hemispheres generates hemifield tuning and that such tuning is dependent on the suppression carried by ITD cues.

While it was originally thought ITDs (the most salient spatial cue) in mammals were coded by a topographical arrangement of coincidence detectors as in the barn owl[34,35], more than a decade of research in multiple mammalian species has revealed a different mechanism[14,36], one in which ITDs are coded by an opponent hemifield code based on neuronal inhibition[14,32,33,36]. Our results here support this view at the cortical level by showing opponent patterns of positive and negative BOLD responses across AC.

In our functional mapping experiments, positive BOLD responses showed a maximum amplitude and spatial spread for contralateral sectors (e.g. near the ear ~ ± 90° - 120°), in agreement with previous lesion[15–17], single-unit[6,7,11,19] and optical imaging data[18]. On average, the BOLD responses on each hemisphere shifted the overall amplitude near the midline, with each cortical field of the same hemisphere showing similar broad tuning to the contralateral hemifield. Although previous single-unit studies of the macaque posterior regions (in particular area CL) have reported sharply tuned neurons to spatial position[3], we found no differences in the spatial tuning curves of CFs, suggesting that neurons in these regions shared similar spatial tuning curves[12]. However, because BOLD signals reflect the average activity across a large neuronal population, it is also conceivable that sharply tuned neurons are highly distributed and can only be detected at the single-unit level. Thus it remains controversial whether posterior regions are the

only regions in AC which processes spatial information[12,27,37], a question that future studies with multiple parallel recordings across fields may resolve.

While single unit studies consistently reported a contralateral bias in the firing rate of cortical neurons[6–8,12], neuroimaging studies in humans have obtained mixed results with respect to the degree of contralaterality[8,27,28]. Here we found a robust contralateral bias in the BOLD contrast to equidistance hemifield sectors in both anesthetized and awake-monkeys, suggesting that the lack of contralaterality in some of the previous neuroimaging studies in humans might be due to differences in sound stimulation, i.e. sounds relying on ITD's[28] or ILD's cues alone and not due to an inherent lack of functional-MRI sensitivity[8]. Our stimulation design consisted of individualized head related transfer functions (HRTFs) sounds and the contralateral bias we obtained in our analyses is in accordance with human neuroimaging studies utilizing individualized HRTFs sounds[9,38,39].

Another important observation is the selective suppression of the right posterior AC during the presentation of sounds sources in the ipsilateral side, despite the overall positive BOLD responses. Particularly, in the right hemisphere of each monkey small activation patches were surrounded by a NBR in posterior and anterior regions of AC. These small ipsilateral patches could correspond to EE regions[40,41] receiving callosal input[42–44], while NBRs could be due to lateral inhibition from EE cells in ITD sensitive regions[45] or due to subcortical inhibition[32]. Importantly, NBRs were dynamic, shifting between hemispheres, which further supported the idea of a neuronal inhibitory process rather than a local vascular phenomena[31].

Given the profound role that inhibition plays in the coding of spatial cues in the brainstem[32], we investigated the effects of removing ITD cues from spatial sounds in the overall BOLD responses in AC. The response patterns no longer exhibited NBRs,

but instead a sustained PBR in the right hemisphere for sounds without ITD cues; primordially sounds based only on ILD and spectral cues. The lack of suppression caused by the removal of ITD cues particularly affected the right hemisphere tuning necessary for contralateral bias. Evidently, the spatial tuning curves showed a lack of contralateral tuning in the right hemisphere, while the left hemisphere still showed some degree of tuning, perhaps due to a similar representation for ILD cues[10,46]. Moreover, the lateralization indexes showed near zero values for NO-ITD sounds around the midline as compared to sounds carrying all spatial cues, indicating that ITD cues were necessary to preserve the shift in hemifield tuning across the hemispheres. These results are in accordance with the role of ITD cues in the right hemisphere of humans in which right hemisphere lesions results in severe deficits in sound localization[47,48] and motion detection[49] for sounds based solely on ITD cues[47].

The dynamics of the concentric activation/deactivation pattern in the right hemisphere indicated that spatial information was greatly emphasized and deemphasized in posterior regions of AC. Such dynamics allowed the right pST to code the population response patterns to a greater degree than any other cortical field. In humans the right pST region is particularly sensitive to spatial auditory motion[4,28] and our results here support this finding by showing that it could segregate the response patterns in an orderly manner similar to a hemifield rate code[8,9]. Our dissimilarity analyses provided partial support to the notion of a specialized posterior region[2,3] for processing auditory space. However, by comparison to the visual system, the format representation is fundamentally different from the spatial topography of the retina in cortex[24]. The lack of topographic organization together with our results showing positive and negative BOLD responses across hemispheres strongly support the opponent-channel model for auditory space in the macaque, rather than a place code format as generally seen in retinotopic space maps in visual cortex. Although similar functional cortical

organization for the representation of visual and auditory space was proposed more than a decade ago[2], the functional analogy is weakened by the consideration of the auditory feature analogous to "where" (position of the sensory periphery) in vision. In the auditory system, position in the sensory periphery is based on sound frequency rather than spatial position as in vision. Therefore, the functional analogy between visual and auditory space representations in primates may not reach that far.

In summary, our results indicate that the representation of auditory space relies on distributed as well as specialized mechanisms in cortical processing. The functional representation was biased towards contralateral space and dependent on the suppression effects brought by ITD cues. Suppression was pronounced in the right pST region, allowing this region to modulate its activity to a greater degree than any other field of AC and thus to code full azimuth space. Such functional specialization of the right pST cortex suggest an evolutionary preserved cortical machinery for processing auditory space in primates. Taken together, our results reconcile seemingly contradictory views of auditory space coding[50] by showing that the representation of space follows a hemifield code and that such representation generates the posterior sensitivity for space commonly seen in spatial studies[2,3,5,49] of primate auditory cortex.

Author contributions: MOR designed the study, programmed stimulus delivery and event design, conducted all recordings all data analyses and wrote the manuscript. FACA and GAK helped with event design, data acquisition for awake-fMRI experiments and revision of the manuscript. MHM helped with anesthetized experiments and revision of manuscript. DB optimized data acquisition for awake-fMRI experiments. MOR and NKL co-design this study and wrote manuscript.

Acknowledgments:

This work was supported by the Max Planck Society and a by the PIRE Grant from the National Science Foundation (OISE-0730255 to Josef P. Rauschecker). We would like to thank Yin Yu for extensive discussion on the nature of BOLD signals. We would also like to thank Thomas Steudel for helping during anesthetized experiments and Mirko Lindig for animal handling and anesthesia. We thank also Vishal Kapoor and Pawel Kusmierek for comments on previous version of this manuscript.

References:

1. Romanski, L. M. *et al.* Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* **2,** 1131–6 (1999).

2. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11800–6 (2000).

3. Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. Functional specialization in rhesus monkey auditory cortex. *Science* **292,** 290–3 (2001).

4. Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H.-J. & Scheich, H. A movement-sensitive area in auditory cortex. *Nature* **400,** 724–726 (1999).

5. Warren, J. D., Zielinski, B. A., Green, G. G. R., Rauschecker, J. P. & Griffiths, T. D. Perception of sound-source motion by the human brain. *Neuron* **34,** 139–148 (2002).

6. Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E. & Recanzone, G. H. Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *J. Neurophysiol.* **96,** 3323–3337 (2006).

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

8. Werner-Reiss, U. & Groh, J. M. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. *J. Neurosci.* **28,** 3747–3758 (2008).

9. Salminen, N. H., May, P. J. C., Alku, P. & Tiitinen, H. A population rate code of auditory space in the human cortex. *PLoS One* **4,** (2009).

10. Magezi, D. A. & Krumbholz, K. Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J. Neurophysiol.* **104,** 1997–2007 (2010).

11. Stecker, G. C., Harrington, I. A. & Middlebrooks, J. C. Location coding by opponent neural populations in the auditory cortex. in *PLoS Biology* **3,** 0520–0528 (2005).

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

13. McAlpine, D. & Grothe, B. Sound localization and delay lines - Do mammals fit the model? *Trends in Neurosciences* **26,** 347–350 (2003).

14. Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of sound localization in mammals. *Physiol. Rev.* **90,** 983–1012 (2010).

15. Jenkins, W. M. & Masterton, R. B. Sound localization: effects of unilateral lesions in central auditory pathways. *J. Neurophysiol.* **47,** 987–1016 (1982).

16. Heffner, H. & Masterton, B. Contribution of auditory cortex to sound localization in the monkey (Macaca mulatta). *J. Neurophysiol.* **38,** 1340–1358 (1975).

17. Nodal, F. R., Bajo, V. M. & King, A. J. Plasticity of spatial hearing: behavioural effects of cortical inactivation. *J. Physiol.* **590,** 3965–3986 (2012).

18. Nelken, I. *et al.* Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *J. Neurophysiol.* **99,** 1928–1941 (2008).

19. Middlebrooks, J. C., Clock, A. E., Xu, L. & Green, D. M. A panoramic code for sound location by cortical neurons. *Science* **264,** 842–844 (1994).

20. Formisano, E. *et al.* Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* **40,** 859–869 (2003).

21. Petkov, C. I., Kayser, C., Augath, M. & Logothetis, N. K. Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* **4,** 1213–1226 (2006).

22. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2,** 4 (2008).

23. Barton, B., Venezia, J. H., Saberi, K., Hickok, G. & Brewer, A. A. Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 20738–43 (2012).

24. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56,** 366–383 (2007).

25. Logothetis, N. K., Guggenberger, H., Peled, S. & Pauls, J. Functional imaging of the monkey brain. *Nat. Neurosci.* **2,** 555–562 (1999).

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

27. Zatorre, R. J., Bouffard, M., Ahad, P. & Belin, P. Where is 'where' in the human auditory cortex? *Nat. Neurosci.* **5,** 905–909 (2002).

28. Krumbholz, K., Hewson-Stoate, N. & Schönwiesner, M. Cortical response to auditory motion suggests an asymmetry in the reliance on inter-hemispheric connections between the left and right auditory cortices. *J. Neurophysiol.* **97,** 1649–1655 (2007).

29. Wilke, M. & Lidzba, K. LI-tool: a new toolbox to assess lateralization in functional MR-data. *J. Neurosci. Methods* **163,** 128–36 (2007).

30. Groh, J. M., Kelly, K. A. & Underhill, A. M. A monotonic code for sound azimuth in primate inferior colliculus. *J. Cogn. Neurosci.* **15,** 1217–1231 (2003).

31. Shmuel, A., Augath, M., Oeltermann, A. & Logothetis, N. K. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat. Neurosci.* **9,** 569–577 (2006).

32. Grothe, B. New roles for synaptic inhibition in sound localization. *Nat. Rev. Neurosci.* **4,** 540–550 (2003).

33. McAlpine, D., Jiang, D. & Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.* **4,** 396–401 (2001).

34. JEFFRESS, L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **41,** 35–39 (1948).

35. Knudsen, E. I. & Konishi, M. A neural map of auditory space in the owl. *Science* **200,** 795–797 (1978).

36. McAlpine, D. & Grothe, B. Sound localization and delay lines - Do mammals fit the model? *Trends in Neurosciences* **26,** 347–350 (2003).

37. Zatorre, R. J. & Penhune, V. B. Spatial localization after excision of human auditory cortex. *J Neurosci* **21,** 6321–6328 (2001).

38. Palomäki, K. J., Tiitinen, H., Mäkinen, V., May, P. J. C. & Alku, P. Spatial processing in human auditory cortex: The effects of 3D, ITD, and ILD stimulation techniques. *Cogn. Brain Res.* **24,** 364–379 (2005).

39. Tiitinen, H. *et al.* Neuromagnetic recordings reveal the temporal dynamics of auditory spatial processing in the human cortex. *Neurosci. Lett.* **396,** 17–22 (2006).

40. Imig, T. J. & Brugge, J. F. Sources and terminations of callosal axons related to binaural and frequency maps in primary auditory cortex of the cat. *J. Comp. Neurol.* **182,** 637–660 (1978).

41. Reser, D. H., Fishman, Y. I., Arezzo, J. C. & Steinschneider, M. Binaural interactions in primary auditory cortex of the awake macaque. *Cereb. Cortex* **10,** 574–584 (2000).

42. Pandya, D. N. & Rosene, D. L. Laminar termination patterns of thalamic, callosal, and association afferents in the primary auditory area of the rhesus monkey. *Exp. Neurol.* **119,** 220–234 (1993).

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

44. Hackett, T. A. Callosal connections of the parabelt auditory cortex in macaque monkeys. *Eur. J. Neurosci.* **11,** 856–866 (1999).

45. Brugge, J. F. & Merzenich, M. M. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J. Neurophysiol.* **36,** 1138–1158 (1973).

46. Grothe, B. & Pecka, M. The natural history of sound localization in mammals – a story of neuronal inhibition. *Front. Neural Circuits* **8,** (2014).

47. Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M. & Clarke, S. Hemispheric competence for auditory spatial representation. *Brain* **132,** 1953–1966 (2009).

48. Bisiach, E., Cornacchia, L., Sterzi, R. & Vallar, G. Disorders of perceived auditory lateralization after lesions of the right hemisphere. *Brain* **107 (Pt 1,** 37–52 (1984).

49. Griffiths, T. D. *et al.* Evidence for a sound movement area in the human cerebral cortex. *Nature* **383,** 425–427 (1996).

50. Belin, P. & Zatorre, R. J. 'What', 'where' and 'how' in auditory cortex. *Nat. Neurosci.* **3,** 965–966 (2000).

ONLINE METHODS

Subjects

Two male rhesus monkeys *(Macaca mulatta*) (M1 and M2, weighing 6-8 kg) were used for anesthetized functional imaging. The anesthesia regiment[25] ensured a bottom-up driven response without the possibility of eye-movement modulations in AC. For experiments in awake-monkeys we used two additional females monkeys (M3 and M4) weighing 8 kg each. In monkeys used for awake-experiments we implanted a custom-made PEEK headpost (polyetheretherketone; TecaPEEK, Ensinger, Inc., Nufringen, Germany) and fixed it to the skull with ceramic screws (zirconium oxide Y_2O_3-TPZ 5x1, Pfannenstiel, Bad Toelz, Germany) and bone cement (Palacos, Merck Biomaterial GmBh, Darmstadt, Germany). Surgical procedures were carry out under general anesthesia with isoflurane (1 - 2%) following pre-anesthetic medication with glycopylorate (i.m. 0.01 mg/kg) and ketamine (13 mg/kg).

For experiments with anesthetized monkeys a head mask was constructed prior to fMRI experiments with thermoplastic polymer material to hold the head in place.

Surgical procedures were approved by the local authorities (Regierungspräsidium Tübingen) and are in compliance with the German law for the protection of animals and guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Binaural recordings and design

Monkeys were lightly anesthetized (ketamine 0.2 ml + domitor 0.4 ml) inside a MRI-chair placed inside a sound-insulated acoustic chamber (Illtec, Illbruck Acoustic GmbH, Germany). In-ear miniature microphones (Danish Pro Audio 4060) were placed at the entrance of the ear canals of the animal. A broadband noise signal (0.12 - 16 kHz, 100 ms in duration) was generated in Matlab (MATLAB 7.10, *MathWorks, Natick, USA*) at a sampling rate of 48 kHz/16-bit and played through a loud speaker (Apple Pro

M653170, 2.2 cm radius) mounted on a custom-made circular frame around the MRI-chair. The signals were played every 5° (72 horizontal angle steps) from −180° to +180° (full circle) at 0° elevation from the interaural plane. A full horizontal plane was recorded from four distances (20, 30, 40 and 50 cm) from head-center for a total of 288-recorded horizontal angles. The signals measured ~82 dB SPL at 20 cm and ~70 dB SPL at 50 cm from the center of the monkey's head (Bruel and Kjaer 4188, 2238 mediator SPL meter).

The recorded signals from the microphones were pre-amplified (Safari Pro, Focuswrite) and recorded using Adobe Audition (AU) CS6 (AU, Adobe, San Jose, CA) on a MacBook Pro (Apple Inc.). The recorded noise bursts were processed in AU for noise reduction using the frequency editing tool and concatenated every 5° to form 12 spatial sectors (**Fig. 2g**). For example, positions referring to a 0° to 30° sector the stimuli were concatenated in the following way: From 0° at a distance of 50 cm to 15° at a distance of 20 cm every 5° to form a looming pattern and from 15° at a distance of 20 cm to 30° at a distance of 50 cm to form a receding pattern; total duration of "motion" pattern = 1200). The same pattern was applied inversely (30°/50 cm to 15°/20 and from 15°/20 to 30°/50). This dynamic/directionality pattern of 2400 ms was repeated 3 times (total time = 7200 ms). Such pattern was used to avoid adaptation in the BOLD responses, to control for directionality (e.g. towards ear/away from ear) and to introduce dynamic "motion" into the perception of horizontal position. For the stimulus manipulation of ITD we calculated interaural delay between left and right microphone signals using cross-correlation and removed the computed lag from either the left or the right microphone signal by subtraction.

Auditory stimulation and presentation

Auditory stimuli were played through electrostatic in-ear headphones (SRS-005S +SRM-252S, STAX, Ltd., Japan), mounted on ear-mold impressions of each monkey's pinna and covered with sound attenuating foam (Tempur-Pedic, KY, U.S.A). The acoustical noise of the MRI scanner (peak intensity, ~105 dB SPL) was reduced by ~32 dB SPL by utilizing earmolds, earmuffs and a custom-made acoustic insulated-helmet covering the entire head. More importantly however, we used sparse-sampling acquisition of 10 seconds intervals to avoid the contamination of scanner noise into the hemodynamic signal of interest (**Fig. 1a**). To match loudness, all sounds were played through the presentation equipment and re-recorded with a microphone probe (Bruel and Kjaer 4188, 2238 mediator SPL meter) positioned inside the MRI scanner.

Stimuli for tonotopic mapping consisted of 250 ms pure tones (PT), 1/3-octave and 1-octave band-pass noise burst (1/3-oct BPN 1-oct BPN) with center frequency every octave from 0.125 to 16kHz. For tonotopy stimuli, the recordings obtained from the microphone were further filtered with an inverted macaque audiogram to simulate the effect of different ear sensitivity at multiple frequencies. The stimuli were finally equalized so that they produced equal maximum root mean square (RMS) amplitude (using a 200-ms sliding window) in filtered recordings. During experiments, all stimuli were played using a QNX real-time operating system (QNX Software Systems, Ottawa, Canada), amplified (Yamaha, AX-496) and delivered at a calibrated RMS amplitude of ~80 dB SPL.

Behavioral training for awake-monkey fMRI

Monkeys assigned to awake-fMRI experiments (M3 and M4) were trained to sit still in an MRI-compatible primate chair placed inside an acoustic shielded box simulating the scanner environment. Inside the box, the animals were trained to be accustomed to wear

headphone equipment and to hear simulated scanner noise, presented by a loudspeaker. Eye movements were monitored using an infrared eye-tracking system (iView, SensoMotoric Instruments GmbH, Teltow, Germany). We considered the monkeys ready to be scanned when the amplitude of motion detected by a custom-made body sensor did not exceed amplitude of breathing in 95% of the scanning time (approximately 2 hours). Typically, while being trained or scanned in the absence of any visual stimulation in darkness, the monkeys kept their eyes closed resembling a light sleep condition.

Anesthesia for fMRI

Anesthesia procedures have been described elsewhere[25]. In brief, anesthesia was induced with a cocktail of short-acting drugs (fentanyl at 3 lg/kg, thiopental at 5 mg per kg, and the muscle relaxant succinyl-choline chloride at 3 mg/kg) after premedication with glyco-pyrolate (i.m. 0.01mg/kg) and ketamine (i.m. 15 mg/kg). Anesthesia was then maintained with remifentanil (0.5^{-2} ug/kg/min) and the muscle relaxant mivacurium chloride (5 mg/kg/h). Physiological parameters (heart rate, blood pressure, blood oxygenation, expiratory CO2 and temperature) were monitored and kept in desired ranges with volume supplements. Data acquisition started approximately ~2 h after the start of animal sedation.

MRI data acquisition

Images for anesthetized experiments were acquired with a vertical 7 T-magnet (Bruker, BioSpin GmbH, Ettlingen, Germany) equipped with a 12-cm custom-made quadrature volume coil covering the whole head. Sparse sampling was implemented to avoid

contamination of the measured stimulus-specific BOLD response by the scanner-noise-evoked BOLD response. Thus, the auditory stimulus was presented without acoustic interference from gradient-switching noise, typical of a continuous fMRI design.

For functional data, gradient-echo echo planar images (GE-EPI) were acquired with 4-segments shots (TR = 500 ms, TE = 18 ms, flip angle= 40°, FOV = 96 X 96 mm^2, matrix= 128 X 128 voxels, slices = 9 - 11, slice thickness = 2 mm, resolution = 0.75 x 0.75 x 2 mm voxel size) with slices aligned for each experimental session along the superior temporal gyrus (STG). Followed by the functional scans, two in-session volumes (FLASH and RARE) were acquired with the following parameters: for RARE (TE= 48 ms, TA= 24 ms, TR = 4000 ms, flip angle = 180°, FOV =96 X 96, matrix= 256 X 256 voxels, resolution= 0.375 x 0.375 mm, slice thickness = 2mm, slices = 9-11. For FLASH (TE= 15 ms, TA= 24 ms, TR = 2000 ms, flip angle = 69°, FOV =96 X 96, matrix= 256 X 256 voxels, resolution= 0.375 x 0.375 mm, slice thickness = 2mm, slices = 9-11. For tonotopic mapping experiments with anesthetized animals, 14 EPI runs (120 volumes) were acquired for M1 over one experimental session (days) and 16 runs (120 volumes) for M2 over one session; while for azimuth space experiments, 14 runs (150 time points each) were acquired for M1 over two sessions and 23 runs for M2 over two sessions.

For overlaying functional images, we acquired a high-resolution structural scan under general anesthesia using a T1-weighted three-dimensional (3D) MDEFT pulse sequence (4 segments, TR = 15 ms, TE = 5.5 ms, flip angle = 16.7 ms, FOV = 112 x 112 x 60.2 mm; matrix = 320 x 320 x 172 voxels, number of slices = 172, resolution = 0.35 x 0.35 x 0.3 5 mm voxel size). A total of 6 scans were acquired with the parameters above to form an average MDEFT high-resolution volume.

Measurements for awake-experiments were made on a vertical 4.7 T-magnet (Bruker, BioSpin GmbH, Ettlingen, Germany) equipped with a 12 cm quadrature volume coil covering the whole head. We acquired functional images with 360 volumes per run for each monkey (GE-EPI sequence: TR = 1000 ms, TE = 18 ms, flip angle = 53°, FOV = 96 X 96 mm^2, matrix = 96 X 96 voxels, number of slices = 18, slice thickness = 2 mm, resolution = 1.0 x 1.0 x 2 mm). For azimuth space experiments in awake-monkeys, 5 runs (360 volumes) were acquired for M3 over one session. Given that 3 volumes were acquired in sparse sampling and the emitted power at any given sequential volume was different but comparable across sparse blocks we separated firsts, seconds and thirds volumes and created 3 separate time courses (120 volumes each) per run. Thus a total of 15 runs per monkey were analyzed.

Similar to the anesthetized experiments, slices were aligned along the STG. In the end of each session, an MDEFT sequence customized for awake-experiments was taken (TE= 15 ms, TA= 840 ms, TR = 2320 ms, flip angle = 20°, FOV =96 x 96 x 80 mm, matrix = 192 x 192 x 80 voxels, slice thickness = 1 mm, resolution= 0.5 x 0. 5 x 1 mm).

FMRI phase-mapping analyses

All data analyses were performed using AFNI (http://afni.nimh.nih.gov/afni), FreeSurfer (http://surfer.nmr.mgh.harvard.edu/*),* SUMA (http://afni.nimh.nih.gov/*)* and Matlab (MathWorks). We used the coherence of the fMRI time series at the stimulus presentation cycle and measured the strength of the BOLD response amplitude in each voxel. Coherence measures the ratio of the amplitude at the fundamental frequency to the signal variance[23,24]. The measure of coherence is

$$C = A(f_0)(\sum_{f_{0-\frac{\Delta f}{2}}}^{f_{o+\frac{\Delta f}{2}}} A\,(f)^2)^{\frac{1}{2}}$$

where f_0 is the stimulus frequency, $A(f_0)$ the amplitude of the signal at that frequency, and Δf the bandwidth of frequencies around the fundamental. For all tonotopy stimuli f_0 corresponds to twelve cycles (1/120 = 0.008 Hz) and Δf corresponds to the frequencies around the fundamental excluding the first and second harmonic (see **Supplementary Fig. 1b** for an example of voxel harmonics). In the case of the spatial mapping f_0 corresponds to twelve cycles (1/180 = 0.007 Hz). Each voxel was given a coherence threshold value of 0.3. The phase response at f_0 encodes the sound frequency or azimuth in degrees in case of the spatial domain (see **Fig. 1e** for an example of three voxels in A1).

GLM analyses

Pre-processing included slice-timing correction, spatial-smoothing (1.5 mm full width at half-maximum Gaussian kernel) and scaling of the time series at each voxel by its mean. For awake-fMRI data, motion correction was used to correct volumes that contained motion shifts > 0.5 mm and/or rotations > 0.5 deg and removed them from further analyses. Lastly, we used linear least-squares detrending to remove nonspecific variations (i.e. scanner drift). Following preprocessing, data were submitted to general linear modeling analyses. The model included 12 stimulus-specific regressors and six estimated motion regressors of no interest for awake-fMRI data. For each stimulus condition (sectors 1 to 12) we estimated a regressor by convolving a one-parameter gamma distribution estimate of the hemodynamic response function with the square-wave stimulus function. We then performed t-tests contrasting each azimuth sector

condition with baseline ("silence" trials) and between equidistance spatial sectors to quantify contralateral biases.

The average anatomical scans (n = 6) were spatially normalized; the head and skull removed and extracted brains were corrected for intensity non-uniformities from the radiofrequency coil. After intensity corrections, the volumes were segmented to obtain the white and gray matter. The white and gray matter were aligned to the in-session FLASH or MDEFT anatomical scans. Whole brain surfaces were then rendered along with the extracted surfaces of the STG using Freesurfer (**Fig. 1b**). Finally, we illustrate the results on a semi-inflated cortical surface extracted with SUMA to facilitate visualization and identification of cortical regions and boundaries.

Laterality index

Significant activations (q FDR > 0.05) from the two hemispheres were used to calculate a laterality index (LI), with a positive index indicating a left-hemisphere bias and a negative index indicating a right-hemisphere bias. Given that LIs show a threshold dependency we measured LI curves to provide a more comprehensive estimate over a whole range of thresholds and to ensured that lateralization effects were not caused by small numbers of highly activated voxels across hemispheres. The LI curves were based on the t-values obtained from each spatial sector condition and were calculated using the LI-toolbox[29] with the following options: $\pm$5 mm mid-sagittal exclusive mask, clustering with a minimum of 5 voxels and default bootstrapping parameters (min/max sample size 5/10000 and bootstrapping set to 25% of data). The bootstrapping method calculates 10,000 times LIs using different thresholds ranging from 0 until the maximum t-value for each condition. For each threshold a cut-off mean value is

obtained from which a weighted mean (LIwm) index value can then be calculated[29]. This analysis returns a single value between −1 and 1 referring to a right- or left-sided hemispheric bias. Indexes between −0.25 and +0.25 were used to exclude a lateralization bias. Indexes higher than +0.5 or bellow −0.5 were designated strongly lateralized.

Spatial tuning curves

Circular statistics and spatial tuning curves were performed using CircStat toolbox for MATLAB (https://philippberens.wordpress.com/code/circstats/). The spatial spread of the BOLD response to each azimuth sector was used to calculate spatial tuning curves. The total number of voxels per CFs was used to calculate the percentage of significant active voxels (q FDR < 0.05) either positive of negative per azimuth sector. The center angles of each sector were converted from degrees to angle in radians using the function *circ_g2rad*. Descriptive statistics, mean, resultant vector length, variance, standard deviation and confidence intervals (see **Supplementary Table 1**) were calculated using the following functions: *circ_mean*, *circ_r*, *circ_var*, *stats* and *circ_confmean* respectively. A Rayleigh test was applied to all circular data with the function *circ_rtest*, to test whether data was uniformly distributed around the circle or had a common mean direction. We confirm that the deviations from circular uniformity were all highly significant (Rayleigh test, $p < 2.3^{-20}$) for all CFs accepting the alternative hypothesis of a non-uniform distribution.

FMRI dissimilarity analyses

Representational similarity analyses (RSA) were performed using the MATLAB toolbox for RSA (http://www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/).

The beta coefficients (β) obtained from the fitted GLM to each stimulus condition (n =12) were subjected to pairwise Pearson's correlation (R) and the distance ($1 - R$) to each spatial sector was ordered into a 12 x 12 representational dissimilarity matrix (RDM) (**Fig. 7**). This analysis was repeated for each CF, AC and hemifield model, providing a total of 11 RDMs (**Supplementary Fig. 5a, b**). For further analyses, we average the RDMs for each session and monkey, resulting in one RDM for each CFs or hemisphere and model. To visualize the geometry of the responses without assuming any categorical-structure we used multidimensional scaling (MDS). MDS arrange the spatial position of sound sources in two-dimensions such that the distance among them reflects the dissimilarities between the response patterns they elicited. Similarly hierarchical clustering was used to visualize the subdivisions in responses patterns. However, unlike MDS this method assumes the existence of some structure but not a particular arrangement.

For the hemifield code RDM (**c**) we used the ITD delay functions for pair-wise correlations (**Supplementary Fig. 5c**) and linearly combined noisy estimates of the ITD RDM with a categorical-model RDM (**Supplementary Fig. 5d**). Because RDMs are symmetric along the diagonal we measured the relationships between the upper triangles of the matrices by calculating the dissimilarity distance (1 − Spearman's R) obtaining a second-order dissimilarity matrix (**Supplementary Fig. 6**). We then used multidimensional scaling (MDS) on the second-order RDM to visualize the similarity distances between cortical representations and hemifield model (**Figure 7e**). Statistical inference among RDMs was performed using stimulus-label randomization and the pair-wise comparisons among them were based on bootstrap resampling of stimulus set. The noise ceiling of the hemifield code model indicated the expected performance given the noise in the modeled RDM.

References:

50. Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R. & Inati, S. Eye position influences auditory responses in primate inferior colliculus. *Neuron* **29,** 509–518 (2001).

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

52. Kusmierek, P. & Rauschecker, J. P. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J. Neurophysiol.* **102,** 1606–22 (2009)

Figure Legends:

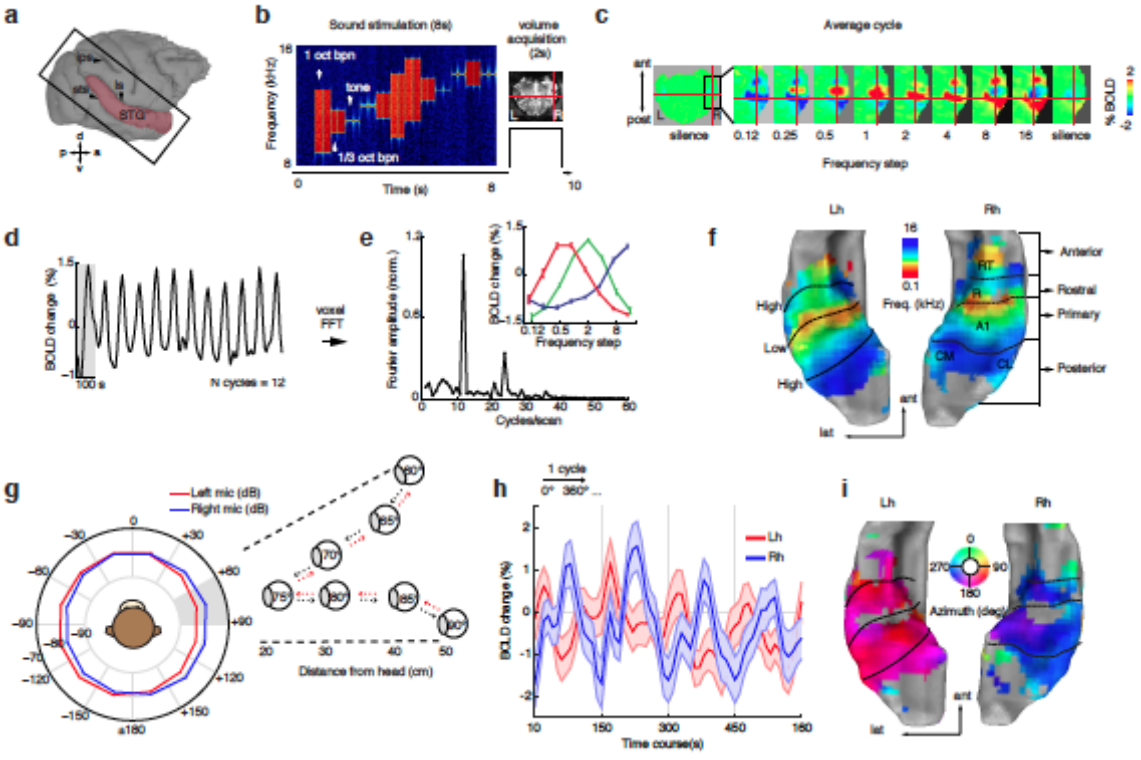


Fig. 1 Phase-mapping for frequency and space. (**a**) Image acquisition plane and extracted surface (red). (**b**) Sparse imaging and stimulation design (e.g. high frequency stimuli, 8-16 kHz). (**c**) Average BOLD response to each frequency step in octaves (labeled frequency refers to the upper range of the frequency presented). (**d**) Time

course of a voxel in A1 (crosshair in c) tuned to high frequency. Gray shading represents one cycle. (**e**) Fourier transform of the same voxel shows a peak at the stimulation rate (0.01 Hz = 12 cycles/1200 s). Inset panel, mean ± SEM of 3 voxels in A1 at the peak stimulation rate. Response peaks were used to calculate the phase of the preferred sound frequency independently at each voxel. (**f**) Tonotopic maps rendered into STG surfaces of each hemisphere. Black dotted lines indicate mirror reversal boundaries between fields. (**g**) Binaural recordings and stimulation design. Polar plot of mean amplitude from sounds (broad-band noise 0.2-16 kHz) recorded at each ear (red and blue). Outset panel illustrates virtual speaker orientations and distances. Sounds were played dynamically every 5° in a leftward, rightward and distance motion pattern (dashed red and black arrows) within a 30° spatial sector (shaded gray, n sectors = 12). (**h**) Mean and ± SEM of all significant voxels (coherence > 0.3) in AC shown for four cycles of the time course to aid visualization of the overall broad amplitude modulation across hemispheres. (**i**) Space map at the stimulation rate (0.008 Hz = 12 cycles/ 1800 s) highlights two phases across hemispheres. STG, superior temporal gyrus; cs, central sulcus; ls, lateral sulcus; ips, intraparietal sulcus; sts, superior temporal sulcus; Lh, left hemisphere; Rh, right hemisphere; ant, anterior; lat, lateral; post, posterior.

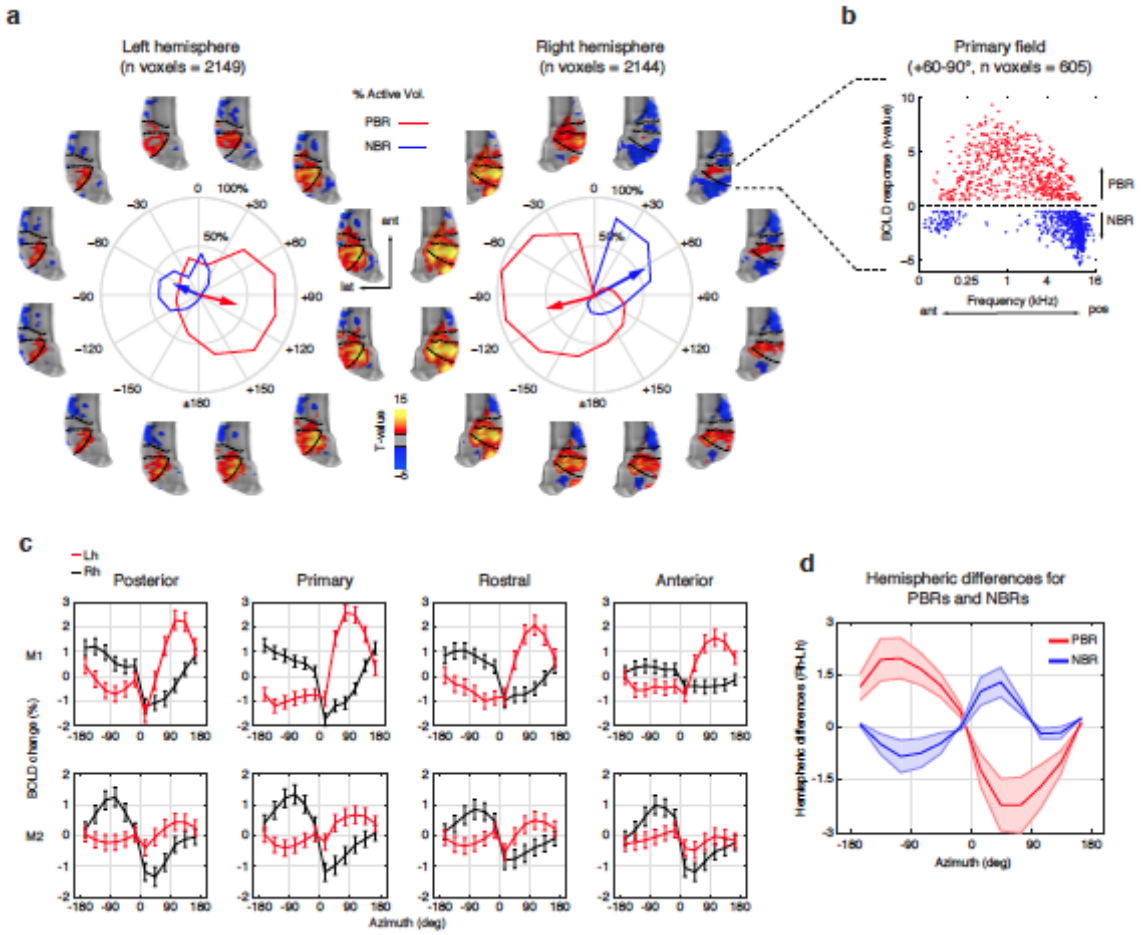


Fig. 2. Positive and negative BOLD responses represent opposite hemifields. (**a**) Activation t-maps with significant positive (red/yellow) and negative (blue) BOLD responses (q FDR < 0.05, $p < 10^{-6}$, cluster size > 10 voxels). Each map is shown around the corresponding spatial sector in polar plots of each hemisphere of monkey M2 (see **Supplementary Fig. 4** for a similar plot in monkey M1). The polar plot shows spatial tuning curves obtained from the spatial spread of the positive (red) and negative (blue) BOLD responses (PBRs and NBRs respectively). Mean resultant vectors (arrows) points towards the preferred angular direction. The length represents the percentage of active voxels around the mean direction. Negative angles (−180° -0°) in polar plot represent the left hemifield and positive angles (+180° - 0°) the right hemifield. (**b**) Scatterplot of voxels in primary field showing PBRs and NBRs to an exemplar spatial sector (+60° - 90°) plotted as function of the frequency tuning of each voxel. (**c**) Mean

and ± SEM of BOLD responses (including both PBRs and NBRs) for cortical fields of each hemisphere (Lh, red; Rh, black) of both monkeys M1 (top) and M2 (bottom). (**d**) Average BOLD amplitude differences across hemispheres for PBRs and NBRs plotted as a function of azimuth. The differential response shows opposite polarity between hemifields with a peak in NBRs for frontal right sectors. Lh, left hemisphere; Rh, right hemisphere; ant, anterior; lat, lateral.

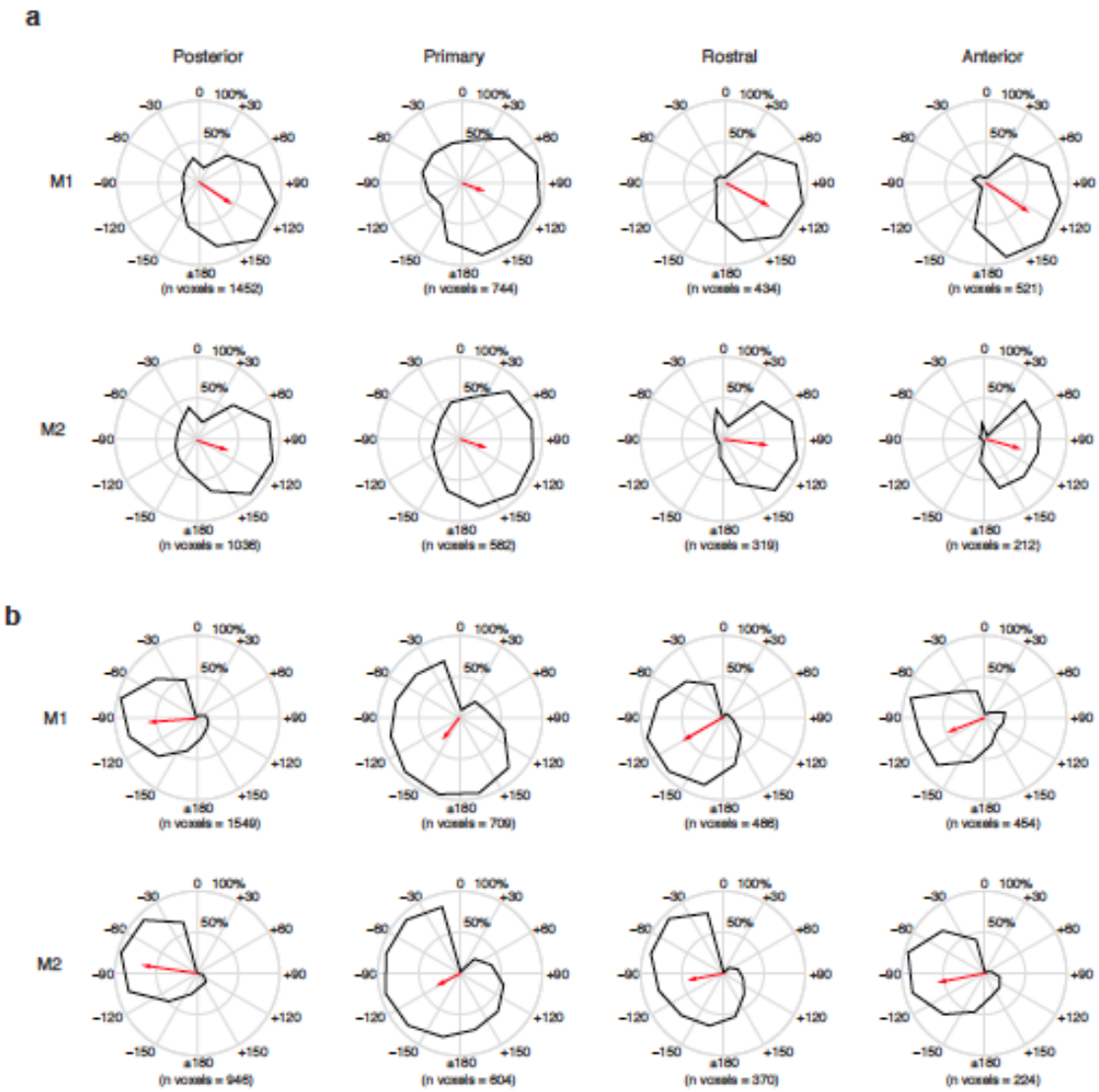


Figure 3 Cortical fields are broadly tuned to contralateral space. The spatial spread of the positive BOLD response was used to calculate spatial tuning curves (black curves) for each cortical field: Posterior, primary, rostral and anterior. (**a**) Left hemisphere for

M1 (top panel) and M2 (bottom panel). (**b**) Right hemisphere for M1 (top panel) and M2 (bottom panel). The mean resultant vectors (red) points towards the preferred angular direction and the length represents the percentage of active voxels around the mean direction. All fields were approximately oriented around ± 90° - 120°. Overall, cortical fields were broadly tuned, with central fields (primary and rostral) slightly broader than anterior and posterior fields (see **Supplementary table 1** for details).

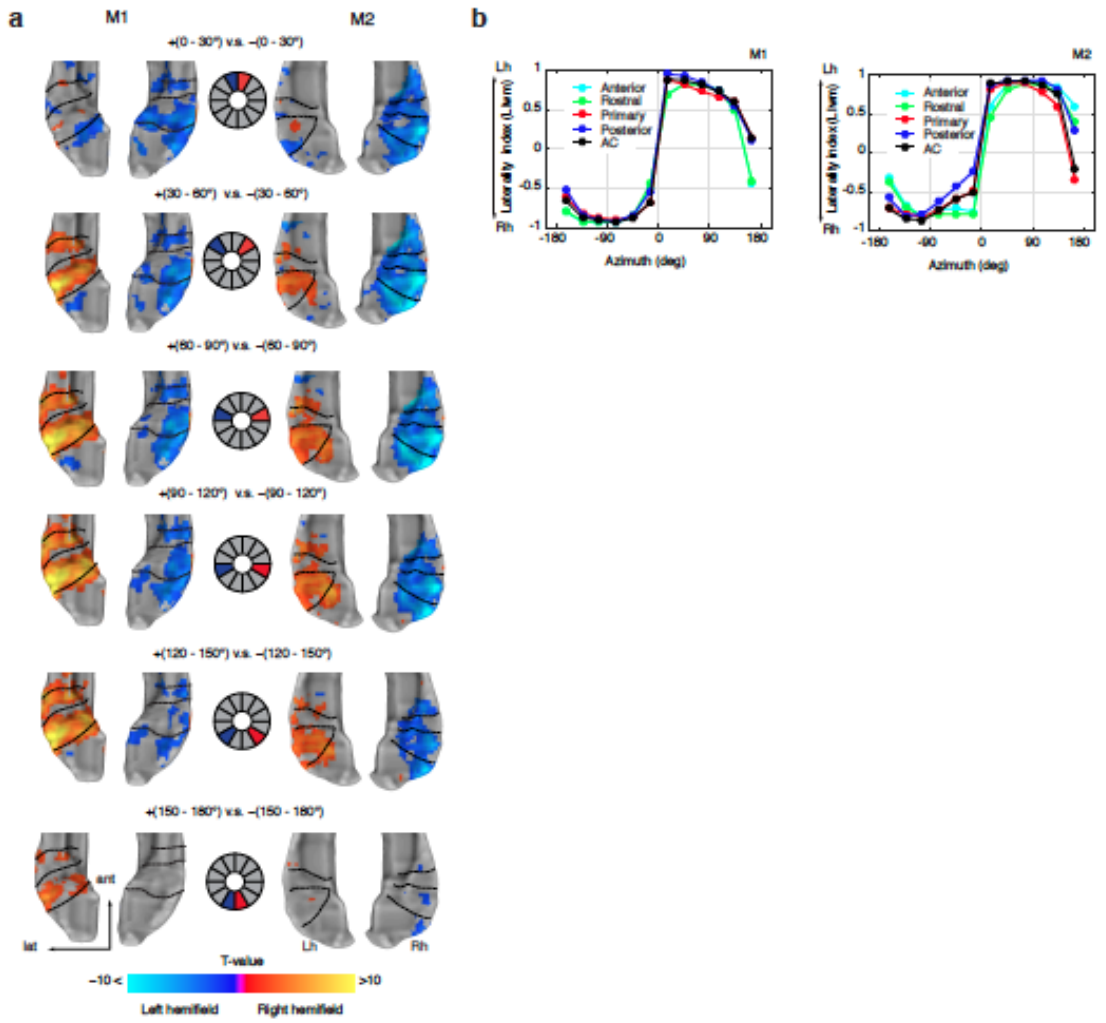


Fig. 4 Auditory cortex represents the contralateral hemifield. (**a**) Contrast t-maps between equidistance sectors for both monkeys. Middle panel illustrates the contrast design between sectors (left hemifield in blue; right hemifield in red). Voxels preferring left hemifield sectors were mapped negatively (blue-to-cyan) while voxels preferring right hemifield sectors were mapped positively (red-to-yellow). The range of t-values (q

FDR < 0.05, $p < 10^{-3}$, cluster size > 10 voxels) in the color bar was scaled according to a maximum t-value of 10 to illustrate the strength of the contrast across sectors and monkeys. (b) Mean-weighted laterality index (LIwm) between hemispheres calculated from the t-value threshold of each spatial sector (see **Online Methods**). Index range between −1 and +1 with a positive value indicating Lh biases and a negative index indicating Rh biases. Indexes curves are shown for each monkey and for each cortical field, including auditory cortex as a whole (all fields combined). Lh, left hemisphere; Rh, right hemisphere.

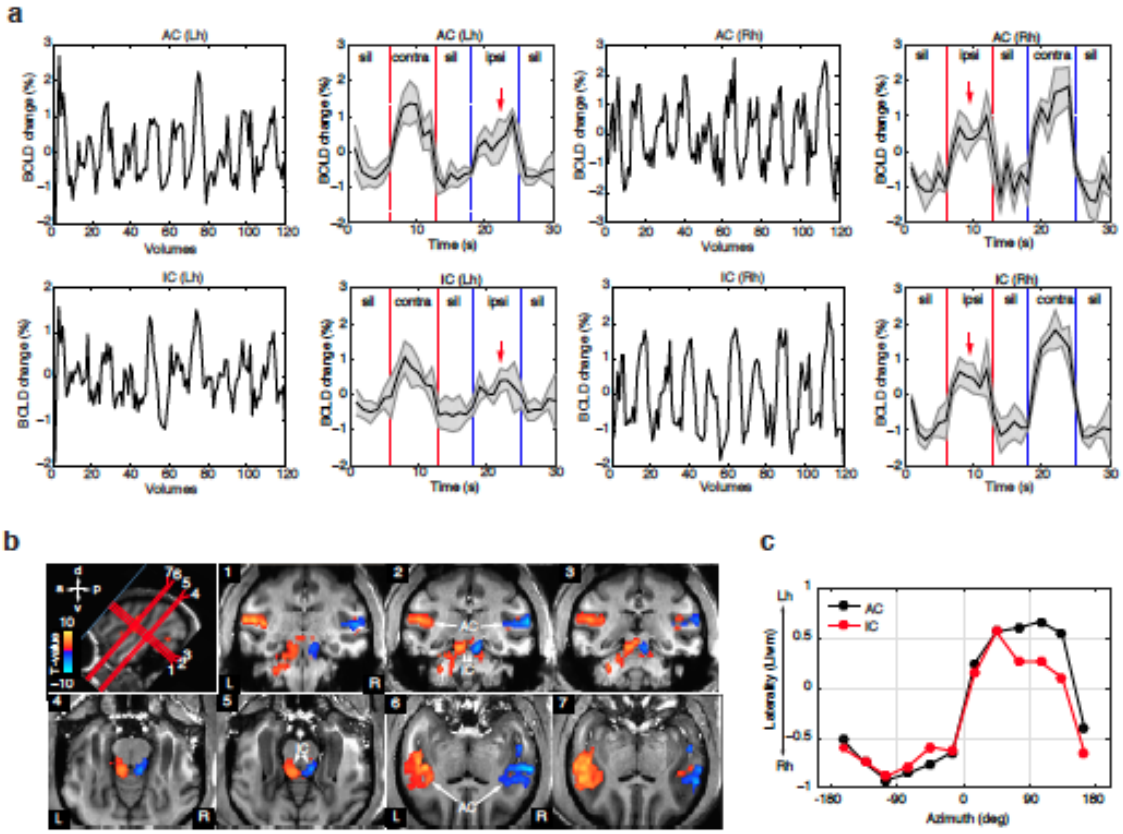


Figure 5 Cortical and subcortical hemifield tuning in the awake-monkey. (**a**) Example time courses and average cycles of voxels in auditory cortex (AC) and inferior colliculli (IC) of each hemisphere. Red dashed lines indicate duration periods of sounds presented in the right hemifield and blue dashed duration periods of sounds presented in the left

hemifield. Notice the amplitude suppression for sounds sources in the ipsilateral side. (**b**) Contrast t-maps (q FDR < 0.05, p < 10^{-3}, cluster size > 10 voxels, t-value range ± 7.8) between all left and all right spatial sectors in monkey M3. Top left image illustrates oblique slice orientations and planes (numbered 1-7) cutting through AC and IC. Voxels preferring the left hemifield sectors were mapped negatively (blue-to-cyan) while voxels preferring the right hemifield sectors were mapped positively (red-to-yellow). (**c**) Laterality index (LIwm) curves for AC and IC of awake-monkey M3.

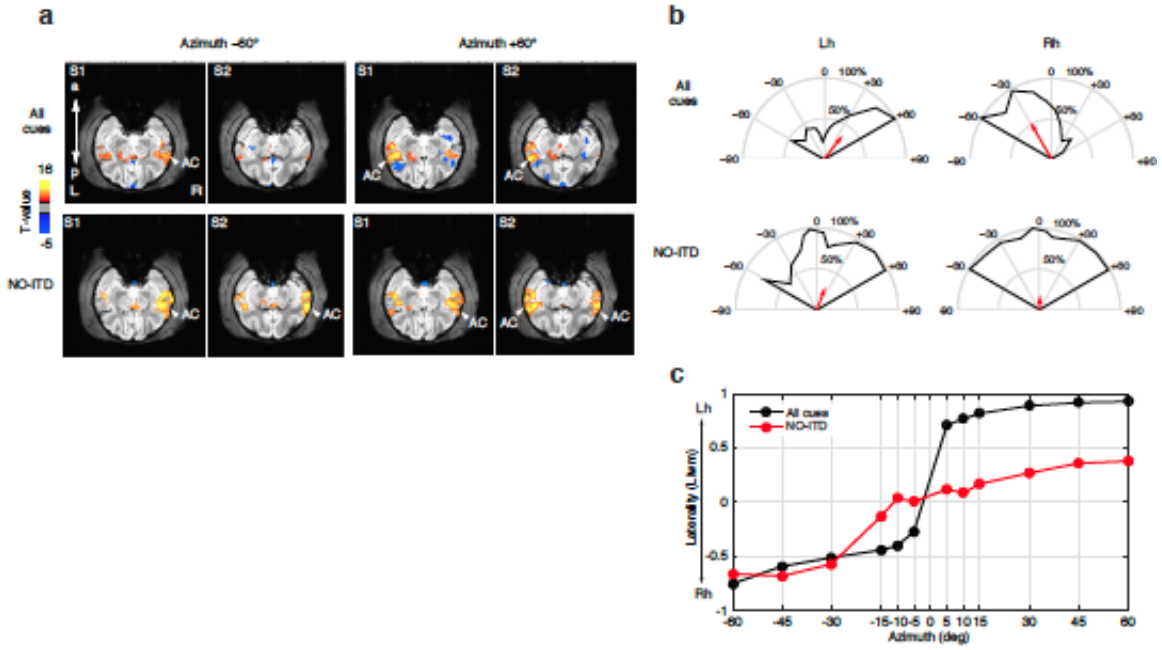


Figure 6 ITD cues are essential for contralateral tuning in auditory cortex. (**a**) Example t-maps with significant BOLD responses (q FDR < 0.05) to spatial sounds presented in left +60° and right −60° hemifields. All cues condition (top panel) and NO-ITD condition (bottom panel) in which ITD cues were removed from the original recorded sounds, sounds mainly based on ILD and spectral cues. Maps are shown for two pair of oblique slices (S1 ventral and S2 dorsal) cutting through the superior temporal gyrus. The response to rightward +60° in the NO-ITD condition was observed in both auditory cortices (e.g. no contralateral tuning). (**b**) Spatial tuning curves for frontal field show a loss of hemifield tuning in the right hemisphere for the NO-ITD condition. (**c**) Laterality

index (LIwm) as a function of frontal azimuth shows a lack of laterality (LIwm near zero) in the midline (± 15°) with only a slightly increase in laterality (LIwm < 0.5) for rightward sounds.

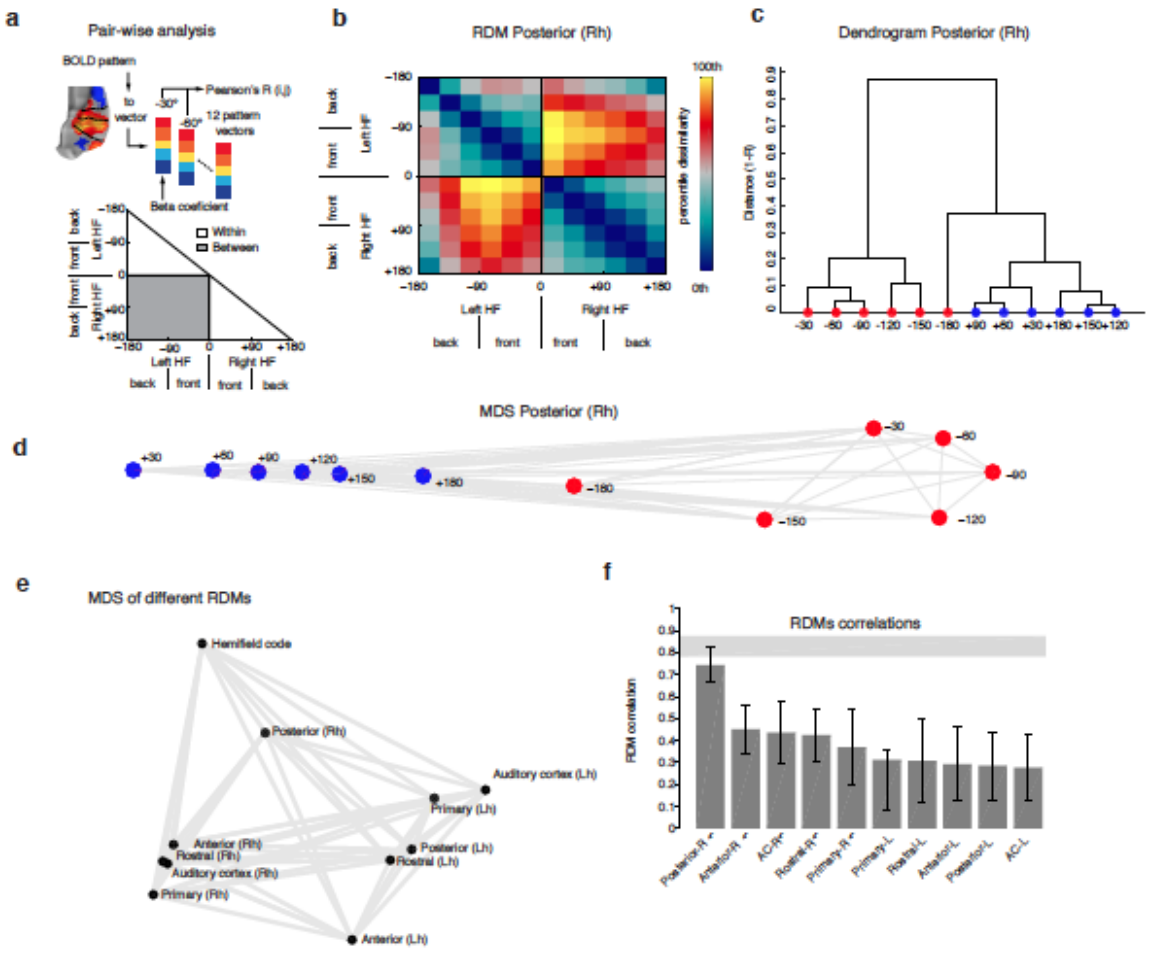
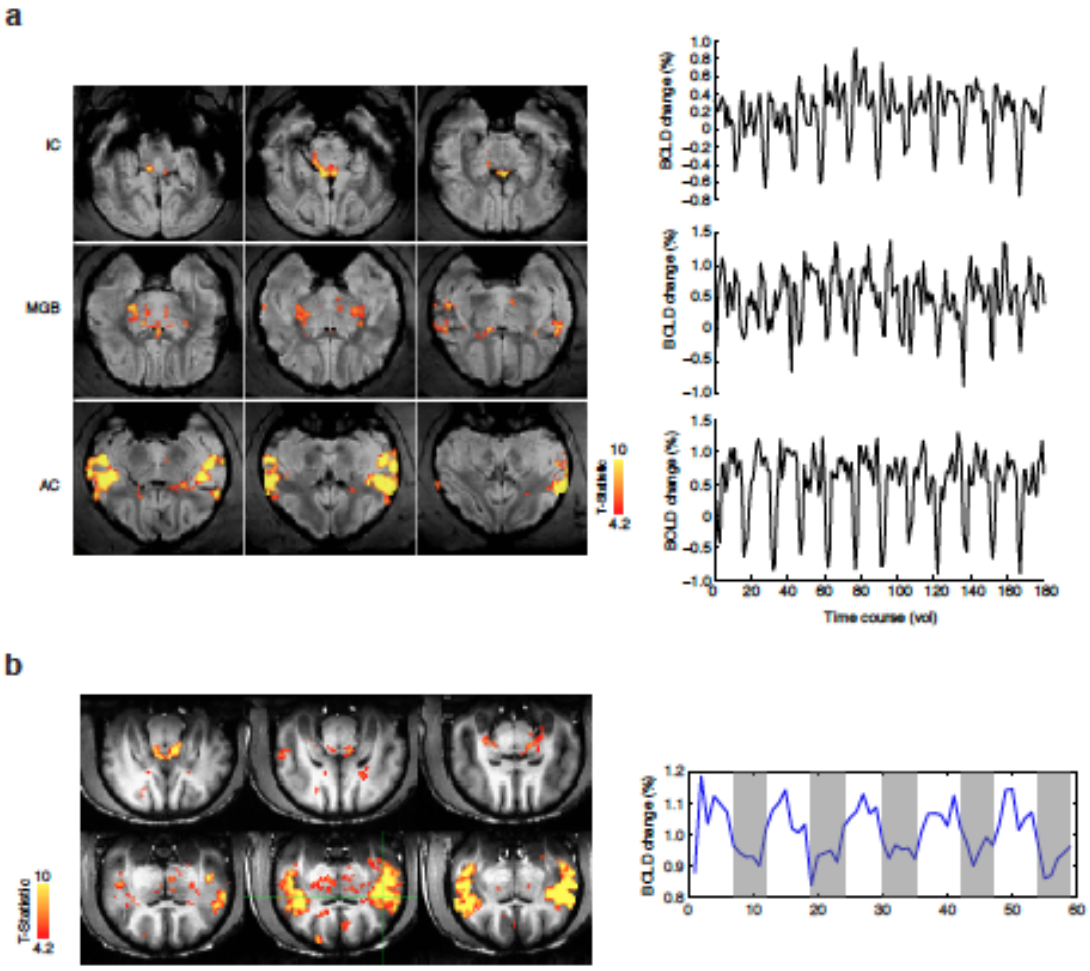
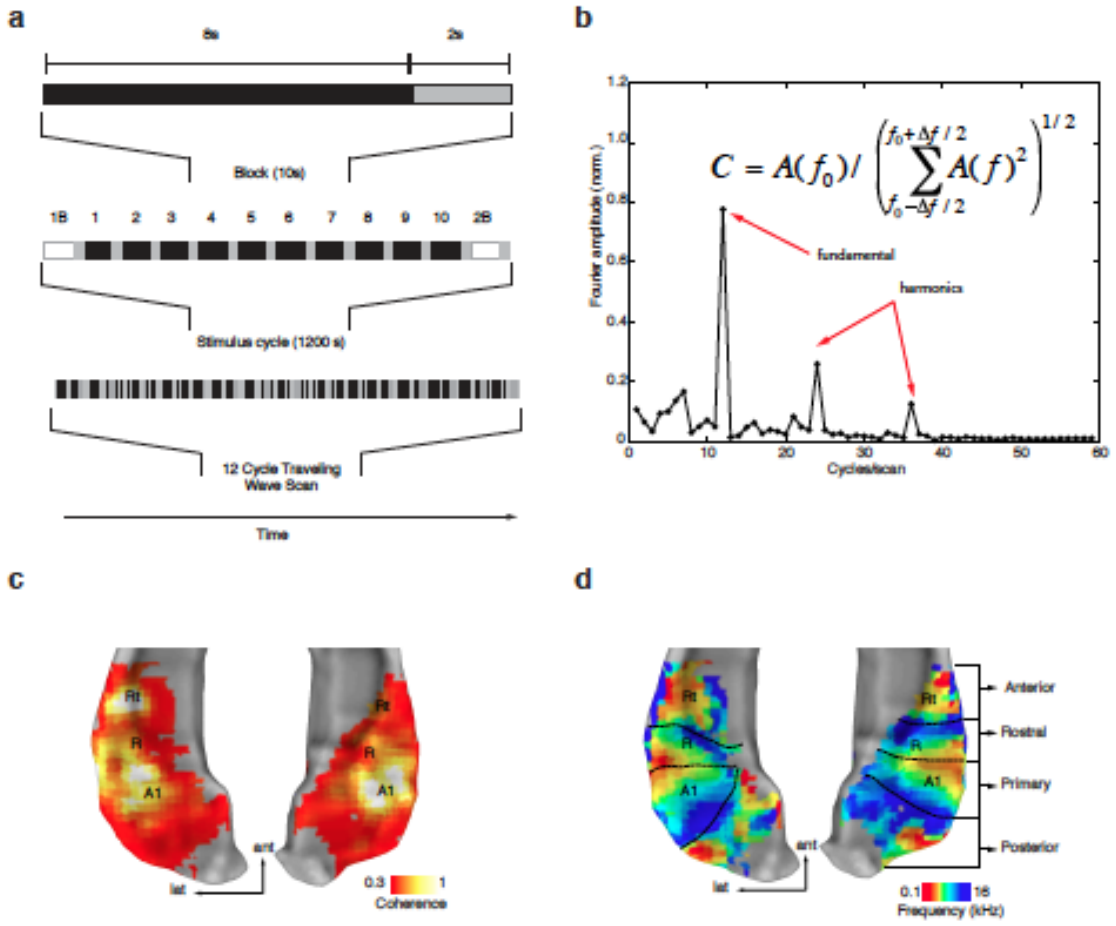


Figure 7 Posterior superior temporal region (pST) codes azimuth space based on a hemifield code. (**a**) For each field we extracted the response patterns to each spatial sector, yielding 12 response patterns. We then calculated pairwise Pearson's correlations (*R*) across all spatial sectors and then assigned the dissimilarity measure (1 − *R*) to a 12 x 12 representational dissimilarity matrix (RDM). This analysis was repeated for each cortical field and hemifield model (see **Supplementary Fig. 5** for all averaged matrices). (**b**) RDM of the right pST region. The color bar reflects dissimilarity in percentiles (low dissimilarity, blue; high dissimilarity, red/yellow). (**c**)

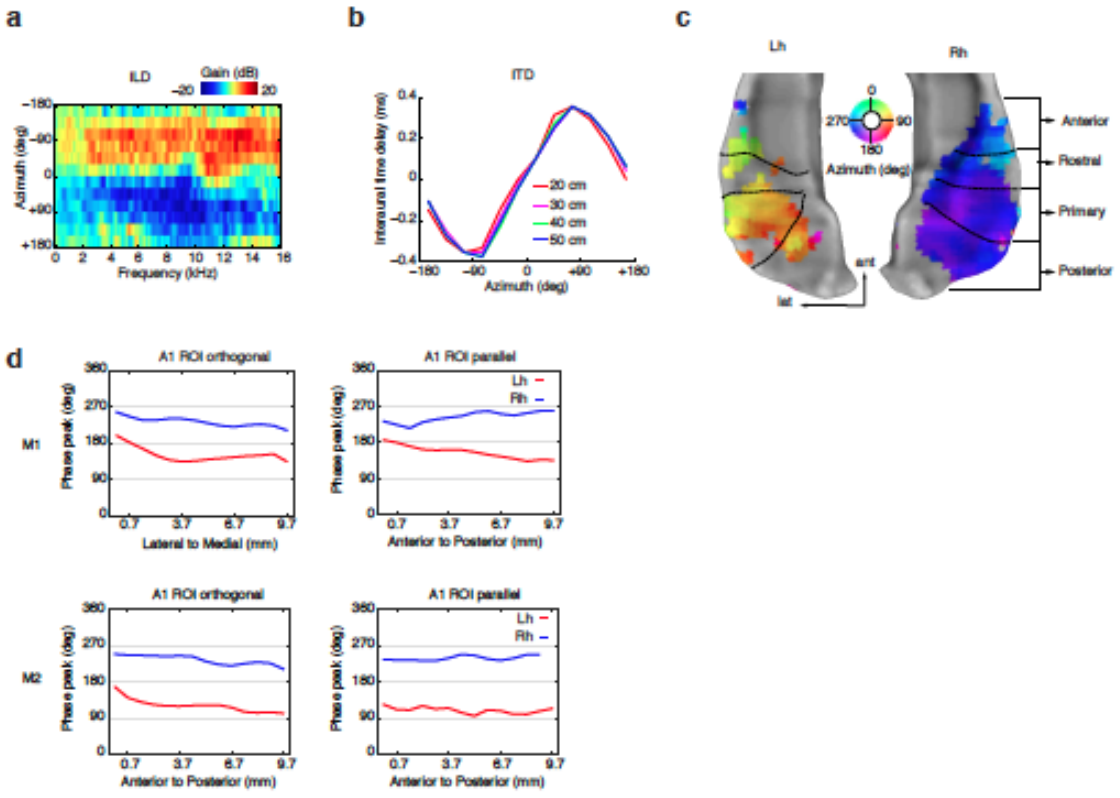
Hierarchical clustering and (**d**) MDS of fMRI responses in right pST. Unsupervised hierarchical clustering (criterion: average dissimilarity) revealed a hierarchical structure dividing left and right hemifields. MDS (criterion: metric stress) showed two groups of dots (red and blue) corresponding to each hemifield. (**e**) We compare the dissimilarity (1 − Spearman's correlation) between RDMs (see **Supplementary Fig.6** for second-order RDM). Visual inspection of the MDS structure reveals that the pST RDM lies closer in distance to the hemifield model than any other cortical region. (**e**) Cortical RDMs were tested and compared to a hemifield model RDM. Relationships were tested using stimulus-label randomization and pair-wise comparisons among cortical RDMs (along with error bars) were based on bootstrap resampling of stimulus set. Shaded gray bar illustrates the noise ceiling of the model, indicating the expected performance given the noise. The statistical comparisons show that right pST relates more to a hemifield code than any other RDM.



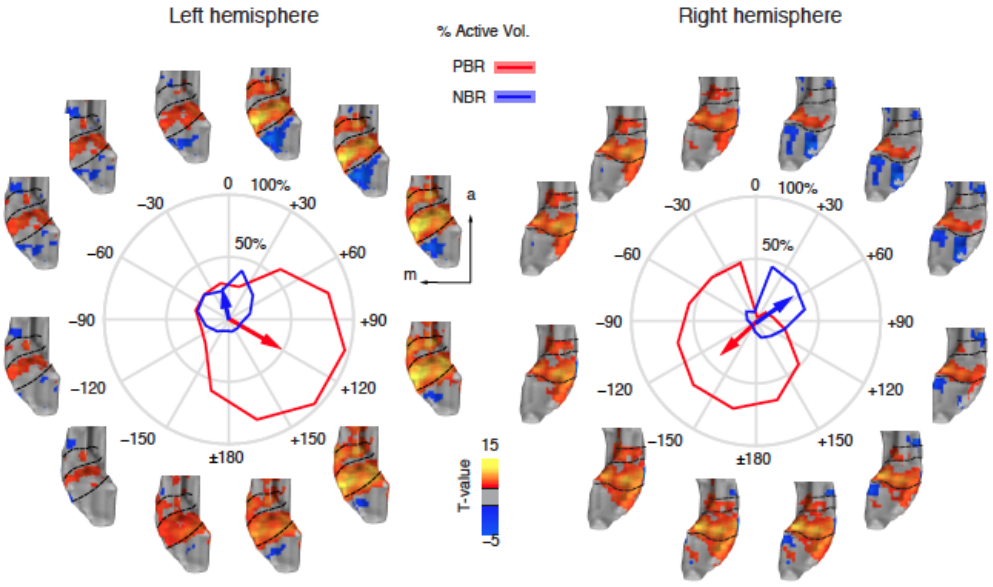
Supplementary Fig.1 Activation of the auditory pathway in anesthetized and awake monkeys. (**a**) Activation maps (q FDR < 0.05, $p < 10^{-7}$, cluster size > 10 voxels) and time course examples of voxels in auditory cortex (AC), medial geniculate body (MGB) and inferior colliculus (IC) of anesthetized monkey M1. (**b**) Overall evoked activation of awake monkey M3 (q FDR < 0.05, $p < 7.8^{-5}$, cluster size > 10 voxels) and an example of time course where the crosshair shows the response modulation of a representative voxel in A1. Gray shaded bars denote silent periods.



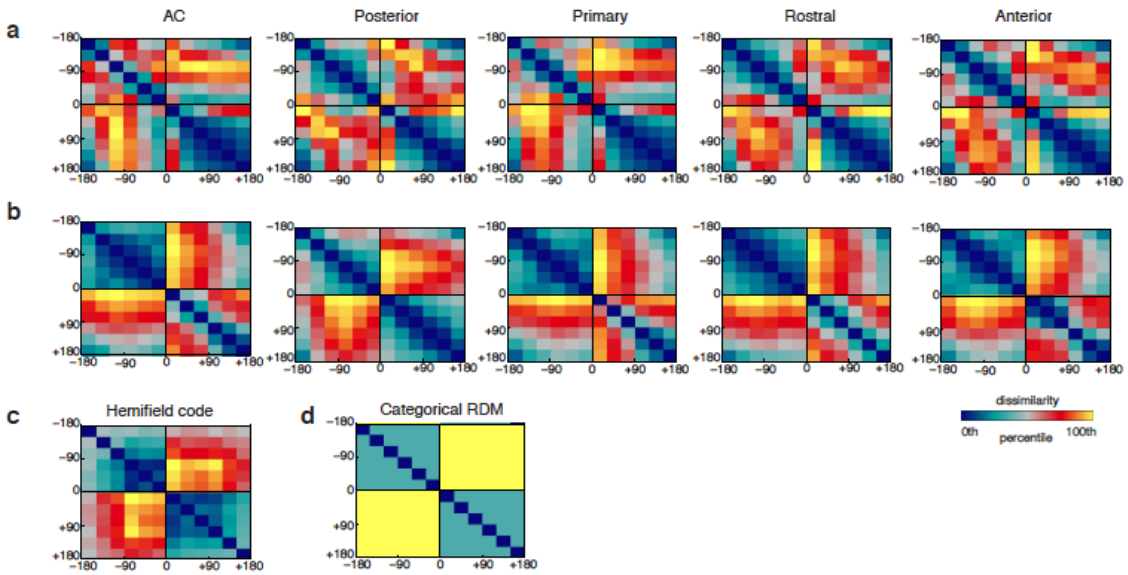
Supplementary Fig.2 Phase-mapping fMRI analyses and frequency maps of monkey M2. (**a**) Traveling wave design and stimulus presentation cycle (12 cycles/run) (**b**) The measure of coherence is equal to the amplitude of the BOLD signal modulation at the stimulus presentation rate (0.01 Hz for tonotopy, 0.08 Hz for space mapping) divided by the square root of the power over all other frequencies except the harmonics. Voxels that exceeded a coherence value > 0.3 were then assigned a phase corresponding to the voxel's peak response to the stimuli presented in the cycle. (**c**) Coherence map used to threshold the phase map. (**d**) Resulting frequency maps and reversal boundaries (black dotted lines) between the four indentified fields. These included: Posterior (Cl, Cm), Primary (Ml, A1, Mm), Rostral (Al, R, Rm) and anterior (Rtl, Rt, Rtm). Lh, left hemisphere; Rh, right hemisphere; ant, anterior; lat, lateral.



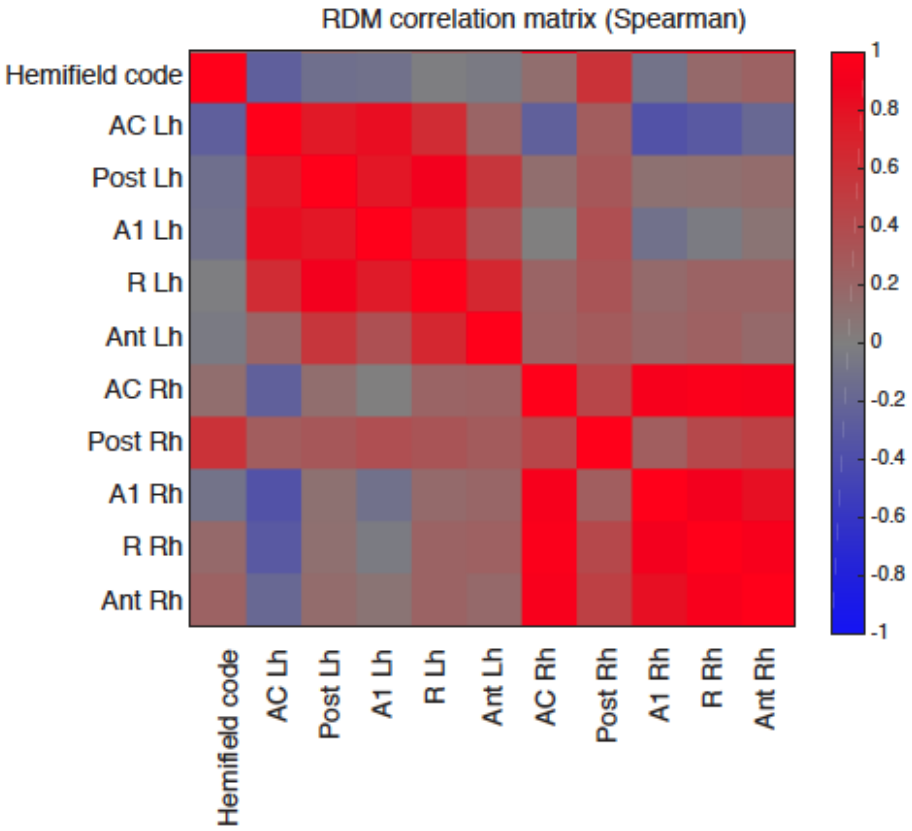
Supplementary Fig. 3 Spatial cues and maps of monkey M2. (**a**) Spectrogram of interaural level differences (L mic − R mic) of the average microphone signals within each sector. (**b**) Plot shows average interaural time delay between left and right microphone signals for each horizontal sector and distance. (**c**) Space map at 0.008 Hz (12 cycles/ 1800 s) showing the two main phases across auditory cortical fields without topographic organization. (**d**) Phase peak along cortical space spanning 10 mm across A1 orthogonal and parallel to the frequency axis shown for M1 (top) and M2 (bottom).



Supplementary Fig.4 Positive and negative BOLD responses represent opponent hemifields. (**a**) Activation t-maps with significant positive (red/yellow) and negative (blue) BOLD responses (q FDR < 0.05, $p < 10^{-4}$, cluster size > 10 voxels). Each map is shown around the corresponding spatial sector in polar plots of each hemisphere of monkey M1. The polar plot shows spatial tuning curves obtained from the spatial spread of the positive (red) and negative (blue) BOLD responses (PBRs and NBRs respectively). Mean resultant vectors points towards the preferred angular direction and the length represents the percentage of active voxels around the mean direction. Negative angles (−180° - 0°) in the polar plot represent the left hemifield while positive angles (+180° - 0°) represent the right hemifield.



Supplementary Fig.5 RDM matrices. We computed pairwise Pearson's correlations (R) between spatial sectors and then assigned the dissimilarity measure ($1 - R$) to a 12 x 12 representational dissimilarity matrix (RDM) indexed by each spatial sector. This analysis was repeated for each cortical field of both left (**a**) and right (**b**) hemispheres (including auditory cortex). For the hemifield code RDM (**c**) we used the ITD delay functions for pair-wise correlations (**Supplementary Fig. 5c**) and linearly combined noisy estimates of the ITD RDMs with a categorical-model RDM (**d**).



Supplementary Fig.6 Matrix of RDM correlations. We calculated the distance (1 − Spearman's correlation) between RDMs. The matrix is symmetric along the diagonal. Notice the Post Rh cortical region correlates best with the hemifield code RDM.

Supplementary table 1

Circular statistics for monkey M1.

CF	*N vox*	*mean*	*var*	*std*	*Upper lim*	*Lower lim*
Lh Post	1452	124	0.60	1.1	2.31	2.03

Lh A1	744	110	0.79	1.26	2.15	1.69
Lh R	434	118	0.47	0.97	2.17	1.96
Lh Ant	521	124	0.46	0.96	2.27	2.07
Rh Post	1549	-94	0.45	0.94	-1.53	-1.77
Rh A1	709	-143	0.74	1.21	-2.31	-2.67
Rh R	486	-120	0.5	1	-1.99	-2.20
Rh Ant	454	-111	0.57	1.07	-1.81	-2.08

Circular statistics for monkey M2.

CF	*N vox*	*mean*	*var*	*std*	*Upper lim*	*Lower lim*
Lh Post	1036	108	0.67	1.16	2.05	1.72
Lh A1	582	108	0.73	1.21	2.08	1.71

Lh R	319	97	0.5	1	1.82	1.58
Lh Ant	212	99	0.47	0.97	1.85	1.6
Rh Post	946	-82	0.38	0.87	-1.33	-1.52
Rh A1	604	-118	0.74	1.22	-1.86	-2.24
Rh R	430	-101	0.61	1.1	-1.62	-1.90
Rh Ant	224	-101	0.48	0.98	-1.65	-1.87