Accessing the encoding of sounds in the auditory cortex using functional UltraSound

Célian Bimbard

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure

Accessing the encoding of sounds in the auditory cortex using functional UltraSound

Soutenue par

Célian BIMBARD

Le 02 Décembre 2019

École doctorale n°158

ED3C

Spécialité

Neurosciences

Composition du jury :

Jennifer BIZLEY Professeur, University College London	*Rapporteuse, Présidente*
Josh MCDERMOTT Professeur, Massachusetts Institute of Technology	*Rapporteur*
Nancy KANWISHER Professeur, Massachusetts Institute of Technology	*Examinatrice*
Pierre POUGET Chargé de Recherche, Institut du Cerveau et de la Moelle épinière	*Examinateur*
Yves BOUBENEC Maître de Conférences, École Normale Supérieure	*Examinateur, Encadrant de thèse*
Shihab SHAMMA Professeur, École Normale Supérieure	*Directeur de thèse*

ENS | PSL

... In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.

— Suarez Miranda, Viajes devarones prudentes,
Libro IV,Cap. XLV, Lerida, 1658
Jorge Luis Borges
(On Exactitude in Science)

Abstract

The world teems with complex sounds that animals have to interpret. To do so, their brain must represent the richness of the sounds' acoustic structure, from simple to high-order features. Understanding how it does so, however, remains filled with challenges. In this thesis, this question was explored through a new technical prism, namely functional UltraSound imaging (fUSi).

First, fUSi was used to investigate with a high fidelity the topographical organization of the auditory system. The tonotopic organization from the inferior colliculus to primary and secondary areas of the auditory cortex was characterized, as well as the laminar organization of cortical fields. This imaging technique unveiled the connectivity scheme between auditory cortex and another brain area thought to affect auditory processing, namely frontal cortex. Thus, I propose that fUSi can be used to investigate the organization of sensory systems in both their entirety and their details.

Second, fUSi was used to explore how the brain represents the richness of natural sounds, in combination with computational tools designed to explore the explanatory power of acoustic models of incremental complexity. It revealed robust response components within ferret auditory cortex, that were amply explained by a canonical model of auditory cortex processing. These components differed fundamentally from human components in their processing of speech and music sounds, as they lacked sensitivity to high-order acoustic features. Even in the context of conspecific ferret vocalizations, such sensitivity was poorly represented, despite the fact that the animals perceptually relied on such features to process vocalizations. This suggests an evolutionary divergence in high-order auditory processing between ferrets and humans.

Last, the complex code underlying spatial localization was investigated across cortical areas. A two-dimensional encoding was revealed, and was consistent across primary and secondary areas within both ventral and dorsal streams, thus challenging the view of a specific 'where' pathway. Moreover, the 3-dimensional cortical organization of azimuth-sensitive voxels was exposed, revealing scattered yet clustered specialization.

Overall, this thesis aimed at developing a systematic approach to studying the functional organization of the brain, through a large-scale yet high-resolution, new imaging technique. It opened up interesting prospects for further studies, that should explore the formation and plasticity of the brain's topography. In particular, it could provide important advances to our understanding of the evolutionary and developmental aspects of speech processing.

Acknowledgements

I would first like to thank the members of the jury, and especially the reviewers of the manuscript, for investing time and energy in reading and discussing it: Jennifer Bizley, Josh McDermott, Pierre Pouget, Nancy Kanwisher.

Thanks to Shihab Shamma, for the opportunities he allowed me to seize, as well as his passion for science and the world in general and his unstoppable optimism. I would like to heartily thank Yves Boubenec, for the amount of energy he accepted to share throughout the years. I would not have learned as much as I did without his brilliant personality, his strength through thick and thin, his integrity, and above all without his friendship and kindness. Un assurage de qualité.

Thanks to all the members of the LSP lab, who have made these four years particularly mind-opening and rich for me on many aspects, through the diversity of their approaches and expertise. The favorable atmosphere of the lab allowed us isolated physiologists not to feel lost in the sea of Psychophysics, but rather enjoy its subtlety and richness. More particularly, I would like to thank the members of the ferret team: Jen Lawlor, for being one of the first stones of the edifice, and having skillfully filled that role, for the shared ups and downs and her always kindly ear; Thibaut Doisy, for his tenacious righteousness, his tolerance to administrative gaffes and his overall good-heartedness; Sam Norman-Haignere, for his enthusiasm and intensity, and his shared interest for the peculiar aspects of the living world; Jonatan Nordmark, for his precious skills and his deadpan humor; Agnès Landemard, for having made this last year much more pleasant and soothing than expected. I'd also like to thank the interns who participated to the lab's life, Constantin and Chloé, whose work has been precious.

Thanks to Mickaël Tanter and Charlie Demené, for sharing the fUS technology and for their support, in the face of technical or intellectual challenges.

I would also like to thank the Ecole Normale Supérieure, for all the opportunities and freedom of thought it provided me with. A particular thank to Mariano Casado, who has been an important figure of support from the very beginning.

Je tenais enfin à remercier tous ceux qui ont été là, indépendamment du travail de thèse, le long de la route ou à chaque pas. Merci à mes amis, quelle que soit l'époque à laquelle je les ai connus ; chaque petit bout de passion que j'ai essayé de transformer en petit bout de science venait en grande partie de vous. Certains ont joué un rôle particulier pour cette thèse. Merci à

Guy, Antonin, Maria, Rémi et Zoé, chacun pour m'avoir montré des qualités que j'ai pu ériger en modèle. À Flo, pour ses passions contagieuses, et infiniment ressourçantes, qui englobent tout à la fois la plus petite *Prosoeca* et le plus gros dôme de granit. À Sophie, pour chaque concept qu'on a essayé de tordre en tous sens, et pour sa présence indéfectible. À Simon, pour ses légumes et ses fleurs qui ont accompagné mes journées, physiquement et métaphoriquement. Un merci particulier à Timothée, mon bon vieux bidouble, sans qui je n'aurais pas pu physiquement écrire cette thèse – toi même tu sais. Enfin, merci à ma famille, pour son soutien et son amour. À Erwan et Gaëlle, pour avoir été tous les jours devant moi des sources d'inspiration. À Dad, pour chacune des enjambées faites à ses côtés. À Domi et à Étienne, pour chaque bouffée d'air que vous avez su partager. À Morgane et à Vincent, et maintenant à Noémie et Olivia, pour apporter d'autant plus de richesse à ce petit cercle.

Merci à Aurore, pour son amour, son énergie et chacune de ses attentions.

Sans tomber dans un lyrisme désuet, je tenais aussi à remercier ceux qui ne m'entendront pas : Nils Frahm, Colin Stetson, Tim Hecker et tant d'autres qui ont à leur manière imprégné cette thèse ; la Loire pour son eau brune qui a porté chacun de mes bouillonnements ; la forêt de Fontainebleau pour ses fougères et son givre les matins d'hiver ; la Bretagne pour ses embruns et ses arméries.

Contents

List of Figures

Chapter 1

Introduction

In order to thrive in the face of evolutionary pressures, animals must acquire knowledge or perform behaviors that are adapted to the external context, i.e., a world teeming with relevant or irrelevant, known or unknown shapes, sounds, colors. They often have to be able to quickly apprehend this huge diversity of incoming sensory signals, flexibly extract relevant information from them and sometimes interpolate missing parts. As humans, a glimpse of introspection reveals the genuine ease with which we can comprehend objects, recognize them, compare them and define their characteristics at both low and high levels. This capacity and this ease must rely on the clarity with which the brain is able to generate representations of the outside world that are both faithful and effortless to navigate.

These representations are thus a cornerstone of the algorithms that the brain uses for sensory processing. Marr's influential theory of computations describes three levels of analysis of complex systems: computation, algorithm, and implementation (Marr, 1982). Because most of what we access of a biological system is its behavior (from which we can derive the computation level) and its physical organization (from which we can derive the implementation level), being able to describe the latter becomes instrumental in allowing us to identify the algorithms used by the brain.

One can often think of many different implementations for a single computation. Starting from the current working hypothesis that what is important for the brain to orchestrate behavior are the bits of information that neurons exchange with a certain structure and dynamics, the brain could have had no physical organization except a complex wiring. Nevertheless, scientists have been unveiling over many years its complex spatial layout at multiple scales.

In this introduction, I will present the overall context of my thesis, which is the encoding of sounds in the auditory cortex. I will raise several outstanding questions, among which some will be directly linked to my thesis, and others will go well beyond its scope and results. However, all will

give the global framework of the direction I tried to follow. In a first part, I will expose the general organization of sensory systems, and what information we can get from this organization concerning the underlying processing. In a second part, I will provide a general commentary on the origins of this organization through three aspects: computational advantage, development, and plasticity. In a third and last part, I will finally evoke the technical challenges one has to face when studying this organization, and will present the recent technical development that enabled the research presented in this manuscript.

1.1 A topographical brain

If phrenologists, led by Franz Joseph Gall, failed to improve the field of neuroscience with any robust theory for reading one's cognitive abilities in the shape of one's skull, they largely contributed to developing the concept of the modularity of mind. The idea that specific body parts house different physical, cognitive or spiritual functions – and especially the senses seating in the brain – has a long-standing history throughout the world, from Hippocrates to Galen and Ibn Sina (Rocca, 2003; Mazengenya and Bhika, 2017). But unlike other organs, the brain contains in itself many different parts that have long remained undescribed. In this section, I will provide a global overview of the different levels of organization in the brain, to finally focus on several aspects that are specifically relevant to this thesis.

1.1.1 From large-scale to small-scale organization

Brain areas

The ideas of modularity of mind and of functional specialization of brain regions were particularly developed with the careful inspection of local brain lesions and their relationship with behavioral disorders, such as the observation by Marc Dax in 1836 of the link between neurological damage to the left frontal lobe and aphasia (Dax, 1863), later made widely known by Paul Broca and Carl Wernicke's complementary work (Broca, 1861; Wernicke, 1874). These studies, amongst others, revealed that certain areas seem to be specifically dedicated to certain brain functions. Lesions in Broca's area provoked expressive aphasia; thus this region was linked to the production of language. Wernicke's area, of which loss or damage provoked receptive aphasia, was rather linked to the understanding of language. Thus became predominent the idea that the brain contained physical 'maps' of cognitive functions.

These discoveries were based on a functional characterization of a brain region. Other large-scale maps of the brain were provided based on other brain features such as cytoarchitecture, the most famous and still in-used being Korbinian Brodmann's for humans (Brodmann, 1909). Brodmann investigated the heterogeneity in cell types and density, cell architecture, layer organization in cortical slices, and noticed that these variables differed across discrete regions of the brain. Some of these areas were actually already identified for their specific functional role. For example, Broca's and Wernicke's functional areas turned out to be Brodmann's physical areas 44/45 and 22. **Generally speaking, four different features can be exploited to explore brain modularity: behavioral or cognitive deficits due to specific lesions; local brain activity in relationship with certain behaviors or cognitive processes; incoming and outcoming connectivity; and cytoarchitecture** (Patel et al., 2014).

Remarkably, the game of finding specific, focal brain regions for specific functions has been, since then, unexpectedly fruitful – up to the discovery of incredibly precise sensory maps at the surface of the cerebral mantle.

Sensory systems

A striking characteristic of sensory systems is that the surface of sensory receptors is systematically mapped onto associated brain structures. This topographic ordering of the senses is a common property across animals. **Understanding this widespread organization scheme could thus reveal recurrent and primordial types of computations across the brain.**

The first evidence for a topographical mapping in the brain came from Hughlings Jackson's observation of the 'march' of epileptic seizures throughout the body in 1886, on which he built the idea of somatotopic representation in the motor system (York and Steinberg, 2011). It was later discovered, mainly through localized lesions studies, that somatosensory, visual and auditory systems contained ordered representation of the surface of the skin (Cushing, 1909), the retina (Inouye, 1909; Head and Holmes, 1911), and the cochlea (Larionow, 1899). These maps are respectively called 'somatotopy', 'retinotopy', and 'tonotopy'. One of the most beautiful demonstrations of the retinotopic organization of primary visual cortex comes from Tootell et al. (1988), that used appropriate stimuli and a C-2-deoxy-glucose infusion that marks activated brain regions (figure 1.1). This experiment elegantly revealed a continuous mapping of the external 2-dimensional space onto a precise, continuous pattern of activity at the surface of the cortex. Thus, brain patches that are close together within this retinotopic organization in the cortex tend to respond to similar regions

of the visual space. In the case of the auditory system, the cochlea naturally performs a frequency decomposition of the incoming sound, that is then transmitted to auditory nuclei in the brainstem up to the auditory cortex. Again, brain patches that are close together within the tonotopic organization will thus respond to a similar frequency range. Compared to the other sensory systems, the main topographic arrangement in the auditory system is single-dimensional.

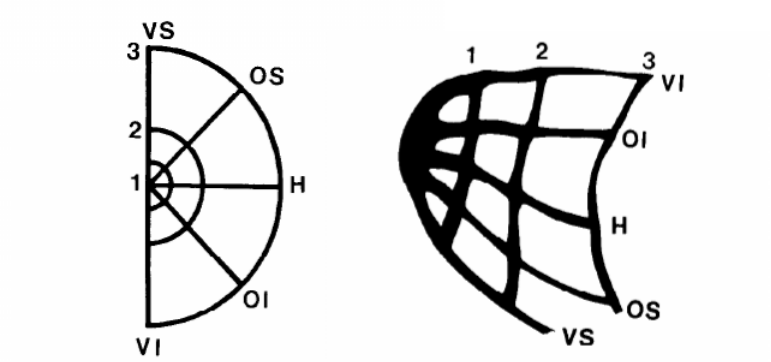

Figure 1.1: **Demonstration of the retinotopic organization of the primary visual cortex of macaque monkey.** Schematic of the retinotopic transformation from stimulus (left) to striate cortex (right). I, 2, 3, Selected regions from fovea to the periphery. VS and VI, vertical superior and vertical inferior rays, respectively. OS and OI, oblique superior and oblique inferior rays, respectively. H, horizontal meridian. – adapted from Tootell et al. (1988)

Interestingly, these maps are distorted, depending on the density of receptors present at the sensory surface (Patel et al., 2014). For example, the fovea is largely over-represented in the cortical retinotopy, and sensitive somatic regions such as hands and lips take up more space at the surface of the somatosensory cortex than an equivalent surface of the back skin. **These distortions suggest that the brain maps' shapes are actually subordinated to behaviorally relevant configurations.**

The use of the functional mapping also revealed the existence of several 'fields' within sensory systems, which are juxtaposed at the surface of the cortex and are usually defined as containing the full representation of the sensory surface (i.e., a full body map, a full visual space, or a full tonotopic axis). These fields are often organized in a hierarchical manner, with primary fields receiving direct inputs from the thalamus, and secondary fields receiving inputs from (and sending feedback inputs to) primary fields. The sensory representation within a field often gets less clear when ascending the hierarchy, as the receptive fields of the cells usually get wider (Guo et al., 2012; Elgueda et al., 2019). The existence of several fields that are topographically mapped allows for serial as well as parallel connectivity (Hackett et al., 1998), that underlies specialized processing. This is well exemplified by the parallel processing of binocularity, depth and color in visual areas 2, 3 and 4

respectively (Kaas, 1989), or for processing spectral and temporal information in primary areas of auditory cortex (Bendor and Wang, 2008). **Nevertheless, the specific computation or set of computations performed by each cortical field is poorly understood.** This is especially true in the auditory system.

Within one single field, several relevant sensory features can be represented. As an example, another fundamental acoustic dimension is the modulation of energy over time, which is determinant in natural sounds processing, such as speech. Several studies investigated how this feature was encoded in different brain regions, and revealed an overlapping, orthogonal map for temporal periodicity in inferior colliculus (Baumann et al., 2011) and auditory cortex (Langner et al., 2009; Baumann et al., 2015; Brewer and Barton, 2016). Brewer and Barton (2016) thus suggested to define auditory field maps using these two fundamental parameters. However, the existence of such a topographic arrangement for temporal periodicity in the auditory cortex is still a matter of debate (Leaver and Rauschecker, 2016). **Thus, the exact organization of the auditory structures' responses to basic stimulus features still remains elusive.**

The link between the functional and the cytoarchitectural organization is often tight. However, recent studies have demonstrated the existence of smooth anatomical gradients at sharp retinotopic borders (Gămănuț et al., 2018) or retinotopic maps that were not aligned with cytoarchitectural markers (Zhuang et al., 2017). In figure 1.2, a clear discrepancy is visible between functional and cytoarchitectural boundaries. Moreover, it is important to stress that there can be large differences across individuals of the same species (Dear et al., 1993; Nelken et al., 2004; Bizley et al., 2005). This makes the characterization of functional maps, and even more the characterization of their relationship with other encoding features more difficult to perform, since averaging across individuals can turn out to blur field maps boundaries and internal structure. From these considerations emerges the idea that **individual, large-scale mapping can be necessary to understand the brain organization at the cortical field scale**.

Cortical columns

The cortex is organized in three dimensions: two 'horizontal' dimensions, along which topography is usually visualized, and a third dimension, 'depth'. Mountcastle (1957) first demonstrated in the somatosensory cortex that along this third direction, cells tended to show less variability in their tuning properties, making cortical columns functionally quite homogeneous. In the visual

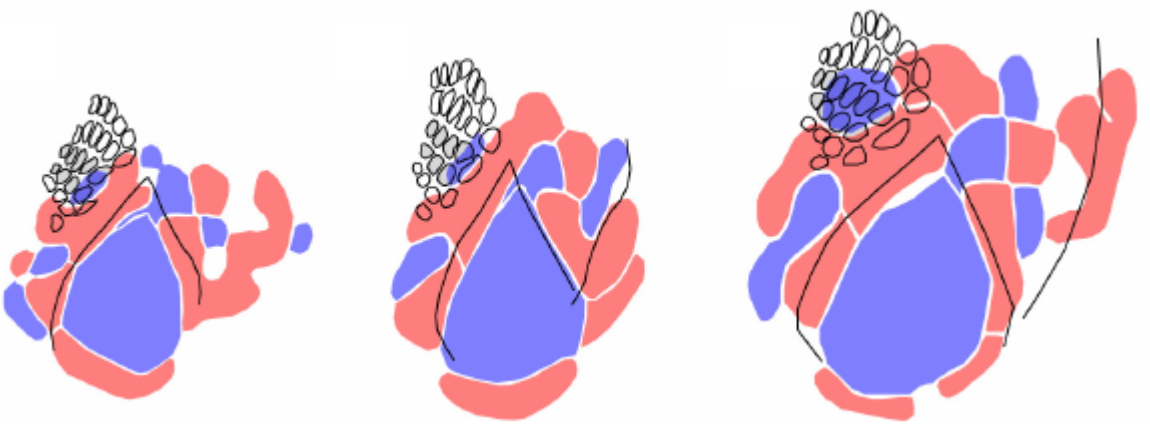

Figure 1.2: **Link between retinotopic organization and cytoarchitecture.** Alignment of functional retinotopic maps and chemoarchitectonic borders for three Emx1-Ai96 mice. Cytoarchitectural borders (black lines) are obtained with cytochrome oxidase-based chemoarchitectonic and functional maps from GCaMP6 fluorescence-based measure of neuronal activity. Functional retinotopic maps (blue and red colored regions) represent positive field sign areas (red) and negative field sign areas (blue). In short, each area contains one full retinotopic map. It clearly shows the mismatch in the location of the lateral border of V1. – adapted from Zhuang et al. (2017)

cortex, seminal work has been pursued by Hubel and Wiesel who showed that spatial selectivity and orientation sensitivity were mainly shared across neurons forming a single column (Hubel and Wiesel, 1962). This observation raises the question of whether specific computations are performed within cortical columns. This second organization of the primary visual cortex sheds light on an important property of sensory cortices: **they allow for an overlap of different sensory maps of different scales**. Interestingly, this second type of mapping does not exist in all species. Particularly, rodent models do not display any orientation preference map like carnivores, but rather a 'salt-and-pepper' organization where neighbouring cells can display very different orientation preferences. A recent study suggested that the orientation selectivity mapping was actually on a much larger scale (Fahey et al., 2019). The origins and the computational significance of such a difference are still under discussion (Kaschube, 2014). **The specific processing occuring within each of these layers is an intensive subject of research in both animal** (Marshel et al., 2019) **and human** (De Martino et al., 2015) **studies**.

Where topography ends

When scaling observations down to finer and finer features, the robustness of the topographic organization starts to crumble. As such, the reliability of the tonotopic organization has long been a subject of debate (Kanold et al., 2014). Indeed, if all researchers agree on the fact that primary auditory cortex is globally tonotopically organized, there has been strong divergence on how

robust tonotopy is at a smaller scale. Low-resolution recording techniques, such as intrinsic signal imaging, have revealed a smooth tonotopic gradient (e.g., Kalatsky et al. (2005) in the mouse, Nelken et al. (2004) in the ferret), while single-cell mappings provided much more nuanced maps (e.g., Bandyopadhyay et al. (2010); Rothschild et al. (2010) in the mouse with 2-photon imaging, Bizley et al. (2005) in the ferret with microelectrodes). These discrepancies could be explained by the differences in resolution across techniques – low-resolution methods averaging over more cells, smoother maps are obtained –, but also because of differences across cortical layers. Indeed, layer 4 in primary sensory cortices is known to receive strong inputs from sensory thalamus. As sensory information is transferred from layer 4 to layer 2/3 and then layer 5, a complex processing occurs, and cells usually get larger receptive fields (Guo et al., 2012). The topographic organization is more robust at the level of layer 4, and then becomes more patchy ('salt-and-pepper') when reaching layer 2/3 or deeper layers (Guo et al., 2012; Hackett et al., 2011; Winkowski and Kanold, 2013). At the single cell level, it is now well established that neurons receive highly heterogeneous signals at the level of single spines in primary auditory cortex (Chen et al., 2011). This heterogeneity of synaptic inputs may be the underlying mechanism for the diversity of single cortical neurons' response properties. Interestingly, the computational advantages of such a heterogeneity remain poorly known. Notably, the functional heterogeneity in layer 2/3 of auditory cortex seems to be more pronounced than the retinotopic organization in V1. It has indeed been argued that this difference could stem from the fact that auditory objects can contain several distant frequencies, while usually visual objects occupy a limited space on the retinal surface (Kanold et al., 2014).

Topography gets lost as one looks for finer details within primary cortices. **Another way to lose track of the straightforward topographic encoding is to climp up the processing hierarchy.** Strikingly, many higher-order areas are still holding back the secret of their organization, as simple topographic mapping fails to apply (Patel et al., 2014). In the auditory system, tertiary areas already lack an apparent tonotopy (Elgueda et al., 2019). Frontal cortices, which have been hypothesized to top this hierarchy (Fritz et al., 2010), have been mainly described as containing a distributed type of encoding.

1.1.2 From sensory to categorical mapping

Sensory systems contain ordered representations of low-level features. Grounded on these foundations, more complex representations can be deployed, extracting abstract features or combining information from different senses. Instead of looking at obvious physical measures of the sensory

world, one then has to look for more 'computational maps' (Knudsen et al., 1987). The aforementioned orientation preference map in the visual system is one of them, that is built upon a straightforward map of space. In the auditory domain, the map of auditory space in the owl's optic tectum is a well-studied case and resembles the retinotopic mapping of the visual system (Knudsen, 1982). This mapping also appears quite clear in the mammal superior colliculus, but gets lost in auditory cortex, where only a distributed code seems to exist (Stecker et al., 2005; Bizley et al., 2009; Wood et al., 2019). Therefore the question of a putative 'computational map' of space in mammal auditory cortex has remained elusive.

Localizing an object is one computation that the brain performs. Another is recognizing an object. In the visual domain, these two computations have been shown to be carried out along two separate processing pathways, namely the 'what' (ventral) and 'where' (dorsal) streams (Mishkin et al., 1983; Wilson et al., 1993). Similar pathways have been suggested to be present in the auditory domain, at least in some species (Romanski et al., 1999; Rauschecker and Tian, 2000; Tian et al., 2001; Lomber and Malhotra, 2008). Object recognition necessitates a complex series of computations to 'untangle' the representation of visual objects, that are thought to be performed in high-order areas (DiCarlo and Cox, 2007). However, the networks that supports such computations and their spatial organization in the brain are not fully described, especially in auditory cortices. While high-order areas often show very heterogeneous selectivity at the single-cell level, domain-like arrangements still seem to prevail over complete random distribution. For simplicity, I will include these high-order domains into the definition of 'topographic organization' for the rest of this manuscript.

Face areas / Speech areas

An extreme form of computational maps can be seen in the existence of domains that are highly selective to very specific features, while being largely invariant to other sensory properties. I mentioned before that even in sensory topographic maps, representations of the stimuli that are of particular behavioral relevance (such as the fovea in vision, or the hands in somatosensation), were over-represented. Extrapolating from here, it would thus sound straightforward that high-order features of interest, or stimuli categories that are determinant in an animal's life, are also well represented in the cortex. In the visual domain, the inferior temporal cortex (IT) has been shown to contain an area selectively responding to faces, thereby called 'fusiform face area' (FFA) (Sergent et al., 1992; Kanwisher et al., 1997). In the auditory domain, much of our use of the auditory system

as humans is for processing speech and music. Belin et al. (2000) discovered voice-specific areas in human auditory cortex, located in the areas surrounding primary auditory cortex. Another set of studies suggested that distinct pathways may exist for speech and music perception, based on their rather low-order acoustic properties such as temporal and frequency richness (Zatorre et al., 2002). It was suggested that these pathways could rely on interhemispheric differences in processing (the left hemisphere being more temporal processing, and the right one more frequency processing). However, most studies were limited by the low number of stimuli used, and univariate (voxel-wise) analysis cannot decipher different, potentially overlapping pathways within secondary areas. In particular, the main quantification used to investigate the organization of these pathways relies on the identification of 'best stimuli'. This technique does not allow one to uncover potential 'subthreshold' components, which spread over potentially large areas while still being, in single voxels, not their favorite. To overcome these limitations, a recent study used new analytical methods and a large stimulus set to extract robust and interpretable responses component from fMRI responses to natural sounds (Norman-Haignere et al., 2015) (figure 1.3-A). They revealed distinct cortical pathways for speech and music, within both hemispheres (figure 1.3-B,C). These separate pathways were spatially organized and located in non-primary areas of the auditory cortex. Interestingly, they relied upon complex, high-order acoustic features, as their responses could not be fully explained from standard parameters of acoustic processing such as frequency, spectral modulation, or temporal modulations contents (Norman-Haignere and McDermott, 2018).

One fundamental question in human neuroscience is to understand whether these specialized pathways are unique to humans, and whether their localization and properties rely on preexisting modules which could be shared across species. As an example, the existence of a speech area in humans is a strong clue for the existence of a specific pathway dedicated to speech processing. However the very particular acoustic niche that speech occupies could well produce *per se* such an artefactual pathway, unspecific of the behavioral relevance of speech in itself, but rather of some of its acoustic properties. Cross-species comparisons can shed light upon the evolutionary origins of these specialized sensory modules. In other words, **investigating the cortical organization of other species can illuminate the unique versus shared features of auditory processing.**

The presence of a face area is now well attested in macaques, in an equivalent of the human FFA (Tsao et al., 2003; Livingstone et al., 2017; Arcaro et al., 2017). It has now been widely studied, and provided interesting perspectives on the common mechanisms for face recognition and

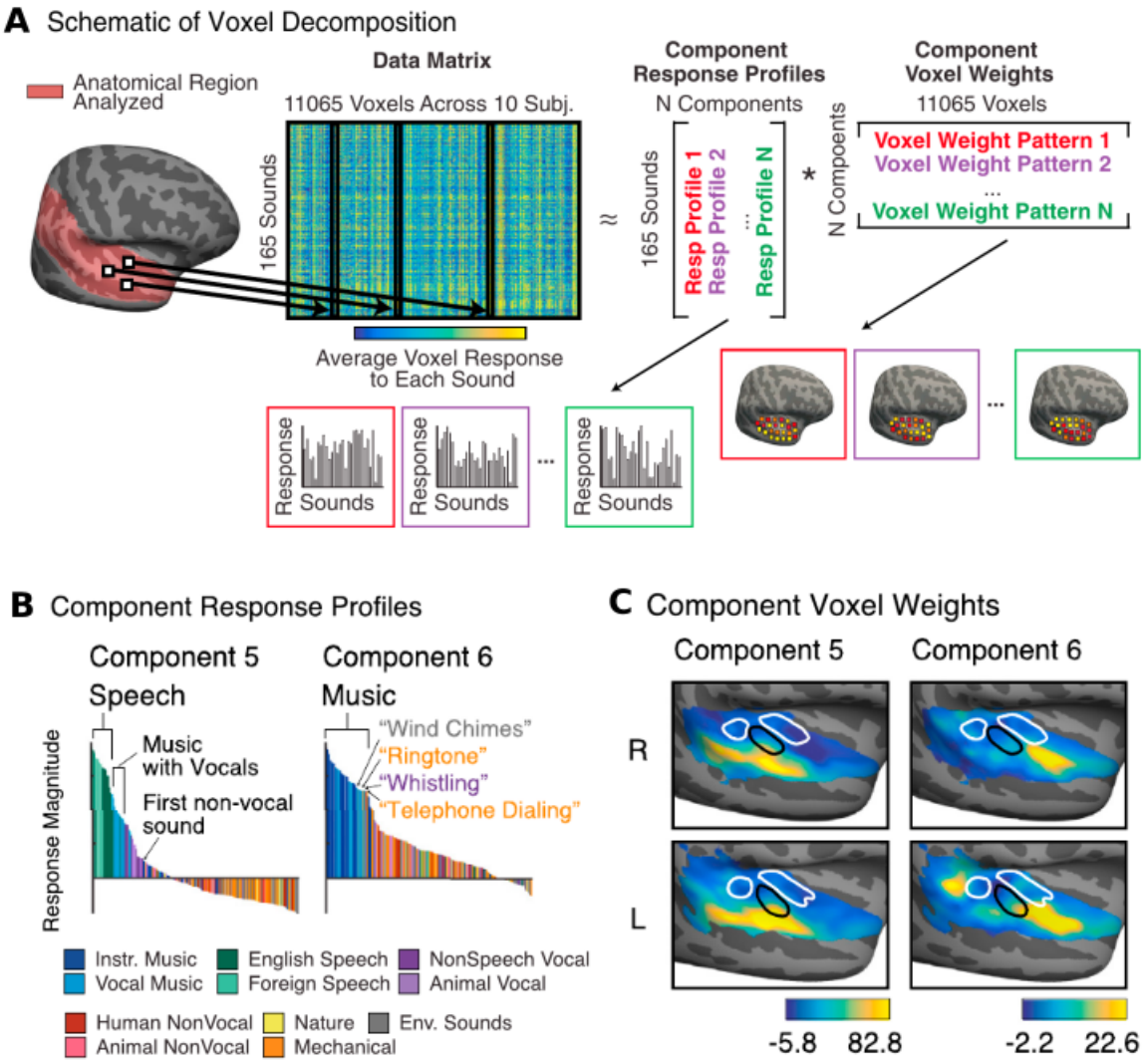

Figure 1.3: **Distinct cortical pathways for speech and music processing.** (**A**) The average response of each voxel to each sound was represented as a matrix (165 sounds 3 11,065 voxels across all ten subjects). Each column contains the response of a single voxel to all 165 sounds. Each voxel's response was modeled as the weighted sum of a set of canonical "response profiles." This decomposition can be expressed as a factorization of the data matrix into a response matrix and a weight matrix. Response profiles and weights were inferred using statistical criteria alone, without using any information about the sounds or anatomical positions of the voxels. (**B**) Response profiles for the inferred components. Each figure plots the response magnitude (arbitrary units) of each component to all 165 sounds tested. Sounds are ordered by response magnitude and colored based on their membership in one of 11 different categories, assigned based on the judgments of human listeners. Components 5 and 6 responded selectively to sounds categorized as speech and music, respectively. (**C**) Voxel weights for the speech (5) and music (6) components. Outlines of high- and low- frequency regions within primary auditory cortex are overlaid. Up: right hemisphere; bottom: left hemisphere – adapted from Norman-Haignere et al. (2015)

brain plasticity in humans and monkeys.

The case of speech and/or vocalization processing in animals is not as clear. Studying how human speech and conspecific vocalizations are processed by animals can provide complementary

information. The former may reveal which part of the speech processing pipeline existing in humans – altogether with its associated acoustic features – are shared across species. The second may clarify, in comparison to human speech processing, whether behaviorally relevant vocalizations are processed along similar pathways, and whether animals rely perceptually and neurally on equivalent acoustic features.

The question of whether speech perception evolved from acoustic processing pathways shared across animals, that were subsequently over-developed or 'hijacked' in humans, or whether it relies on totally different mechanisms ('speech is special') is still under debate (Trout, 2003; Kriengwatana et al., 2015). Notably, **whether human speech-specific domains are specific to humans is unclear.**

Conspecific vocalization processing has been further explored. In the monkey, caudolateral areas of the auditory cortex encode the highest level of spatial information and more anterolateral areas show the highest degree of specificity for monkey calls (Romanski et al., 1999; Rauschecker and Tian, 2000). Similarly to the human-voice specific area, a putative conspecific voice-specific area has indeed been uncovered in macaque monkeys, on the upper bank of the superior-temporal sulcus (Petkov et al., 2008). Vocalizations areas have also been detected in the frontal cortex of the macaque monkey (Romanski and Goldman-Rakic, 2002). However, the role of earlier, upstream areas (such as primary and secondary auditory cortices) in the processing of vocalizations is still elusive (Recanzone, 2008), despite their functional role established by lesions studies of temporal cortex, especially the left hemisphere (Heffner and Heffner, 1984). Interestingly, a fair amount of processing seems to occur in the auditory cortex, since voice-identity sensitive cortex relies on processes taking place in the adjacent temporal lobe, before gaining access to frontal cortex (Petkov et al., 2015). This specific type of processing, however, remains poorly understood. In particular, **the exact acoustic features upon which animals rely to process (perceptually and neurally) conspecific vocalizations are still largely unknown**.

Overall, categorical representations seem to top the computational hierarchy of (auditory or visual) object recognition. Understanding how they are implemented across brain areas, across modalities and across species will be key for understanding perception.

Non-sensory maps

Sensory systems, both at the level of primary and secondary areas, also contain extensive information about non-sensory variables, such as behavioral state, decision, or motor movements (Stringer et al.,

2019c; Musall et al., 2018). The spatial organization of the encoding of such variables just starts to be investigated. Minderer et al. (2019) recently found that neural activity encoding behaviorally relevant variables formed smooth gradients across large parts of the posterior cortex in the context of a visuo-spatial task (figure 1.4). Notably, the gradients of task encoding contrasted with the sharp boundaries suggested by retinotopic mapping (figure 1.4-B). The exact origins of such mappings, and how they unfold in relationship with existing sensory maps, is unknown.

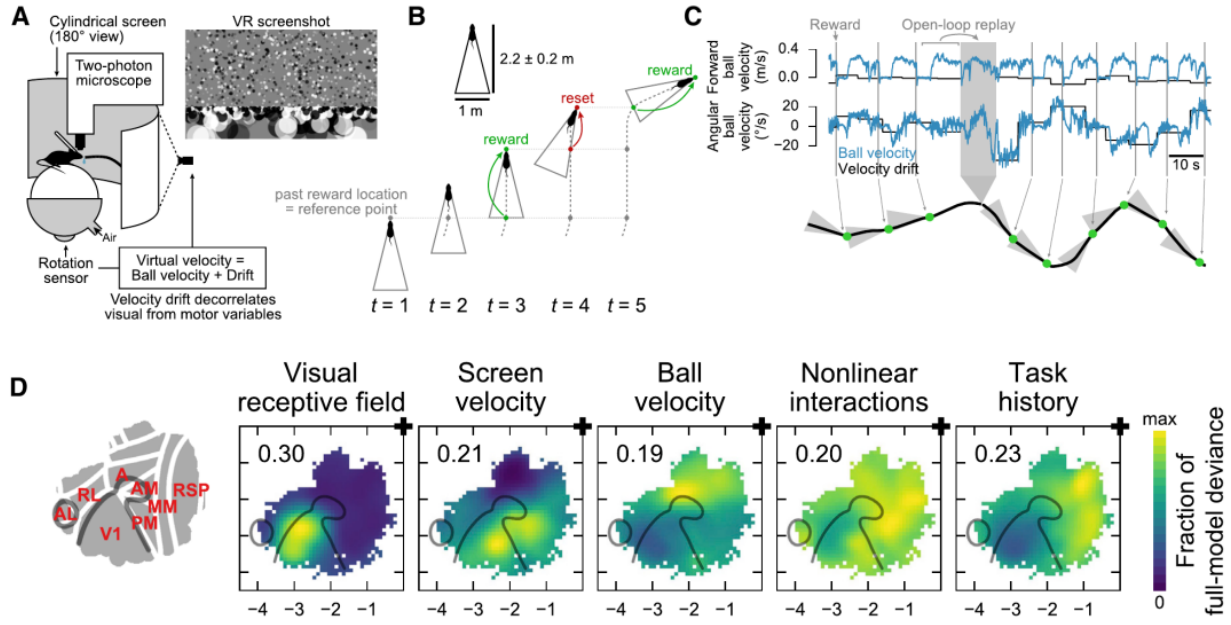

Figure 1.4: **Non-sensory maps across the mouse's posterior cortex.** Mice were trained on a task engaging the visual and navigation-related networks in the dorsal posterior cortex. In short, the task consisted in running approximately two meters straight forward, in virtual world coordinates, from an invisible reference point to obtain a reward. To obtain rewards, mice therefore continuously needed to adjust their running to compensate for an induced drift in ball's velocity and run straight in the virtual world. (**A**) Experimental setup and screenshot of the virtual reality environment. (**B**) Schematic of the reward condition in the task (top-down view). Dashed lines, path taken by the mouse. Solid gray triangle, invisible boundaries used to determine reward delivery. (**C**) Top: example velocity traces. Bottom: corresponding top-down view of the path taken by the mouse. Green dots indicate reward times. (**D**) Maps of encoding strength. Left: Anatomy of the mouse's posterior cortex Right: maps of unique contributions of different features to the full model fit. Intense yellow thus represents a large contribution of this feature to the activity of a pixel. All maps are scaled from zero to the value indicated in the top-left corner – adapted from Minderer et al. (2019)

A distributed processing

The existence of such a precise organization within brain areas, and the sometimes very precise and local effect of brain lesions, have fed the idea that the brain is organized in highly specialized modules,

each encoding one specific brain function. However, the reality of the existence of such modules at different levels should not overshadow more nuanced theories, which argue for a more distributed encoding. Complex cognitive processes, such as conscious perception for example, have been shown to rely on a global ignition of multiple brain areas at once (Van Vugt et al., 2018). Moreover, Poeppel et al. (2008) suggest that speech perception is actually much more distributed than the sole Wernicke's area. **Thus, beyond brain modularity, understanding how brain areas interact with one another for specific cognitive functions has become a fundamental challenge in neuroscience.**

Summary. In this section, I discussed the topographical organization of sensory systems, at several scales. I argued that this topography can be either based on very simple sensory features, or more computational and categorical properties. **Overall, the precise arrangement of features encoding in cortical areas remains poorly understood. The specific computations performed by each cortical field, despite being a major question in neuroscience, have been surprisingly elusive.** Notably, **the existence of speech specific domains along the auditory pathway in humans raises the question of whether such domains exist in other animals, and whether they share a common socle.**

1.2 The origins of topography

In this section, I will tentatively expose three main aspects of the origins of topography: the advantage they might confer under evolutionary pressures, their developmental unfolding, and their refinement through life experience.

1.2.1 A topographical code?

Topographic maps are prevalent all over the brain, across sensory systems, and are quite faithfully preserved across many nuclei and cortical areas along the processing hierarchy. This raises the question of whether this organization conveys any computational advantages, and/or has been selected *per se* through evolution. **The question of whether they have a specific 'meaning' for the brain, or are rather a byproduct of evolutionary pressures at the local level, remains unsolved.**

An efficient wiring

Brain computations are thought to rely mainly on the connectivity scheme that exists between neurons. A coarse estimation of the number of connections per computational unit gives about 30,000 synapses per neuron in the human neocortex (De Felipe et al., 2002). Theoretically, in a physically non-limited world, the spatial position of the cells, as long as they verify this connectivity scheme, could be totally random; the brain would still perfectly perform its job. However, the way these connections are organized spatially has a huge influence on their physical cost to the brain, both in terms of the time required for information to travel along axons (linked to some computational cost), and energy (for the creation of axons, dendrites and synapses). Moreover, the space taken up by the axons in an unorganized brain would considerably swell its size – a fully interconnected brain would take the size of a bathtub (Cherniak, 1990).

It would consequently be a 'good design' to spatially group together neurons that connect together (Kaas, 1997; Chklovskii and Koulakov, 2004). In that direction, it is now well established that a simple distance - connectivity probability rule emerges in the statistics of cortico-cortical connectivity (Buzsáki et al., 2004). Since neurons that wire together often fire together, topographic maps could well emerge from simple, cost-limiting rules.

Spatially-constrained computations

In terms of actual computational advantage, topographic maps could be one way to easily perform common computations used by the brain across short, local connections (Kaas, 1997; Thivierge and Marcus, 2007). For example, biological systems often have to compare stimuli that are close together in space, so as to assess the context of an object (its color differences, its relative movement, etc.). Thus, center-surround comparison, which is of most biological importance, is relatively easy to perform with local rules of connectivity and a topographic organization (Kaas, 1997). More generally, any comparison between adjacent stimuli along the represented space, as in local facilitation, lateral inhibition, averaging, interpolating, motion direction detection, are encoded in a straightforward way in topographic organization (Kaas, 1997; Knudsen et al., 1987). These types of computations are prevalent in sensory processing (both visual, auditory, and somatosensory) for the characterization of objects, edges, or surrounds.

Along this line, having multiple maps of different sizes could allow parallel processing at different scales, using the same computational local rules (Kaas, 1997, 2015). For example, a small map could allow for center-surround comparison across large distance, while large maps would allow

for more details in the processing.

Finally, the role of certain topographic organizations seems to be rather faint, as are orientation preference or ocular dominance columns. Indeed, if they are present in some species, they remain absent in other species that do not show any visual deficiencies in comparison (Horton and Adams, 2005; Thivierge and Marcus, 2007).

Is the brain actually reading the topography?

Thus, if some arguments are quite reasonable with regard to the evolutionary pressures that lead to topographic organization, its overall role in brain function is still largely disputed. In particular, the relationship between topographic maps and perception and behavior remains unclear.

Single neurons usually encode only little information, and it is thought that most of brain processing occurs through the dynamics of the activity of large neuronal populations. However, it is sometimes observed that the brain itself contains much more information than what seems to be used for the production of behavior. For example, Stringer et al. (2019a) found that V1 was able to discriminate orientation of gratings with much more accuracy than what the mice can behaviorally do. If there are many reasons for the brain to lose information on the way from sensory stimuli to decision, one can wonder whether the high degree of information contained in sensory cortices is read with less precision than what could be optimally done. One hypothesis, which puts the topographic organization of these cortices at the center of the stage, would be that higher-order areas actually read topographic mappings at a coarser scale than the single neurons. **This would suggest that a relevant scale to look at in order to understand brain computations would be a meso-scale, rather than a single-neuron scale.** A few studies have set out to explore this idea and go beyond simple correlations, with mixed results and interpretations.

Michel et al. (2013) used a model-based illusion to investigate whether the global extent and shape of neuronal activity at the surface of the brain was determinant in the perception of object shape. Interestingly, they found that stimuli that had objectively similar spatial extent but were designed to evoke different global spatial patterns of activity in the retinotopic map of V1 in macaque monkey (figure 1.5-A) were perceived by humans as having different shape (figure 1.5-B). This first study thus suggests that the topographic pattern of neural population responses in visual cortex contributes to visual perception. However, the lack of a perceptually robust definition of what shape is for such stimuli can cast doubts on the overall conclusion.

In a second study, Benvenuti et al. (2018) revealed a large-scale component of population

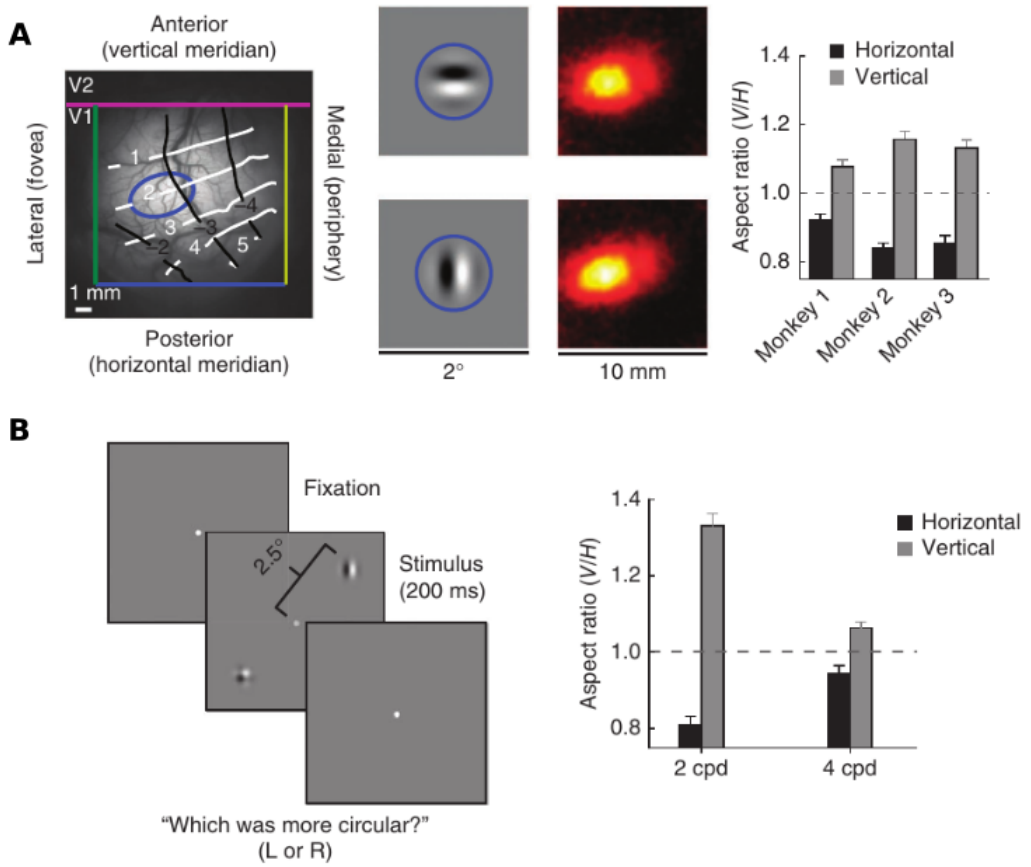

Figure 1.5: **The implication of cortical topography in shape perception.** (**A**) Physiological stimuli and spatial distributions of V1 responses measured with VSDI in monkeys. Left: The visual stimuli, with solid blue circles overlaid to indicate the 3σ contour of the stimulus envelopes for the horizontal (top) and vertical (bottom) Gabor stimuli. Center: Amplitudes of cortical responses in a 10 x 10 mm2 patch of cortex to the horizontal (top) and vertical (bottom) Gabor stimuli, averaged over an experimental session. Right: Normalized aspect ratios (ARresponse/ARretinotopic) of responses to 2-cpd horizontal and vertical Gabors measured in three different monkeys, demonstrating both the robustness of the orientation-dependent elongation effect and the variability of its magnitude across different individuals (n = 5 experiments for monkey 1, 8 experiments for monkey 2, and 4 experiments for monkey 3). (**B**) Psychophysical results averaged across ten human subjects. Left: Schematic of the visual stimulus. Human observers briefly (200 ms) viewed a display consisting of a plaid standard and an oriented comparison stimulus whose vertical-to-horizontal aspect ratios were selected randomly, and were asked to decide which stimulus had a more circular envelope. Right: Psychophysically determined perceptual aspect ratios and 95% confidence intervals for 2-cpd and 4-cpd Gabor stimuli (n = 10 human subjects). – adapted from Michel et al. (2013)

responses to low-frequency spatial patterns, whereas in the hypothesis of A1 encoding local contrast, the small aperture of V1 neurons' receptive field should have yielded a globally flat activity pattern. This distributed representation, visible at the retinotopic scale, seems to be linked to perception, as humans readily infer the orientation of low spatial frequency gratings. The authors thus suggest that two overlapping encodings exist in V1, for both local and global scales. Overall, they conclude

that the topographic pattern of neural population responses at the retinotopic scale in visual cortex contributes to visual perception.

To test the hypothesis that perception is based on topography at a coarse level, one would need to generate artificial patterns of activity with different scales of details, and observe whether realistic perception can be elicited. Electrical stimulations of the cortical surface can elicit phosphene or sound perception (Borchers et al., 2012). However, the overall effect of electrical stimulation remains poorly understood, and quite counter-intuitive effects can occur both at the physiological and behavioral levels. Such an approach, enriched with more recent techniques like optogenetics (Marshel et al., 2019), could be of interest in trying to decipher what is the appropriate spatial scale at which brain computations occur. The use of mutants deprived of topographic maps (Lokmane et al., 2013) could also be of special interest in studying the perceptual and behavioral importance of maps.

Despite the large prevalence of topographic mappings across the brain, the evidence for their computational advantage, if any, remains largely speculative, and their role in perception poorly understood. **The development of tools to access the representation and encoding of sensory stimuli at multiple scales can thus be of foremost interest to understand the link between topography and perception.**

1.2.2 Developmental emergence of topography

Three main forces usually prevail to organize brain structures: genetic patterns, spontaneous patterns, and experience-dependent patterns.

Intrinsic factors: genetic determinants and spontaneous activity

Brain development is largely governed by genetic and molecular rules; the set up of topographic maps is no exception. A very simple experiment to prove this was performed by Sperry (1963), on the development of the retinotopic map in the tectum. The optic tract was severed and the eyes were rotated by 180°. Axonal pathways regenerated, and a retinotopic map in the tectum is still present, but also rotated by 180°. This suggests that axons from the retina, despite being spatially rotated, found their way back to their original connection place in the tectum. This experiment and others led Sperry to formulate the 'chemoaffinity hypothesis', which has been well explored in other animal models, such as mammals (mouse), and relies on genetically encoded gradients of molecular

tags (Thivierge and Marcus, 2007). The same principle applies to the formation of somatotopy and tonotopy. This gradient-based principle usually leads to the formation of the smooth, large-scale maps present in the brain.

However, other local factors are at play. Notably, local activity patterns (spontaneous or triggered) might refine the local tuning connectivity schemes through Hebbian plasticity, and locally rearrange cortical maps (Kaas, 1997). The development of ocular dominance columns well reflects this principle, since eye-preferring columns are formed within primary visual cortex before birth, possibly by both genetic factors and intrinsic, correlated activity coming from spontaneous retinal waves in each eye (Katz and Crowley, 2002).

Extrinsic factors: the role of early experience

Following on this example, another factor seems to deeply influence mappings in the brain. Namely, experience shapes ocular dominance patterns during a critical period, after birth, during which patterns of activity triggered by the stimuli on the retina will rearrange the cortical mapping. Deprivation of an eye indeed makes ocular dominance columns associated to the corresponding eye shrink dramatically (Katz and Crowley, 2002). Similarly, a visual environment in which vertical lines are over-represented at an early critical period will substantially modify the orientation preference of visual neurons, and thus the global orientation preference map (Blakemore and Cooper, 1970). Thus, **the statistics of objects' features in the external world seem to affect and shape sensory cortices computational maps, in a way that more common features in the environment will occupy a more prominent space in cortical representations**.

Do natural statistics in the environment also influence the formation of larger-scale maps, such as retinotopy or tonotopy? Tonotopy is known to be present very early in life. Zhang et al. (2001) explored the role of the early acoustic environment in shaping the adult tonotopic organization of primary auditory cortex. By overexposing developing rats to specific tones, they revealed an exaggerated representation of the exposed tone-frequency and its facilitated emergence in the tonotopic organization of A1. These results suggest that **early life experience can also shape well-established topographic frameworks at large scales.**

These results put forward the role of the input's nature and statistics to shape cortical organization. Then, one can wonder whether different cortical areas contain fundamentally different network and processing 'tools', or whether their role can be switched and shaped according to their inputs. Sur et al. (1988) revealed that functional visual projections can be routed into the

auditory thalamus and cortex of infant ferrets, suggesting that the modality of a sensory cortical area may be specified by its inputs during development. The portion of cortex reached by these new inputs displayed a retinotopic organization. Notably, the representation of a two-dimensional sensory epithelium, the retina, in cortex that normally represents a one-dimensional epithelium, the cochlea, suggests that the same cortical area can support different types of maps (Roe et al., 1990). These results go even further since Sharma et al. (2000) demonstrated the existence of orientation preference columns in the rewired cortex. These rewiring experiments suggest overall that **cortical representations are highly dependent on the structure of their inputs, and might rearrange during development according to those**.

One can then wonder what the role of early-life experience in the development of more complex, category-related brain patterns is. Would a feral child, raised with no exposure to language, develop a 'speech' or a 'voice' area? Would he/she develop a human 'face' area? And, if not, why does the 'face' area, or the 'speech' area, always develop in a stereotyped location in normal children? Overall, the contribution of evolutionary processes versus experience is still under debate for face- and speech-selective domains (McKone et al., 2012).

Livingstone et al. (2017) explored the development of the face patch system in the high-order inferior temporal cortex of macaque monkeys. Their study suggests the existence of a 'shape-biased retinotopic proto-map that is refined by experience', as an initial substrate for the formation of a face area. The creation of this domain was mainly driven by a decrease of responses to non-face stimuli. Interestingly, face-orienting behavior in young monkeys is present even before the appearance of a fMRI-detectable face patch (Sugita, 2008), suggesting that the emergence of this specialized area might be triggered by a looking behavior. In that direction, the same group further showed that seeing faces is actually a necessary condition for the development of the face patch system (Arcaro et al., 2017) (figure 1.6). Instead, face-deprived monkeys developed domains specific of body parts such as hands, with which they had extensive, behaviorally relevant experience. These results also hint at a model in which **experience builds upon a proto-organization to create self-organized domains, in which the representations of co-occurring features are reinforced until the appearance of very specific – and behaviorally relevant – domains**.

The development of voice/speech areas has been much less studied. Grossmann et al. (2010) showed that 7-month old infants displayed a voice-sensitive area, whereas 4-month old infants do not. This observation suggests that voice processing is built up in the brain during a critical period

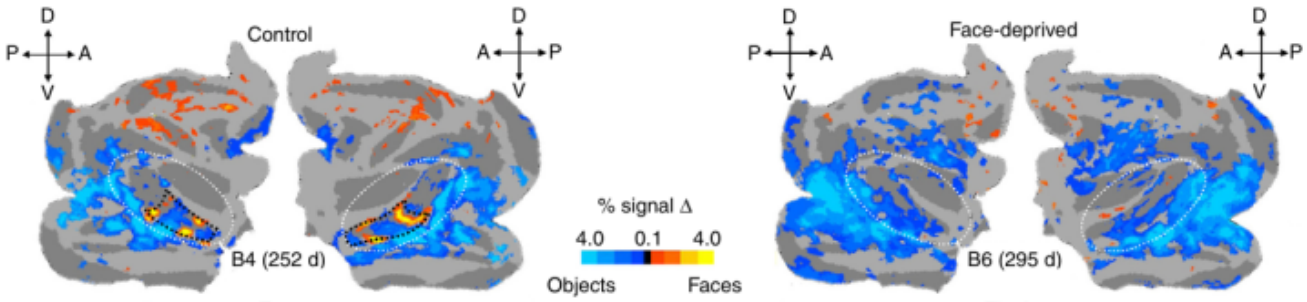

Figure 1.6: **Seeing faces is necessary for face-domain formation.** Faces>objects and hands>objects activations in control and face-deprived monkeys. Representative maps for the contrast faces-minus-objects aligned onto the standard macaque F99 flattened cortical surface (light gray, gyri; dark gray, sulci). These examples show percent signal change (beta coefficients) thresholded at $P < 0.01$ (FDR corrected) for one session each for control (left) and face-deprived (right) monkeys at 252 and 295 days old, as indicated. Dotted white ovals indicate the STS region shown in the bottom half for all scan sessions for all seven monkeys. – adapted from Arcaro et al. (2017)

in childhood. How the different areas engaged in speech processing come together while a child is learning language is still unknown. **Furthermore, whether animals can develop voice or speech sensitive regions (which would suggest that auditory mechanisms similar to humans could be at stake) has not been explored.**

Overall, a combination of genetic and spontaneous activity-related factors seems to provide an initial scaffolding for the formation of sensory maps, on top of which early experience preserves or rearranges topographic mappings, or initiates the formation of category-specific domains. Whether these domains are formed as clusters because they build up on a proto-topographic mapping of some sort, and whether such topographic arrangements are computationally advantageous, is still unclear.

1.2.3 Plastic changes in topography induced by learning

Development generates organized brain structures, shared across all individuals, which have been selected under evolutionary pressures. However, each individual encounters different obstacles in different ecological surroundings, and is able to learn from its own experience and adapt its behavior, or even sharpen its senses. The question of whether – or rather how – this experience modifies brain structure at the adult stage has now been largely investigated. I will first examine the role of perceptual training in shaping primary sensory cortices, and then explore its role in the formation of higher-order, categorical domains.

Modifying the organization of sensory systems

It is often said that blind people get better hearing, as their auditory system has taken some space from the now unused visual system. More generally, our senses can get sharper with training, in a quantifiable way through psychophysics (Watanabe and Sasaki, 2015). The neural correlates of such behavioral improvements have slowly been unveiled over the past decades.

The plasticity of cortical maps has now been widely studied in adults, mainly in auditory and somatosensory systems (Weinberger, 1995; Buonomano and Merzenich, 1998). Early studies primarily investigated the modifications of somatotopic maps following a deafferentation of a certain part of the skin, and showed that the unresponsive associated area came to be excited by inputs from neighboring skin surfaces within weeks after the intervention (Buonomano and Merzenich, 1998). A less intrusive approach (and more relevant to behavior) was to train owl monkeys on a task that produced cutaneous stimulation of a limited sector of their fingers (Jenkins et al., 1990). The study revealed that after a few months of training, the cortical representation of the specific surfaces of the digit tips that were stimulated during training displayed several-fold magnification compared to control animals. On the same note but in the auditory domain, classical conditioning (a tone followed by a shock) has been shown to trigger changes in the representation of the acoustic conditioned stimulus in the auditory cortex (Weinberger, 1995). Of course, these maps reorganizations are coupled with changes of the tuning of cells that can be finer and more diverse. Whether these modifications were due to the statistics of the bottom-up sensory inputs to auditory cortex or to a more top-down, task-dependant control, however, was not clear at the time. Using an elegant protocol consisting of a constant set of stimuli but varying task demands, Polley et al. (2006) showed that **only the task-relevant stimulus feature displayed an enhanced representation in the brain** (figure 1.7).

Other studies investigated the neural basis of such reorganization, and found a way to bypass the conditioning or learning part at the behavioral level. Indeed, a pairing of tone presentation and Nucleus basalis stimulation has been shown to trigger similar map reorganizations (Kilgard and Merzenich, 1998), associated with synaptic changes (Froemke et al., 2007). These long-term modifications have been shown to enhance behavior (Froemke et al., 2013). Nucleus basalis activity is associated to acetylcholine release in auditory cortex and could be controlled by top-down influences. Another pathway, at least in the mouse, could be that (orbito-) frontal cortex influences auditory cortex map through direct connections (Winkowski et al., 2013).

Interestingly, the behavioral relevance of these modifications remains a subject of debate,

as map plasticity could participate to perceptual learning but not overall performance (Reed et al., 2011). Map modifications have thus been shown to disappear after learning. More generally, the stimuli and behavioral paradigms used in these studies might be far from ecological situations in which animals have to learn. A more natural behavior which involves changes in perception is motherhood, when e.g., mice learn to recognize and discriminate their pup's ultrasonic calls (Liu et al., 2006; Shepard et al., 2016). When recording the representation of pup calls frequency in the tonotopic arrangement of primary auditory cortex of naive vs. mother mice, Shepard et al. (2016) revealed an enhancement in the contrast between how neurons tuned to and away from ultrasonic frequencies respond to those calls, rather than an increase in the domain occupied by these frequencies. Thus, **natural behaviors during which specific sounds become behaviorally prevalent may trigger subtle changes in topographic maps**.

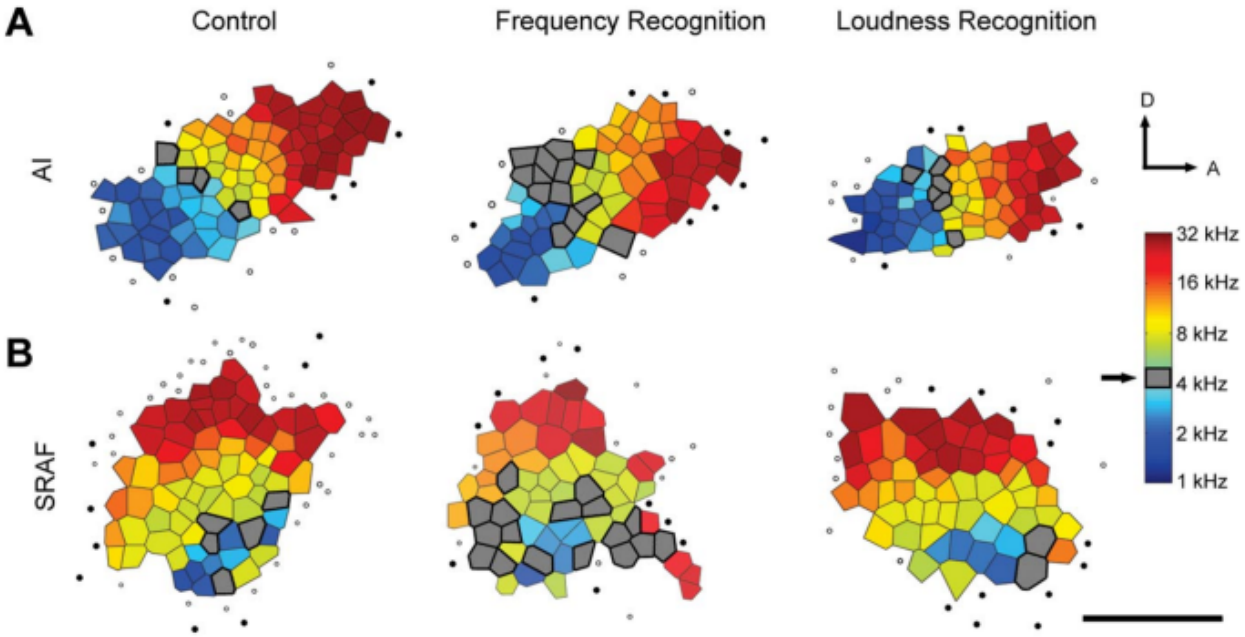

Figure 1.7: **Perceptual learning directs auditory cortical map reorganization through top-down influences.** Task-specific reorganization of cortical maps in the frequency domain. Representative tonotopic maps from primary auditory cortex (AI) (**A**) and suprarhinal auditory field (SRAF) (**B**) were delineated with fine-grain microelectrode mapping. The color of each polygon in the tessellated map represents the CF associated with neurons located in the middle cortical layers at that position in the map. Gray shaded polygons indicate recording sites with CF values within the trained frequency range (5 kHz 0.375 octaves). Filled circles indicate unresponsive sites. Open circles represent sites with sound-driven responses that did not meet the criteria for inclusion in AI or SRAF. Scale bar, 1 mm. The arrows indicate dorsal (D) and anterior (A) orientations – adapted from Polley et al. (2006)

Creation of more complex patterns with training

The biological preeminence and importance of certain categories, such as face or speech, could well suggest that evolutionary pressures have progressively shaped brains to naturally develop these domains, or at least a proto-architecture on which they can unfold. In that direction, their location and extent is very conserved across individuals. However, the fact that humans display cerebral domains associated with cultural, recent (at the scale of evolution) categories, such as written words (McCandliss et al., 2003; Dehaene and Cohen, 2011), suggests that some of these regions recycle part of the cortex and develop strictly under the drive of experience. The question of how, and where they develop is still open.

Notably, whether inferior temporal cortex (IT) is specialized in face perception, or whether it encodes highly-familiar objects' representations has long been unknown. Early work by Gauthier et al. (1999) used novel objects ('greebles') to test whether forming new perceptual, high-level categories would trigger in IT the formation of a specialized domain. Their results indeed suggested that perceptual training (to discriminate greebles identity) led to increased activation in the right hemisphere FFA. However, it was not known how such domain formation was link to pre-existing architecture. Subsequent studies nuanced the effects of category training, showing that changes could occur in complex patterns and outside of FFA (Op de Beeck et al., 2006). Interestingly, the strength of training effects in the object-selective cortex was correlated with behavioral improvements on a task involving those new objects. This result suggests that those changes are relevant for behavior. However, the spatial distribution of training effects could not be predicted from the spatial distribution of either pretrained responses or face selectivity. This observation is not consistent with the idea of a proto-architecture for domain creation in IT. The same team performed similar experiments in monkeys and showed that in IT cortex, topography of selectivity for these novel object classes was stable across time and training, task, and position of object (Op De Beeck et al., 2008). These patterns of activation indeed mainly reflected shape. Whether new category representations entrain changes in high-order sensory areas in animals was thus unsure.

Because the formation of abstract category-related domains could occur only during a critical period in life, Srihasam et al. (2012) trained juvenile monkeys on an abstract symbol manipulation task. Intensive training led to the formation of novel specialized cortical domain, in stereotyped locations of IT, detected using fMRI. Adults took longer to learn the task, and displayed no symbol-specific domain after training. The same team further demonstrated that the cerebral localization of training-induced changes does not depend on function or expertise, but rather on some kind of

proto-organization (Srihasam et al., 2014) (figure 1.8). Training on different symbol sets led to different specialized domains, of which location did not depend on training order, nor the global performance of the animal. This observation led the authors to formulate the hypothesis that **plasticity for abstract symbols is constrained by some native organization in cortex (here based on curvature and eccentricity), similarly to the appearance of the face-patch system.** We note here that these studies relied on a rather conservative measure which is domain formation (i.e., maximal response) rather than finer grained analysis, as what was observed in humans (Op de Beeck et al., 2006).

Such investigations in the auditory domain have been scarce, and the question of whether there exists an auditory equivalent of IT, i.e., an area containing specific domains for behaviorally-relevant auditory objects, is unclear. Moreover, most of these studies in the visual system focused on high-order areas, with a domain selectivity that is strong enough to be uniquely quantified by looking at the stimuli which evoke the maximum response on a single-voxel basis. What happens in earlier areas has been poorly explored. In particular, the role of experience in the creation of the specific pathways for speech and music that have been uncovered in parabelt auditory areas (Norman-Haignere et al., 2015) is still unknown. **Would the discoveries made in IT, which contains face- and symbol- specific domains, extend to these speech- and music- specific pathways?** A recent study suggested that training humans to discriminate monkey calls actually sharpens neural selectivity to auditory features in left auditory cortex and induces auditory category selectivity in lateral prefrontal cortex (Jiang et al., 2018). Thus, **both early and late areas of the auditory cortex might be modified by a training on new auditory objects.**

Finally, the lack of results of domain formation for higher-order stimuli in adults seems to be dissonant with the easiness with which early areas, that one could have thought most stable since they are the foundation of most sensory processing, can be extensively modified. **One can then wonder whether the scale at which it happens is smaller than the one actually tested (fMRI), or if more generally those modifications could be more subtle than just domain expansion.**

Summary. The origins of high-order domains and processing pathways in the sensory cortices are far from being well understood, especially in the auditory domain. In particular, speech and

music processing domains might develop during childhood. **How are these pathways shaped by experience? Are there computational limitations, specific to each species, for the development of such domains?**

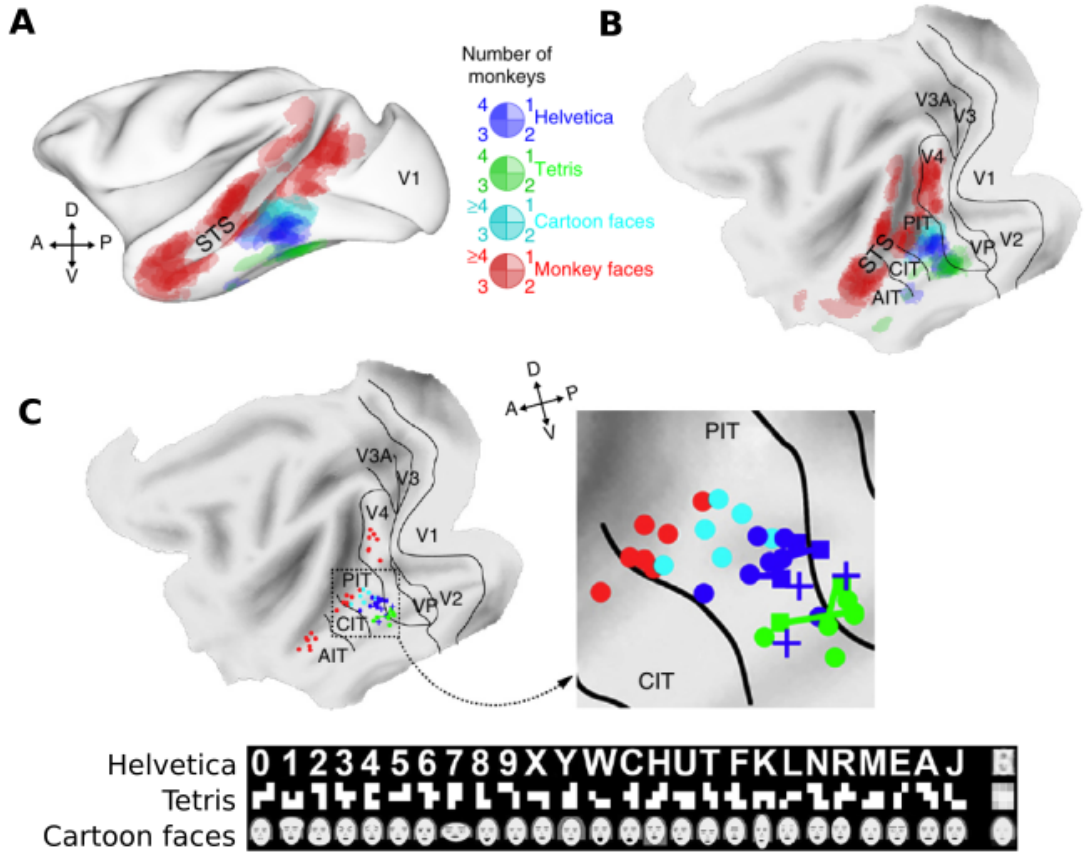

Figure 1.8: **Overall organization of selective responsiveness to trained symbol sets and face patches in seven monkeys.** (**A**) Patches of significant activations for all three symbol sets and monkey faces (each contrasted with its control) from each of the seven multiple-symbol-set trained monkeys projected onto a standard macaque brain shown in a semi-inflated lateral view. (**B**) Same data shown on a flattened standard map of macaque cortex with areal borders. (**C**) Centers of mass for different selective patches. Dots indicate the center of mass of each of the three main monkey faces > shapes patches and each of the trained- symbol-selective patches, indicated by color, in the monkeys trained in this and their previous study. For monkeys who learned both Helvetica and Tetris, the squares indicate centers of the first-learned symbol set region immediately after learning that symbol set and circles indicate centers of the same, first-trained, symbol set but after learning the second symbol set; the two patches for the first-learned symbol set for each monkey are linked by a line of the same color. +'s indicate the centers of the Helvetica patches for the three monkeys from their previous study who were trained as juveniles on Helvetica only. The indicated dorso-ventral and antero-posterior axes for the flat maps are meaningful only for the lateral surface of the brain. – adapted from Srihasam et al. (2014)

1.3 Reading the topographical brain

These numerous scientific breakthroughs have been enabled by parallel developments of brain recording techniques. The large diversity of possible spatial and temporal resolutions, portability, as well as different types of recorded signals, have provided scientists with a variety of approaches, each revealing shared or specific facets of brain organization. In this section, I will first highlight a few technical aspects that are relevant when studying brain organization. I will then present a newly developed technique that I used throughout my work, namely functional UltraSound, and expose its key characteristics.

1.3.1 Technical aspects of brain imaging

The holy grail of experimental neuroscience could be to record at once all neurons of a brain. This has been so far obviously impossible, and most single-cell techniques rely on sub-sampling neuronal populations within small parts of the brain. If population analysis provides priceless information about the representation and processing of sensory stimuli in the recorded area, it can sometimes be difficult to replace them back in a more general – large-scale – context, or even to understand their spatial arrangement in the brain tissue. I will not propose a catalogue of comparisons between available techniques, but rather highlight several technical aspects of imaging that shape our understanding of sensory processing.

On the one hand, low spatial resolution is in itself detrimental to precise mapping. This is the case of fMRI, which usually uses voxels large of at least one to several millimeters-cube, thus averaging neuronal activity over thousands of units. Such a technique is of unquestionable use in primates, or more generally large-mammals, but has rarely been determinant in small animals' research. As a comparison, only a few 2-millimeter-sided voxels would approximately fill up the whole auditory cortex of a mouse, leaving only little room for topographic mapping. It is worth noting that some very high-resolution techniques are being developed. Notably, they were used to map the tonotopic organization of the inferior colliculus of the mouse with success (Cheung et al., 2012). Further, a few studies investigated the nature of fMRI signals at the cortical layer level, in rats (Silva and Koretsky, 2002), cats (Harel et al., 2006), monkeys (Chen et al., 2013) and humans (De Martino et al., 2015). However, fMRI is in no way a portable tool (and magnets for high-resolution imaging are rare and even less portable), and behavioral experiments in animals are harder to perform, due to the restricted size of the head, body movements, and scanning noise.

On the other hand, the use of medium-resolution but high-sampling, large-scale techniques can provide crucial information on many aspects. First, very high-spatial resolution but low-sampling techniques (such as with electrodes) might provide very different results from low-spatial resolution but exhaustive sampling techniques (such as with fMRI). If the presence of face-selective units in IT of macaques has been reported in early studies, the existence of a larger scale 'face-area' has only been discovered through the use of fMRI, as solely a lower-resolution, averaging technique could reveal a more global architecture (Tsao et al., 2003). In the same spirit, IT has long been seen as a high-order area with no or little overall spatial topographic organization related to low-level (i.e., non-categorical) features. The use of fMRI in monkeys has recently started to unveil more subtle topographies, such as a 'large-scale spatial organization for some dimensions of shape' (Op De Beeck et al., 2008), or a 'correlation between curvature and eccentricity' in both IT and earlier areas (Srihasam et al., 2014). Finally, the lack of organization of the orientation selectivity in V1 in rodents, compared to the orientation preference columns, has led to the idea that the encoding was 'salt-and-pepper'. A recent study, using large-scale two photon imaging, challenged this view and actually observed a very large-scale orientation selectivity map in the mouse V1 (Fahey et al., 2019). These findings deeply modified our vision of the processing occurring in those regions. Thus, **the use of 'coarser' techniques can reveal global organizations better than single-cell resolution, because they average (and thus sample) the activity of virtually all neurons in each voxel**.

Second, brain shape and topographic organizations vary considerably across individuals. The overall shape, the position of the sulci, the extent of each area, but also the whole functional mapping, such as the map structure and its borders, contain a significant amount of variance. Averaging across individuals can hide specific organization, especially for loosely topographic and highly variable mappings such as in frontal cortex (Michalka et al., 2015). In order to avoid such averaging effects, and increase overall the reproducibility of our experiments, **having a larger framework can a be a necessary step for an individualized approach**. This framework could be based on anatomical and functional landmarks. In this direction, it seems reasonable that such landmarks are especially necessary in hardly accessible regions, such as the depth of sulci.

Third, a large field of neuroscience now tends to show that many brain processes are actually distributed over large portions of the brain, such as for speech processing (Poeppel et al., 2008), or conscious perception and report (Van Vugt et al., 2018). **Understanding how different brain regions interact during the course of these phenomena requires large-scale (potentially**

whole-brain) imaging.

Fourth, **whole brain techniques allow for an easier and more exhaustive exploration of brain activity**, to dig out regions or patterns that could be linked to certain cognitive processes. On the opposite, low-sampling high-resolution techniques such as electrodes tend to reinforce the well-known and well-feared 'looking for one's keys under the lamp-post' issue.

Some techniques, such as intrinsic imaging (Nelken et al., 2008), voltage-sensitive dyes (Michel et al., 2013; Benvenuti et al., 2018), or large-scale calcium imaging (Zhuang et al., 2017) allow for a rather high spatial resolution mapping while still providing pretty large-scale images. However, their use remains undermined by the fact that they can not penetrate deeply in the brain. 2- and 3- photon imaging, using color-shifted dyes, can provide images down to layer 4, but have had so far a limitated use and field of view. As many mammals have convoluted brains, those surface techniques rapidly show their limitations for cortex studies, not speaking about subcortical structures.

In short, **whole-brain, medium-resolution and high sampling imaging in small animals has remained, so far, particularly difficult**.

1.3.2 functional UltraSound

Current state-of-the-art

Functional UltraSound (fUS) has just started to be used to tackle fundamental questions in neuroscience (Deffieux et al., 2018). In short, fUS technology uses plane wave transmissions and measures the ultrasonic energy backscattered from red blood cells or the Doppler shift they induce in each pixel of the image, to obtain a proxy for blood volume and speed in brain tissue. After fundamental changes in paradigm that allowed to overcome previous difficulties for functional imaging with ultrasound, such as frame rate (Tanter and Fink, 2014), Macé et al. (2011) provided a first proof of concept of ultrafast fUS in anesthetized rats. Activity of large portions of the brain (full coronal slices) were dynamically imaged with an unprecedented high-spatial resolution for ultrasound (100µm). Strikingly, the high sensitivity of the technique allowed the experimenters to obtain dynamical imaging of epileptic seizures. The technique was further tested with stimulations of the whisker pad, which triggered neural responses in somatosensory thalamus and cortex. This study thus set the ground for a new way of imaging brain activity in small animals. Since then, fUS range of application has been widened, as illustrated in figure 1.9. Indeed, it has been applied to

a diverse range of species, such as rats (Macé et al., 2011), mice (Macé et al., 2018), monkeys (Dizeux et al., 2019), or human neonates (Demene et al., 2017) and adults (Imbault et al., 2017); a diverse range of situations, such as in anesthetized animals (Macé et al., 2011; Urban et al., 2014; Gesnik et al., 2017), head-restrained behaving animals (Dizeux et al., 2019; Macé et al., 2018), or freely-moving animals (Urban et al., 2015; Sieu et al., 2015); and a diverse range of brain systems, such as somatosensory system (Macé et al., 2011; Urban et al., 2014), olfactory system (Osmanski et al., 2014a) or visual system (Dizeux et al., 2019; Macé et al., 2018). Its use for studying brain activity at the level of whole structures, such as cortical areas or subcortical nuclei, has now been well established (Macé et al., 2011; Urban et al., 2014; Gesnik et al., 2017). Only a few, very recent studies are starting to address the question of the organization within such areas (Dizeux et al., 2019; Macé et al., 2018), while this opportunity seems very promising.

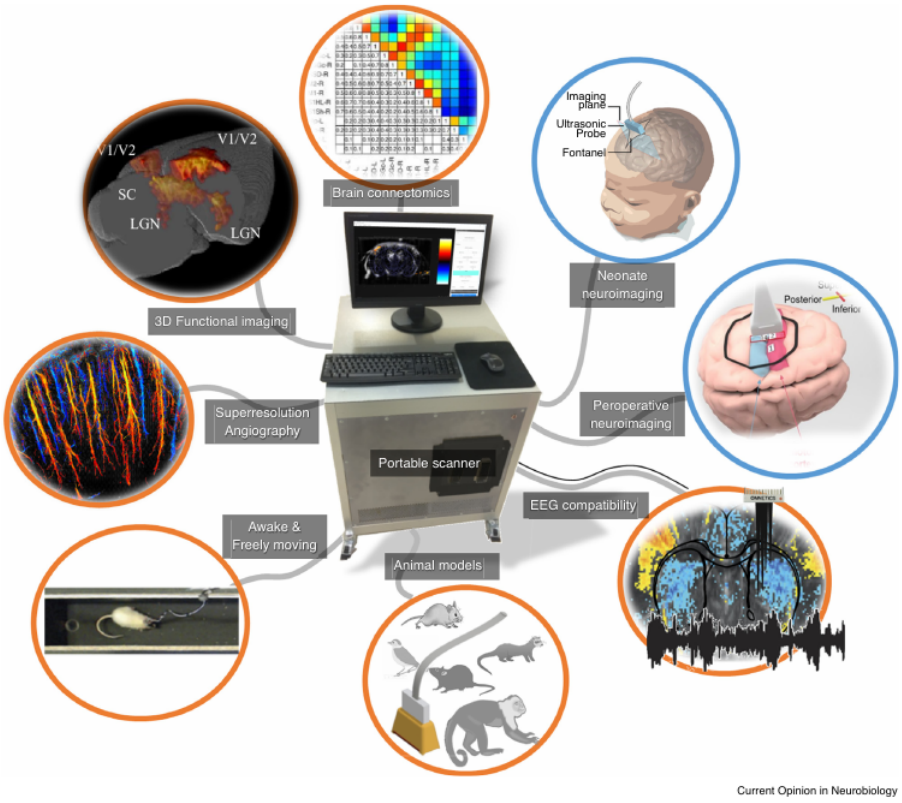

Figure 1.9: **The main applications and features of functional ultrasound (fUS) imaging.** fUS imaging provides (i) a compatibility with a wide range of animal models for preclinical studies, (ii) the ability to image awake and freely moving animals, (iii) the possibility to combine with super-resolution ultrasound localization microscopy, (iv) a possible extension to 3D imaging, (v) functional connectivity mapping for brain connectomics, (vi) translation to clinical neuroimaging in human neonates or (vii) peroperative neuroimaging during brain surgery and (viii) EEG compatibility for EEG-fUS recordings. – adapted from Deffieux et al. (2018)

To position fUS in the large field of neuroimaging, we can look at several criteria, that

are determinant for tackling fundamental questions of neuroscience (figure 1.10). In particular, its spatial resolution is much higher than what is usually possible with fMRI, while its portability is closer to optical imaging or electrodes. Furthermore, several techniques have been used to increase the spatial resolution of fUS (Deffieux et al., 2018), such as injected microbubbles (Errico et al., 2015). Nevertheless, its temporal resolution still relies on the speed of neurovascular coupling. Dizeux et al. (2019), for example, used this aspect to explore the propagation of information across layers in the cortex. This aspect of fUS, despite its special relevance in the encoding/decoding of dynamical stimuli (such as sounds), will not be discussed in this thesis.

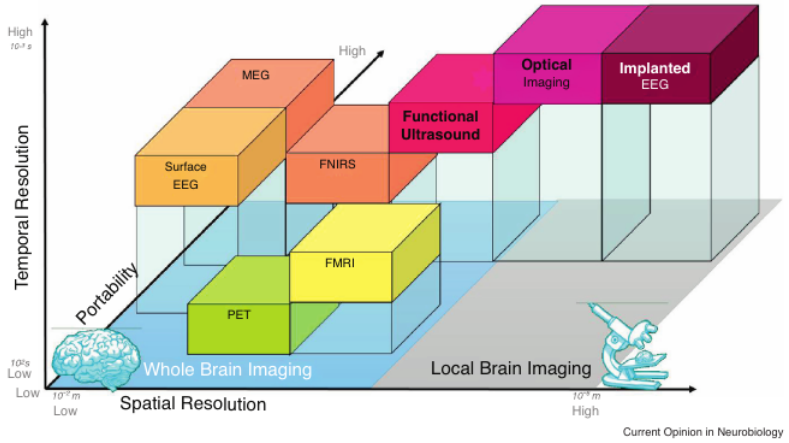

Figure 1.10: **Main brain functional imaging techniques on a three-axis chart (temporal resolution, spatial resolution, portability).** Techniques were separated between local and whole-brain imaging. Functional ultrasound fills a gap between whole brain imaging and microscopy, as well as between fMRI and Optics. – adapted from Deffieux et al. (2018)

Thereby, functional UltraSound is a promising tool to tackle some of the fundamental questions raised throughout this introduction. The study by Macé et al. (2018), which consists in an exhaustive look-out for a definite link between neural activity and specific cognitive processes with fUS, before digging into the details of single-neuron responses with electrodes, perfectly exemplifies some of the opportunities fUS might provide. **A question that now remains open is whether fUS could be used to assess fine topographic mappings *within* subcortical and cortical areas.**

The nature of the fUS signal

fUS is based on neurovascular coupling. Despite its fundamental importance for fMRI and other blood-based techniques, the link between vascular modifications and neural activity remains poorly understood and modeled, especially at small scales. Understanding this link is an essential step to

capture the fundamental limits of such a technique.

Boido et al. (2019) tried to bridge the gap between fMRI, fUS and electrical activity. They showed that vascular signals in glomeruli of the olfactory bulb, irrespective of the recording technique (local microscopy, fUS, fMRI), were linearly correlated with the local changes in local synaptic changes, across a large range of odor concentrations. This study thus suggests that fUS signal is a very good proxy for local brain activity, and one that is much more sensitive that fMRI signals for single trials as well as low odorconcentration trials.

It is not yet really known what the smallest scale at which blood vessels respond to surrounding neuronal activity is, nor how the spatial organization of blood vessels and capillaries shapes the overall hemodynamic response we eventually record. By coupling functional intrinsic imaging of acoustic stimulation-evoked activity in auditory cortex and corrosion casts of the blood vessels in the same brain, Harrison et al. (2002) showed that the observed spatial pattern of intrinsic signals directly correlates with the physical position and density of capillary beds, as well as with the myogenic valves that actually control blood flow within these capillary beds. Importantly, this study points to the fact that brain areas can have very different capillary density, and the less vascularized ones could provide a lesser signal. Notably, the authors mention the high level of vascularization in primary sensory areas, and suggest that the functional resolution (i.e., the lowest spatial scale at which one can functionally separate two signals) might be of the order of 100-150µm based on the structure of the capillary beds. This resolution, of course, can vary considerably across brain regions and animals. While fMRI spatial resolution is most of the time much lower than this, **the high resolution for fUS can be particularly useful to explore the minimal scale at which brain signals can be extracted from hemodynamic-based studies.**

More recently, O'Herron et al. (2016) investigated the tuning of individual arterioles in the visual cortex of cats, and suggested that blood flow in parenchymal vessels is driven by local neural activity as well as by an additional global component arising from adjacent functional columns. Single vessels' tuning to the orientation of visual gratings was thus much less precise than a global average of synaptic or spiking activity within up to 600µm radius. This study also shows a clear difference between pial and parenchymal arterioles, and suggests that the orientation selectivity of parenchymal vessels is actually an order of magnitude higher than what is obtained using intrinsic signal optical imaging. Thus, **the in depth view enabled by fUS could provide much higher sensitivity to the functional organization of cortex, by avoiding pial vessels' contamination of the signal**.

As of today, **the extent to which hemodynamic-based techniques can teach us about brain organization remains uncertain**. All these considerations are to be kept in mind throughout the current thesis.

Summary. Functional UltraSound is emerging as a very good alternative to fMRI, especially for small animals, offering a large field of view and a high resolution while still being portable and easy-to-use. Its use is under rapid expansion, and both technical and analytical advancements should provide a new light to understand the brain's organization.

1.4 Outstanding questions and project summary

In this introduction, I first exposed the global, modular organization of the brain, and highlighted the interesting attributes of topography in its diversity and complexity. Topography seems ubiquitous in the brain, in both low- and higher-order areas, and many aspects of its organization are still poorly understood. In particular, overlapping processing streams seem to be at play in primary and non-primary regions, but their organization in animals remains unclear. This is especially true in the auditory system when it comes to complex natural sounds. I then unwrapped several aspects of the origins of such an organization, to provide global insights on the computational advantages it might carry, on its developmental scheme and on the plasticity dynamics which might shape it. Both the basic topographic organization and the more categorical domains in sensory cortices seem to be highly shaped by behavioral experience. However, the mechanisms underlying the emergence and plasticity of these domains is not yet fully understood. Finally, I exposed more technical aspects of brain imaging, in an attempt to link it conceptually to the fundamental questions hinted at in the previous sections, and discuss how the imaging techniques that one uses also shape one's perception of brain organization and computations. I argued that the emergent imaging technique functional UltraSound provides an opportunity to explore brain organization across multiple scales.

A few **outstanding questions** arise from these considerations. From a technical point of view, one could ask:

- *Can we use functional UltraSound to explore the topographic organization of sensory systems at large- and meso-scales?*
- *Can we access encoding patterns that are not obviously topographically organized?*

Based on the responses to these questions and turning to a more fundamental point of view, several interrogations can then be raised:

- *How are the different layers of complexity of sound processing organized in the auditory cortex?*
- *Are the domains selective for speech processing, that exist in humans, specific to humans?*
- *How do these domains emerge in the cortex?*

In the present manuscript, I investigated several of these questions, while the others propose a longer-term direction to this research. The project is presented in three quasi-independent parts, that all revolve around different aspects of sound encoding. I will finally integrate these results and prospects in a general discussion.

chapter 2: Mapping the auditory hierarchy. We first set out to explore the possibilities of fUS imaging in the awake ferret. We used the well-known tonotopic organization of the auditory system as a benchmark to prove that fUS could be used to inspect the fine organization of sensory systems – from small and deep nuclei to non-primary cortical areas – as well as connectivity patterns between areas. This project has been published in eLife (Bimbard et al., 2018).

chapter 3: Natural sounds processing. We then explored how natural and complex sounds are encoded in the ferret brain. We investigated the predictive power of models of auditory cortical responses of various complexities, and showed that a canonical model could explain auditory responses across both primary and non-primary areas. Our results reveal fundamental differences in the processing of sounds between ferrets and humans, especially for high-order acoustic features such as those found in speech and music. Finally, we explored the encoding of conspecific vocalization in auditory cortex, at both the perceptual and neural levels. Our results suggest that ferret do rely on high-order features for vocalization processing, despite lacking a specific pathway within auditory cortex to process them.

chapter 4: Space encoding in auditory cortex. Finally, we explored the organization of azimuth encoding in the auditory cortex, of which nature is debated. We investigated whether the meso-scale accessible with fUS could reveal broader properties of encoding, that are spatially fragmented and non-continuous. Our study paves a way for future research on this subject.

Chapter 2

Mapping the auditory hierarchy

Contributors: Charlie Demené, Constantin Girard, Suzanne Radtke-Schuller, Shihab Shamma, Mickaël Tanter & Yves Boubenec.

2.1 Abstract

A major challenge in neuroscience is to longitudinally monitor whole brain activity across multiple spatial scales in the same animal. Functional UltraSound (fUS) is an emerging technology that offers images of cerebral blood volume over large brain portions. Here we show for the first time its capability to resolve the functional organization of sensory systems at multiple scales in awake animals, both within small structures by precisely mapping and differentiating sensory responses, and between structures by elucidating the connectivity scheme of top-down projections. We demonstrate that fUS provides stable (over days), yet rapid, highly-resolved 3D tonotopic maps in the auditory pathway of awake ferrets, thus revealing its unprecedented functional resolution (100/300μm). This was performed in four different brain regions, including very small (1–2 mm^{-3}size), deeply situated subcortical (8 mm deep) and previously undescribed structures in the ferret. Furthermore, we used fUS to map long-distance projections from frontal cortex, a key source of sensory response modulation, to auditory cortex.

2.2 Introduction

Functional ultrasound imaging (fUS) based on Ultrafast Doppler (UfD) was first introduced in neuroimaging in 2011 (Macé et al., 2011). Using ultrasonic plane wave emissions, this system exhibits a 50-fold enhanced sensitivity to blood volume changes compared to conventional ultrasound Doppler techniques (Macé et al., 2013), with a very high acquisition rate (ms) enabling unambiguous discrimination between blood flow and motion artifacts (breathing motion, tissue pulsatility,...) (Demené et al., 2015). Relative to fMRI, it also presents substantially higher spatial resolution for cerebral blood flow imaging at the expense of non-invasiveness, greater portability and lower cost, and versatility for awake animal imaging. However, most fUS studies thus far have investigated its sensitivity in capturing coarse-grained sensory responses (Tiran et al., 2017; Osmanski et al., 2014b; Gesnik et al., 2017; Urban et al., 2014, 2015), or used it to explore indirect in-plane brain connectivity (Osmanski et al., 2014a; Rideau Batista Novais et al., 2016). Also, while the theoretical spatial resolution of Ultrafast Doppler for high sensitivity mapping of microvascularisation has been shown to be 100 µm for whole brain imaging in rats (Macé et al., 2013; Demené et al., 2016), the ability of the fUS technique to measure independent information on functional brain activity from the cerebral blood volume (CBV) variation maps at such a small scale, that is the truly informative fUS imaging resolution, has remained to date unproven.

Here, we demonstrate fUS imaging capability in capturing a fine-grained 3D functional characterization of sensory systems and direct, long-distance connectivity scheme between brain structures. Our first goal was to provide such 3D high-resolution functional mapping in the auditory system. However the limited richness of stimuli previously applied in state-of-the-art fUS imaging together with their long duration (typically 10 to 30 s) constituted an obstacle as they would require several days of acquisitions incompatible with in vivo investigations. Moreover, most studies used physiological stimuli (Macé et al., 2011; Gesnik et al., 2017; Urban et al., 2015) or direct electrical stimulations (Urban et al., 2015) specifically designed to activate at most the entire sensory structures. We therefore drastically reduced the durations and repetitions of presented stimuli while increasing their diversity to push the sensitivity limits of fUS imaging. Consequently we show that this technique can rapidly produce highly-resolved 3D in vivo maps of responses reflecting precise tonotopic organizations of the vascular system in the almost complete auditory pathway of awake ferrets. We further demonstrate that fUS imaging can provide voxel to voxel independent information (with a functional resolution of 100 µm for voxel responsiveness, 300 µm for voxel frequency

tuning), indicative of its high sensitivity. These measurements are repeated over several days in small (1–2 mm^{-3}size) and deep nuclei (8 mm below the cortical surface), as well as across various fields of the auditory cortex. On a broader scale, we describe how fUS can be used to assess long distance (out-of-plane) connectivity, with a study of top-down projections from frontal cortex to the auditory cortex.

Therefore, fUS can provide a multi-scale functional mapping of a sensory system, from the functional properties of highly-resolved single voxels, to inter-area functional connectivity patterns.

2.3 Material and methods

2.3.1 Animal preparation

Experiments were approved by the French Ministry of Agriculture (protocol authorization: 01236.02) and strictly comply with the European directives on the protection of animals used for scientific purposes (2010/63/EU). To secure stability during imaging, a stainless steel headpost was surgically implanted on the skull and stereotaxis locations of the dorsolateral frontal cortex (FC) and the auditory cortex (AC) were marked (Atiani et al., 2014). Under anaesthesia (isoflurane 1%), four craniotomies above the auditory cortex were performed on three ferrets (V_{right} and $V_{left,\ right}$, and S_{right}), using a surgical micro drill, yielding a ~15×10 mm window over the brain. After clean-up and antibiotic application, the hole was sealed with an ultrasound-transparent TPX cover, embedded in an implant of dental cement (Sieu et al., 2015). Animals could then recover for one week, with unrestricted access to food, water and environmental enrichment.

For fUS imaging, animals were habituated to stay in a head-fixed contention tube. The ultrasonic probe was then inserted in the implant and acoustic coupling was assured via degassed ultrasound gel. Experiments were conducted in a double-walled sound attenuation chamber. All sounds were synthesized using a 100 kHz sampling rate, and presented through Sennheiser IE800 earphones (HDVA 600 amplifier) that was equalized to achieve a flat gain. Stimulus presentation were controlled by custom software written in Matlab (MathWorks) and available on a bitbucket repository at this link: https://bitbucket.org/abcng/baphy/branch/abcng; copy archived at https://github.com/elifesciences-publications/baphy-branch-abcng/.

2.3.2 Ultrafast doppler imaging

We used a custom miniaturized probe (15 MHz central frequency, 70% bandwidth, 0.110 mm pitch, 128 elements) inserted in a four degree-of-freedom motorized setup. The probe was driven using a custom fully-programmable ultrasonic research platform (PI electronics) and dedicated Matlab software. Ultrasound codes are all are available within the framework of research collaboration agreements between academic institutions.

2.3.3 3D vascular imaging

Vascular anatomy of the brain portion accessible from the craniotomy was imaged in 3D using the Ultrafast Doppler Tomography (UFD-T) strategy described in (Demené et al., 2016). Briefly, this method acquires 2D Ultrafast Power Doppler (UfD) images at a frame rate of 500 Hz. Each frame is a compound frame built with 11 tilted plane wave emissions (-10° to 10° with 2° steps) fired at a PRF of 5500 Hz, combined with mechanical translation and rotation, and then post-processed via a Wiener deconvolution to correct for the intrinsic out-of-plane loss of resolution, so that we ultimately recover an isotropic 100 µm 3D resolution. In the end, a 3D (14 × 14 × 20 mm) blood volume reconstruction of the vasculature is obtained (voxel size: 50 µm, isotropic resolution 100 µm). This 3D vascular imaging was performed on each craniotomy, and was used as a local reference framework, specific to the craniotomy, where recording planes could be repositioned over days using correlation methods.

2.3.4 fUS imaging

fUS imaging relies on rapid acquisition (every 1 s) of ultrasensitive 2D Power UfD images of the ferret brain. For each Power image, 300 frames are acquired at a 500 Hz frame rate (covering 600ms, that is one to two ferret cardiac cycles), each frame being a compound frame acquired via 11 tilted plane wave emissions (-10° to 10° with 2° steps) fired at a PRF of 5500 Hz. Image reconstruction is performed using an in-house GPU-parallelized delay-and-sum beamforming. Those 300 frames at 500 Hz are filtered to discard global tissue motion from the signal using a dedicated spatio-temporal clutter filter (Demené et al., 2015) based on a singular value decomposition of the spatio-temporal raw data. Although the ultrafast 2ms temporal resolution is available for the CBV image generation, they are in fact averaged into one CBV image every second to capture the dynamics of the cerebral blood physiological response. Nevertheless, it should be noted that this rapid sampling rate is a key

asset to unambiguously cancel any respiratory or tissue pulsatility artifacts Demené et al. (2015) in the final averaged images. Blood signal energy (called Power UfD) is then computed for each voxel (100 × 100 × ~400 µm, the latter dimension, called elevation, being slightly dependent of depth) by taking the integral $\frac{1}{T}\int_0^T s(t)^2 dt$ over the 300 time points (Macé et al., 2013). This power Doppler is known to be proportional to blood volume (Rubin et al., 1994). A certain band of Doppler frequencies can be chosen before computation of the power using a bandpass filter (in our case a fifth order low-pass Butterworth filter), enabling the selection of a particular range of axial blood flow speeds, that is roughly discriminating between capillaries and arterioles (slow blood flow) and big vessels (fast blood flow). In our study, we set the filtering to better focus on small vessels with axial velocity lower than 3.1 mm/sec when indicated in the text. Power UfD signal was normalized towards the baseline to monitor changes in Cerebral Blood Volume (%CBV).

2.3.5 Protocol for sensory response acquisition

Auditory responses were studied by playing different sounds through animal earphones during recording of the brain activity via fUS imaging. The protocol for sound presentation is as follows: 10 s of silence (baseline), then 3 s of sound followed by 8 s of silence (return to baseline). Trials were following each other with only a little random jitter in time of about 1 to 3 s, and fUS acquisitions were synchronized with the beginning of each trial.

Visual responses were obtained by playing a flickering red-light stimulus instead of sound, with the same durations of different epochs.

2.3.6 Localization of the auditory structures

In order to find the boundaries of the auditory structures in the imaged portion of the brain, white noise sound was played (70 dB).

2.3.7 Mapping of the tonotopic organization of the auditory structures

Auditory structures are known to exhibit tonotopic organization based on extensive physiological and structural studies (in the ferret, see (Bizley et al., 2005; Moore et al., 1983; Pallas et al., 1990; Versnel et al., 2017; Nelken et al., 2004). To image these tonotopic maps, we played unmodulated pure tones while recording fUS images at five equally spaced frequencies on a logarithmic scale (602 Hz, 1430 Hz, 3400 Hz, 8087 Hz, 19234 Hz, covering the auditory hearing spectrum of the ferret,

at 65 dBSPL). The tones were played in random order, 10 trials/frequency (20 in the animal S.). To obtain the whole tonotopic organization in a 3D volume, this process was repeated in different slices in order to build a 3D stack from successive 2D slices (spaced by 300 μm). Each slice was acquired in ~15 min, thus allowing us to map in 3D the whole auditory cortex within a few hours.

We note that these tone stimuli elicited large and reliable responses in the whole auditory tract despite being unmodulated. This suggests that a variety of other auditory stimuli (such as natural sounds) can be used to elicit stronger responses and hence reveal more organizational properties.

2.3.8 Frontal cortex stimulation

Frontal cortex (FC) electric stimulations were adapted from previously described protocols (Logothetis et al., 2010; Tolias et al., 2005). Platinium-iridium stimulation electrodes (impedance 200-400kOhms, FHC) were positioned in the region in between the anterior part of the anterior sigmoid gyrus and the posterior part of the proreal gyrus using stereotaxic coordinates, obtained from functional recordings in behaving animals (AP: 25.5–28.5 mm (0 to 3 mm on Figure 2.2d) from caudal crest, caudal crest antero-posterior position being defined at 5 mm lateral from the medial crest/ML: 2 mm (Radtke-Schuller, 2018)). Each trial consisted of 10 s of baseline, then 6 s of monophasic stimulation at 100 Hz and 200 μA (2 ms pulses, 200ms-long train, repeated at 2 Hz), after a return to baseline of 10 s. The %CBV was computed as the mean response between 3 and 6 s after stimulation onset. 30 trials were performed for each A-P position of the electrodes. In these connectivity experiments, the animal was slightly sedated using a small dose of medetomidine (Domitor 0.02 mL at 0.08 $\mathrm{mg/kg}$) to reduce movement artifacts. Stimulation experiments were performed in one ferret, and each of the four experiments presented (Figure 2.3 and its figures supplements) was done once, on different days.

2.3.9 Anatomical tracers

A one year old female ferret weighing 620 g received a 2 μl injection of pAAV2.5-CaMKIIa-hChR2(H134R)-EYFP (PennCore) as anterograde tracer into left FC. Six months later the animal was perfused and the brain was cryoprotected, shock frozen and cut on a cryostat into 50 μm thick frontal sections into parallel series of which one was counterstained with neutral red. For overview images, combined B_{right}field and fluorescence images were taken with a Hamamatsu slide scanner 2.0HT (Institut de la Vision) (Figure 2.2e, left). For details, fluorescence images were taken with a virtual slide microscope (VS120 S1, Olympus BX61VST) at 10× magnification (Figure 2.2e, right).

Anatomical structures were reconstructed in accord with the ferret brain atlas (Radtke-Schuller, 2018).

2.3.10 Signal processing, analysis and statistics

Tonotopic maps

Power UfD signal normalized towards the baseline was used to monitor changes in Cerebral Blood Volume (%CBV). The %CBV varied after stimulus presentation (Figure 2.1c) and we quantified voxel responses with the mean of %CBV in a time-window 3 to 5 s after sound onset. Tonotopy of the imaged structures was mapped as follows: for each voxel this mean vascular response across the five tested frequencies was used to determine its best frequency (BF). Statistical differences of the responses to different frequencies in an individual voxel (Figure 2.1c, tuning curve) were assessed using a Wilcoxon rank sum test (post-hoc test after significant ANOVA $p<1e\text{-}3$). For visualization purpose, maps were thresholded by showing only voxels that had (i) a minimal 15% response and (ii) a mean response at their BF highly correlated ($p<1e\text{-}3$) with the mean hemodynamic response. This thresholding method was used to highlight sound-responsive voxels (disregarding of frequency tuning), and thus allows for the display of zones that were poorly tonotopic (such as AEG). Note here that this thresholding was used only for visualization purposes. Maps constructed with a threshold based on frequency tuning gave similar qualitative results. The mean hemodynamic response was used to approximate the typical vascular response to stimulus (as the Hemodynamic Response Function does for fMRI) and was computed in each structure as the average response over all the voxels showing a response to sound with z-score >3. Note that thresholds could be adjusted depending on the overall responsiveness of different structures and different animals, for illustration purpose. Intriguingly, two additional ferrets did not show any reliable response to sound (responses below 10 %CBV), for unknown reasons. They were not used in the experiments.

Last, maps were spatially smoothed with a $3 \times 3 \times 1$ voxel gaussian filter (std $= 0.5$), and a 3D median filter ($3 \times 3 \times 3$) was applied to the significance map to remove isolated voxels. The view of the brain surface (Figure 2.1c) was computed as the mean BF averaged from 5 to 10 voxels from the auditory cortex surface delimited manually. For 3D reconstructions of the cortex only, manually adjusted masks were used in order to show only tonotopic regions, and avoid crowdy representations caused by voxel transparency in the 3D visualization. Cortical depths were obtained by manually tracing the surface (just below the pia's blood vessels) and depth limits of the cortex. The 10 different depths were then automatically extracted by a custom-made algorithm (Figure

2.2a and Figure 2.2—figure supplements 1 and 2). The number of voxels at each depth was then equalized for the decoding analysis.

For the single slice analysis presented in Figure S2.5, the protocol was designed to speed up tone-responses acquisition (2 s tone, and random interval of 4 to 6 s - uniformly distributed - between two tone presentations). We then used a General Linear Model (GLM) to compute impulse responses of individual voxels to each tone frequency, without any predefined hemodynamic response function. This allowed us to present more stimuli (75 per frequency) in a relatively shorter time (~45 min).

Decoding

Frequency selectivity of the auditory cortex was assessed using a 5-class linear classifier and a leave-one out strategy: for each frequency pair, vascular responses of the two frequencies (%CBV averaged over 4 to 5 s after sound onset) were separated in a voxel-based space via a linear boundary optimized on 9 of the 10 trials in a learning set. No thresholding procedure was used in this analysis. Overall, pseudo-populations were built by grouping, across all slices recorded within the same structure, trials with identical frequency labels. The decoder was run over 100 shuffles of these pseudo-populations, where train and test sets were randomly chosen. In single slice analysis (Figure S2.5), we used a Fisher decoder (normalized by covariance) in order to take into account the noise correlation between voxels in decoding analysis. This was doable thanks to the higher number of tone presentations that allowed us to have a stable estimation of the covariance matrix.

In order to prove the significance of the obtained accuracy, we used a permutation procedure in which we shuffled the labels (i.e., which frequency was played during each trial) across trials, and performed the same decoding analysis, thus obtaining the chance distribution for decoding accuracies. We used 100 permutations, and considered that the real decoding accuracy was significantly out of the chance distribution (trial frequency labels shuffled) when above the 95th percentile. All the actual decoding accuracies were above the chance decoding accuracies. Our p-value resolution is limited by the number of permutations (100) and therefore our obtained p-values are all below 0.01.

To evaluate whether cortical depth had an effect on decoding accuracy (Figure 2.2a), we performed a one-way repeated-measure ANOVA over the four different craniotomies, with depth as the factor.

Resolution quantification

In order to quantify the minimal spatial scale at which fUS can provide independent information from two neighbouring voxels, we focused on sharp edges of functional transition and performed 2-way (voxel and frequency as factors) ANOVA on the tuning curves (%CBV averaged over 4 to 5 s after sound onset) of each pair of voxels within a certain contour (example transect and contour shown in Figure 2.2b, left panel). The voxel factor quantified the dissimilarity in the average responses for two voxels, being thus representative of an overall responsiveness dissimilarity when significant. The interaction term (frequency x voxel) quantified how dissimilar the tuning curves were for two different voxels, independently of their overall responsiveness. This term therefore represented our ability to discriminate between different functional voxel tuning. Pairs of voxels were considered to be 'dissimilar' (in responsiveness or tuning) when the associated p-value was $<5.10^{-2}$. Importantly, these values depend on the smoothness of the underlying functional neuronal map (the sharper the better) and on the number of trials used in each experiments (the higher the better). Here, we show that using only 10 trials per frequency, we could go down to a functional resolution comparable to the voxel size (100 µm) for the overall responsiveness, and of 300 µm for the tuning.

We randomized 50 times the responses over all voxels and all frequencies and performed the same analysis to find the average distribution expected by chance for both responsiveness and tuning dissimilarity percentages. We determined the spatial resolution as the shortest distance between two voxels at which the actual number of dissimilar pairs was above the 95th percentile of the randomized distribution. Distance between voxels defined by coordinates (x_1,y_1) and (x_2,y_2) was computed as the rounding of $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

Finally, we performed this analysis in different regions (AC and IC) and different animals (B_{right}, V_{left}, V_{right}, S_{right}) in order to generalize this result (Figure S2.6).

2.4 Results

2.4.1 Mapping the tonotopic organization of auditory structures

Physiological experiments were conducted in three awake ferrets (*Mustela putorius furo*, thereafter called V, B and S). After performing craniotomies over the temporal lobe, chronic imaging chambers

were installed (both hemispheres in one animal, and right hemispheres in the other two) to access a large portion of both the auditory (middle and posterior ectosylvian gyri - resp. MEG and PEG) and visual cortex (in caudal suprasylvian and lateral gyri) (Figure 2.1a). The 3D scan of the craniotomy via Ultrafast Doppler Tomography Demené et al. (2016) revealed the in-depth vasculature of the Auditory Cortex (AC) surrounded by the supra-sylvian sulcus (Figure 2.1a and b). In addition, we were able to detect and image deep auditory-responsive structures such as the Medial Geniculate Body (MGB), the Inferior Colliculus (IC) and the dorsal nucleus of the Lateral Lemniscus (DNLL), as well as visually-responsive nuclei such as the Lateral Geniculate (LGN) (Figure S2.1).

In order to reveal the tonotopic organization of the auditory structures, we recorded in each voxel the evoked hemodynamic responses to pure tones of 5 different frequencies by computing the %CBV, defined as the percentage of variation in CBV. We then computed the resultant 3-dimensional tonotopic map (Figure 2.1c–e, Figure S2.2). Within a relatively short time (10 to 15 min per slice), we could accurately reproduce the known tonotopic organization of the primary (A1 and AAF in the middle ectosylvian gyrus) and secondary auditory cortex (PPF and PSF in the posterior ectosylvian gyrus) (Bizley et al., 2005; Mrsic-Flogel et al., 2006; Nelken et al., 2008), with a high- to low-frequency gradient in A1, reversing to a low- to high-frequency gradient in the dorsal PEG (Figure 2.1c). We note that the fUS enabled us to map within the challenging deep folds of the ferret auditory cortex, such as the supra-sylvian sulcus (sss) and pseudo-sylvian sulcus (pss). Recordings could be performed in the same slice across days, with a high repositioning precision (error $<$1 slice, 200 µm in that case), which was within the range of the out-of-plane point-spread function for fUS (Figure S2.3). Interestingly we were able to capture inter-individual variability along the transect going from the pss to the sss, consistent with previous work in the ferret (Bizley et al., 2005).

Large-scale, 3D functional maps were also recorded in the deep and smaller structures of the auditory thalamus (MGB, Figure 2.1d), the inferior colliculus (IC, Figure 2.1e) and the DNLL (Figure 2.1e). The 3D views obtained in fUS allowed us to describe for the first time the tonotopical organizations of the ferret ventral division of the MGB and DNLL. This is particularly remarkable in the latter structure in which we characterize a precise tonotopic map despite its small size (~1 mm-long) and subcortical position (8 mm deep below brain surface). Moreover, such a large field of view allows one to measure simultaneously the functional organization of any coplanar structure (such

as A1 and the MGB here), thus opening the door to precise, frequency specific (thalamo-cortical) connectivity studies. In this respect, future development of high frequency fUS matrix-probes for 3D UfD imaging (Provost et al., 2015) will extend this capability to any brain structure.

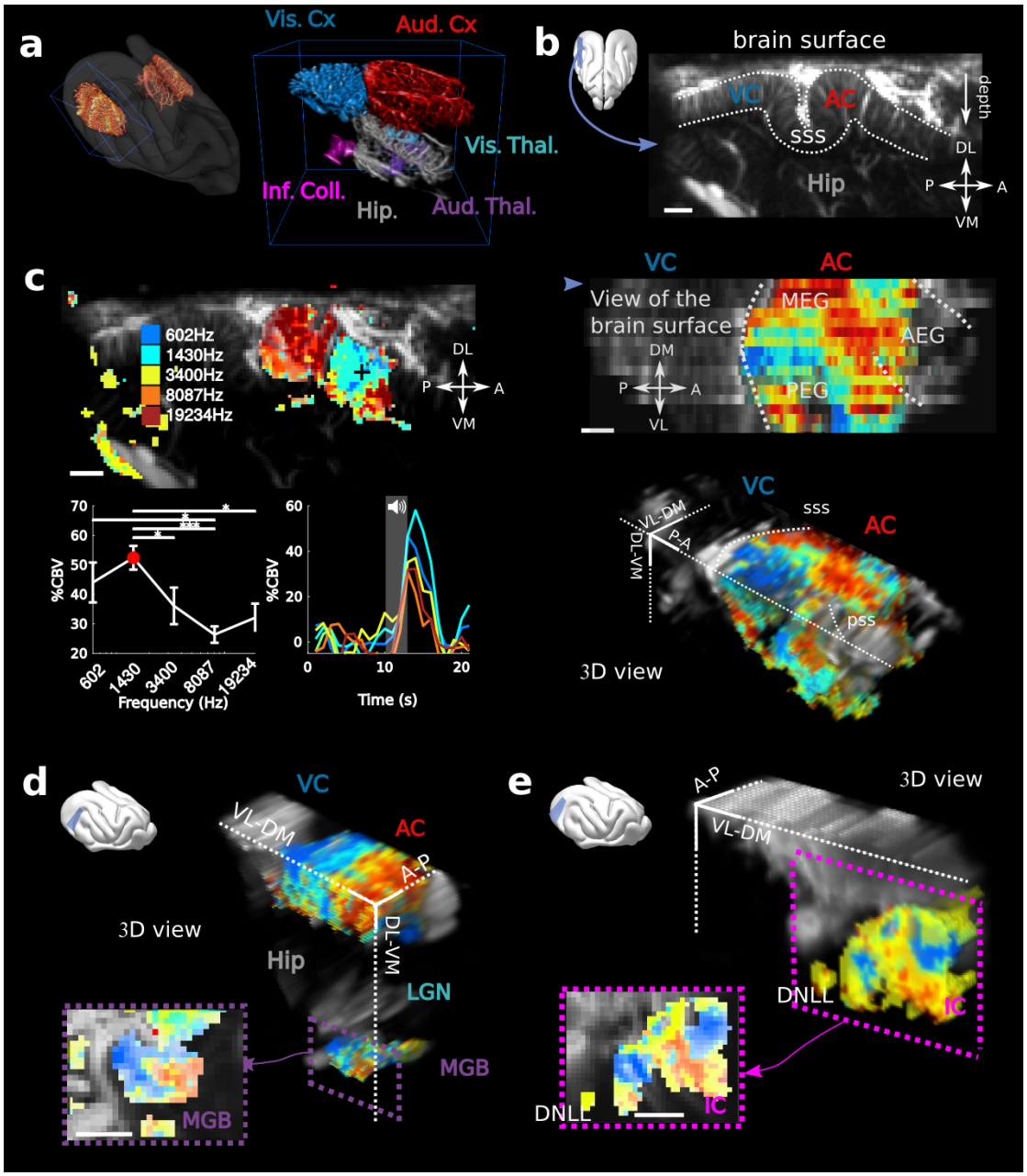

Figure 2.1: **fUS imaging reveals the tonotopic organization of cortical, sub-cortical, and intracortical auditory structures in the awake ferret.** (**a**) Left: UFD-T of the left and right craniotomies, superimposed on an MRI scan of a ferret brain. Right: magnification of the blue bounding box (left). Auditory structures: auditory cortices (AC), medial geniculate body (MGB), inferior colliculus (IC). Other structures: hippocampus (Hip), visual cortex (VC). (**b**) Structural view of a tilted parasagittal slice (~30° from D-V axis) of the visual and auditory cortices (represented as a blue plane on the 3D brain). Lining delineates the cortex. (legend continued on next page)

Figure 2.1: (**c**) Upper left: Tonotopic organization of the slice described in (b). Lower left: tuning curve (mean ± sem) and average responses in %CBV (see Materials and methods) for the voxel located in the upper panel (black cross). Upper right: combination of 16 similar slices over the surface of the AC, arrow depicts slice of (b). AEG/MEG/PEG: anterior/middle/posterior ectosylvian gyrus. Lower right: 3D reconstruction of the whole AC's functional organization. (**d**) 3D reconstruction of both the auditory cortex and auditory thalamus (non-tonotopic areas were masked on this reconstruction for clarity of the representation). Inset: single slice centered on the MGB. Its tonotopic axis runs along the PL-AM axis. Note that (b–d) were extracted from the left side of the brain, but flipped for visual clarity and coherence. (**e**) 3D reconstruction of the inferior colliculus and the dorsal nucleus of the lateral lemniscus (DNLL). Inset: single slice centered on the IC. Both (d) and (e) are tilted coronal slices (~30° from D-V axis). Their tonotopic axis runs along a ~20°-tilted D-V axis. All individual and converging scale bars: 1 mm. D: dorsal, V: ventral, M: medial, L: lateral, A: anterior, P: posterior.

2.4.2 Decoding

Single-trial analysis is essential for understanding brain dynamics and behavioral variability. However, it remains a challenge as it necessitates to record high-quality signal from a large number of neurons/voxels at the same time. In order to estimate the reliability and selectivity of fUS single-trial responses, we used MultiVoxel Pattern Analysis (MVPA) to decode the stimulus frequency from the hemodynamic signal. Using a simple linear decoder, we attained high decoding accuracy in the auditory cortex (from 0.46 to 0.63 probability, with chance at 0.2) which was even more striking in the IC and DNLL (from 0.72 to 0.98), despite their smaller size and subcortical location (Figure 2.2a). These results suggest that single trials show reliable and significant activity across all structures.

On a different scale, we sought to demonstrate whether fUS could also reveal encoding differences across cortical layers. We focused on imaging the small vessels in the cortex (keeping only data corresponding to an axial projection of blood flow lower than 3.1 mm/s) and defined cortical layers using an unfolding algorithm providing a flattened version of the AC (Figure S2.4). A linear decoder yielded a significantly higher decoding accuracy when using only measurements at intermediate cortical depths ($p<1e\text{-}3$), peaking around 400–500 µm below the surface (up to 0.83, mean 0.67), consistent with it being granular. As a control, we note that baseline blood volume and response magnitude did not show a similar depth-dependent profile (Figure S2.4), suggesting that the observed decoding accuracy may be due to variations in capillaries structure within cortical layers (Adams et al., 2015). An alternative explanation would be that the improved accuracy at the intermediate depths reflects the underlying neuronal activity, and more specifically the sharper frequency tuning observed in granular layers (Guo et al., 2012). Importantly, all these results could be confirmed in single slice recordings, and over several days (Figure S2.5), showing that the

hemodynamic signal imaged in fUS is reliable enough to decode brain activity on a single-trial basis within a single experiment.

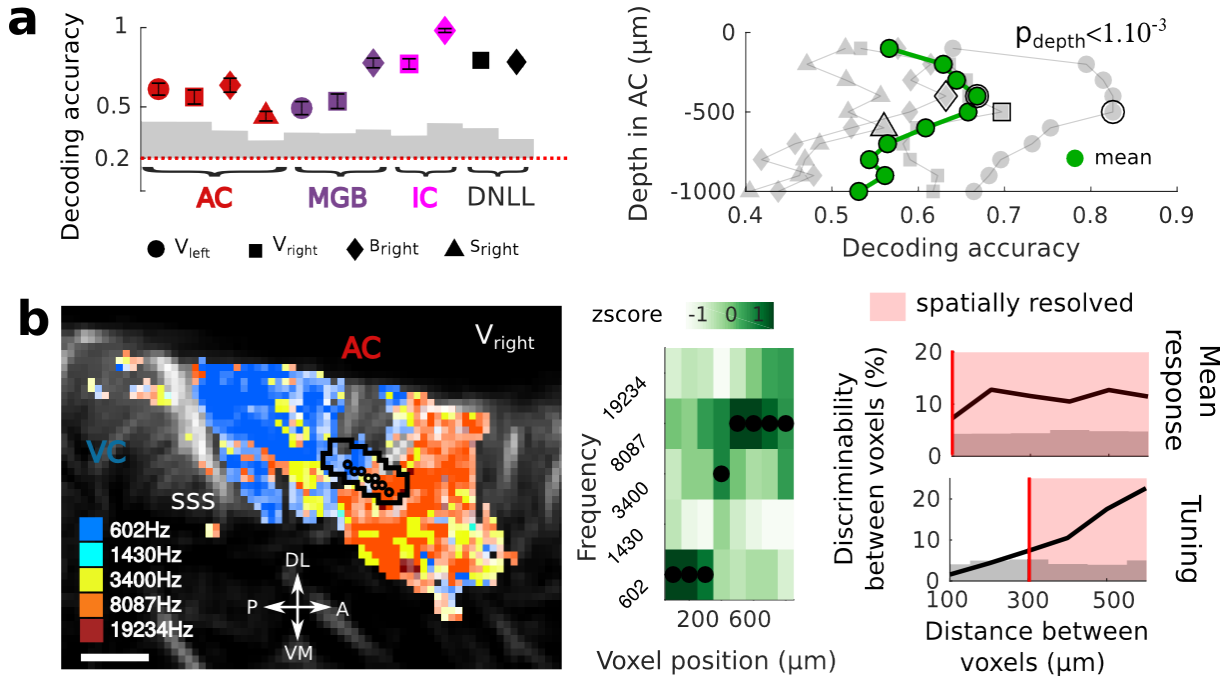

Figure 2.2: **Key features of fUS in awake animals: decoding accuracy, layer effect, and effective spatial resolution.** (**a**) Left panel: Decoding accuracy over the five frequencies, in different structures and different craniotomies (see legend). Grey histogram shows the upper limit for chance ($p<10^{-2}$, mean $\pm$ 2 sem computed over 100 randomized decoding sessions). All structures showed significant decoding ($p<10^{-2}$). Right panel: decoding accuracy over depths, computed from the activity in the AC of 3 different animals (grey plots). All showed a similar profile, with the accuracy peaking between 400 and 500 µm. The green plot shows the average trend (repeated-measure ANOVA over depth, $p<10^{-3}$). (**b**) Left panel: example of a sharp tonotopic transition from low to high frequency, in the auditory cortex of Vright (map not smoothed). Scale bar: 1 mm. Middle panel: heatmap of the z-scored tuning curves of the consecutive voxels (shown by circles in left panel), with the best frequency indicated by a black dot, showing a shift from low to high frequency preference. Right panel: quantification of the lower spatial limit at which one can significantly find differences in the responsiveness (upper) or tuning (lower) of two voxels, with respect to their distance. Grey histogram shows the upper limit for chance ($p<5.10^{-2}$, 5% percentile over 50 randomizations). In that specific case, it was respectively 100 µm and 300 µm. The voxels used in this analysis are the ones within the black contour in left panel, centered on the sharp transition.

2.4.3 Functional resolution

Next, we took a closer look at the tonotopic organization in different structures to examine how tuning curves in neighboring voxels change abruptly. This finding exemplifies the ability of fUS imaging to measure independent information at a very small spatial scale. To quantify the minimal functional spatial resolution of the technique, we defined a discriminability index between voxels,

and focused on sharp transition areas (Figure 2.2b left panels). We found that fUS can discriminate responsiveness of neighboring voxels, with a functional resolution as fine as 100 µm (Figure 2.2b). Furthermore, we were able to discriminate voxels based on their tuning curves within a distance of 300 µm in as little as 10 repetitions per frequency (Figure 2.2b and Figure S2.6). Importantly, this is a conservative measure of functional resolution, since it largely depends on the smoothness of the underlying functional organization itself (tonotopy) and of the number of trials. The functional resolution described here is thus a lower limit, and could be improved by increasing, for example, the trial number. These results suggest that fUS can be useful to assess the fine organization of vascular domains within brain structures and to better understand the functional coupling between local neuronal activity and the dynamics of surrounding blood vessels, two important questions for hemodynamic-based techniques (O'Herron et al., 2016; Harrison et al., 2002).

2.4.4 Assessing connectivity between structures

Another fundamental view of brain function and functional organization is revealed by mapping brain connectivity among various structures. Localizing and quantifying such connections in awake animals, however, remains technically challenging since tracer injections are not an option, and fMRI gives only access to indirect, spatially diffuse measures of connectivity strength. Here, we demonstrate that fUS can be used to probe the functional connectivity between two brain structures that are far apart: the frontal and the auditory cortices. The frontal cortex (FC) is a region that has been shown to be involved in top-down modulation of early sensory areas, and in particular of the auditory cortex (Fritz et al., 2010; Winkowski et al., 2013). To reveal its potential links to the auditory areas, we electrically stimulated at different points within the FC while recording evoked hemodynamic responses in the auditory cortex of an awake (slightly sedated) animal (Figure 2.3a). Importantly, this technique does not require any precise priors on the location and nature of the terminal projections. By imaging widely in the auditory cortex, we observed evoked activity in the insular cortex of the pseudosylvian sulcus (PSSC/insula), which was maximal for a certain depth and position of the stimulating electrode (Figure 2.3b, Figure S2.7). By contrast, there was no evoked activity recorded in secondary auditory areas such as the PEG (Figure S2.8). We also observed a decrease in blood volume in the MEG, possibly originating from polysynaptic connections between FC and A1 (Logothetis et al., 2010; Klink et al., 2017).

From these recordings, we cannot disentangle orthodromic versus antidromic activation. We therefore anatomically confirmed the existence of such descending projections from FC to

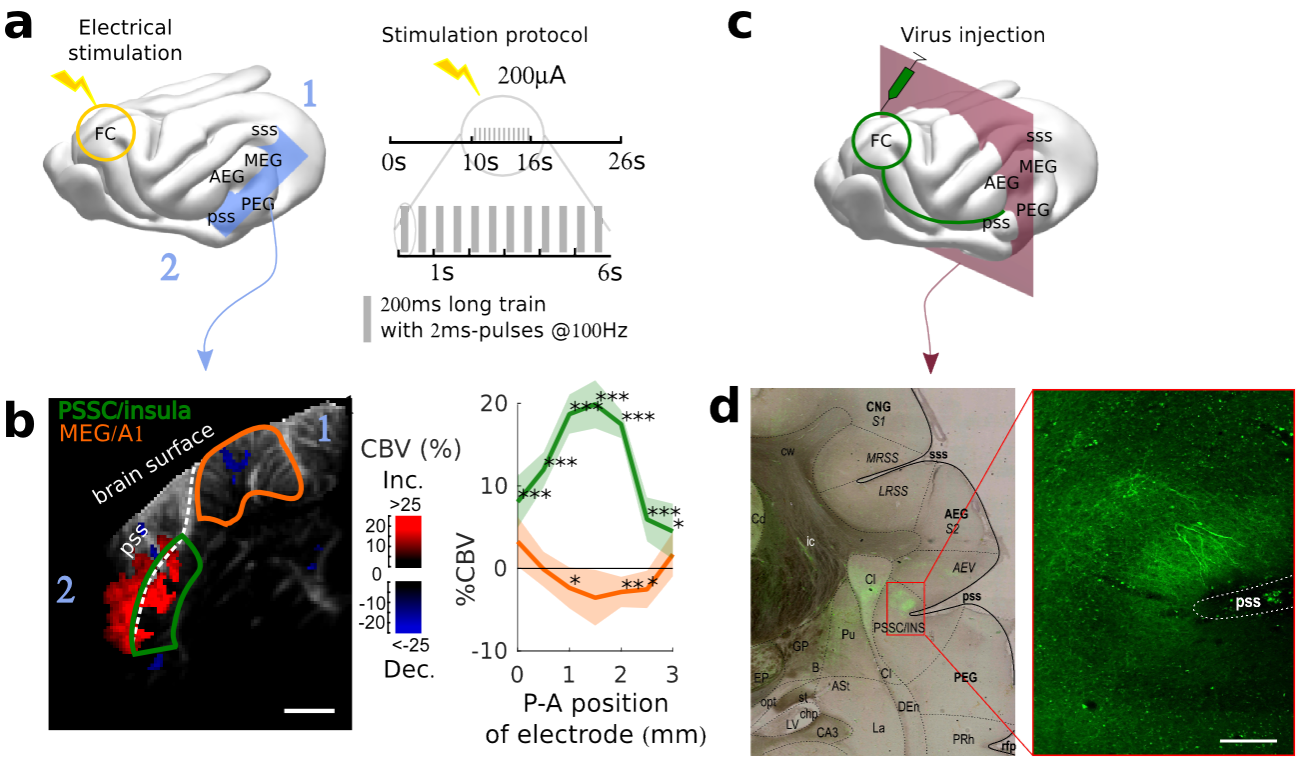

Figure 2.3: **Exploring long-distance connectivity: the example of top-down projections from dlFC to the auditory system.** (**a**) Ferret brain with localization of electric stimulation (lightning) and site of fUS imaging shown in (b) (blue plane). A schematic of the electrical stimulation protocol (details in Materials and methods) is also shown in right panel. (**b**) FC-AC direct projection patterns revealed in fUS. Left: fUS imaging plane along the PSSC/insula, showing modulations of hemodynamic activity in MEG (orange delimitation) and PSSC/Insula (green delimitation) evoked by FC stimulation (map thresholded at $+4$ sem). The numbers 1 and 2 are here to help orientation. Right: %CBV in the 2 regions of interest after FC electric stimulation (highlighted in the left panel) with respect to the postero-anterior position of the stimulation electrode (0 represents 25.5 mm from caudal crest, 3 represents 28.5 mm), revealing a hot-spot of connectivity at about 1 mm (i.e 26.5 mm from caudal crest) (mean ± 2 sem). ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$.. Scale bar: 1 mm. (**c**) Ferret brain with localization of virus (tracer) injection site (green circle) with symbolized projections, and coronal slice represented in (d) (red plane). (**d**) Anatomical confirmation of connectivity. Left: bright field combined with fluorescence imaging, showing green fluorescent FC projections concentrated in the depth of the PSSC/insula and delineated anatomical structures (scale bar: 200 µm). Right: close-up of the labelled FC projection terminals in the PSSC/insula.

PSSC/insula with independent anterograde virus injections in FC. These injections revealed monosynaptic projections that targeted the PSSC/insula (Figure 2.3c–d), consistent with a contribution of direct projections from FC to A1 to the functional connectivity pattern revealed by the fUS approach. We also observed FC projections in the Claustrum (Cl in Figure 2.3c), ventro-medial with respect to the PSSC/insula. Because the neighboring regions have been reported to be multimodal (Bizley et al., 2006; Bizley and King, 2008), we subsequently explored the responsiveness of the FC-targeted PSSC/insula to acoustic and visual stimuli. We found this region to be less responsive to broadband

noise than A1 (~5% instead of 15%), and not driven by visual stimuli (Figure S2.9). Altogether, this experiment offers a proof-of-concept of how fUS can serve as a tool to characterize large-scale functional connectivity without sacrificing any resolution. We can point out two key applications building up on such experiments. First, one may explore connectivity changes in animals, for example during different brain states (e.g., sleep vs. awake), or during the course of learning. Second, and maybe even more importantly, the use of optogenetics can allow a precise mapping between brain structures, targeting for example specific neuronal subpopulations, or projection patterns. The development of such tools has just started, but has been so far limited to fMRI (Lee et al., 2010).

2.5 Discussion and conclusion

In this chapter, we have shown that fUS imaging can serve as a technique to record in awake animals a very stable (over days), high-resolution and simultaneous tonotopic mapping of various brain regions, be they large, small, superficial, or deep. This was done over multiple scales, from functional tuning of individual voxels to large-scale connectivity between brain regions. The amplitude of the fUS responses (~20% in the ferret, and close to 50% in neonates (Demene et al., 2017)) is quite large compared to typical auditory cortex BOLD responses in fMRI (~5%). This makes mapping both rapid, compared to the electrophysiological approach with multiple penetrations (Bizley et al., 2005; Mrsic-Flogel et al., 2006), and precise, as illustrated by the ease with which single-trial information can be decoded from its high-sensitivity signal, a key feature when it comes to recording in behaving animals. Furthermore, fUS can be a valuable tool in acquiring broad, yet accurate views of the functional organization of unmapped brain regions and their connectivity with the rest of the brain. Finally, fUS imaging can be readily adapted to mobile and highly stable configurations (Sieu et al., 2015), which will make it ideally suited for behavioral cognitive neuroscience studies requiring extended observations, as in the characterization of the neural correlates of learning.

2.6 Supplementary figures

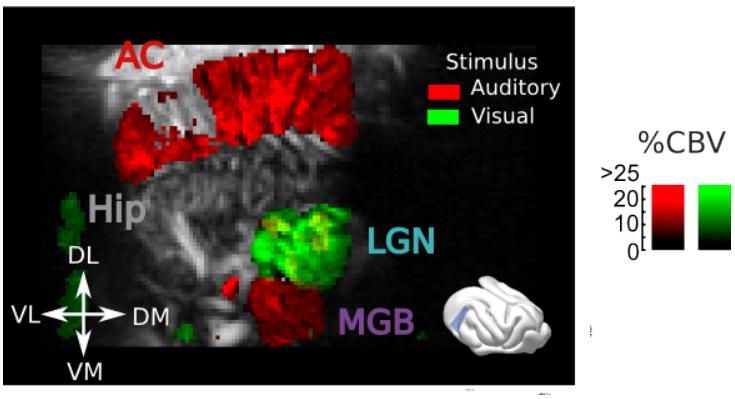

Figure S2.1: **Responses to visual and auditory stimuli in the cortex and thalamus.** Tilted coronal slice (30° from D-V axis) over the AC and thalamus, showing hemodynamic responses evoked by a flickering light (green) or a broadband auditory noise (red) (map thresholded at +4 sem). Note that sound evoked activity in the most anterior part of the LGN can be visible (yellow color). Scale bar: 1 mm.

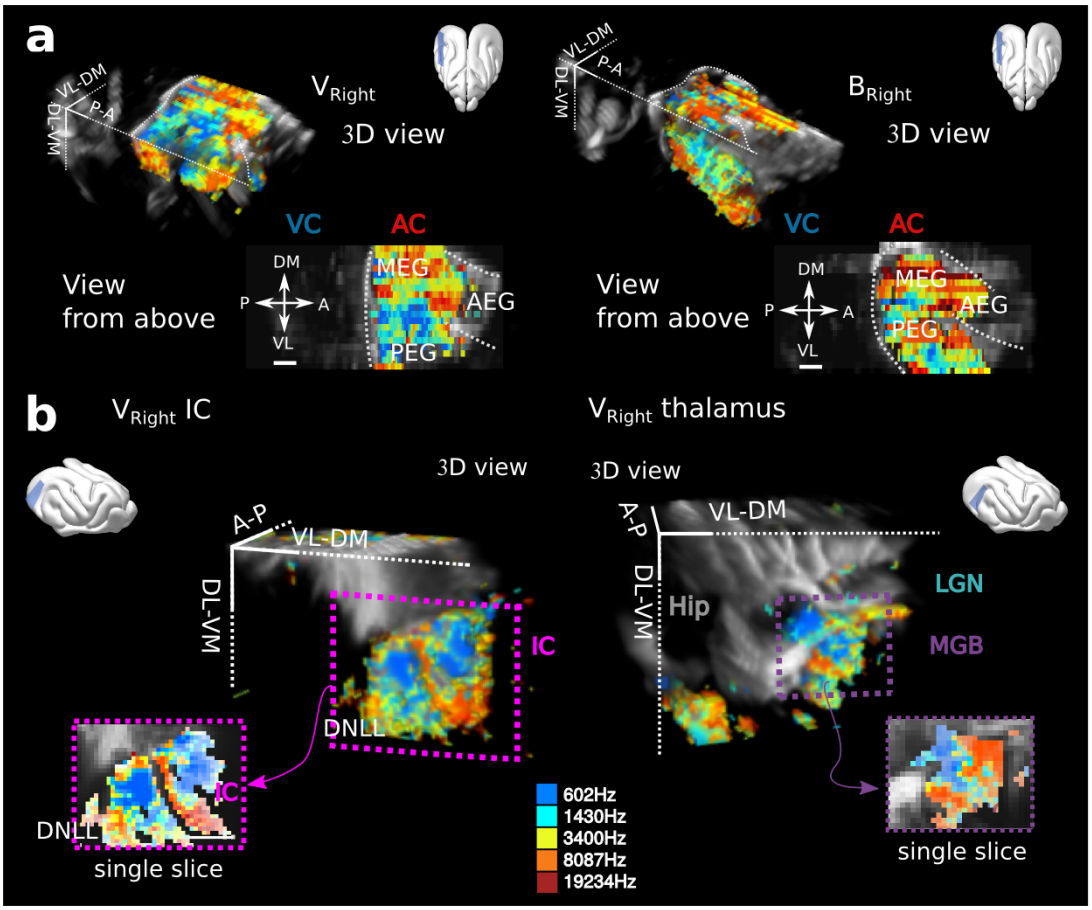

Figure S2.2: **Tonotopies in AC, IC and MGB for other animals.** (**a**) 3D reconstructions and views from above for two other craniotomies, the right side of the one (named V) presented in Figure 1c, and another animal (B). Note the clear double reversal from MEG to PEG to VP in Bright. (**b**) Tonotopy for the IC in Vright (left), in which both IC and DNLL are visible, and the MGB in Vright (right). All tonotopic axis are consistent across craniotomies, even if substantial anatomical differences can be seen across animals, especially illustrated in the AC. Presented structures are oriented as tilted coronal sections (30° from D-V axis). All individual and converging scale bars: 1 mm.

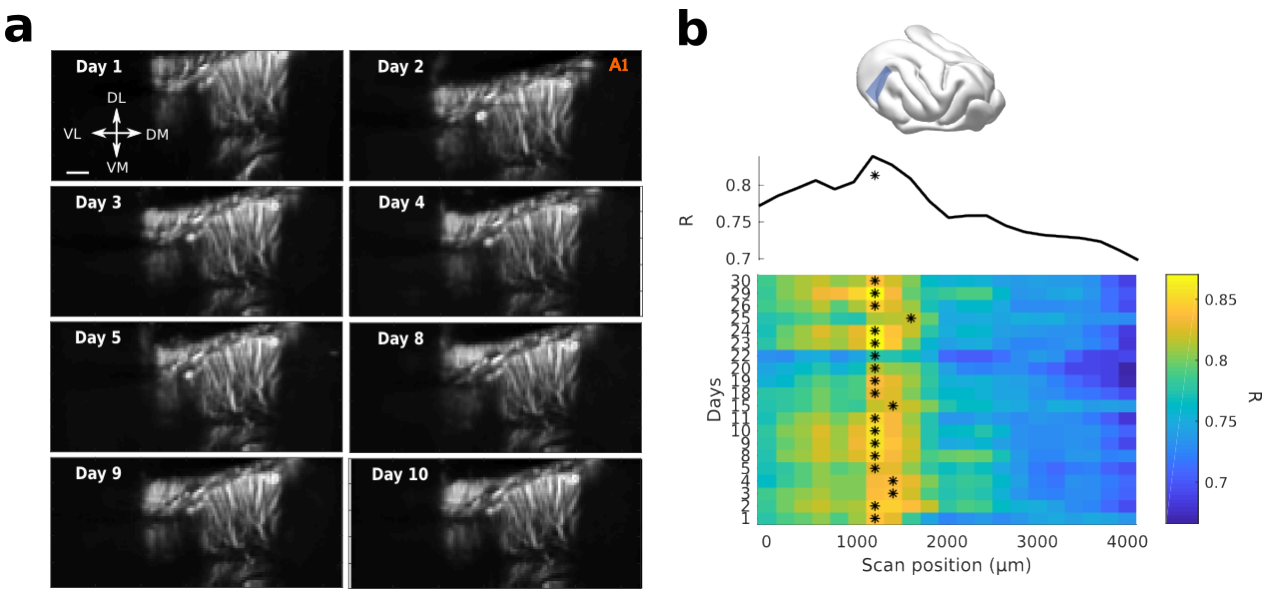

Figure S2.3: **fUS allows for high recording stability and repositioning over days.** (**a**) Structural slices (tilted coronal slice, 30° from D-V axis, right hemisphere) over days, by repositioning the probe with a stereotaxic apparatus. Scale bar: 1 mm. (**b**) Recordings from the same slice were performed everyday for a long period of time. Each daily slice was repositioned in a vascular atlas previously obtained in the same animal, same craniotomy. The position is obtained by maximizing the correlation (R) between the new slice and the previous vascular atlas. Here the heatmap of R for different days (y-axis) correlated to different A-P regions of the atlas (x-axis) is shown. The star shows the maximum correlation for each day, and so the repositioning of the new slice (~1200 µm in that case). The upper panel shows R averaged over days.

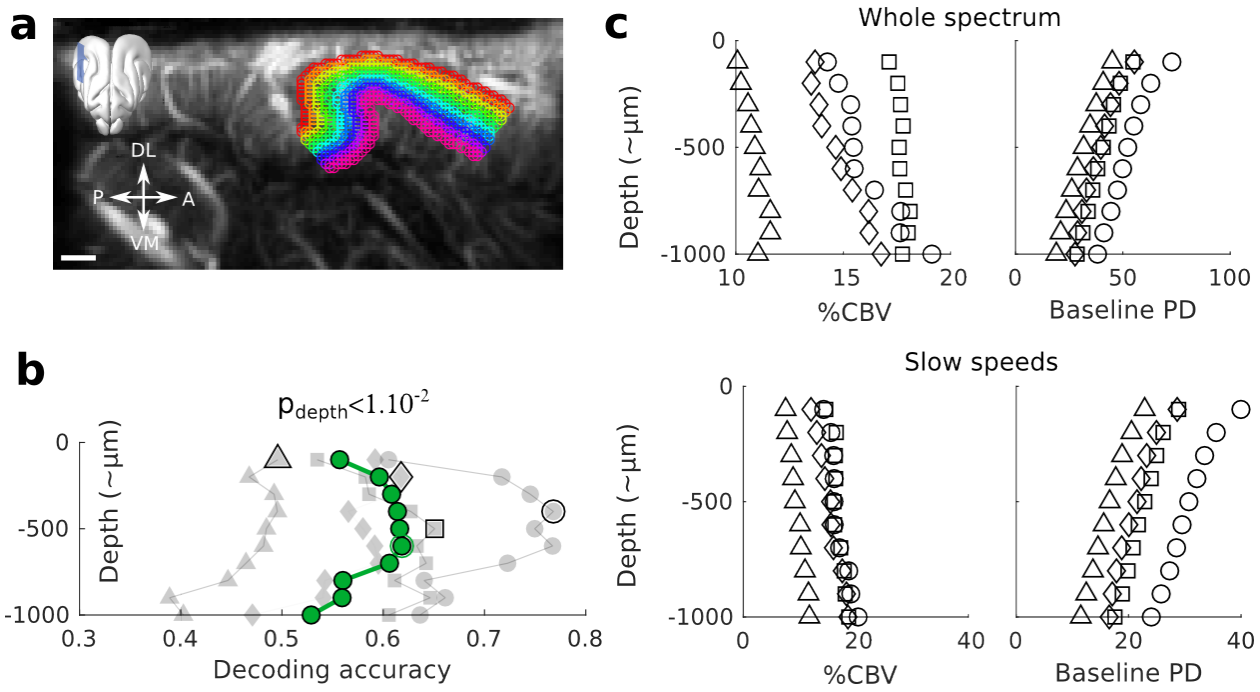

Figure S2.4: **Controls for the decoding across depths.** (**a**) Example slice where the different depths are superimposed on the structural image. The upper part of the cortex was identifiable by the high density of vessels, while the lower part was approximated based on the end of vertical blood vessel, and distance to the surface (~1 mm). Note that with this definition, the first upper layer could accidentally contain some voxels within the pia. Scale bar: 1 mm. (**b**) Decoding across depths without focusing on the capillaries (whole spectrum). The same trend ($p<10^{-2}$) than in Figure 2.2a is visible, but less peaked and with lower accuracies. (**c**) Control measures of the %CBV (average maximum response over all frequencies) and baseline Power Doppler (PD, arbitrary unit) as a function of depth, indicating respectively the average responsiveness of each depth and its average baseline CBV. Upper panel: whole spectrum (no filtering). Lower panel: blood speeds above 3.1 mm/sec were filtered out, as in the main Figure 2.1f.

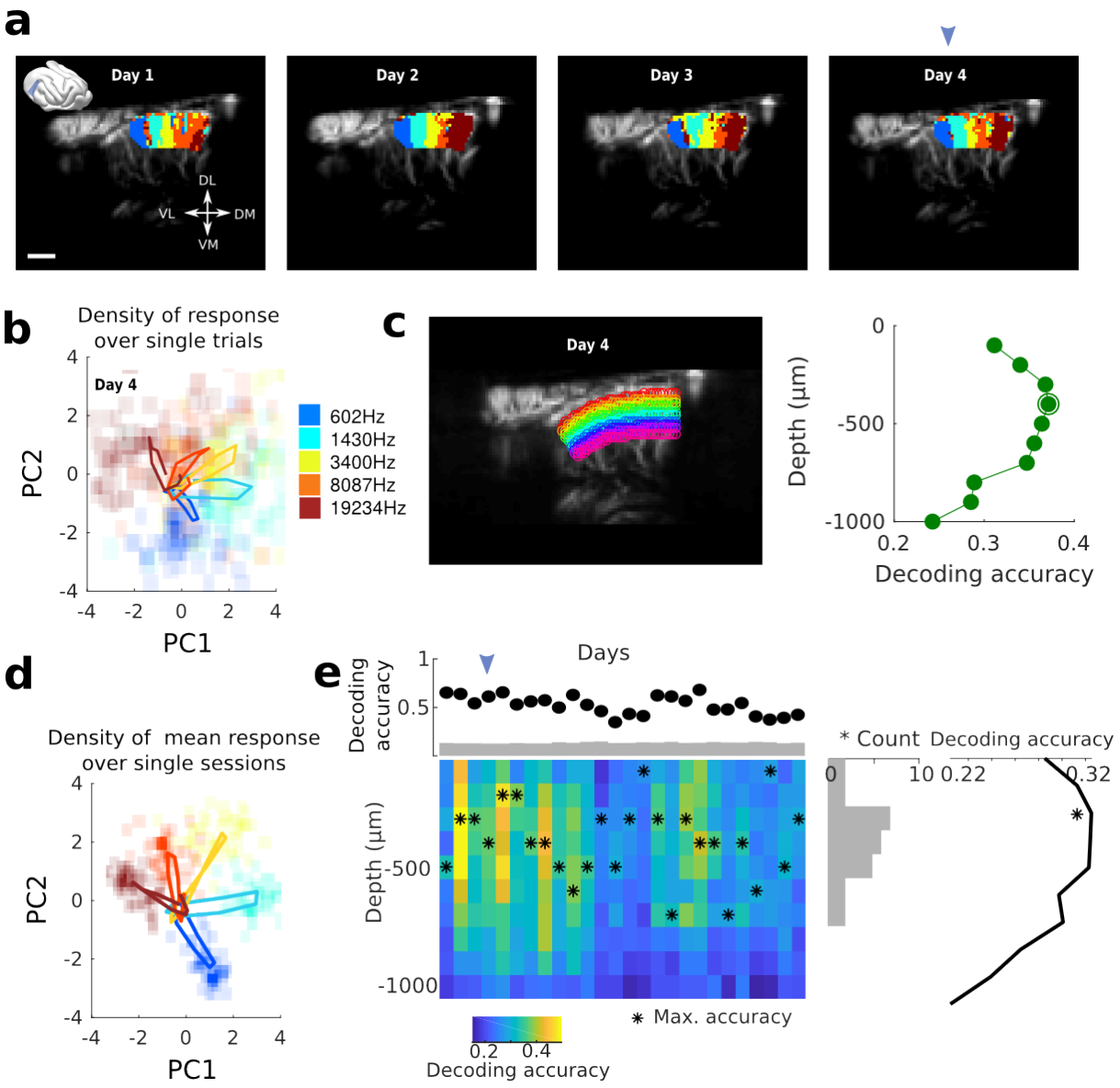

Figure S2.5: **Single-slice recordings show high decoding possibility on an actual single-trial basis.** (**a**) Tonotopic organization of a tilted (30° from D-V axis) coronal slice of A1, over four consecutive days. A mask has been applied to focus on the tonotopic area. Scale bar: 1 mm. (**b**) PCA analysis over the averaged response for each frequency and all the voxels highlighted in (a), for the single slice designed by an arrowhead. Plotted here as plain lines are the mean hemodynamic response (starting from the center at sound onset, and increasing in all five directions), superimposed on the density of the peak responses at the single trial levels (N = 75 trials per frequency). Each frequency is designed by its color, and the intensity of the colored shading shows the density of trials displaying a response at this location. We can clearly see a separation of the different frequencies on a single-trial basis (reflected in the decoding analysis). (legend continued on next page)

Figure S2.5: (**c**) Decoding accuracy as a function of depth (slow speed vessels only). Left: different depth for the specific slice. Right: decoding accuracy peaks again around -400 μm, thus confirming that this effect could be observable on a single-slice basis. (**d**) Similar analysis as in (b), but this time the PCA is computed over the mean responses averaged over trials and daily sessions. The density map shows here the density of the peak responses at a single session level, averaged over all trials. The fact that densities are quite centered around the mean response suggests that all sessions have similar patterns of activity, and that tonotopic organization is relatively stable. (**e**) Decoding analysis as a function of depth, over days. Upper panel: mean decoding accuracy for all sessions (all blood vessel speeds). Heatmap shows the dependence of decoding accuracy on cortical depth. Stars show the peak of accuracy for each day, which is summed up in the right histogram showing the distribution of peak accuracy position. It clearly peaks around -300/400 μm, thus confirming our results over multiple recordings in the same slice. The blue arrowhead indicates the position of the slice shown in (c). Far right: mean decoding accuracy as a function of depth, averaged over days. The heterogeneity in decoding accuracy can be due to many parameters, such as small sample size and real biological variations.

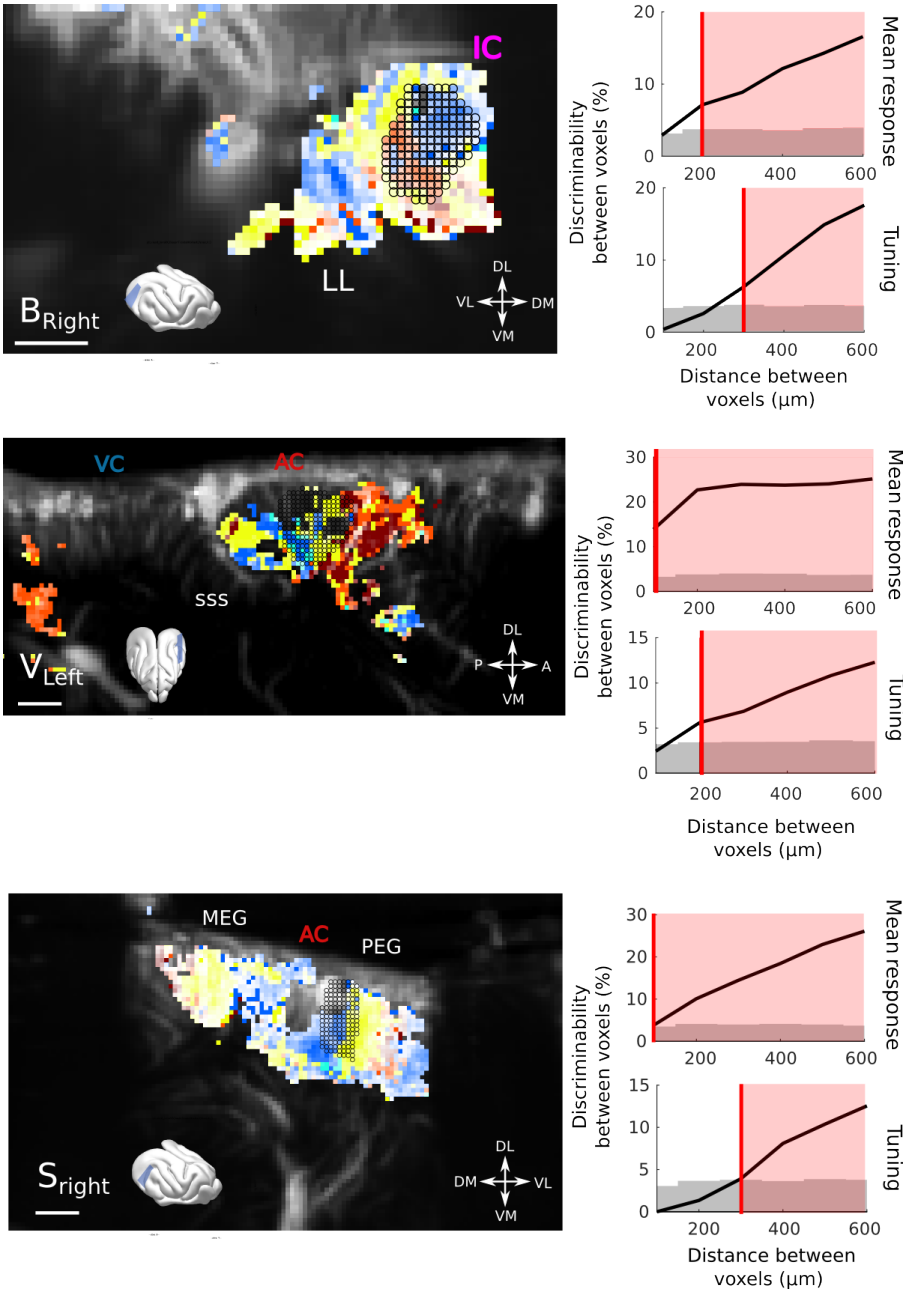

Figure S2.6: **Resolution quantification in other regions of the brain, and other animals.** We performed the same analysis as shown in Figure 2b, in other regions and different animals. Overall, the obtained resolution are similar, that is, 100µm for responsiveness and 200-300 µm for tuning, within only 10 trials. (**a**) Quantification in the IC of (B_{right} (10 trials). (**b**) Quantification in the AC of (V_{left} (10 trials). One can note here the heterogeneity of tunings within a small distance range. (c) Quantification in the AC of (S_{right} (20 trials, coronal slice). All scale bars: 1 mm.

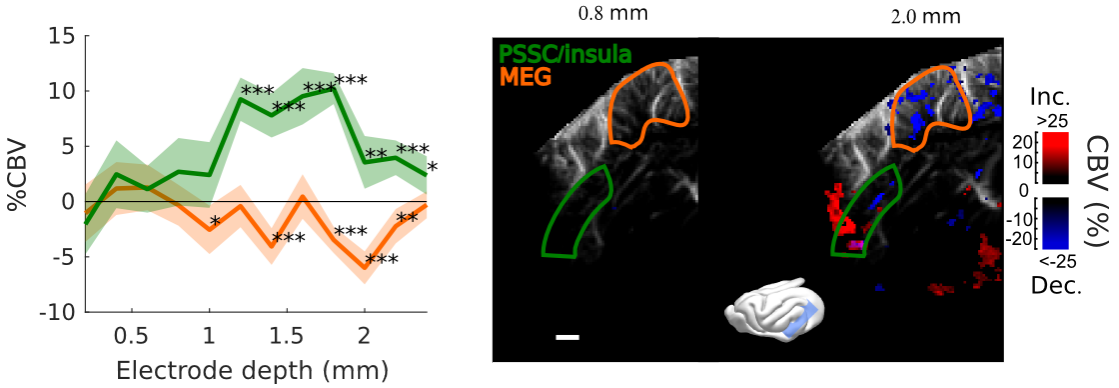

Figure S2.7: **Frontal Cortex - Auditory cortex connectivity explored further: cortical depth.** Evoked responses in MEG/A1 and PSSC/Insula as a function of the vertical depth of the stimulation electrode (mean $\pm$ 2 sem, map thresholded at $+4$ sem). Again, a hot spot of activation is found, suggesting that the bolus of activation triggered by our electrode does not exceed ~500 µm of a radius. Here, the 0 is set at the surface of the tissue covering the brain, that can be up to 1 mm thick. ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$. Scale bar: 1 mm.

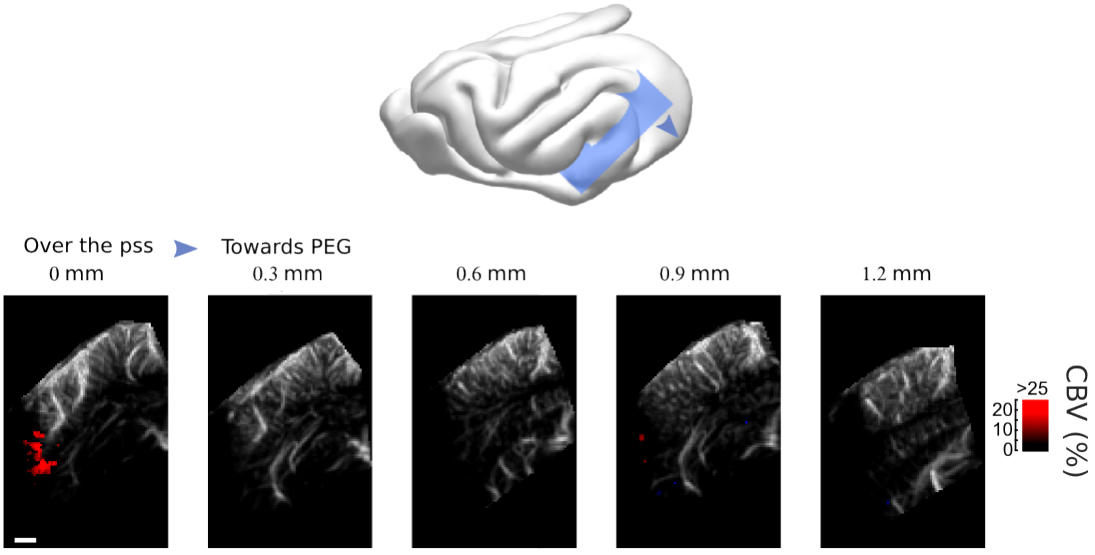

Figure S2.8: **Frontal Cortex - Auditory cortex connectivity explored further: secondary areas.** Scanning over the whole auditory MEG and PEG, showing that responses were evoked only in the fundus of the sulcus (map thresholded at $+4$ sem). Scale bar: 1 mm.

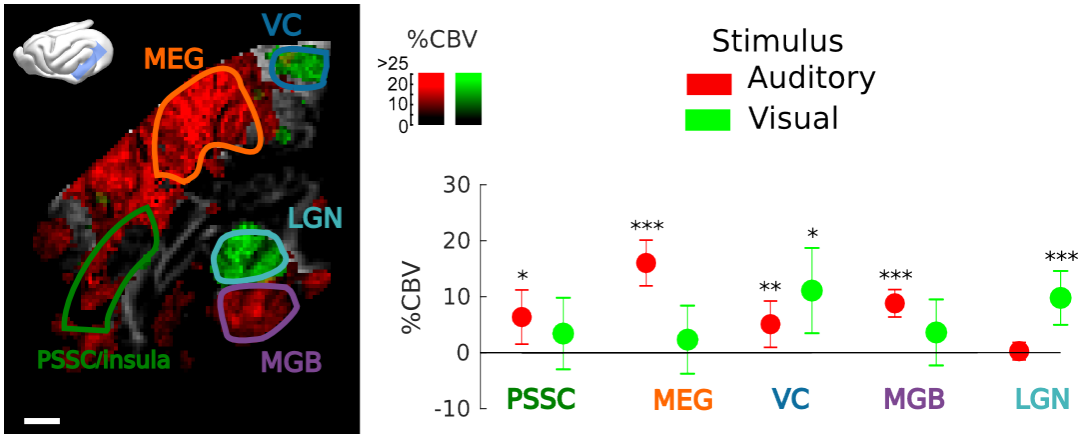

Figure S2.9: **Frontal Cortex - Auditory cortex connectivity explored further: sound and vision.** Exploration of the multimodal responsiveness of the area. We played broadband noise (red) or flickering light (green) while recording the evoked %CBV in the same imaging plane. Scale bar: 1 mm. Left: overall responses for both visual and auditory stimulations (map thresholded at $+2$ sem). Anatomical regions of interest used for quantification are outlined. Right: mean evoked responses in these different regions. Note that the PSSC/insula was only weakly activated by sound, compared to MEG/A1. The part of the visual cortex shown here also presented bimodal responses, suggesting that this could be part of higher association areas such as area 21a of visual cortex or posterior parietal cortex. These experiments were performed on the same animal, but different days. Errorbars show mean ± 2 sem. ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$.

Chapter 3

Natural sounds processing

Contributors: Agnès Landemard, Sam Norman-Haigneré & Yves Boubenec.

3.1 Abstract

How have speech and music shaped the human brain? Many signatures of speech and music processing have been observed in non-human animals, raising the question of whether there exist uniquely human mechanisms for processing these two categories of sounds. Humans have non-primary neural populations that respond selectively to speech and music compared both with other natural sounds and with synthetic sounds that have matched spectrotemporal modulation statistics ('model-matched' sounds), suggesting selectivity for higher-order structure. Using functional ultrasound (fUS) imaging, a cutting-edge high-resolution neuroimaging technique, we tested if similar regions are present in ferrets.

We measured responses from the auditory cortex of passively listening head-fixed ferrets to natural and model-matched sounds tested previously in humans. Ferret cortical responses recapitulated many of the response patterns observed in humans. Interestingly, we observed speech selective regions in the ferret auditory cortex. However, and contrary to the speech- and music-selective response components observed in human non-primary regions, ferret auditory cortex did not show selective responses to natural vs. model-matched sounds. These findings suggest that human cortical organization differs from ferrets' in non-primary auditory cortex due to the need to represent higher-order structure in speech and music.

Because speech and music are not ecologically relevant sounds for ferrets, we tested whether

ferret auditory cortex could discriminate between ferret vocalizations and their corresponding model-matched versions. We observed differences in animal motor activity for original compared to model-matched vocalizations, indicating that the animal is able to perceptually discriminate these two classes of sounds. Our data provided only weak evidence for the existence of brain areas specifically contrasting model-matched and original vocalizations. This suggests that ferret brains do not represent high-order acoustic features as strongly as humans do, even for ecologically relevant stimuli. Further studies should inspect the neural pathways which underlie such processing in ferret brains.

3.2 Introduction

Sensory systems are adapted to extract and represent precise and subtle information from natural stimuli despite their complexity. However, how auditory cortex encodes this richness of acoustic features into spatially organized patterns of activity remains poorly understood.

The most frequent and ethologically relevant natural sounds for humans are speech and music, of which our extensive use is commonly referred to as an illustration of the uniqueness of the human species. Yet, there are surprisingly few perceptual and neural signatures of speech or music processing that are known to be specific to humans (Kriengwatana et al., 2015). Perceptually, it was shown that phoneme category discrimination, a supposedly classic signature of speech processing, could be performed by trained chinchillas (Kuhl; and Miller, 1975). Moreover, a trained chimpanzee was shown to be able to discriminate human speech to a certain extent, and its capabilities seem to rely on acoustic cues similar to the ones humans use (Heimbauer et al., 2011; Fitch, 2011). Neurally, ferret primary auditory cortex responses are sufficiently rich to encode and discriminate phoneme classes (Mesgarani et al., 2008). Thus, human speech processing could rely upon general auditory mechanisms that predated the evolution of spoken language and could be shared across a large group of mammals.

The neural mechanisms for speech and music perception in humans have begun to be unraveled. Distinct regions in secondary areas of the auditory cortex have been shown to be selective to speech and music over other natural sounds (Norman-Haignere et al., 2015). Furthermore, Norman-Haignere and McDermott (2018) showed that this specialization might rely on high-order acoustic features, beyond simple frequency or spectro-temporal modulation tuning (for example, the presence of phonemic or melodic structure). The question of whether either property is unique to humans

remains unknown (questions (1) and (2)).

Beyond the focus of speech or music perception and more generally, natural sounds processing is based on a hierarchy of acoustic features (Theunissen and Elie, 2014). Overall, the different stages or pathways in auditory cortex that subserve such different levels of processing have proved difficult to decipher in both human and non-human animals. Indeed, if the large-scale tonotopic organization of primary areas is well established, how spectral and temporal modulations are encoded within cortex is not well understood, despite their crucial behavioral importance. Several studies have suggested that an orthogonal map for temporal modulation encoding in the auditory cortex of cats (Langner et al., 2009), macaques (Baumann et al., 2015), and humans (Brewer and Barton, 2016). However, this view, in spite of its elegance, is undermined by a high inter-study and inter-individual variance and is still a matter of debate (Leaver and Rauschecker, 2016). On the other hand, the encoding of spectral modulations has remained elusive (see Read et al. (2001) however). Overall, whether spectrotemporal modulations are topographically encoded across auditory cortex is yet to be characterized (question (3)).

Furthermore, the question of how higher-order features are represented in the auditory cortex has received little attention, especially in animal research. It is now well established that sensory cortices have been shaped by the statistics of natural stimuli (Theunissen and Elie, 2014). For example, the spectrotemporal receptive field of single neurons differs depending on whether it has been computed on artificial vs. natural sounds (Theunissen et al., 2000). Investigating the processing pathways within auditory cortex using natural sounds might thus be crucial. However, acoustic features in natural sounds can be highly correlated, which limits our possibility to use them to test models of responses. Indeed, a model feature might explain neural responses while not actually driving them if it is correlated with a second, hidden variable that is itself causally linked to neural activity. For this reason, most studies have used artificial sounds, where correlations between acoustic features can be controlled. To tackle this problem with natural sounds, a new computational method has recently been developed (Norman-Haignere and McDermott, 2018). This approach consists in designing, for each natural sound, an associated 'model-matched' stimulus that has the same average distribution of spectrotemporal modulation statistics, and thus should yield the same time-averaged response if spectrotemporal tuning underlies the response. Doing so, Norman-Haignere and McDermott (2018) revealed that human primary auditory cortex responses could be well explained

by a simple spectrotemporal model (Chi et al., 2005), in sharp contrast with secondary areas that were only poorly predicted (and thus sensitive to high-order acoustic features). How much of a simple spectrotemporal model can explain responses across the auditory cortex of animals remains unclear (question (4)). Furthermore, whether high-order features are processed topographically in auditory cortex is not known.

Finally, the sensitivity of brain responses to high-order features in humans was particularly strong for speech and music sounds. It remains unclear whether animals rely on such high-order features to perceptually discriminate ethologically relevant sounds, such as conspecific vocalizations; and if so, where in the brain those are represented (question (5)).

These questions have been, so far, hard to address. In particular, accessing human auditory cortex has essentially been possible with fMRI, while non-human animal studies have focused either on electrophysiological recordings in small animals (e.g., Read et al. (2001)), or fMRI in macaques (e.g., Petkov et al. (2008)). Electrophysiology brings fundamental information to the question of sounds processing; however, it lacks a global view of the encoding throughout the cortex, and comparison with fMRI studies can be more difficult. If fMRI studies in macaques have provided important advances for our understanding of vocalization encoding (Petkov et al., 2008; Rauschecker and Scott, 2009), the exact acoustic features on which these processing streams are based is still to be characterized, and the field would greatly benefit from a larger variety of species to draw comparisons with. However, in smaller animals, the use of fMRI has been limited by its low spatial resolution. No study, to our knowledge, has managed to examine the large-scale encoding of natural sounds in small animals, and perform cross-species comparisons.

The main goal of this study was to provide new clues for the aforementioned questions, which are reformulated below:

(1) Is selectivity for speech and music compared with other natural sounds specific to humans?

(2) Is the selective response for natural speech compared with the synthetic controls specific to humans?

(3) How are spectrotemporal modulations represented at the surface of the ferret auditory cortex?

(4) How much can a spectrotemporal model explain neural responses across cortical areas in non-human animals?

(5) Do non-human animals also rely on high-order features for vocalization processing?

We used functional UltraSound, a newly developed technique based on blood flow imaging (Macé et al., 2011; Bimbard et al., 2018), to investigate the encoding of speech, music and environmental sounds, as well as conspecific vocalizations in the ferret auditory cortex. In order to decipher the specific contribution of different acoustic features to the encoding of natural sounds, we used the computational approach developed by Norman-Haignere and McDermott (2018), that contrasts the brain responses to original and model-matched stimuli that match part or all of a set of acoustic features. We then used voxel decomposition in order to identify a low number of reliable and interpretable 'components', shared across animals, whose weighted combination explained voxel responses. In a first study, we presented the same set of sounds as in Norman-Haignere and McDermott (2018) to awake, passively listening ferrets, and recorded the activity over their auditory cortex using fUS imaging. Doing so, we were able to show that ferret brain actually showed a complex set of components, among which we could surprisingly find one selective for speech. Overall, human brain responses were largely predictable by and similar to ferret brain responses. However, our study revealed that only in humans were the responses to speech and music dependent on high-order features. In a second study, we explored the organization of the encoding of ferret vocalizations. We showed that ferrets were able to perceptually discriminate most model-matched and natural stimuli, and especially ferret vocalizations. Interestingly, however, the activity of auditory cortex showed little, if any, difference of response between artificial and natural vocalizations. The magnitude of the differences we observed, in addition to the fact that strong movement-related activity was observed, were far from the magnitudes observed in humans for speech and music.

Thus our study suggests that the auditory cortex of humans but not ferrets strongly relies upon high-order features for speech and music processing. Furthermore, ferret brains were showing little specificity for high-order acoustic features even for ferret vocalizations, despite a perceptual discrimination capacity.

3.3 Material and methods

3.3.1 fUS imaging

This section was developed in chapter Mapping the auditory hierarchy. Brains were scanned in the coronal plane, with a spacing of $\sim$400 µm between slices. Craniotomies could be performed several times on the same side and animal, when tissue and bone regrowth were shadowing brain areas of

interest. Experiments were performed in two ferrets (A. and T.), across three hemispheres in both studies (A_{left}, A_{right} and T_{right} in Study 1, A_{left}, T_{left} and T_{right} in Study 2).

3.3.2 Evaluating the tonotopic organization using pure tones

Prior to all experiments, the tonotopic organization of the auditory cortex was assessed as in chapter Mapping the auditory hierarchy (Bimbard et al., 2018). In short, the responses to 2-s long pure tones of 5 different frequencies were recorded in coronal slices, spaced by 400µm, that spanned the whole craniotomy. We then used these landmarks to establish the delimitations between primary and secondary areas in all hemispheres, as well as to compare them to those obtained with natural sounds.

3.3.3 Protocol for sensory response acquisition

Auditory responses were evoked by playing sounds through calibrated earphones (BRAND, 65 dB) while recording hemodynamic responses via fUS imaging. Sounds were presented in random order, and each sound was presented 4 times. The protocol for sound presentation was as follows: 7 s of silence (baseline), then 10 s of sound followed by 3 s of silence (return to baseline). Trials were following each other with only a little random jitter in time of about 1 to 3 s, and fUS acquisitions were synchronized with the beginning of each trial.

Study 1: Full natural sounds experiment

In this first study, the full sound list comprised 36 sounds ranging from speech and music, to environmental sounds (the same as in Norman-Haignere and McDermott (2018)), with the addition of 4 ferret-related sounds: fights calls, infants calls, fear vocalization (shocks), and play calls (dooking). Thus, we used 40 different original sounds, and their 4 model-matched counterparts (200 sounds in total). Contrary to Norman-Haignere and McDermott (2018), sounds were presented for 10s and were not chopped. This was made possible by the fact that fUS imaging produces no noise while recording, contrary to fMRI. Each slice was acquired on a single day.

Study 2: Ferret vocalizations experiment

In this second study, we presented 60 sounds mixing speech extracts (14), music extracts (16), ferret fight calls (5), single-pup calls (17) or multiple (overlapping) pup calls (8). Because what

we were interested in was mainly the encoding of high-order features, we presented only the full spectro-temporally matched models and the original sounds, thus making 120 sounds in total. Ferret vocalizations were obtained by gathering datasets from different labs (our own, Stephen David's and Jennifer Bizley(s laboratories). Due to the lower duration of the total experiment, two sessions could be performed on a single day (thus recording activity from two brain slices).

It is important to note that ferret A. had pups before, while ferret T. has not known motherhood.

3.3.4 Video analysis

In Study 2 as well as part of Study 1, videos of the head of the animal were recorded. Global movement was obtained by taking the average over the full image (across voxels) of the absolute value of the derivative of the intensity value in each voxel. The movement amplitude for each sound was computed as the mean movement during the response period (3 to 11s after sound onset), from which we subtracted the baseline movement and which we renormalized by this baseline.

A higher-dimensional encoding of movement was obtained by performing PCA on the raw image and examining the 30 first PCs.

3.3.5 Signal processing, and main analysis

The procedures used to analyse the data in this chapter differs substantially from chapter Mapping the auditory hierarchy. Here, the richness of the stimulus set (200 stimuli in Study 1, 120 in Study 2) allowed us to perform more complex analysis, as well as more advanced denoising procedures to ensure that activity patterns are optimally accessible.

Interpreting the activity patterns in a large number of voxels, sometimes across individual subjects or animals, can be a difficult exercise. Different approaches have been proposed to reduce the dimensionality of such datasets, and aim at providing a low number of interpretable and robust *components*, defined by their response profiles as well as their anatomical organization in the brain. Our procedure combines different approaches in order to obtain such components. First, we deploy a denoising procedure, that aims at focusing on the smallest number of dimensions possible, these being selected as being the most reliable, and the most shared across animals (subsection Denoising procedure). Then, we orient the data in this reduced space so as to obtain the most interpretable components (subsection Independent Component Analysis).

All analysis were performed on a hand-designed region delimitating the auditory cortex.

Denoising procedure

Our denoising procedure combines two approaches: focusing on the components, within each recorded hemisphere, that are the most reliable across repetitions of the same stimulus set, and focusing on the components that are mostly shared across recorded hemispheres.

To focus on the most reliable components of the data, we used a method inspired from Denoising Source Separation (DSS) to clean up the data (de Cheveigne and Simon, 2008). In short, this method provides a way to decompose the brain signal in a set of components ordered in terms of decreasing susceptibility to a certain *bias*, which will be in our case the reproducibility of the responses. A threshold then allows us to remove the components that are the least reproducible.

In order to extract the components that were shared across animals, we coupled the DSS method with Multiway Canonical Correlation Analysis (MCCA) (de Cheveigné et al., 2019). In short, this method has been developed initially to the analysis of brain responses to stimuli presented only once, but on many subjects. In that case, the heterogeneity of the spatial arrangements of responses across electrodes or voxels between different subjects reduces a lot the quality of the signal when averaging coarsely across subjects. MCCA provides a way to focus on the components that are common to all subjects.

We performed these analysis using the full timecourse of the responses, since fUS allows us to record during the full duration of the sounds. The idea behind this choice was to possibly extract and differentiate components also based on their temporal structure.

In a first step, we centered the data the following way. For each sound presentation, we subtracted the baseline for each voxel, before dividing by the value over the full experiment (we thus obtain a normalized value for each voxel, that can be expressed in % of change in cerebral blood volume: %CBV). In order to account for global variations of blood perfusion across time in single slices, we then subtracted the mean %CBV over all voxels in each slice at each timepoint, and re-subtracted the baseline for each voxel. This also allowed to control for possible differences of overall brain response across days and across hemispheres.

In a second step, we excluded part of the data (~25% of the sounds). The left out sounds were equally shared across categories. Testing set was then left untouched.

In a third step, we performed the DSS *per se* on the training set. We whitened the data using SVD on a single slice basis (keeping all components) before averaging across sound repetitions (*bias filter* for DSS). We then performed SVD on the window of response to the stimulus in order to extract preferentially the components that showed maximally reliable responses.

In a fourth and last step, we concatenated the reliable components from each animals, and performed SVD on this concatenated matrix to obtain the components that were mostly shared across animals. Finally, we selected only the top K components in the full dataset to obtain a DSS denoised dataset. K is then the number of components implemented in the ICA.

Independent Component Analysis

After denoising the data, we performed Independent Component Analysis (ICA) decomposition across all stimuli, using the same method as in (Norman-Haignere et al., 2015). The timecourse of the response was also included in the ICA. Thus, we could obtain Independent Components (ICs) that were defined by the overall timecourse of their response to each sound, as well as their spatial patterns across voxels.

The optimal number of ICs (K^*) to implement was obtained using a crossvalidation method (figure S3.2-). For different values of K, an ICA model was trained on the training set, and the testing set was used to obtain the goodness of the model. Explicitely, half of the testing set (defined as the odd repetitions on all test sounds) were averaged and denoised using only the components obtained with the ICA. Then, we computed a measure of the goodness of the model prediction (R^2) across all sounds, timepoints and voxels by comparing this prediction to the other half of the testing set (defined as the even repetitions on all test sounds). This analysis was performed 4 times in order to use all the sound in training and testing sets. The average R^2 was then examined to obtain the K yielding the optimal reconstruction (then called K^*).

To evaluate with a single value the magnitude of the response of each component to each sound, we projected, for each component, the response timecourse of each sound on the mean response timecourse over all sounds to this component. This approach relies on the hypothesis that, on each component, each sound evokes a response with a timecourse that is a scaled version of a mean timecourse. This looked, by eye, true for most of the components.

In order to identify the components that were linked to movement, we simply correlated this single-value response and the averaged movement of the animals for each component.

Normalized Squared Error and NSE maps

The Normalized Squared Error (NSE) and the noise corrected NSE were computed as in Norman-Haignere and McDermott (2018). In short, the NSE take a value of 0 if the response to natural and model-matched sounds is identical, and 1 if there is no correspondence between responses to natural

and model-matched sounds. The noise-corrected version takes into account the possible difference in overall response reliability across voxels (or components).

To obtain the NSE maps in the ferret (fig. 3.1), we first denoised the data using the full DSS/MCCA procedure (see subsection Denoising procedure). By looping over all the different folds and by keeping only the first K^* components (see subsection Independent Component Analysis - the movement-related component was also projected out), we obtained a denoised pattern of response for each and every sound presentation. Finally, we computed the time averaged response of each voxel to each sound over the response window (3s to 11s after sound onset). We used these time-averaged responses to evaluate the noise-corrected NSE over different models.

Estimating the tuning of single voxels

In order to explore the topographical organization of spectrotemporal modulation encoding in the cortex, we estimated the prefered frequency, rate (temporal modulation) and scale (spectral modulation) of each voxel. To do so, we first computed the correlation coefficient of the voxel's response and the frequency content across all sounds for each frequency bin. The best frequency was selected as the one yielding the maximal correlation coefficient. Then, after regressing out the contribution of frequency to the voxel's response, we computed a similar correlation coefficient between the voxel's response and the rate and scale content across all sounds. The best rate and scale were chosen as the coordinates in the rate/scale space yielding the maximal correlation coefficient.

Predicting human components from ferret responses

In order to investigate how much of the human responses were comparable to ferret responses, we performed a linear prediction from ferret components to human components.

To do so, we selected the optimal K^* components obtained with the DSS/MCCA/ICA analysis across all ferrets. The component in the ferret that was mainly due to movement (figure S3.6) was discarded from the analysis. The human components used as prediction targets were the same as in Norman-Haignere and McDermott (2018).

The prediction from ferret components of human components was performed using a 9-fold, 3-way cross-validated regression. The data is first split into 9 folds, one being left out for testing and the rest used for training. The weights of the regression analysis were estimated from the training set, and then applied to the features of the test data. The weights were themselves chosen using 9-fold, 2-way cross-validation applied to the training data. Regression weights were obtained by

minimizing least-square error. All the sounds were used as test sounds. The final predictions used to be contrasted with the original responses regroup the sounds from the testing set of different runs, in a cross validated approach.

Finally, we defined the shared variance as the part of human responses that were explained by the ferret components (i.e., the prediction itself), and the unique variance as the residuals (i.e., the original from which we subtracted the prediction).

Estimating the encoding pattern

We wanted to estimate the timecourses with which the representation of the sounds evolved (figure S3.3). To do so, we computed the cross correlation of the activity pattern evoked by each sounds. We thus obtain a representational dissimilarity matrix (RDM) (Kriegeskorte et al., 2008). To evaluate the stability of the representation across time, we then computed a second-order RDM based on the correlation of the activity-pattern based RDMs across time.

Decoding sound identity

We wanted to explore how much information did brain responses contain about stimulus identity, and how the representation of these stimuli evolved over time.

To estimate the discriminability between sounds, we used a basic correlative decoder. To maximize decoding accuracy, we used the optimal ICA components as the decoding space, and estimated how well we could identify left out sounds amongst all the different sounds. In order to cross-validate the decoding procedure, we split each tested sound into an equal number of repetitions (2 and 2). We used the average over the 2 first repetitions to compute an activity pattern in the ICA space associated with each sound in the testing set. Combined with the activity patterns of the sounds that were part of the training set, this provided us with a bank of possible activity patterns in response to all sounds. Then, we used the average for the 2 last repetitions as a 'unknown' pattern that we wished to identify. By computing the correlation of the 'unknown' activity pattern with all the other activity patterns (for each other sound, both in the training set, and the testing set), we were able to identify the activity pattern with which it correlated the most. We thus obtained a decoding accuracy defined as the number of true identification over the total number of choice (i.e., of sounds). By looping over all the possible combinations of train and test, all the sounds were going through this decoding procedure. Further studies should use a bootstrap procedure in order to evaluate the overall significance of the decoding, and the significance of differences between

conditions. We provide this figure to illustrate the overall trend in the data.

An attempt at movement cancellation (Study 2)

In order to cancel out the artefactual contribution of movement to the fUS signal, we used a MCCA approach to uncover the components common to both the fUS signal and the movement of the animal. This step was performed before any denoising. In short, the goal of MCCA was to find a subset of components that were shared between the global movement of the animal, and the signal in the brain.

To do so, we whitened the data by performing SVD on both the raw video and the full brain Power Doppler throughout the whole session on a single slice basis. Then, we selected the first 30 PCs of each and performed SVD on the concatenation of the two sets of components ($D_{mcca} = [U_{video}, U_{brain}]$, U_{video} (resp. U_{brain}) containing the timecourses of the first principal components in the video (resp. the brain signal)), to retrieve the components that were shared between brain signal and the video data. The timecourses of these shared components are U_{mcca} (from the decomposition $D_{mcca} = U_{mcca}S_{mcca}V'_{mcca}$). In order to examine the quality of the common subspace, we projected independently both U_{video} and U_{brain} on their respective weights in V_{mcca} (V_{mcca} can be split in two, one set of weights corresponding to the video, the other to the brain signal). We then computed the NSE between the two projections and selected only the components that had a NSE below 0.8, which we interpreted as the transition from reliably shared to unreliable components. Finally, we projected out all these shared dimensions from the data.

3.3.6 Display

Views from above were obtained by computing the average of the variable of interest in each vertical column of voxels from the upper part of the manually defined cortical mask.

3.4 Results

3.4.1 Study 1: Natural sounds processing in ferrets vs. humans

The main goal of our first study was two-fold. First, we wanted to explore how much of the brain responses of awake ferrets, as assessed by functional UltraSound (fUS), could be explained by a standard bio-inspired model of acoustic processing (Chi et al., 2005) (question (4)). Doing so,

we wanted to uncover the different pathways that underlie the hierarchical processing of acoustic features within ferret auditory cortex (question (3)). Second, we wanted to draw a comparison between this organization and the human auditory cortex's, in order to reveal the features that were either shared across species, or unique to humans (questions (1) and (2)).

We measured fUS responses in the ferret brain to 40 different stimuli and their model-matched counterparts, across primary and secondary areas of the auditory cortex. Stimuli were drawn from different categories, such as human speech, music, or environmental sounds (Norman-Haignere and McDermott, 2018), with the addition of ferret calls. In order to decipher the contribution of each acoustic features, we hierarchically matched different aspects of the model: frequency content (*cochlear*), frequency and spectral modulation content (*specmod*), frequency and temporal modulation content (*tempmod*), or full frequency and spectro-temporal modulation context (*spectempmod*).

Full spectro-temporal model explains most of the fUS responses across ferret auditory cortex

In a first approach, we studied the global similarities of the responses between original and model matched sounds. To do so, we used a single-voxel approach and contrasted the responses to the original sounds and their associated model-matched in each voxel. Because single voxel responses showed a low reliability at first, we used a denoising approach to extract reliable components from the data and reduce its dimensionality (see section Normalized Squared Error and NSE maps in Material and methods, and de Cheveigne and Simon (2008)).

Figure 3.1-a shows the denoised response to original sounds of two single voxels, taken either in the high-frequency primary auditory cortex or in the low-frequency area between primary and secondary areas of one ferret (coordinates shown in 3.1-b). FUS imaging allowed us to record brain activity during the presentation of the sounds – an information that is not available with fMRI. Alongside, we show the average response of these same voxels to the original vs. the synthetic sounds matching different levels of complexity. Adding levels of complexity to the model yielded better and better comparison, up to fully well predicted responses in the case of the full spectro-temporal model, for both voxels. In order to quantify the extent with which each models could explain the responses to natural sounds across the whole auditory cortex, we compute in each voxel the noise-corrected Normalized Squared Error (ncNSE) between original and model-matched sounds on the denoised dataset. In short, the NSE takes a value of 0 when the two variables to

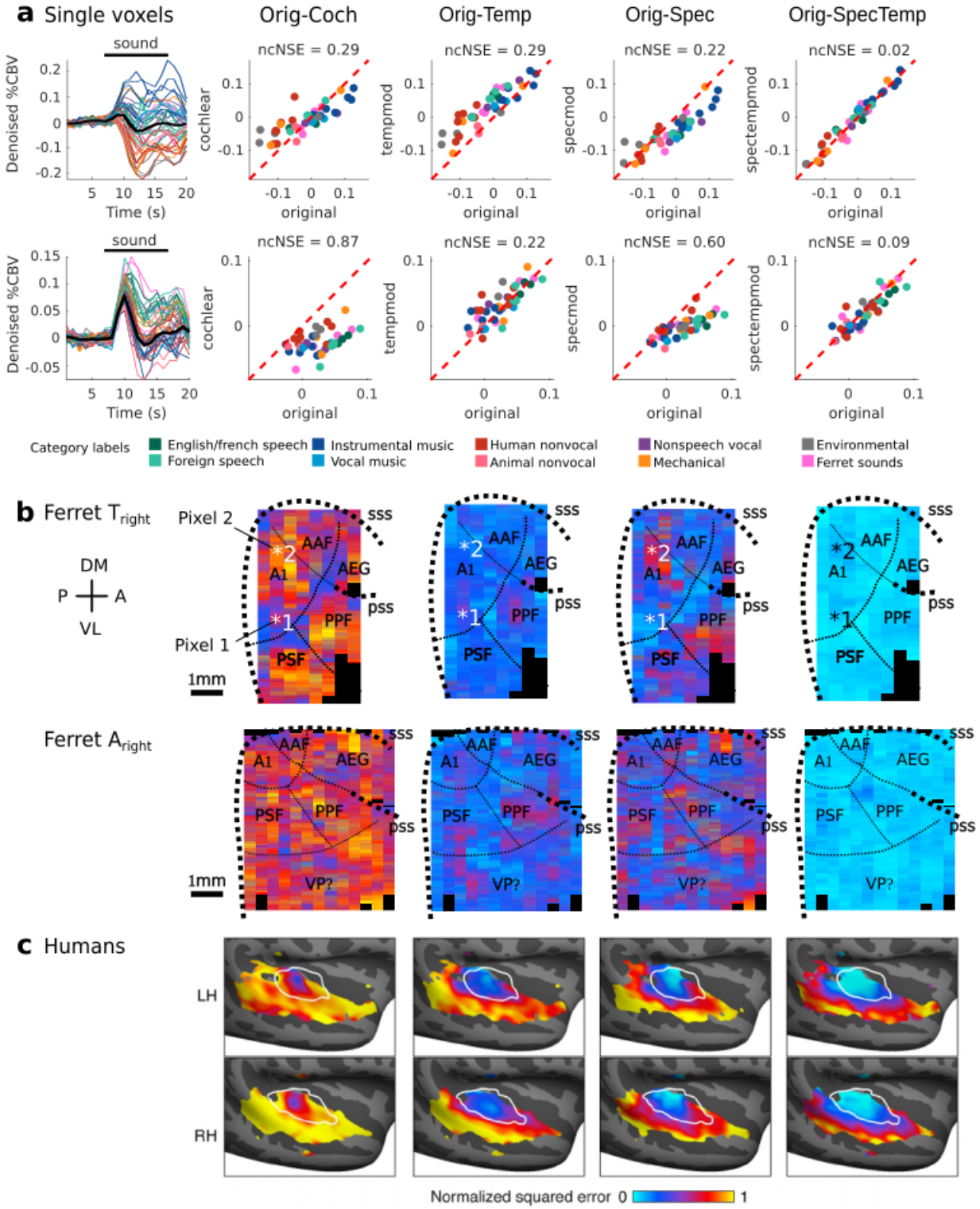

Figure 3.1: **Original and model-matched natural sounds evoked similar responses across the ferret auditory cortex (Study 1).** (**a**) Responses for two example denoised voxels (their position is shown in (**b**)). Left: Timecourse of the denoised responses (mean response over all sounds (thick black line) and to each individual original sound (thin colored line). Black bar shows sound presentation. (legend continued on next page)

Figure 3.1: The actual shape of the timecourse may be modified by the denoising procedure and might not represent exactly the actual response timecourse of the voxels. The relationship between the different sound-evoked responses, however, is preserved. Right: Responses to original vs. model-matched sounds. The response to each sound was obtained as the average response in the window 3 to 11s after sound onset. Noise corrected Normalized Squared Error (ncNSE) is shown. (**b**) ncNSE for each model (cocheal, temporal, spectral, spectrotemporal) for 2 animals. It represents a view from above of the sylvian gyri which contains auditory cortex. Dashed thick lines represented the supra-sylvian sulcus (sss) and the pseudo-sylvian sulcus (pss) that delimit auditory cortex. Boundaries for the different functional areas (dashed thin lines) are based on the tonotopic organizations. A1 and AAF are primary areas, while AEG, and PEG (PPF and PSF) are secondary areas. (**c**) ncNSE maps for a human (taken from Norman-Haignere and McDermott (2018)). LH: left hemisphere; RH: right hemisphere. White line delimits primary auditory cortex.

compare (here model-matched vs. original sounds evoked responses) are identical, and 1 if they are uncorrelated. The ncNSE takes into account the experimental noise (i.e., test-retest reliability) for correcting NSE values. In the two example voxels, the ncNSE well quantified this increase in predictability by the model (thus shown by a decrease in ncNSE). We computed ncNSE values for all voxels throughout the auditory cortex (figure 3.1-b). As in humans (figure 3.1-c), adding more features progressively decreased the NSE values across the map (thus showing better predictions of the models). Nevertheless, and contrary to humans, ferret responses could strikingly be largely predicted by the full spectrotemporal model in both primary and secondary areas (question (4)). This suggests that ferret brain responses might rely on simpler acoustic features than humans.

The topographic organization of spectrotemporal modulation tuning is poorly known (question (3)). By taking advantage of the large number of sounds in our study, we computed the tuning of each voxels for frequency, spectral modulation and temporal modulation (see Material and methods) (figure S3.1). The maps obtained for frequency tuning were largely coherent with those obtained with pure tones (data not shown). Strikingly, the encoding of both spectral and temporal modulations was highly topographic. Notably, we observed that high frequency regions were also tuned to high rates, as in Santoro et al. (2014). The relationship with the scale maps was harder to interpret. Moreover, the temporal modulation tuning map did not appear to be orthogonal to the frequency tuning map.

Ferret auditory cortex is highly organized and contains speech-selective components

This difference in model predictability between humans and ferrets could stem from the fact that ferret auditory cortex responds very differently to the set of sounds we used compared to humans.

Indeed, Norman-Haignere and McDermott (2018) have shown that most of the responses that were unexplained by the cortical model were actually responses to speech and music. These two categories are behaviorally fundamental to humans, and appear as fundamental processing streams in the human auditory cortex (Norman-Haignere et al., 2015). We thus wondered whether such streams were unique to humans, or could be observed in non-humans animals.

To uncover such pathways, one has to make sense of the overly complex, high-dimensional structure of the brain response patterns. Moreover, because fUS is a low-resolution high-sampling method, spatially overlapping yet functionally different neural populations can be hard to retrieve using standard, univariate methods.

Here, we used a computational method to infer the latent structure of the data, and reduce it to only a few reliable and meaningful components that can be linearly combined to explain the voxels' responses (figure 3.2). We obtained these components with a two-step method. First, we used a Denoising Source Separation (DSS)- and Multiway Canonical Correlation Analysis (MCCA)-based approach to denoise the data (see section Denoising procedure in Material and methods). In short, DSS allowed us to extract the components that showed most reliable responses to each sound. MCCA allowed us to extract the components that were mostly shared across ferrets. Second, we performed Independent Component Analysis to obtain a minimal set of meaningful components, as in (Norman-Haignere et al., 2015). Each component was thus defined as a certain set of responses to every sound, and a certain set of weights that represented the contribution of each component to explaining the response of each voxel. The optimal number of components (K=13) was found using cross-validation (figure S3.2). Importantly, this method is hypothesis-free: it does not rely on any assumption, e.g., on the difference between model-matched and original sounds, or on the categories of sounds. It should thus reflect natural structure in the data.

Unlike Norman-Haignere et al. (2015), we included response timecourse in this method, in order to possibly disentangle components that could show different temporal dynamics, and possibly exclude artefactual components (ICA is also often used as a denoising procedure in fMRI).

Figure 3.2 shows a selected set of components that were obtained through this method. The components obtained were highly reliable, as estimated by the low NSE values for test-retest comparisons on the left-out set of sounds (figure 3.2-a, inset). Interestingly, their timecourses was highly similar, showing an onset- and an offset peak, flanking a plateau (figure 3.2-b).

In order to determine what drove these components, we measured the correlation between the

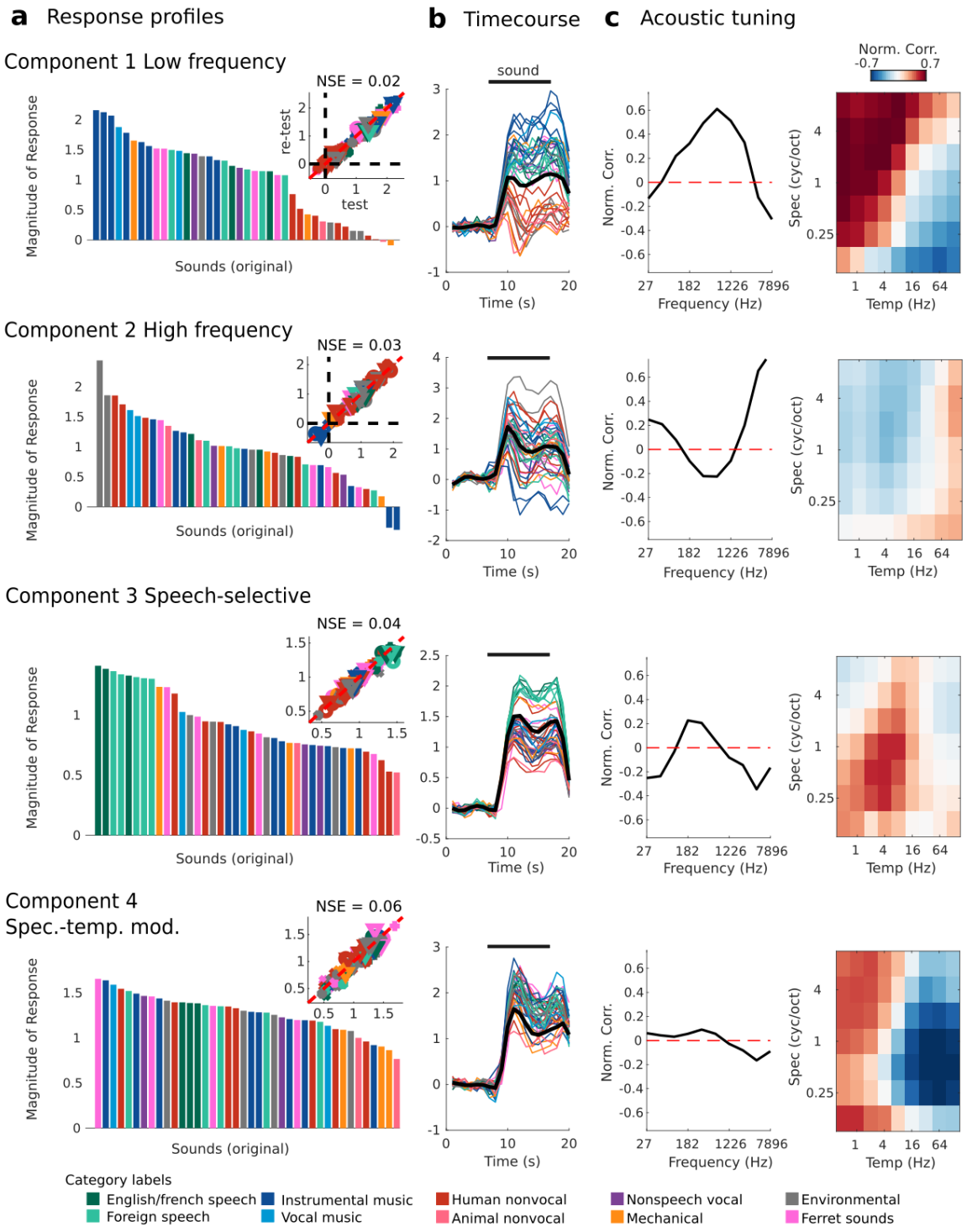

Figure 3.2: **Ferret brain exhibits organized responses based on acoustic features.** Four selected ICs are represented. (**a**) Timecourse-projected response profiles for each component, ordered by response magnitudes (arbitrary unit). Only the 40 original sounds are represented. Colors correspond to the different categories of the sounds, similarly to Norman-Haignere and McDermott (2018) (fuschia corresponds to the added ferrets sounds). Inset: test-retest response amplitude taken from the testing set of sounds, and quantified with NSE (all models and original). (legend continued on next page)

Figure 3.2: (**b**) Timecourse of the responses. Thick black line: average over all the responses (both original and model-matched). Thin colored lines: response for each individual sound (only the original sounds are represented). Black bars corresponds to the sound presentation. (**c**) Correlation of the response with basic acoustic properties of the sounds. Left: Correlation of component response profiles with energy in different frequency bands. right: Correlation of component response profiles with spectrotemporal modulation energy in the cochleograms for each sound.

response and several low-level acoustic features, such as frequency and spectro-temporal modulations (figure 3.2-c). Several components (mostly the most reliable ones) showed responses profiles highly driven by these acoustic features. Component f1 (*f* for *ferret*), for example, was highly driven by low frequency sounds, while Component f2 was mainly driven by high-frequency sounds. Their spatial weights were moreover concentrated in low- and high frequency regions, as assessed by pure tone responses (figure 3.3-a, b). Their response profile bore lots of resemblance to human Component h1 (low frequency) and h2 (high frequency) from Norman-Haignere et al. (2015). Strikingly, Component f3 showed a high selectivity for speech sounds, in a similar way as human speech Component h5 (question (1)). It was mainly tuned to the specific properties of speech, for a temporal modulation of approximately 3Hz and spectral modulation of 0.5 cycle/octave (Singh and Theunissen, 2003). This component was found in both primary and secondary areas (figure 3.3-b). Note here that we use loosely the term 'selectivity', in the sense of maximally responding to a certain category of sounds among all sounds that were presented. Finally, Component f4 was highly tuned to spectro-temporal modulations, being suppressed by high (>8Hz) rates. It was mainly located in PEG's PPF in two out of the three hemispheres.

We did not display here all the components, because not all of them made sense at first sight – and detailing them all will not serve the purpose of this study. In short, among the rest of them, some were finely tuned to other ranges of acoustic features (such as medium frequency) while others participated to explaining the overall timecourse of the responses or its differences across regions. One component was showing high responses to ferret calls, and was mainly explained by the animal's stereotyped movement (see section Ferrets behaviorally discriminate model match vs. original sounds in Results, and figure S3.6).

We then set out to explore the responses to original vs. model-matched sounds along each of these functional components (figure 3.4). Again, increasing the level of complexity of the model improved the prediction for all components. Among components, the high frequency component f2 was already well predicted by the cochlear model (NSE = 0.17), but the full model still significantly

improved the prediction (to NSE = 0.06, p-value$<10^{-3}$, paired Wilcoxon test) (figure 3.4-b). On the contrary, Component f4 was well predicted only by the full spectro-temporal model, with a large gap of accuracy between the partial and full models (from NSE >0.5 to NSE = 0.03, p-value$<10^{-3}$, paired Wilcoxon test) (figure 3.4-d). In humans, the responses to speech along the speech-selective component could not be well predicted by model-matched stimuli, suggesting that processing along this stream relies on high-order acoustic features (Norman-Haignere and McDermott, 2018). In the ferret, we observed that speech Component f3 was fully explained by the full spectro-temporal model (NSE = 0.08, similar to the test-retest NSE = 0.04) (figure 3.4-c). It thus suggests that this component relies on the specific basic acoustic features of speech.

These observations suggest that our method allowed us to uncover components of the ferret

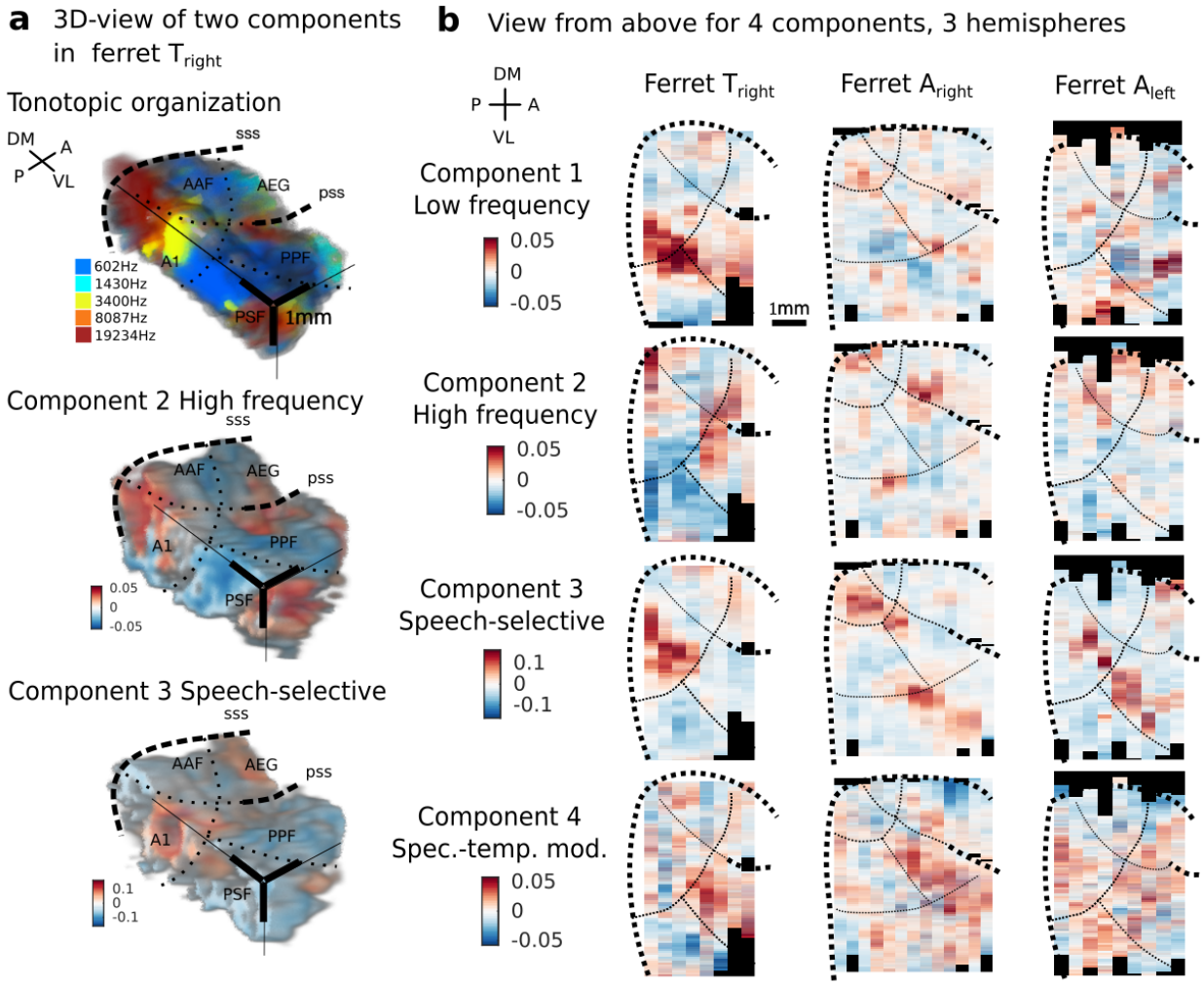

Figure 3.3: **Components are spatially coherent and clustered, and distributed throughout cortex.** (**a**) 3D view of the tonotopic organization of one hemisphere, as assessed with pure tones (left), as well as the spatial organization of two components, the high-frequency component (middle) and speech-selective component (right). Scale bar: 1 mm. (**b**). View from above for the 4 selected components and the three different hemispheres.

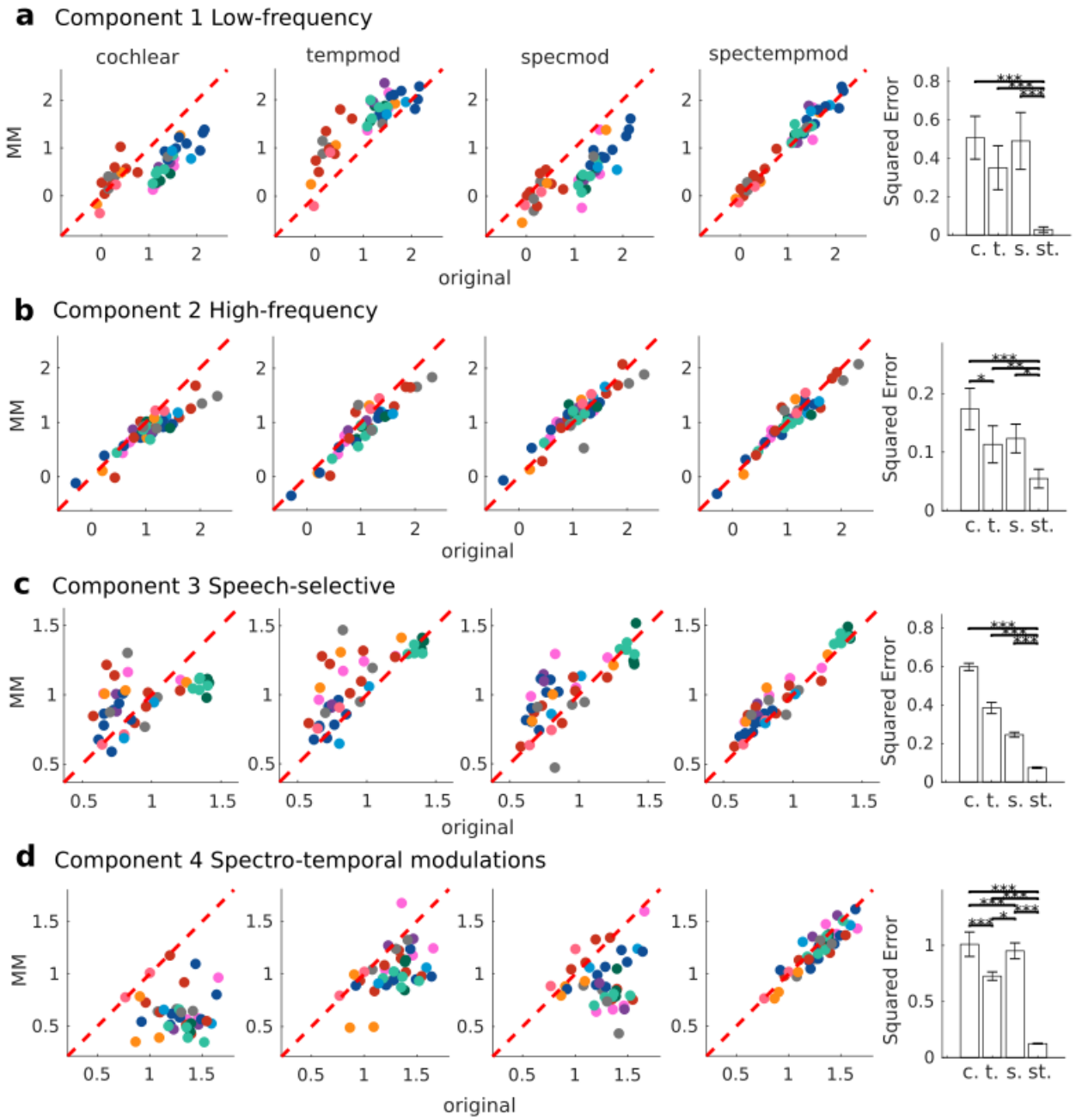

Figure 3.4: **The full model can explain a large part of the ferret brain responses.** In each plot is represented the relationship between the timecourse-projected responses to original stimuli versus each of the different model-matched stimuli. If a model perfectly predicts the brain responses, then all sounds should be aligned on the diagonal (red dashed line). Far right: quantification of the Squared Error between the responses to original and each of the model-matched sounds. We did not use NSE here so as to statistically compare the prediction error across models. Errorbars show mean ± 2 sem. Paired Wilcoxon test, ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$. (**a**) Component 1, low frequency specific. (**b**) Component 2, high frequency specific. (**c**) Component 3, speech selective. (**d**) Component 4, spectro-temporal modulation selective.

data that were highly structured, and that bore lots of similarity with human components. Indeed, ICA decomposition naturally revealed components tuned to frequency as well as specific spectrotem-

poral modulation contents (question (3)). This suggests that approaches using only natural sounds could well reveal speech or music 'specific' streams in animal auditory cortex (question (1)). However, we could not find components for which responses to speech and music sounds could not be predicted by the full spectrotemporal model. Thus, only a stimulus controlling for several low-order acoustic features (full model-matched stimuli) allowed us to clearly establish the differences in processing between ferrets and humans.

Inter-species comparison of acoustic space representation

The component analysis observed in the previous section suggests that ferret brain, as seen through our methods, does not display processing streams for high-order features such as those exhibited by speech and music sounds (question (2)). In order to further test this hypothesis, we directly tried to predict humans components from ferret data. Contrary to the previous analysis that relied on analogies in the intrinsic structure of the human and ferret brain responses, this new analysis tested more directly for a possibility to find the patterns of response of human components in the ferret data.

We thus linearly predicted each human component from the optimal set of ferret components, using cross-validation (figure 3.5). The part of the human responses that was predictable by ferret data was considered as being *shared* across species, while the residuals were considered as being the *unique* part of the response specific to humans.

Overall, ferret cortical responses recapitulated most of the human component's response profiles. Cross-validated predictions yielded high accuracy, as exemplified in figure 3.5-a, suggesting that a large part of the auditory responses in humans were, in a way, similar to ferret auditory responses. In order to identify the specificity of human responses, we computed the reconstruction error for each subgroup of sounds, either original or model-matched. Several aspects can be highlighted from this analysis. The most striking error in prediction came from human Components h5 and h6, of which specific responses to original speech and music respectively could not be predicted from ferret data, whereas predictions were accurate for model-matched speech and music sounds (p-value$<10^{-2}$, paired Wilcoxon test). This suggests that real speech and music specific high-order features could thus not be found in any linear combination of the ferret components to the same extent as in humans (question (2)).

Another way to look at this is to investigate what part of the variance is shared across species (i.e., what is actually predicted of the human responses by ferret data) vs. what part is unique to

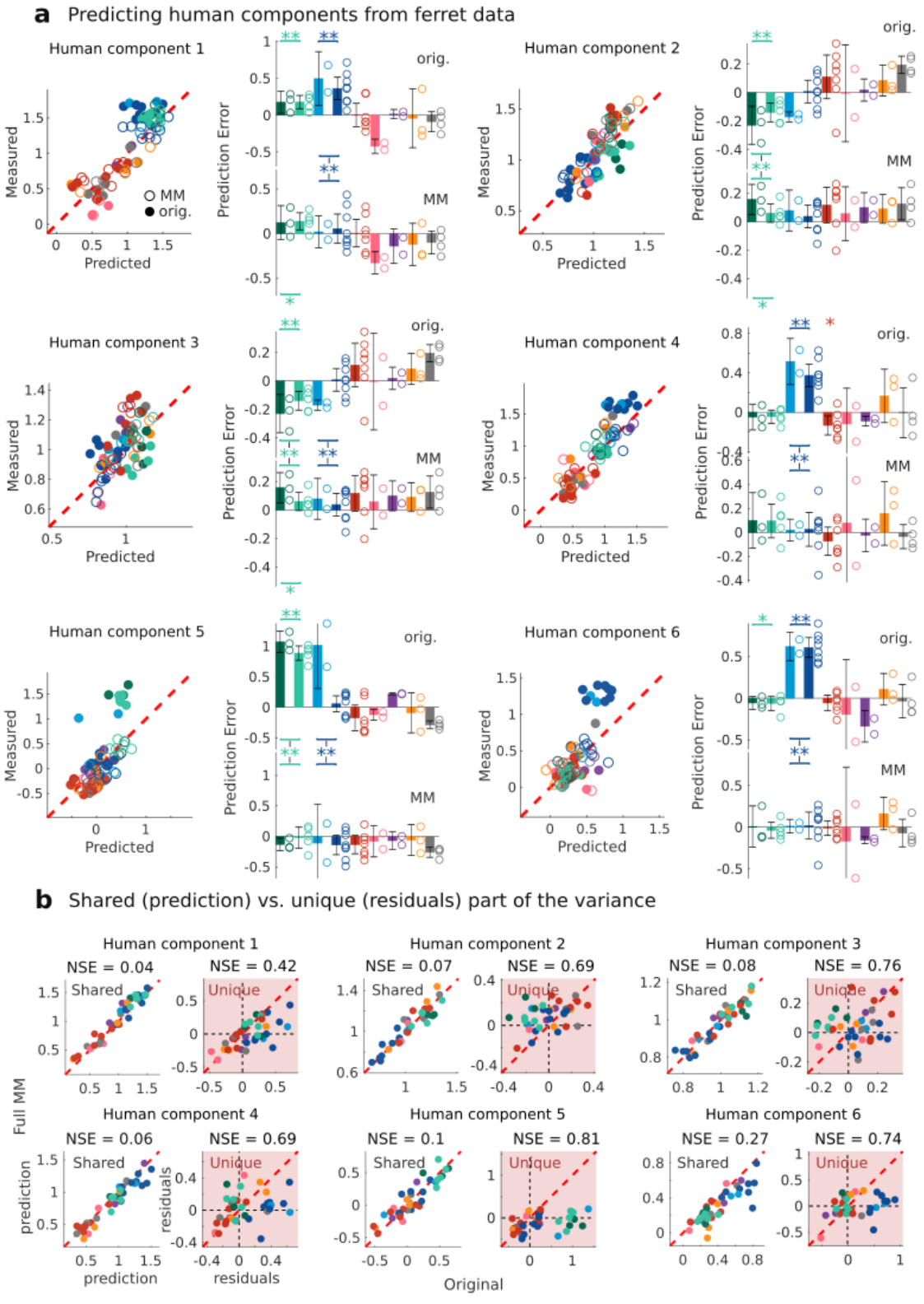

Figure 3.5: **Ferret brain responses recapitulate most of the human brain responses, except high-order specialization for speech and music.** (**a**) Predictions obtained through cross-validation on all the sounds. Six panels represent the six human components. (legend continued on next page)

Figure 3.5: Within each panel: Left: Relationship between the original magnitude of response and the ferret data-based predicted responses for each human component. Right: Quantification of the error in prediction for each category of sounds, for original (top) and full model-matched (bottom) sounds. The significance for pooled speech (all languages) and pooled music (vocal and non-vocal) sounds are shown with a colored bar above the concerned bars. The significance of the difference in prediction error between model-matched and original sounds across categories is shown by transverse lines linking the categories across the two subplots. Errorbars show mean ± 2 sem. Paired Wilcoxon test, ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$. (**b**) Magnitude of response for original vs. model-matched sounds for both the part shared across humans and ferrets (white background, defined as the prediction itself), and the part unique to human (light red background, defined as the residuals of the predictions).

humans (i.e., the residuals of the prediction) (figure 3.5-b). When separating those two aspects, we observed that within the part of the variance that was shared, original and model-matched sounds showed similar responses, even on human Components h5 and h6 (left panels for each component in figure 3.5-b). Furthermore the part of the variance unique to humans, quantified by the residuals of the predictions (right panels for each component in figure 3.5-b), contains both the error of predictions due to noise in the cross-validation procedure and true human-specific responses. Speech and music stimuli exhibited a distinct pattern of residuals on human Components 5 and 6, with large differences of response amplitudes between original and model-matched versions. This indicates that the large selective response to the high-order features of speech and music was mostly specific to humans and not shared with ferrets. We note here a small contribution of the ferret prediction to the difference between original and model-matched sounds for Component h6. This could stem from either true responses or artefactual signals, and is further discussed in the Discussion.

Several more marginal observations can also be drawn. First, an error in prediction was also visible for original music sounds along Component h1, which mainly loads in human primary auditory cortex (and not for model-matched sounds, the difference in error prediction being significant p-value$<10^{-2}$, paired Wilcoxon test). This suggests that high-order processing is actually already happening in human primary auditory cortex (this is also already visible in figure 6 of Norman-Haignere and McDermott (2018)), while being mostly absent from ferret auditory cortex. Along these same components, speech sounds were inaccurately reconstructed for both their original and model-matched. This observation was also true for Component h4, which is distributed near the borders of the primary auditory cortex. Second, errors in prediction could also be observed for model-matched and original speech sounds along Component h1, as well as along Component h1, h4, h5 and h6 for model-matched speech or music when predicting only model-matched sounds. These could reflect the differences in global tuning of the cortex of humans vs. ferrets to the spe-

cific low-level acoustic properties of speech and music, or stem from hidden speech- or music-like high-order structure of full model-matched sounds.

3.4.2 Study 2: Encoding of ferret vocalizations in ferret auditory cortex

Because speech and music are sounds that are relevant to humans, but not to ferrets, we wondered whether the lack of evidence for higher-order processing observed with our analysis reflected the actual behavioral relevance of the sound set (question (5)). We thus probed auditory cortex responses with a more diverse set of ferret vocalizations and their associated full-model-matched sounds.

Ferrets behaviorally discriminate model match vs. original sounds.

Model-matched and original stimuli sound really different to humans ears. We first explored whether ferrets behaviorally discriminate the two.

To investigate the natural responses of the animal to different sound categories, we filmed the behavior of the animal while listening to the sound sets that comprised both speech, music, and ferret vocalizations (figure 3.6). We then quantified the behavioral response of the animal simply as the normalized quantity of movement during the presentation window of the sound. Strikingly, ferrets displayed much larger movement in response to ferret vocalizations (and especially fight calls) than to speech or music (p-value $< 10^{-3}$, Wilcoxon test). Stereotyped movement was still present on those two categories.

When contrasting the behavioral response to original vs. model-matched vocalizations, we observed a much stronger movement on the original versions (for both fight and kit calls). This observation suggests that animals can behaviorally discriminate original and synthetic sounds, even when fully matched (question (5), perceptual aspect). Their natural reaction thus gives us a lower bound on their discrimination threshold. Interestingly, this was also true for music sounds, were ferrets responded more to natural than synthetic versions of the sounds. This is in contrast with the fact that only a small difference, if any, could actually be observed in the functional responses of the brain. We thus wondered whether this was also the case for ferret vocalizations.

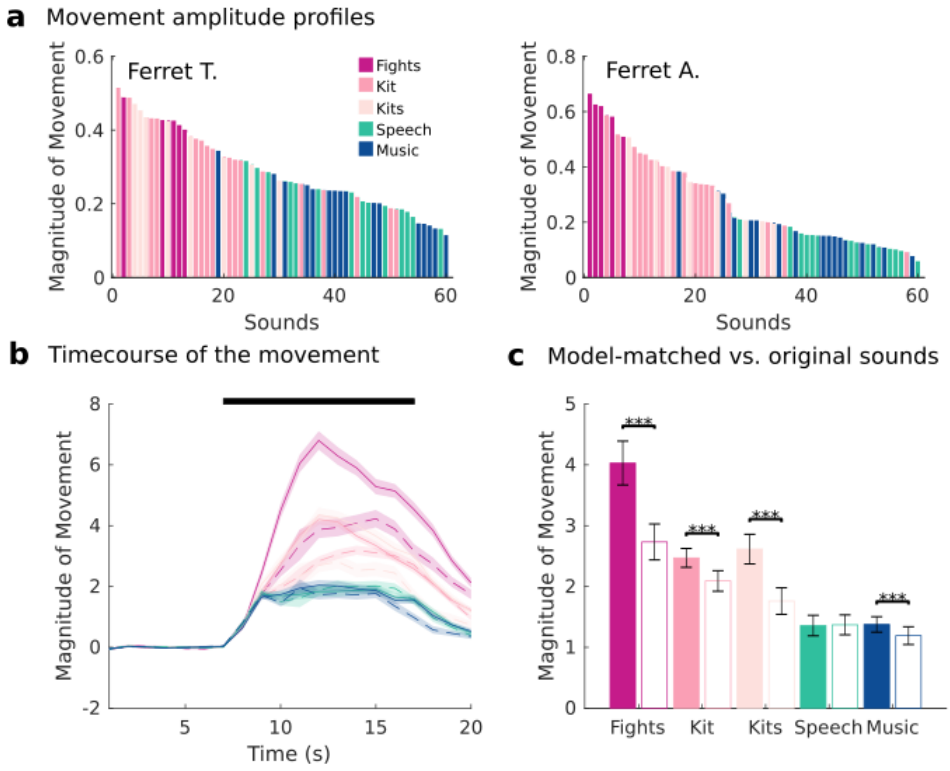

Figure 3.6: **Ferrets naturally discriminate between sound categories and original vs. model-matched sounds.** (**a**) Profile for the amplitude of movement (arbitrary units) across original sounds, ordered by response size. The profile for the two ferrets, averaged across all sessions, is shown. (**b**) Timecourse of movement (arbitrary units) averaged over sounds for each category and type (original: solid lines, full spectro-temporal model-matched: dashed lines), and averaged over animals. The movement during each sound presentation was baseline-normalized. Errorbars show mean $\pm$ sem. (**c**) Average movement amplitude for each sound category, for both original (filled bar) and full model-matched (empty bar). Significance between original and model-matched was assessed with a paired Wilcoxon test. Significance across categories is not shown here (but was strong for ferret sounds vs. speech and music: p-value$<10^{-3}$, Wilcoxon test.). Errorbars show mean ± 2 sem. ***: p-value$<10^{-3}$.

Auditory cortex hemodynamic responses showed little differences to model-matched and original conspecific vocalizations.

Animal's movement generated artefactual signals, which were stereotyped and reliable across trials. We indeed observed that one of the components observed in Study 1 displayed a response profile highly correlated with the movement profile of the animal (figure S3.6). The actual weight pattern of this component, instead of reflecting local brain regions, was mainly highlighting differences between small and large vessels, or anatomical edges (data not shown). Moreover, its timecourse was substantially different from the acoustically driven ones, and resembled the actual movement temporal profile. In Study 1, only one component picked-up the movement-related changes in Power Doppler.

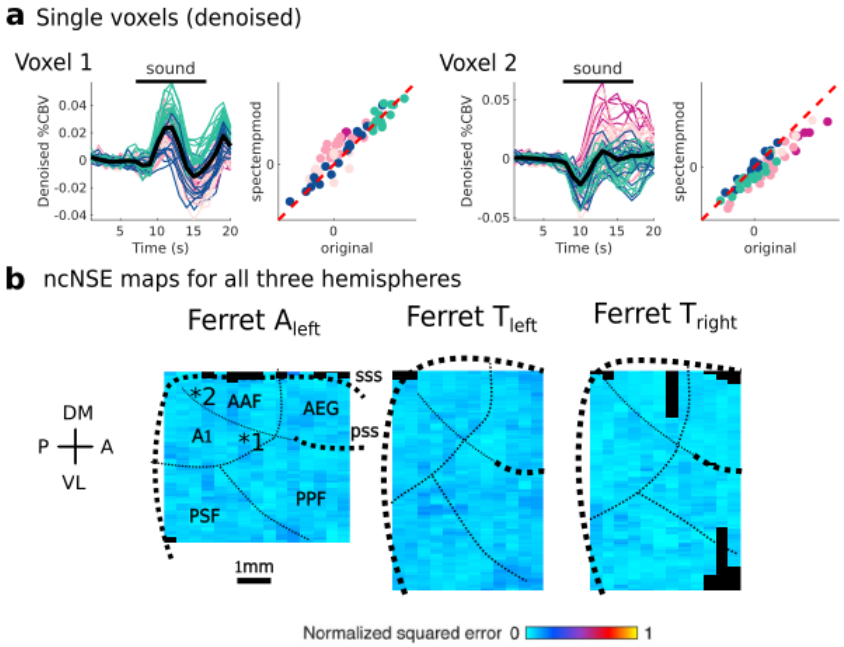

Figure 3.7: **Original and model-matched natural sounds evoked similar responses across the ferret auditory cortex (Study 2).** (**a**) Responses for two example denoised voxels (their position is shown in (**b**)). Left: Timecourse of the denoised responses (mean response over all sounds (thick black line) and to each individual original sound (thin colored line). Black bar shows sound presentation. The actual shape of the timecourse may be modified by the denoising procedure and might not represent exactly the actual response timecourse of the voxels. The relationship between the different sound-evoked responses, however, is preserved. Right: Responses to original vs. full model-matched sounds. The response to each sound was obtained as the average response in the window 3 to 11s after sound onset. (**b**) Noise corrected Normalized Squared Error (ncNSE) for the full model-matched for all three hemispheres. It represents a view from above of the sylvian gyri which contains auditory cortex. Dashed thick lines represented the supra-sylvian sulcus (sss) and the pseudo-sylvian sulcus (pss) that delimit auditory cortex. Boundaries for the different functional areas (dashed thin lines) are based on the tonotopic organizations. A1 and AAF are primary areas, while AEG, and PEG (PPF and PSF) are secondary areas.

In order to limit these effects in Study 2 (where ferret sounds were much more numerous), we used a MCCA approach to evaluate the components in the fUS data that were maximally explained by movement (see section An attempt at movement cancellation (Study 2) in Material and methods). Having identified those components, we then projected them out of the data. This method allowed us to remove substantial part of movement participation to brain responses (figure S3.8). We then used the same method as in Study 1 to denoise the data, and focus on components that were both reliable and shared across hemispheres. In that case, 10 components were needed to model the data (figure S3.7).

Single-voxel analysis revealed that model-matched and original stimuli evoked similar responses across the auditory cortex (figure 3.7). NSE values were overall very low, and no difference

could be detected between primary and secondary areas. This is in sharp contrast with what one could have expected under the hypothesis that ferret vocalization were processed in the ferret brain the same way that speech is processed in the human brain, i.e., in a specific subregion of auditory cortex which would be sensitive to high-order features found in vocalizations.

In order to uncover the latent structure in the data and possibly reveal latent vocalization-selective components, we then used the same component approach as in Study 1. Figure 3.8 shows the different components that could be obtained from this new dataset, amongst the 10 that were retained through cross-validation. Again, some components were clearly driven by basic acoustic properties (3.8-c). For example, Component fv1 and fv2 (*fv* as *ferret vocalization experiment*) were mainly driven by low- and high-frequency sounds, and their weights resembled the ones obtained on the low- and high-frequency components in Study 1 (figure 3.9). We note here that the weight patterns across animals was even clearer and more consistent in Study 2 compared to Study 1, despite the fact that less (and less diverse) sounds were presented.

When investigating the difference of response magnitude between model-matched and original sounds along these components, we could not identify any component revealing a difference for vocalizations as large as the one observed in humans for speech or music. Overall, the difference, if any, was small, and uncertain because of the persistance of movement artefacts on two of the components (e.g., Component fv4 in 3.8 - correlation quantified in figure S3.8). Thus, it seems that ferret vocalizations may not be processed selectively within a pathway as seen in the speech pathway in humans. Further investigations are needed to explore this question.

3.5 Discussion

In this study, we showed that fUS imaging could be used to investigate the large-scale organization of ferret auditory cortex responses to natural sounds. By extracting meaningful and reliable patterns across three hemispheres, we were able to identify several strong components that were driven by basic features, such as frequency gradient, or spectrotemporal modulations (question (3)). Strikingly, a simple spectrotemporal model could explain most if not all of the ferret responses, across both primary and secondary auditory cortices, and within both ventral (PEG) and dorsal (AEG) pathways (question (4)). This was in contrast to human auditory cortex organization, where this same model failed to explain auditory responses in non-primary areas. In particular, responses to speech and music relied on high-order features in humans, and such processing pathways could not be found in

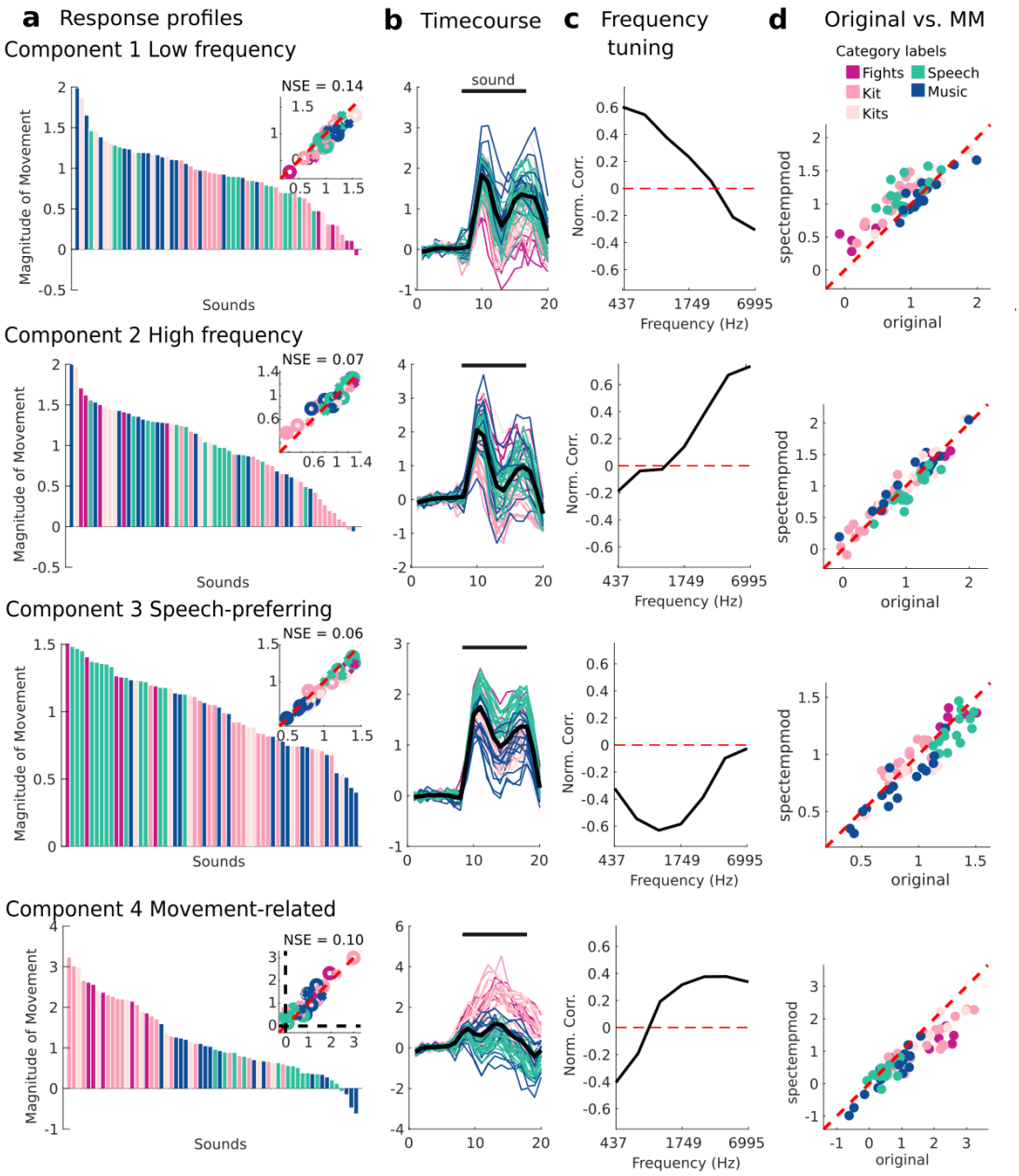

Figure 3.8: **Ferret vocalization and speech and music sounds revealed latent organization based on acoustic properties.** (**a**) Timecourse-projected response profiles for each component, ordered by response magnitudes (arbitrary unit). Only the 60 original sounds are represented. Colors correspond to the different categories of the sounds. Inset: test-retest response amplitude taken from the testing set of sounds, and quantified with NSE (all models and original). (**b**) Timecourse of the responses. Thick black line: average over all the responses (both original and model-matched). Thin colored lines: response for each individual sound (only the original sounds are represented). Black bars corresponds to the sound presentation. (**c**) Correlation of component response profiles with energy in different frequency bands. (**d**) Relationship between the timecourse-projected responses to original stimuli versus each of the different model-matched stimuli. If a model perfectly predicts the brain responses, then all sounds should be aligned on the diagonal (red dashed line).

ferret data (question (2)), despite a strong speech-selective component (question (1)). Furthermore,

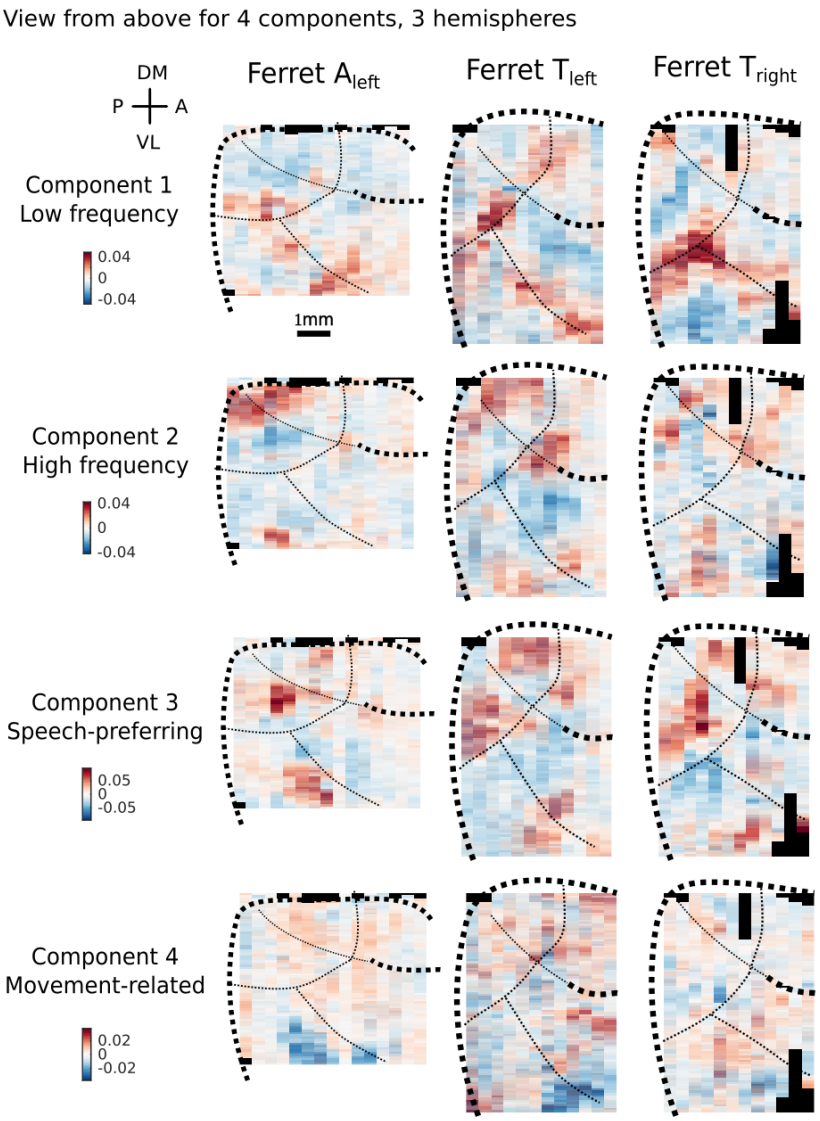

Figure 3.9: **Components are spatially coherent and clustered, and distributed throughout cortex.** View from above for the four selected components and the three different hemispheres.

we used ferret vocalizations to explore whether such dependence of the brain response on high-order features could be found with ethologically relevant sounds (question (5)). Again, evidence for such a vocalization specific pathways were scarce, despite ferrets being behaviorally capable to discriminate original and model-matched sounds. Our study suggests that ferrets do not display processing streams for high-order acoustic features as strongly as humans do. It opens up several interesting perspectives.

Methodological advancement

Functional UltraSound imaging is increasingly being used for tackling fundamental neuroscience questions, and recent papers are starting to investigate brain responses to finer behaviors or more

diverse stimuli (Macé et al., 2018; Dizeux et al., 2019). Along this line, our study was the first to show that fUS can give access to the complex encoding of natural stimuli throughout primary and non-primary sensory cortices. To do so, we adapted recent computational tools, such as DSS/MCCA, and ICA to uncover patterns of activity that can be hard to observe using single-voxel analysis (Norman-Haignere et al., 2015).

From the point of view of modelling, Norman-Haignere and McDermott (2018) showed that the cortical model from Chi et al. (2005) was actually a good predictor of fMRI responses in the human primary auditory cortex. However fMRI voxels contain a far larger number of neurons than fUS voxels, and average hemodynamic responses across larger portions of blood vasculature. With fUS, single vessels are actually observable, and the functional resolution of the technique reaches really low spatial scales within few presentations (see chapter Mapping the auditory hierarchy, or Bimbard et al. (2018)). Thus, this first observation confirms that the model-matching approach can provide valuable information on neural encoding, even at finer scales than fMRI. Importantly, the use of Independent Component Analysis was crucial to reveal the underlying organization of the responses, making use of the large number of sounds used in our study.

Interestingly, fUS can also provide access to changes in blood flow throughout the presentation of the sound. Despite its slow timecourse, a change in blood flow can contain information about fast modulations of incoming stimuli, as we and Norman-Haignere and McDermott (2018), amongst others, have demonstrated. Thus, the evolution of blood flow throughout the sound presentation could provide useful information about the changes in statistics of the sound. Recent studies have tried to reconstruct the spectro-temporal modulation content of natural sounds using fMRI (Santoro et al., 2017). This study was based on short sound durations, in order to separate sound presentation from the fMRI acquisition noise. It thus provided no access to the timecourse of the fMRI responses across time. Moreover, it was suggested that early hemodynamic response contains less information than late responses (Berwick et al., 2008). In our data, a stable pattern of activity was obtained 2 to 3s after sound onset (figure S3.3). Thus, the first two seconds during which CBV starts to increase were not yet representing sound identity in a stable manner. However, during the full sound presentation, %CBV then contained rich information about sound identity. Further investigations along this line could provide important advances in potential blood-based decoding and sound reconstruction.

Behavioral discrimination and animal's movement

The observation of the animal's movement in response to different sounds raises two important remarks. First, it brings strong evidence to the fact that vocalization processing in non-human animals can rely on high-order features. Second, a simple quantification of movement allowed us to estimate a lower bound on the capacity of the animals to behaviorally discriminate sounds, and especially original vs. synthetic versions. We insist here on the fact that the animals were head-fixed and not trained in any manner. Notably, such important behavioral responses to high-pitched sounds have already been reported in domestic ferrets (Boyce et al., 2001). This resonates with a recent study that showed that observing the behavior of 'passively listening' animals can actually reveal their latent knowledge about the task (Kuchibhotla et al., 2019); and in their case even higher-performance than during active behavior. While this sounds obvious once said, the natural behavior of the animals is rarely recorded during passive sessions. Such information could be hard to obtain under training, which is often more about limiting the animal's impulsiveness. Third, recent studies have demonstrated that large-scale activity linked to the animals' subtle (or large!) movement could be present all over the cortex (Stringer et al., 2019c; Musall et al., 2018). Especially, oro-facial movement elicits substantial responses in sensory cortices. Thus, a careful monitoring of any movement during natural sound presentation might be crucial to actually understand the encoding of behaviorally relevant sounds.

The global artefacts (i.e., changes in Power Doppler) that we observed in relation with movement could come from several sources. First, since the probe was not attached to the animal's head, any small movement of the latter could generate Doppler shifts and be interpreted as signals. Such artefacts were often observed in our data, and movement-related Power Doppler changes could be observed in regions of the image that were outside of the brain. Other fUS studies have shown that recording fUS signals in freely moving animals was possible (Sieu et al., 2015; Urban et al., 2015). However, the level of precision of the encoding patterns of interest were much scarcer, and were thus less sensitive to noise. Also, the fact that the probe was coherently moving with the animal's head could have limited the differential motion and thus possible artefacts. Finally, the use of optimized insonification sequences could help reduce such effects (Tiran et al., 2017). Second, a global excitation in response to sounds could provoke a global blood rush in the brain. As mentioned above, animal's movement could also globally modify neuronal activity and thus generate local changes in blood volume. Third, top-down signals could specifically modulate auditory cortex activity, as behavioral engagement has been shown to provoke substantial changes in neural activity

across the auditory cortex (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al., 2019). These two last effects are much harder to disentangle, but offer interesting prospects for further studies. For example, the topographical organization of top-down signals is poorly known, despite its crucial function in adaptive behavior and perceptual decision-making.

Finally, we note here that the movement related component (figure 3.8) displays some structure, especially a negative weight region in the very lateral part of the craniotomy. While this could represent a real pattern in brain activity, we note that bone destruction and tissue regrowth on the most lateral side of the craniotomy sometimes created physical links with the jaw muscles. Thus, the most lateral part of the craniotomy was the most sensitive to physical movements of the animal.

Through both the analysis of Study 1 and Study 2, one could detect some components or regions of which responses were discriminating model-matched and original stimuli (e.g., in figure 3.8). Furthermore, some parts of the shared variance showed a difference for music sounds (figure 3.5-b), which we further showed to be perceptually disciminable by our ferrets (figure 3.6). First, these differences were really small compared to the ones observed in humans. Second, the problem of movement was not fully addressed in Study 1 (the movement-related component might not have fully caught movement-related signals) nor in Study 2 (where even after denoising we could still observe clear movement-related components). This casts a doubt on the true functional reality of such discriminative signals in the brain. We thus interpreted our results as showing only little proof for the existence of high-order acoustics selectivity in the ferret responses, and even less so for a specific pathway. Further analysis and studies should focus on movement cancellation and on determining to what extent these signals still hold, before confirming of disconfirming our suggestion.

Where in the brain is high-order acoustics processing happening?

In the ferret, the responses to complex sounds have been rarely explored (Nelken et al., 2008), and the response to ferret vocalizations only approached using sparse, single-cell recordings in primary auditory cortex (Schnupp et al., 2006). Thus, only few points of comparison can be drawn with the literature. In our hands, ferrets seemed to rely on high-order acoustic features to interpret ferret vocalization, yet their auditory cortex did not show strong signatures of such processing. The small differences that we observed between model-matched and original vocalizations were distributed across the whole portion of cortex that we imaged, and were not strongly identifiable along any of our components. Our conclusions are fully dependent on several technical aspects of our experiments: the nature of the signal, the field of view over auditory cortex as well as the species used here.

First, high-order acoustics processing streams could well be hidden (or made less visible) by the way we monitor brain activity. Indeed, each voxel might contain a very heterogeneous set of neurons, with some of them showing differences in their responses to model-matched vs original stimuli. However, in humans, the differences observed were really large for speech and music. Even if single neurons could discriminate the two, our conclusion on cross-species comparisons would still be unchanged since similar approaches yielded very different effect sizes in the two different species. Moreover, after large-scale recordings with fUS imaging, a more careful inspection of single neurons could be an interesting path to follow.

Second, our imaging window focused on primary and secondary areas of the auditory cortex. One could thus reasonably argue that the reason why we are not seeing strong evidence for high-order acoustics processing is because we did not look in the right place. In humans, the regions specific to speech and music observed in Norman-Haignere et al. (2015) were present in non-primary areas of the auditory cortex (suprerior temporal gyrus, planum temporale and planum polare) that were adjacent to primary auditory cortex. The field of view that we managed to obtain was beyond what was done in our previous study (Bimbard et al., 2018), to reach areas that have been only recently functionally described in the ferret auditory cortex, like ventral posterior auditory field (Elgueda et al., 2019). We identified these areas using the pure-tone and/or natural-sound proofed tonotopic organization, and by direclty observing the brain folding ventro-medially on fUS images. These areas delimit the most lateral and ventral part of the ferret auditory cortex.

Some vocalization areas have been described in more frontal areas in macaques (Romanski and Goldman-Rakic, 2002). Investigating the existence of such areas would be directly in the line of our study. Importantly, the use of model-matched stimuli allows for very efficient controls, with only a limited set of sounds, to understand the acoustic properties to which brain areas or components are actually specific.

Third, the lack of a specialized pathway for vocalizations' high-order acoustic features could be due to a lack of exposure to their own conspecific vocalizations. However, our animals were housed in pairs or more, were allowed to play throughout the week, and vocalized when playing. It could also be argued that ferrets are actually not a *vocal* species. On the contrary, we argue that they display a large range of vocalizations (fight calls, pup calls, play calls, ...) (Boyce et al.,

2001), and that these vocalizations are after all sufficiently relevant to trigger significant and reliable movement upon hearing it (figure 3.6). Finally, most of our stimuli were kit calls. It has been shown in the mouse that motherhood actually provoked large changes in pup vocalization encoding even within primary auditory cortex (Shepard et al., 2016), and one could argue that high-order acoustic sensitivity to these sounds would be present only in mothers. In our case, one of the animal had experiences motherhood while the other had not. However, we could not observe clear differences between the two animals.

Another hypothesis would be that vocalization processing in ferrets mostly relies on much simpler acoustic features. Most processing could thus come from a global tuning of the auditory cortex to the specific low-level acoustic niche of vocalizations, as what was observed with the rates to which macaque monkeys vs humans are most sensitive (Erb et al., 2018). Specific pathways to process higher-order features would thus be less needed than human speech. As a matter of fact, model-matched vocalizations (as well as other natural sounds that resembled pup calls in terms of basic acoustics, such as baby crying), also evoked very large behavioral responses (higher than to any other sounds, excluding natural vocalizations), despite their lack of vocalization-specific high-order structure. Performing the same experiments (i.e., presenting ferret vocalizations) in humans could provide insight on the specificity of ferret auditory cortex to process their conspecific vocalizations.

Prospects: effect of exposure and training

Our results raise the question of how such high-order acoustics specific domains appear in humans. Are they already built-in? Is it based on early experience? The robustness of their location in the brain across subjects raises the question of the original substrate upon which it is built. Is it based on a proto-architecture already present in the processing streams, upon which music and speech selectivity develop? Could we generate such domains in animals' brains, which this proto-architecture is actually shared (even if differently exploited) across mammals? Where would they appear?

One could draw comparisons from visual research, that focused on the origins of category-specific domains within inferior temporal cortex (IT) of macaques (Op De Beeck et al., 2008; Srihasam et al., 2012; Livingstone et al., 2017; Arcaro et al., 2017). In particular, it was suggested that plasticity for abstract symbols is constrained by some native organization in cortex (Srihasam et al., 2012). This turned out to be also true with the development of face-sensitivity, that necessitates the animals to actually see faces during development (Arcaro et al., 2017), and that might

rely on the existence of a native shape-based retinotopic proto-map (Livingstone et al., 2017). Such questions are yet to be addressed for the development of speech and music specificity in the human brain. Our experimental framework could be particularly interesting to examine such questions in animals. Indeed, the use of the model-matched approach allows us to identify rapidly possible regions that could be specific for high-order acoustic features, like the ones observed in humans. Exposing animals to original vs. model-matched speech or music during their early development, or intensively training them to actively discriminate the two with a reward system, could be a way to examine the development of category-specific areas in different species. Earlier studies have shown that training can lead to large reorganization of basic features such as tonotopic organization (Polley et al., 2006), up to more complex patterns in response to more complex sounds such as cross-specific vocalizations (Jiang et al., 2018) already in primary auditory cortices. The effect of such training in secondary areas would be expected to be even more drastic and more category-specific, especially during behavioral performance (Atiani et al., 2014; Elgueda et al., 2019). It would thus be interesting to explore such reorganization with our new methods, that allow for a large-scale yet high-resolution mapping of the whole sensory cortex with a high sensitivity. This could be used to test the hypothesis that speech processing relies on a native organization that is specific to humans, or shared across species. In our study, we identified a specific component in the ferret auditory cortex that was selective for speech's acoustic properties, amongst other natural sounds. An appealing hypothesis would be that *real* speech selectivity (i.e, sensitivity to high-order features of speech) would emerge along such a component; another hypothesis being in another area dedicated to plastic learning.

Identifying a region (or a 'component', that can be spatially distributed) that displays differences in model-matched vs. original vocalizations could be a necessary first step to explore these questions, as it would provide one neural substrate for encoding high-order acoustic properties that are clearly extracted at the behavioral level. However, pathways for vocalizations processing could be highly specific and rely on already built-in wiring.

3.6 Summary

In this chapter, we explored the different levels of encoding of natural sounds in the ferret auditory cortex. We adapted and developed new computational methods, for both denoising and analysing fUS signals, that allowed us to uncover structured components of responses in the ferret brain. By contrasting the response properties between ferrets and humans, we were able to make progress

on deciphering the unique contribution of human auditory cortex to the processing of speech and music, which is its dependence on high-order features of speech and music. Finally, we applied these methods to study the encoding of conspecific vocalizations in the ferret brain. After showing that ferrets perceptually rely on such high-order features to interpret their conspecific calls, we explored the underlying neural mechanisms of this capacity and failed to identify brain regions or components strongly supporting such processing. Overall, this study provides perspectives on the evolution of auditory processing between humans and ferrets, and paves the way for future studies of vocalization processing.

3.7 Supplementary figures

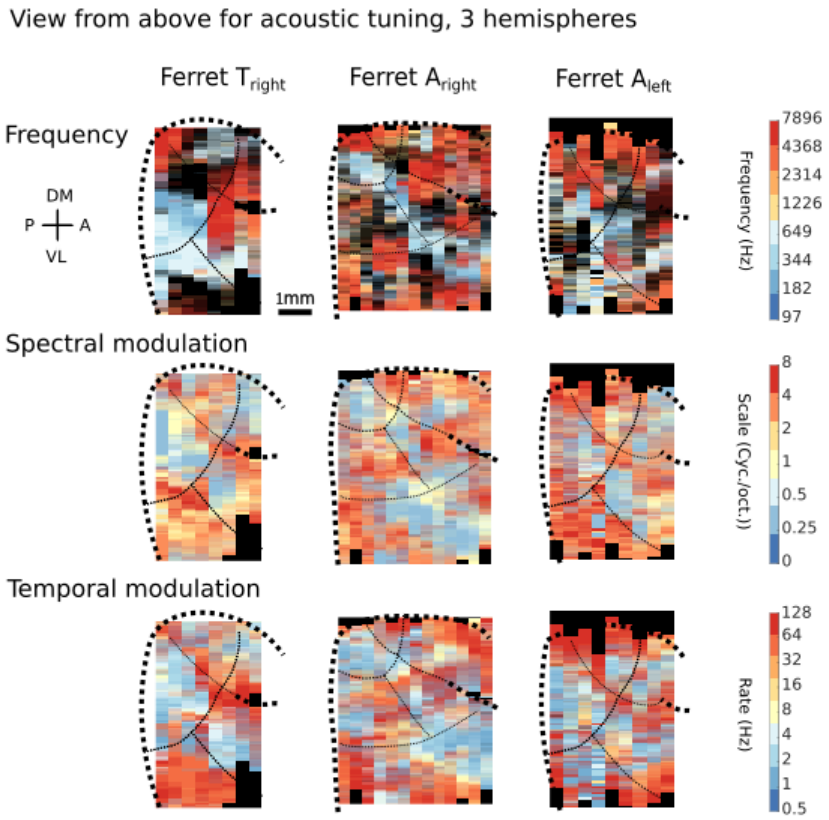

Figure S3.1: **Spectrotemporal modulations are topographically encoded at the surface of the auditory cortex.** Tuning maps for frequency, spectral modulations and temporal modulations tuning for three different hemispheres.

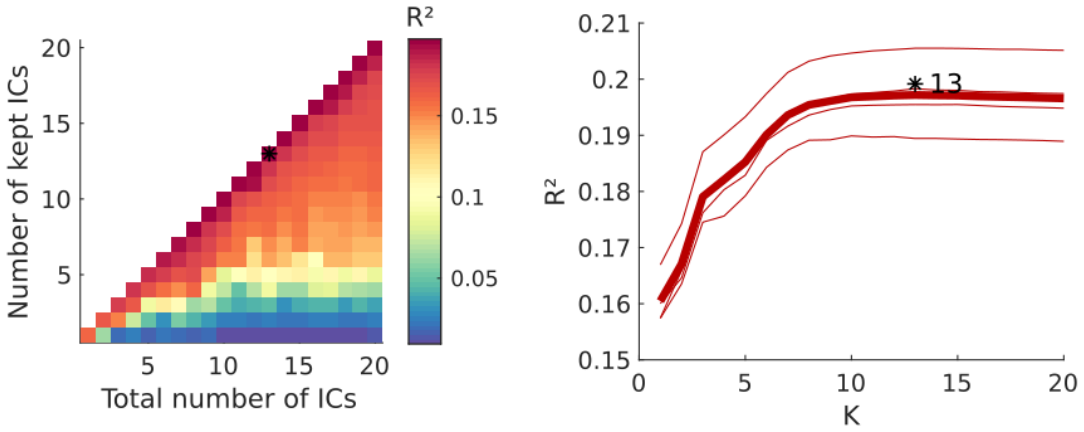

Figure S3.2: **Brain responses to model-matched and natural sounds can be summed up with 13 components (Study 1).** (**a**) R^2 landscape across two parameters, the number of Independent Components (ICs) implemented in the model, and the number of ICs actually kept for reconstruction. Maximal response is shown by a star (13 ICs implemented, 13 ICs kept). (**b**) R^2 values for different values of ICs implemented in the model (corresponds to the diagonal in (**a**). Large line corresponds to the average over all 4 difference testing sets, while smaller lines correspond to each of the 4 different test-sets.

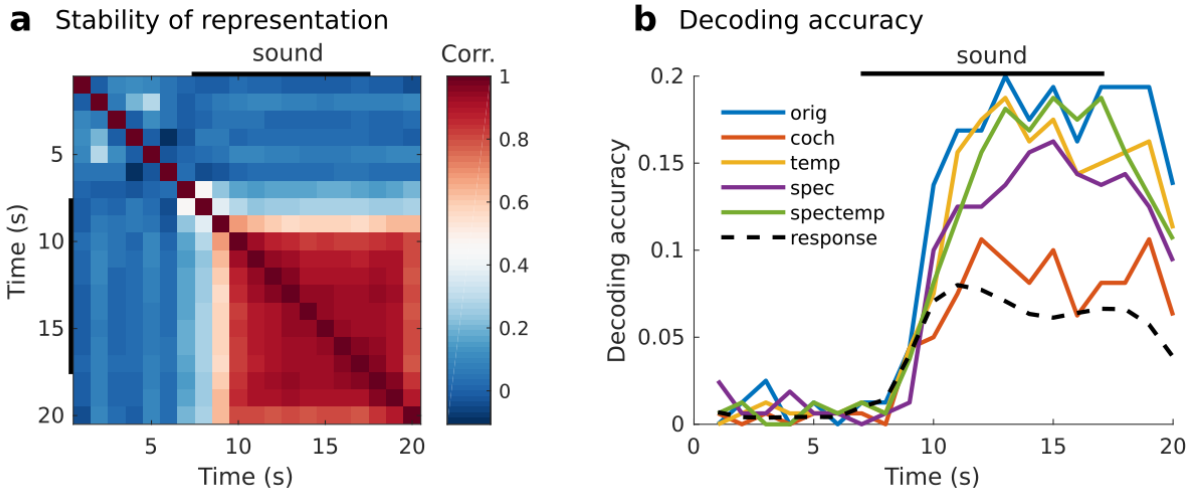

Figure S3.3: **The encoding patterns is stable throughout sound presentation.** (**a**) Correlation across time bins between the encoding patterns (obtained by computing the representational distance matrix over all sounds). Sound presentation is indicated by the thick black lines. (**b**) Decoding accuracy across time, when decoding sound identity, depending on sound complexity. Sounds were taken from the testing set (and thus did not participate to the ICA model), and had to be identified out of the 40 sounds of their model type. This procedure was repeated across all models, and original categories.

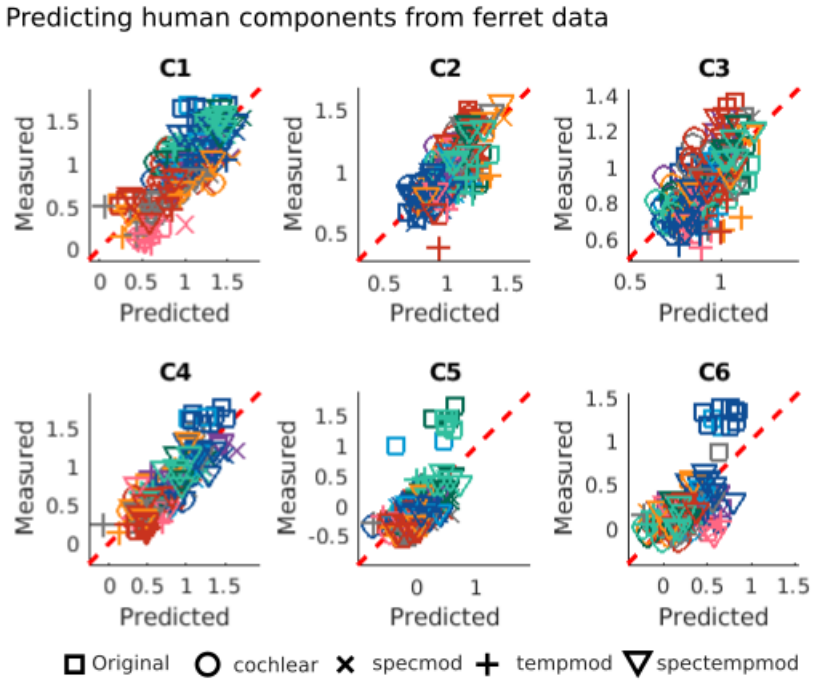

Figure S3.4: **Predicting human components from ferret data (prediction on all sounds, display all).** Predictions obtained through cross-validation on all sounds. Relationship between the original magnitude of response and the ferret data-based predicted responses for each human component.

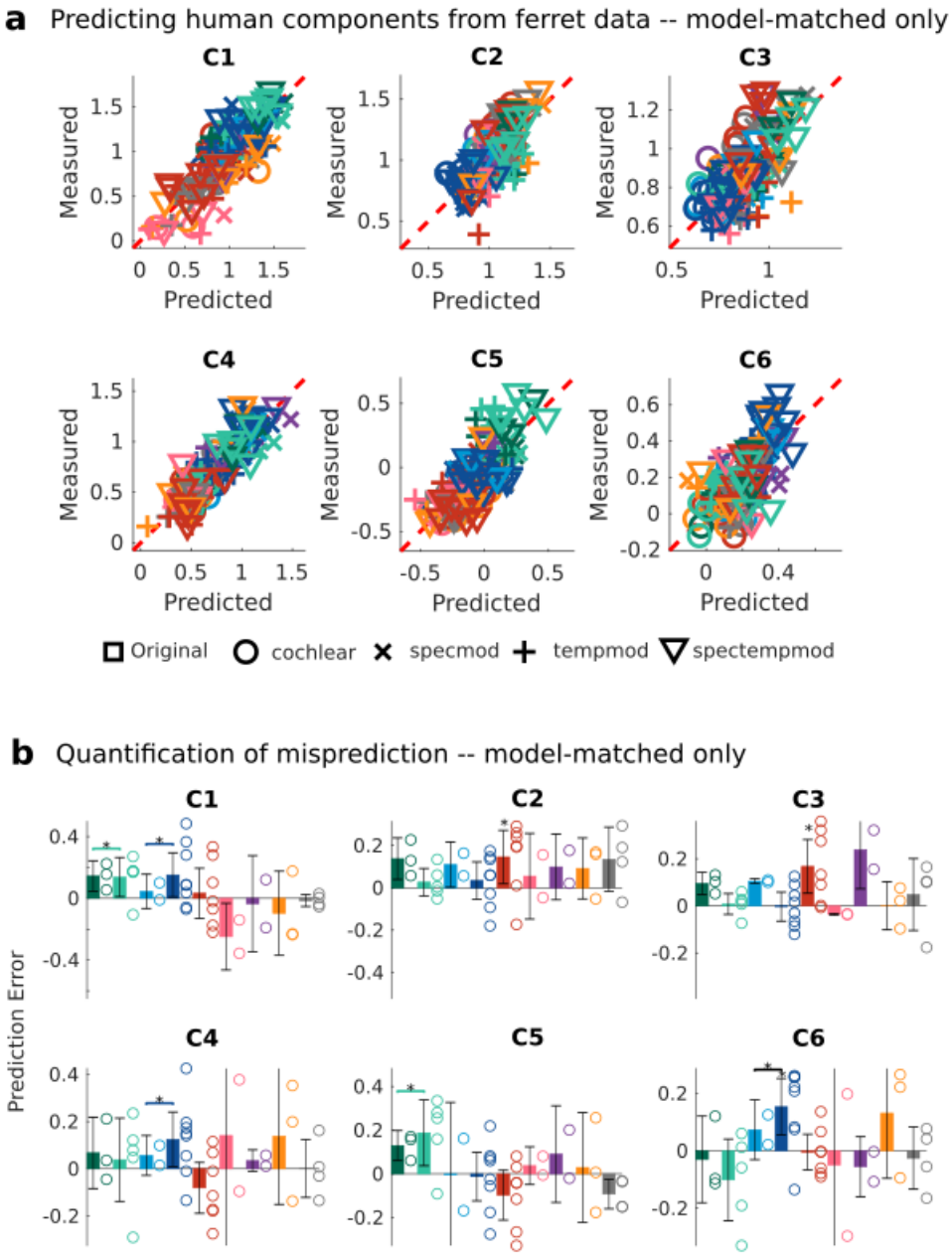

Figure S3.5: **Predicting human components from ferret data (prediction on MM sounds only, display all MM).** Predictions obtained through cross-validation on model-matched sounds only. (**a**) Relationship between the original magnitude of response and the ferret data-based predicted responses for each human component. (**b**) Quantification of the error in prediction for each category of sounds, for full model-matched sounds. Errorbars show mean ± 2 sem. Paired Wilcoxon test, ***: p-value$<10^{-3}$, **: p-value$<10^{-2}$, *: p-value$<5.10^{-2}$.

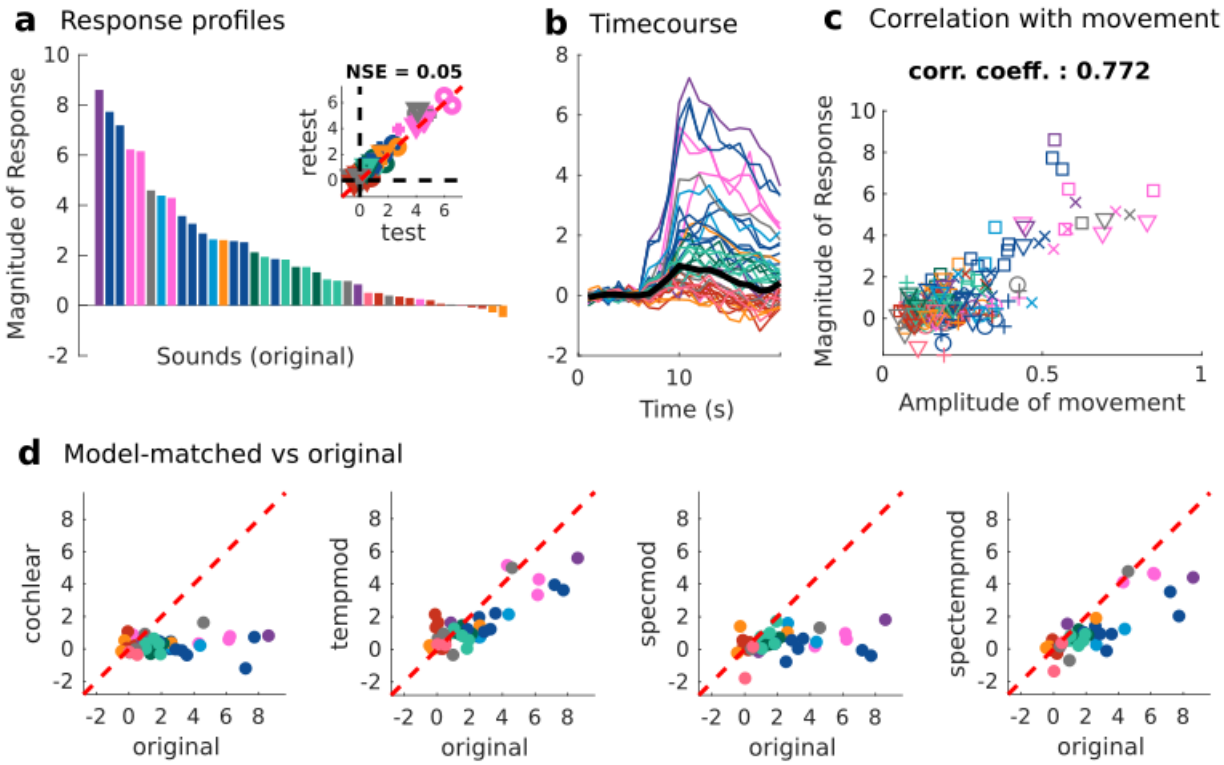

Figure S3.6: **Components linked to the animal's movement (Study 1).** (**a**) Timecourse-projected response profiles for the movement-related component in Study 1, ordered by response magnitudes (arbitrary unit). Only the 40 original sounds are represented. Colors correspond to the different categories of the sounds. Inset: test-retest response amplitude taken from the testing set of sounds, and quantified with NSE (all models and original). (**b**) Timecourse of the responses. Thick black line: average over all the responses (both original and model-matched). Thin colored lines: response for each individual sound (only the original sounds are represented). Black bars corresponds to the sound presentation. (**c**) Correlation of component response profiles with the average evoked movement profile. (**d**) Relationship between the timecourse-projected responses to original stimuli versus each of the different model-matched stimuli. If a model perfectly predcits the brain responses, then all sounds should be aligned on the diagonal (red dashed line).

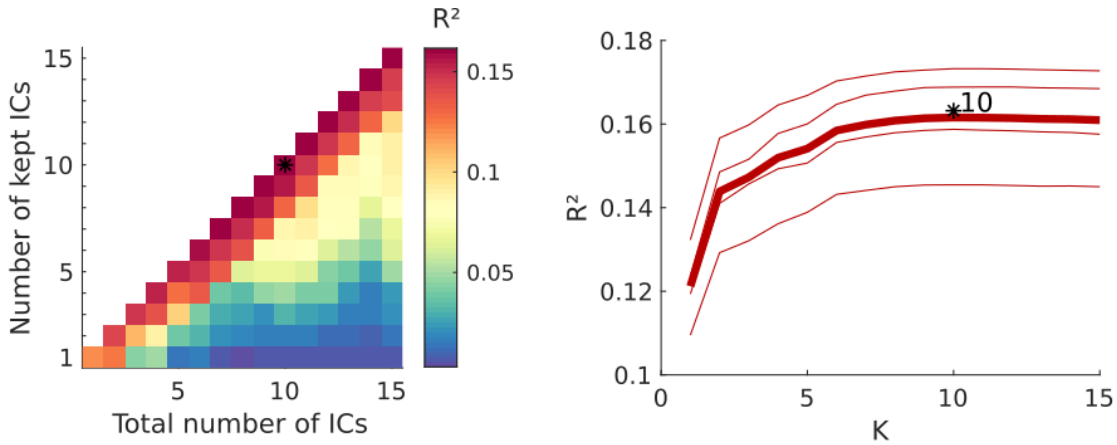

Figure S3.7: **Brain responses to model-matched and natural sounds can be summed up with 10 components (Study 2).** (**a**) R^2 landscape across two parameters, the number of Independent Components (ICs) implemented in the model, and the number of ICs actually kept for reconstruction. Maximal response is shown by a star (10 ICs implemented, 10 ICs kept). (**b**) R^2 values for different values of ICs implemented in the model (corresponds to the diagonal in (**a**). Large line corresponds to the average over all 4 difference testing sets, while smaller lines correspond to each of the 4 different test-sets.

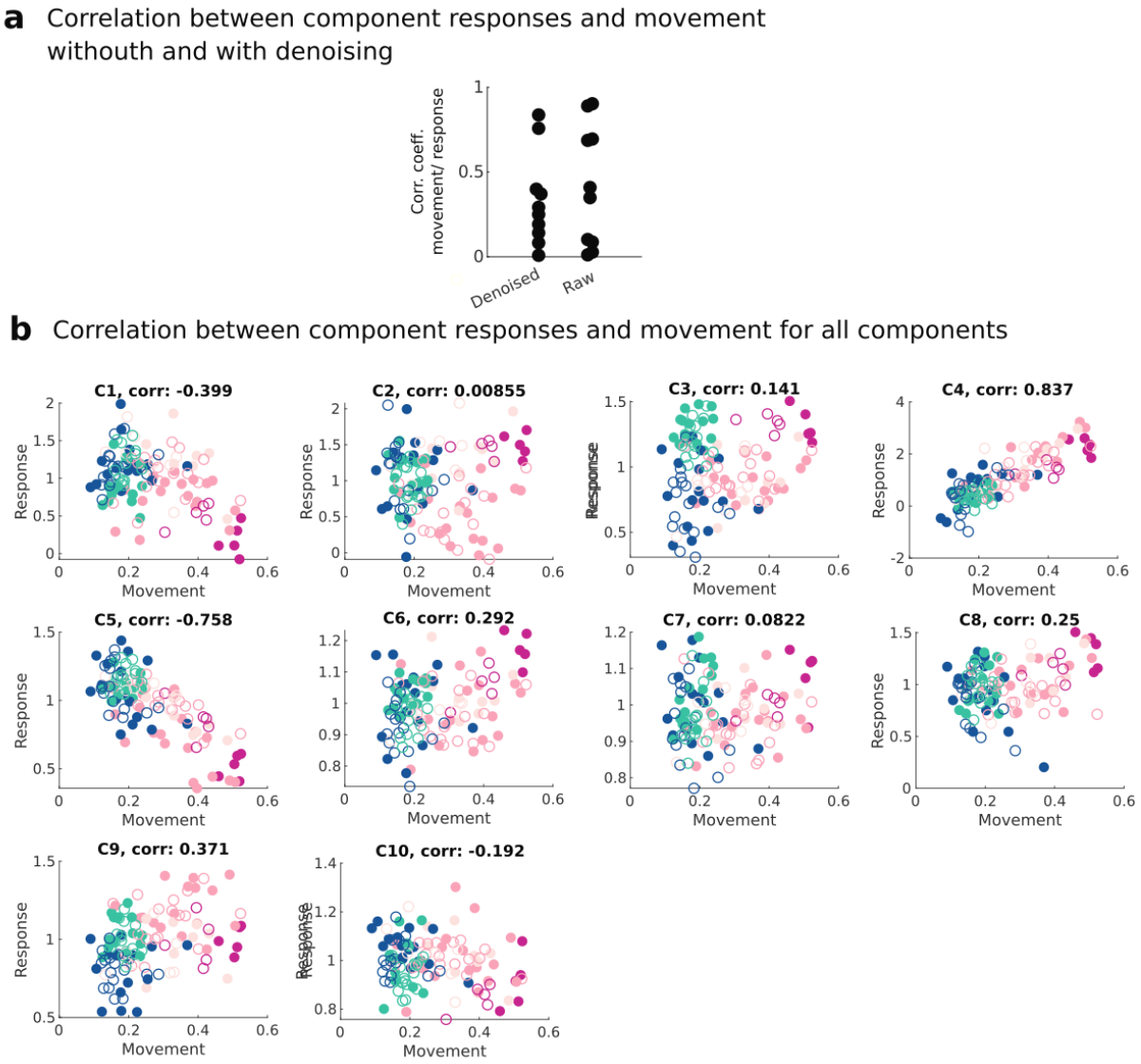

Figure S3.8: **Denoising helps reducing the contribution of movement.** (**a**) Correlation between movement amplitude of the animal and magnitude of responses on each component, for both denoised and non-denoised (raw) data. The difference between the two distributions was not significant. However, it was easier to identify movement-contaminated components in the denoised data (2 of them were clearly identified by their correlation coefficient and their peculiar timecourses - one is showed in figure 3.8), as opposed to non-denoised data (where 4 components showed high correlation with movement). (**b**) Correlation between movement amplitude of the animal and magnitude of responses for each component. Filled dots: original sounds, empty dots: full model-matched sounds.

Chapter 4

Space encoding in auditory cortex

Contributors: Jonatan Nordmark, Agnès Landemard, Yves Boubenec.

4.1 Abstract

Auditory cortex is necessary for accurate sound localization. Nevertheless, the neural mechanisms for space processing as well as its cortical functional specialization remain debated. Previous studies have suggested carnivore auditory cortex is divided into two distinct functional pathways; a dorsal stream devoted to space encoding and a ventral stream dedicated to object recognition. This view has been challenged by the idea that azimuth position is indiscriminately encoded across auditory fields at the single-neuron level. Here, we performed large-scale comparisons of azimuth encoding in primary and secondary auditory fields of awake ferrets using functional ultrasound imaging (fUS).

We imaged high-resolution hemodynamic brain responses to noise bursts coming from spatially separated azimuthal locations. We report hemispherical spatial tuning to both hemifields in distinct regions across auditory cortex, both in primary and secondary regions. Further investigations will help understand the dimensionality of encoding across brain regions, and challenge existing models of spatial processing.

4.2 Introduction

Sound source localization is one of the best examples to illustrate the elegance of brain computations, and how to (elegantly as well) study and understand them (Krakauer et al., 2017; Knudsen et al., 1987). It indeed relies on a simple computational problem that can be based on simple algorithms which provide strong physiological predictions, some of them having encountered quite a success (Jeffress, 1948), at least in some species (Grothe et al., 2010). The representation of space is straightforward in the visual system, where the surface of the retina is directly mapped onto subsequent brain structures, up to visual cortex (e.g., Tootell et al. (1988)). In the auditory domain, such maps have been found in early areas, such as the optic tectum of the owl (Knudsen, 1982), or in the superior (Palmer and King, 1982; King and Hutchings, 1987) and inferior (Binns et al., 1992) colliculi of mammals. In these structures, azimuth and elevation were encoded in continuous maps, reminiscent of the retinotopic organization in the visual system. However, despite the importance of auditory cortex in sound localization (Wood et al., 2017), the underlying neural representation of space within cortical streams has remained a subject of debate.

Most studies investigating single neurons encoding have failed to find a topographic ordering of space representation within auditory cortical fields, thus suggesting that neurons exhibit a highly inhomogeneous encoding (Stecker and Middlebrooks, 2003; Stecker et al., 2005). Results suggested that auditory space is represented within the cortex by a population of broadly tuned neurons, each of which being mainly tuned to one hemifield. This heterogeneity led to the formulation of the 'two-channel' hypothesis (Stecker et al., 2005), according to which auditory cortex would represent space through the relative sum of two main subpopulations, each representing one hemifield. However, recent studies have challenged this view in primary auditory cortex, showing that each hemisphere contained multiple channels tuned to locations in contralateral space, rather than only two channels representing left or right space (Wood et al., 2019).

How spatial information is encoded across cortical fields has remained elusive. In the visual domain, two processing pathways have been identified: a ventral stream dedicated to object recognition and a dorsal stream involved in spatial vision (Mishkin et al., 1983; Wilson et al., 1993). Inspired by these observations, anatomical (Romanski et al., 1999), physiological (Rauschecker and Tian, 2000; Tian et al., 2001) and behavioral (Lomber and Malhotra, 2008) studies suggested that auditory object vs. space processing was organized in a similar manner. As an illustration, deactivation of the *posterior* auditory field in cats resulted in behavioral deficits in sound localization task, while deactivation of the *anterior* auditory field resulted in deficits in a pattern-discrimination

(Lomber and Malhotra, 2008). Moreover, despite a non-topographic code, spatial information was found to be more robust in cat's posterior than anterior auditory fields (Stecker and Middlebrooks, 2003; Harrington et al., 2008). In contrast, single-unit recordings in ferret auditory cortex have revealed that information about the azimuthal position of a stimulus was distributed across all cortical fields (Bizley et al., 2009). Thus, the organization of processing pathways in auditory cortex is yet to be uncovered.

In this study, we tried to tackle these questions using functional UltraSound (fUS) imaging, a newly developed technique that allows high-resolution, large-scale mapping of vascular responses in the brain. Does ferret brain contain specific pathways for space processing? Do primary and non-primary areas encode space through a multi-channel algorithm? We saw in chapter Natural sounds processing that fUS provided a detailed view of the topographic encoding of natural sounds throughout the auditory cortex. Here, previous studies seem to indicate that acoustic space is encoded in a non-topographic way. How much of this complex and debated spatial code could be grasped with a large-scale, high-resolution and high-sampling technique?

In a first approach, we exposed head-fixed, awake ferrets to noise-burst that differed in their azimuthal positions, and recorded hemodynamic responses over the auditory cortex. We observed that single hemispheres were encoding the full auditory space mainly along two dimensions of activity.

4.3 Material and methods

4.3.1 fUS imaging

This section was developed in chapter Mapping the auditory hierarchy. Brains were scanned in the coronal plane, with a spacing of $\sim$600 μmbetween slices. Experiments were performed in two ferrets (A. and T.), across three hemispheres (A_{left}, A_{right} and T_{right}).

4.3.2 Protocol for sensory response acquisition

Head-fixed awake ferrets were presented 3s-long broadband noise bursts (65dB) sequentially presented from a 12-loudspeaker array positioned 60 cm from the head of the animal. Loudspeaker azimuthal spatial resolution was 10°with a 5°offset from midline. Each sound was presented 20 times. Sound presentations were randomized. The protocol for sound presentation was as follows: 10 s of silence (baseline), then 3 s of sound followed by 6 s of silence (return to baseline). Trials were

following each other with only a little random jitter in time of about 1 to 3 s, and fUS acquisitions were synchronized with the beginning of each trial.

4.3.3 Signal processing, and main analysis

Single-voxel analysis

Single voxel tuning curves were obtained by averaging the response of the voxel on a window 2 to 7s after sound onset. Tuning curve significance was estimated using a one-way anova.

Denoising procedure

Responses across a large number of voxels, and across animals, are often hard to interpret. To do so, one can look for a low-dimensional representation of brain activity, that would reveal in an interpretable manner the underlying encoding of the presented stimulus.

We used a similar method as the one presented in chapter Natural sounds processing (see section Denoising procedure in Material and methods). In short, we used a combination of Denoising Source Separation (DSS) and Multiway Canonical Correlation Analysis (MCCA) to extract the components which were both reliable across trials and shared across animals. We did not include the timecourse in the procedure, but focused on the average response to the sounds on the response window (2 to 7s after sound onset). In order to evaluate the encoding in single hemispheres, the data from A_{left} was flipped so that the loudspeaker order could be defined as ipsi- to contralateral systematically in all hemispheres. At the end of the procedure, we obtained 12 components (limited by the number of sound locations), ordered by their reliability across trials and animals.

In order to evaluate the goodness of the obtained components, we trained the component model on half of the trials (10), and left out the other half (10) as a test set. By measuring the correlation of the projection between train and test set trials, we could measure the reliability of the components. We repeated this procedure many times over different combinations of trials across slices (which were arbitrarily paired, since they were recorded over different sessions), which we call pseudo-populations. In order to estimate the significance level of these correlations, we used a conservative method by randomizing the labels 5 times across 50 pseudo-populations (thus yielding 250 control curves). This randomization method is conservative because we do not average across pseudo-populations. Two components showed significant correlation between train and test set projections, and were thus selected as the *reliable encoding subspace*.

Decoding sound azimuth

In order to estimate the accuracy of sound azimuth encoding, we used a decoding approach. We projected even and odd trials on the reliable encoding subspace (estimated as in section Denoising procedure) and computed, for each pair of sounds, the euclidean distance between evoked brain patterns. Doing so, we were able to get an estimate of the relationship between distance between speakers and distance between brain patterns. We looped this procedures over 50 pseudo-populations so as to have more robust estimation of the distances between brain pattern.

4.4 Results

We recorded fUS responses over large portions of the brain (mainly medio-, posterior and anterior ectosylvian gyri) in three hemispheres, while playing broadband noises coming from 12 different locations, spanning a radial range of 110°(figure 4.1-a). In a first approach, we explored the responses of single voxels and their tuning to sound azimuth (figure 4.1-b). Single-voxels showed strong responses and tuning to the contralateral hemisphere, with mainly monotonic tuning curves. Among the pixels having a very significant tuning curve (one-way anova, p-value$<10^{-4}$, n=196), 95% had their peak response on the contralateral side. Furthermore, 45% of the voxels were tuned to the most contralateral speaker (speaker number 1, -55°). The significantly tuned voxels were organized in small clusters distributed throughout the brain (figure S4.1). Interestingly, these clusters could not be explained by the overall responsiveness of the voxels, since the global response maps were mostly uncorrelated with the spatially tuned clusters (figure S4.2).

Single-voxel analysis can sometimes hide underlying large-scale structure and are biased by low statistical power. We used a denoising approach similar to the one developed in chapter Natural sounds processing (see section Denoising procedure in Material and methods). In short, we extracted the components that were the most reliable, and shared across hemispheres. Because we recorded both left and right hemispheres, the speaker order was defined so as to align contra- vs. ipsilateral sides between hemispheres. Doing so, we were able to extract two significant components (figure 4.1-c). The first component (PC1) encoded in a monotonic fashion the contra- to ipsilateral position of the speaker, while the second component represented the distance of the speaker to the midline (figure 4.1-d). The weights of these components were distributed in small clusters (figure 4.1-e), and PC1 weights resembled the ones obtained with the single-voxel tuning analysis (figure S4.1). Finally, we quantified the smallest angular distance that could be discriminated between speakers

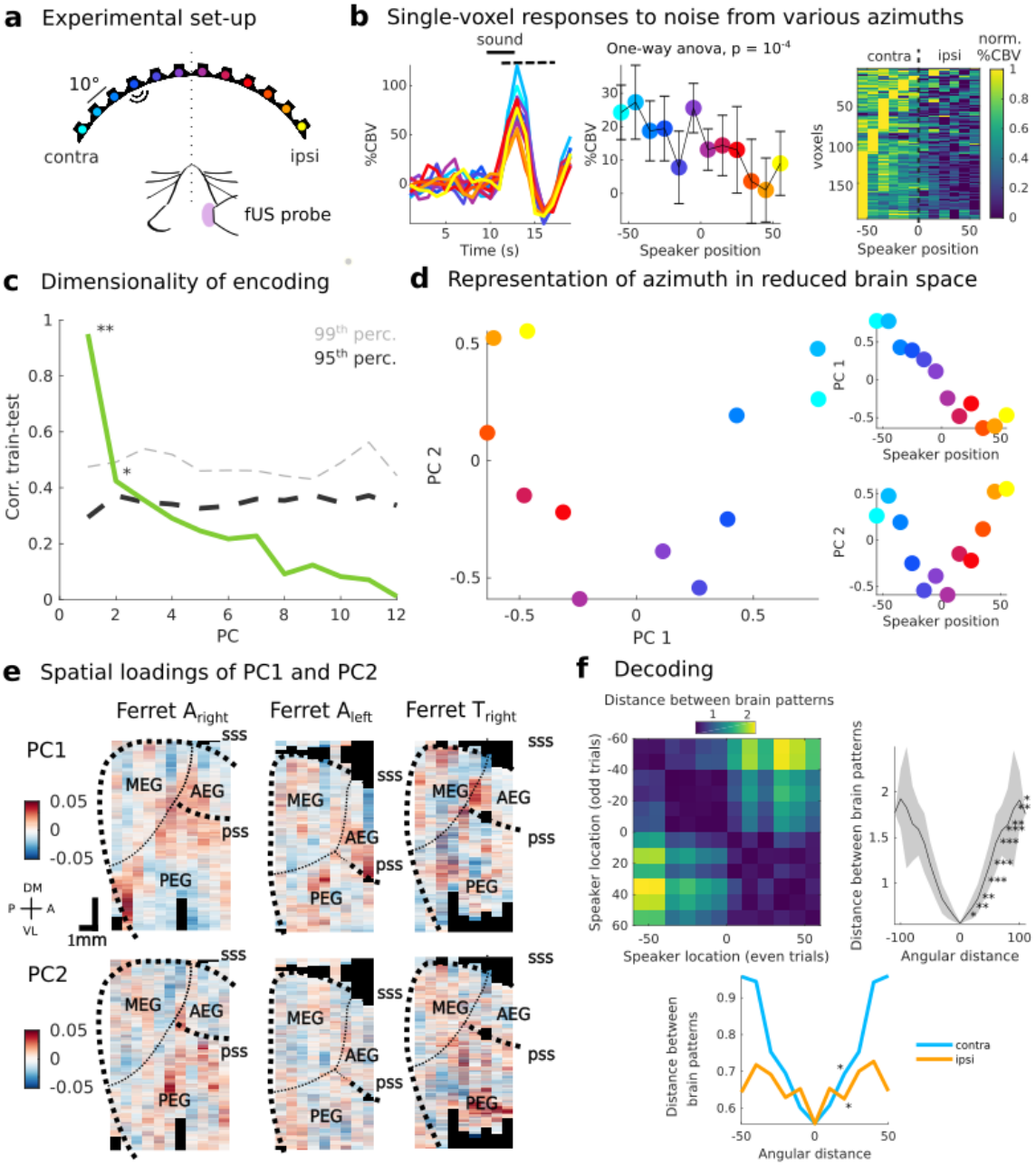

Figure 4.1: **Single hemispheres provide a complete representation of azimuth.** (**a**) Experimental set-up. 12 speakers were positioned 60cm from the head of head-fixed awake ferrets, with a resolution of 10°, while recording the activity over the auditory cortex using fUS imaging. (**b**) Single voxels responses. Left: Timecourse of the response for an example single voxel. Dashed line shows the response window used. Center: Azimuth tuning curve for the same voxel. Errorbars show mean ± 2 sem. Right: Tuning curves of all significantly tuned voxels (one-way anova, p-value $< 10^{-4}$), ordered by the peak of their tuning curves. (**c**) Dimensionality of the underlying brain subspace for azimuth representation accessible in our experiments. It relies on the measure of the correlation between the projection of the train and the test sets on each component. Conservative randomization procedures (shuffling sound positions across trials) were used to estimate the significance threshold. Two dimensions were significantly encoding spatial position. *: p-value$<5.10^{-2}$, **: p-value$<10^{-2}$. (legend continued on next page)

Figure 4.1: (**d**) Low-dimensional brain representation of azimuth. Left: Position of the sound-evoked brain responses within the space spanned by PC1 and PC2. Right: Azimuth tuning curves of PC1 (up) and PC2 (bottom). (**e**) Spatial maps of the weights of each of the two components, for each animal. Scale bar: 1mm. (**f**) Decoding approach. Top left: distance matrix representing the euclidean distance between evoked brain patterns between each pair of sounds. Top right: quantification of the relationship between speaker distance and distances between evoked brain patterns. Errorbars show mean ± 2 sem. *: p-value$<5.10^{-2}$, **: p-value$<10^{-2}$, ***: p-value$<10^{-3}$. Bottom: quantification of the relationship between speaker distance and distances between evoked brain patterns for both contra- and ipsilateral hemifields. Errobars are not represented for clarity.

(figure 4.1-f). We found that within the reliable encoding subspace, speakers could be discriminated even at the smallest angular distance (10°, p-value$<5.10^{-2}$; top panels in figure 4.1-f). Within both contra- and ipsilateral fields, brain responses could significantly discriminate different speakers (p-value$<5.10^{-2}$; bottom panel in figure 4.1-f). Our data tends to show that the encoding was sharper within the contralateral field. However, our low statistical power did not allow us to show this trend as significant.

In analogy with visual processing, previous studies have suggested that auditory processing could be divided in two parallel streams, one for object recognition and one for sound localization (Rauschecker and Tian, 2000). In the ferret auditory cortex, an analog of the dorsal stream goes along the anterior ectosylvian gyrus (AEG), while the ventral stream would fit along the posterior ectosylvian gyrus (PEG). Primary auditory cortex is located in the middle ectosylvian gyrus (MEG). To explore the specialization in spatial processing of these different regions, we looked at the distributions of the weights of the two encoding components that we identified, PC1 and PC2, across those different areas (figure 4.2-a). Within hemispheres, we observed significant differences in the distribution of the weights between auditory fields (two-sample Kolmogorov-Smirnov test, p-value$<10^{-2}$). However, these differences were not consistent across animals. Thus, we interpret this result as the auditory cortex being organized in rather clustered azimuth-sensitive regions, that are not strictly bound to a specific region of the auditory cortex.

These different areas could rely on different underlying codes. To explore this idea, we applied the same pipeline of analysis to each area separately (yet combining different hemispheres) (figure 4.2-b). Through this prism, we observed that the strong monotonic component (PC1) was present in all areas. However, the evidence for the presence of a second component was scarce in MEG and AEG, but unexpectedly stronger in PEG. This difference could stem from the fact that more voxels were present in PEG (n=13453) vs. MEG (n=8056) and AEG (n=7218). All areas encoded space

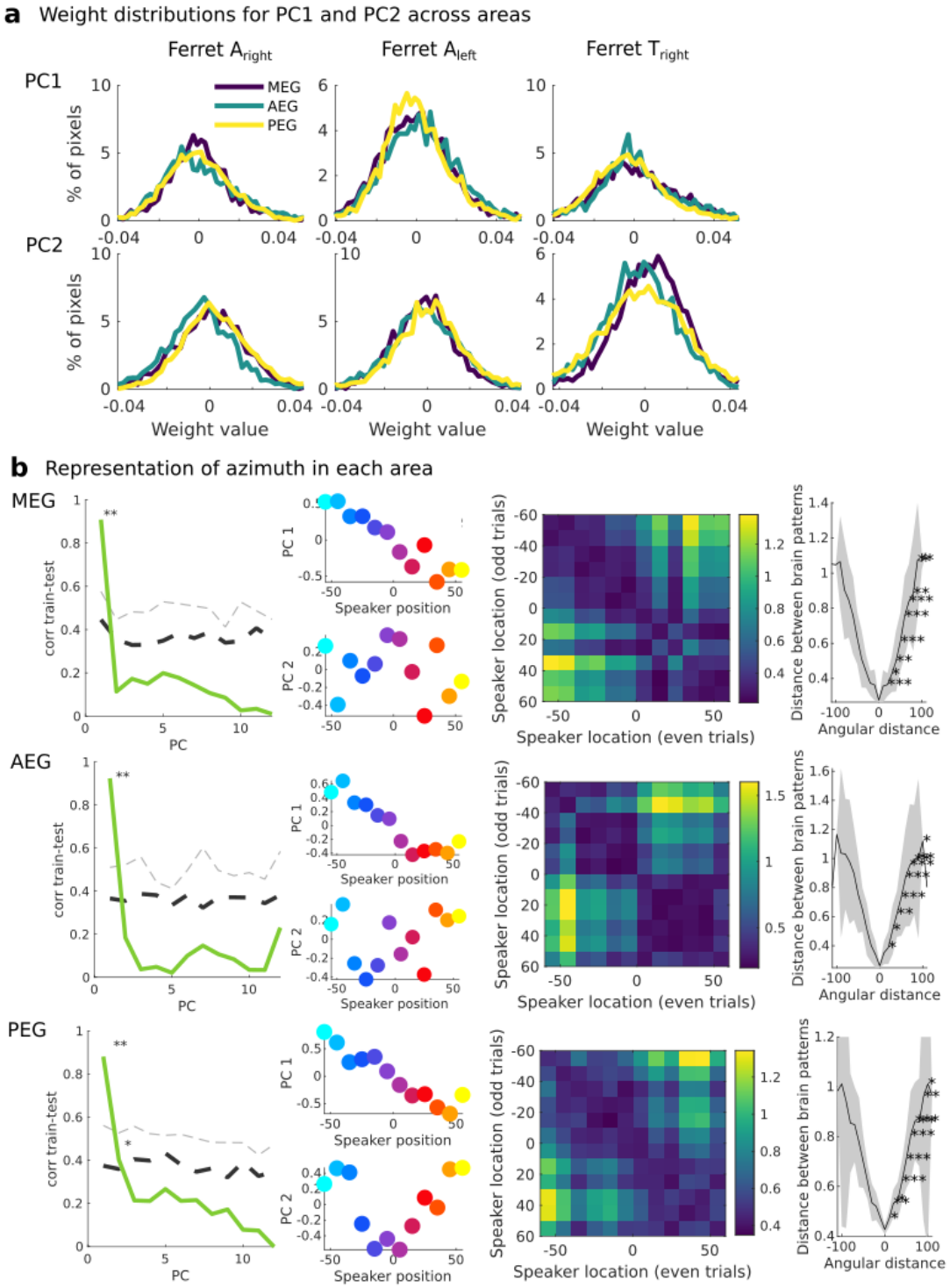

Figure 4.2: **Azimuth does not seem to be specifically encoded within a single stream.** (**a**) Weight distributions for PC1 and PC2 across areas for individual hemispheres. (**b**) Same quantification as in figure 4.1 for each area independently (top: MEG, center: AEG and bottom: PEG). Far left: Dimensionality of the underlying brain subspace for azimuth representation accessible in our experiments. Left: Azimuth tuning curves of PC1 (up) and PC2 (bottom) Right: Confusion matrix representing the euclidean distance in distances between evoked brain pattern between each pairs of sounds. Far right: Quantification of the relationship between speaker distance and distances between evoked brain pattern.

with a high accuracy, with an angular resolution of at least 20°in all areas (10°in PEG).

4.5 Discussion

In this chapter, we explored how auditory cortex encodes the spatial location of sounds. Within the prism of fUS imaging, auditory cortex displayed azimuth-sensitive patches distributed throughout cortex. Single hemispheres contained information about the full 120-degree auditory space, with an encoding seemingly sharper for the contralateral hemifield. The encoding subspace could reliably be summed up by two main components, one linearly encoding azimuth from the most contra- to most ipsilateral speaker, and the second representing the distance to midline. Finally, our preliminary results suggest that there is no specialization between ventral and dorsal streams in azimuth encoding in the ferret.

Thus, fUS can also give access to cortical codes that are more distributed than the more well-known topographic organization of acoustic features such as frequency tuning. Our recordings recapitulate a large part of what has been observed with single unit recordings. In addition, it allowed us to rapidly map, in 3 dimensions, the localization of azimuth sensitive clusters, opening up interesting experimental and theoretical perspectives. For example, fUS could be used, in single animals, to identify the regions that seem to be mostly sensitive to space, before deciphering more precisely the underlying mechanisms with electrodes in the target areas. Moreover, it can potentially provide interesting clues to the ongoing debate on the representation of space throughout auditory cortex.

Azimuth tuning

Using single-pixel analysis, we could observe that an outstanding majority of the voxels were tuned to the contralateral side, and especially to the most contralateral speakers (figure 4.1-b). This is in accordance with what has been observed in single units in the auditory cortex of cats (e.g., Stecker et al. (2005)), rats (e.g., Yao et al. (2013)) or monkeys (e.g., Woods et al. (2006)). Moreover, it has previously been suggested that one hemisphere encoded sound location only in its contralateral hemifield (Stecker et al., 2005; Wood et al., 2019). Our results tend to show that both hemifield seem to be represented, with a discrimination that is weaker but still significant for the ipsilateral hemifield.

Behavioral studies have shown the highest behavioral sensitivity in azimuth discrimination was

found around the midline, where auditory cortex neurons' tuning displays its sharpest slope (Stecker et al., 2005). Our recordings are consistent with this observation, since the distances between activity patterns were indeed increasing the most sharply between sounds around the midline (figure 4.1-f). This is also visible in the higher slope of the encoding along PC1 for a 0°azimuth (figure 4.1-d). Further quantification and experiments could be performed to explore this direction.

Spatial organization of azimuth-sensitive responses

Previous studies have also investigated the spatial organization of azimuth sensitive units. In some species, azimuth sensitive units were distributed throughout the cortex, and organized in small clusters that were scattered throughout auditory cortex (Bizley et al., 2009; Panniello et al., 2018). On the contrary, primary auditory cortex of the pallid bats is organized in two clusters of 'peaked' vs. 'binaurally inhibited' cells (i.e., encoding one hemifield) sensitive to interaural level differences, within which azimuth was topographically encoded (Razak, 2011). Our results are consistent with the observation of scattered clusters. Interestingly, we could investigate this organization in 3 dimensions. In the ferret, azimuth-sensitive clusters have been shown to be scattered throughout A1, or located at the tip of pseudo-sylvian sulcus (Bizley et al., 2009). Consistently, we show that most of the loadings for our main component (PC1) were located mostly throughout primary auditory cortex, PEG, or around the pseudo-sylvian sulcus (figure 4.1-e). Azimuth-sensitive voxels' location was animal dependent, and the ones located around the tip of the pseudo-sylvian sulcus were either on the dorsal lip, or at the fundus of the sulcus. We also observed azimuth-sensitive clusters within the folding of the superior-sylvian sulcus in the most medial part of primary auditory cortex. Single-voxels analysis were consistent with this topographic organization (figure S4.1). We could not reveal any topographical organization for response type nor preferred azimuth (figure 4.1-e), which is coherent with previous reports (Bizley et al., 2009; Panniello et al., 2018). As mentioned in Bizley et al. (2009), we could not observe noticeable differences between brain regions, especially PEG and AEG, despite the reports of spatial sensitivity for visual, auditory and bisensory stimulation being highest in AEG (Bizley and King, 2008).

Overall, the origins of this organization remain poorly understood. Our approach, combining both precise and large scale mapping could help investigate this question. Notably, the connectivity of such areas with other azimuth sensitive brain structures could be assessed in order to determine whether this organization is actually inherited from upstream structures, or a specific property of auditory cortex. Moreover, early studies have suggested that sound localization processing in the

cortex was frequency-channel dependent (Jenkins and Merzenich, 1984). It is not clear if this fits with the type of clustering that we see, since some of the frequency channels might be out of the azimuth-sensitive areas we observe. Thus, understanding how these two representations combine at the surface of the cortex might require experiments in which more sophisticated sounds are played. This would clarify the interaction that might exist between 'what' and 'where' within auditory cortex.

Dimensionality of the representation and current models of space representation

It is still unclear whether auditory cortex relies on a very simple encoding mechanism for space, i.e., a two-channel model (Stecker et al., 2005), or a more complex, distributed encoding (Wood et al., 2019). The granularity (sampling and spatial resolution) of the recording technique can influence the observed type of encoding – coarser techniques favoring low-dimensional representations. For example, Higgins et al. (2017) mainly showed unidimensional coding in single hemispheres using fMRI in humans, with hints (yet not demonstrated) in favor of the opponent two-channel model. Thanks to the combination of a higher-resolution method and advanced analytical tools, we could observe at least two significant encoding dimensions in single hemispheres, with a full representation of space. This observation is consistent with two out of the three encoding models explored in Wood et al. (2019), namely the distributed code model (where neurons exhibits heterogeneous spatial tuning) and the opponent two channel model (where two populations encode left and right hemifields). The third model, where each hemisphere responds maximally to its contralateral field, would be unidimensional when recording in a single hemisphere. More precisely, our data favor a model where both contra- vs. ipsilateral hemifields are contrasted, in parallel to the distance to the midline. Our results are in line with the two-channel model demonstrated with positive and negative BOLD responses across the auditory cortex of macaques (Ortiz-Rios et al., 2017). However, we did not observe positive and negative %CBV, but rather a bidimensional encoding through positive responses (data not shown).

The method used here to estimate the underlying dimensionality might not be the most appropriate to resolve these issues, nor the most sensitive. Other methods have been developed that rely on a much larger number of stimuli presented (Stringer et al., 2019b). Ongoing work will help improve this estimation and extract the underlying subspace in a more robust way.

Effects of attention

Processing pathways can be highly modified by the state of the animal, or its engagement in the task it is performing (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al., 2019). In that direction, Ahveninen et al. (2006) have suggested that attention can actually play an important role in the specific modulation of the 'what' and 'where' pathways in humans. Moreover, task engagement has been shown to sharpen spatial tuning in primary auditory cortex of cats (Lee and Middlebrooks, 2011). Thus, a behavioral task could be crucial in actually revealing the specificity of the different processing streams at stake.

4.6 Summary

Our study demonstrates how mesoscale techniques can provide advantageous information on complex, spatially non-continuous encoding. We believe further work in that direction will clarify the algorithm underlying space processing in auditory cortex, and how it unfolds within the different cortical regions. Our approach can also provide a global yet rapidly accessible view of azimuth sensitive patches, to target more specific, single-neuron recordings.

4.7 Supplementary figures

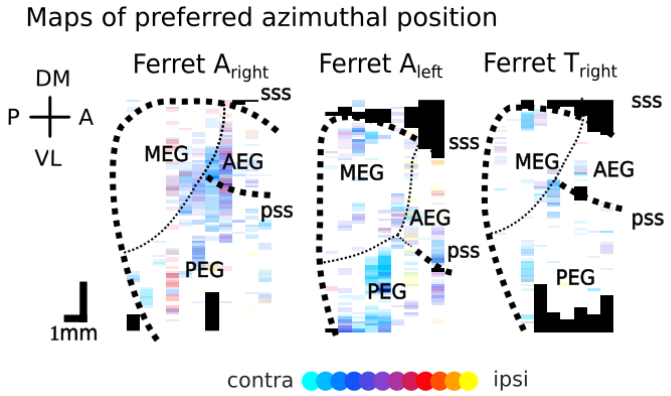

Figure S4.1: **Maps of azimuthal tuning for each hemisphere.** Location of the speaker that elicited highest responses was computed for each loosely significant voxel (one-way anova, p-value $< 10^{-2}$), and plotted in a *view from above* for each hemisphere.

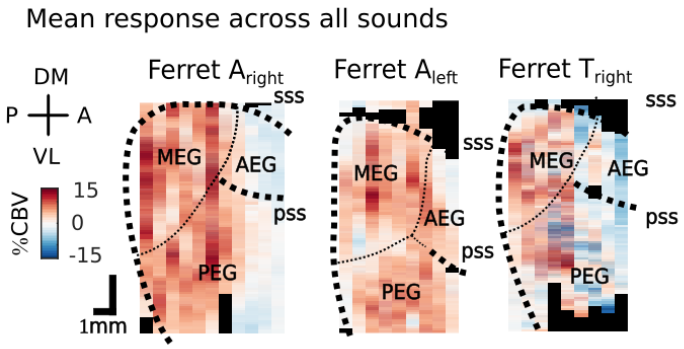

Figure S4.2: **Average response (%CBV) for each hemisphere.** Average response over all sounds for each hemisphere, viewed from above. A consistent, sharp transition is visible when reaching more anterior secondary areas.

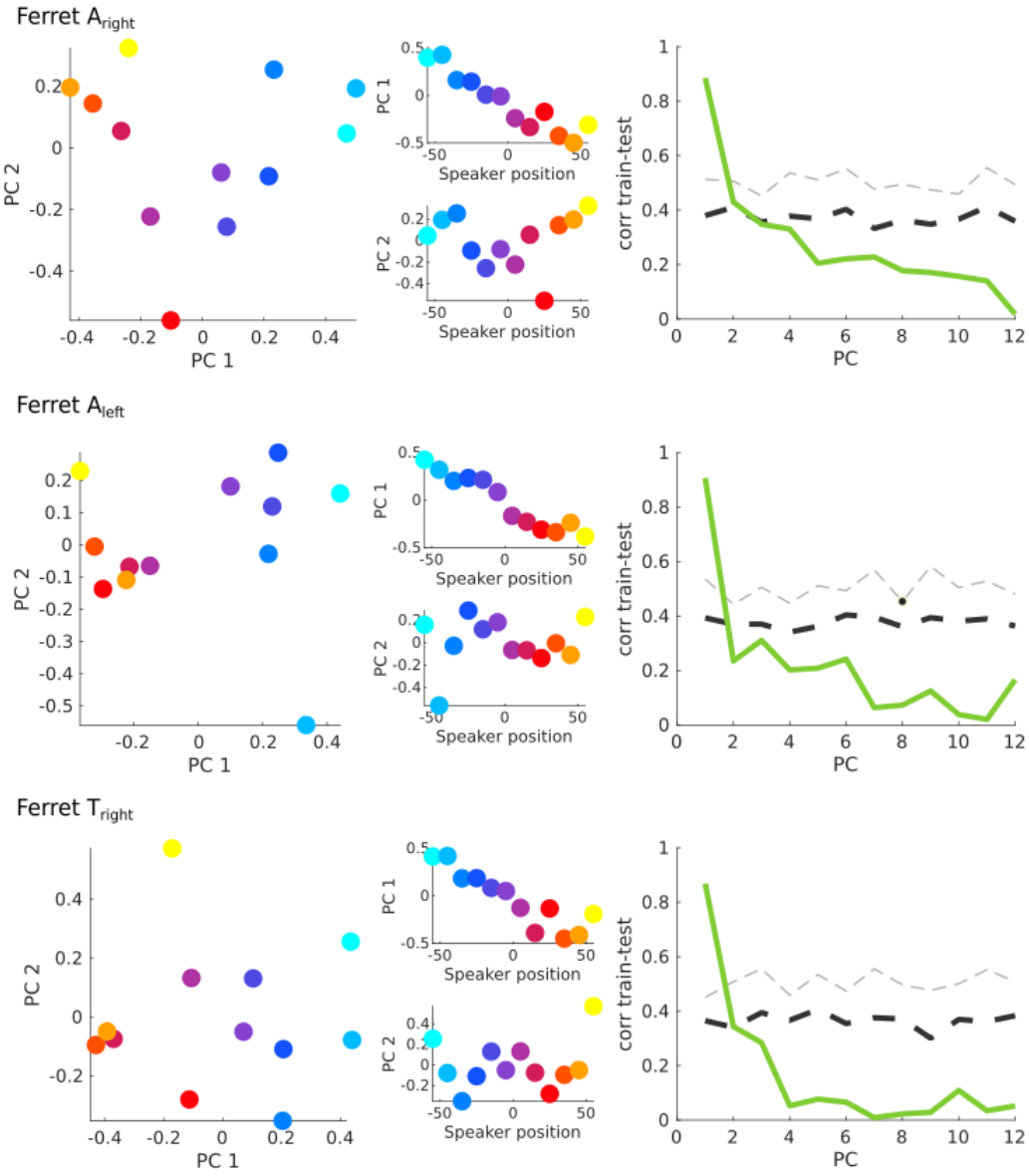

Figure S4.3: **Low-dimensional space representation for each hemisphere.** Same plots as in figure 4.1, for each animal individually.

Chapter 5

General discussion

5.1 Results summary

Throughout this thesis, we explored the encoding of sounds within the auditory system using a newly developed technique: functional UltraSound imaging. This thesis aimed at answering both technical and fundamental questions.

chapter 2: Mapping the auditory hierarchy. In this chapter, fUS imaging was used to characterize the global organization of a sensory system, in our case the auditory system. Sensory systems are often defined by their internal organization, along multiple brain structures, as well as by their connectivity with other areas. Our study examined those two aspects. Taking advantage of the well-known tonotopic arrangement across auditory structures, we showed that fUS could be used to inspect with a high resolution the functional organization of both small and deep nuclei, and across primary and secondary areas of the auditory cortex. Moreover, we explored the top-down connectivity between frontal and auditory cortices, previously undescribed in the ferret. With other groups, we thus argued that fUS imaging could provide a new prism to look at sensory systems organization, which would confer specific advantages: full brain recording, high sampling, portable, silent acquisition.

chapter 3: Natural sounds processing. Building up on the technical demonstration developed in chapter 2, we set out to tackle a more fundamental question in auditory neuroscience, which is how natural sounds are encoded in the brain. We showed that ferret auditory cortex responses to natural sounds were highly organized, and a computational approach allowed us to extract struc-

tured components sensitive to specific acoustic features, among which some were specific to human speech. Strikingly, these responses could well be explained by a simple auditory processing model in both primary and secondary areas of the cortex. This was in contrast with similar experiments performed in humans, where responses to speech and music in non-primary areas were based on high-order features that were not contained in the model. These results suggest that human and ferret auditory cortex might contain different levels of processing, thus demonstrating rigorously a functional specialization of human brain to speech and music compared to ferrets. Because speech and music are not ecologically relevant sounds for a ferret, we also investigated the responses of the ferret brain to ferret vocalizations. Ferrets were behaviorally able to discriminate the different categories of sounds (vocalization vs. speech or music) as well as whether they were original or model-matched versions. However, we could not observe large, specific fUS responses in their brain to original vocalizations. This suggested that even in the context of ecologically relevant sounds, ferret auditory cortex displays only little, if any, signature for high-order acoustics processing.

Our study raises several technical remarks. First, the existence of speech-'selective' areas can be found in other species than humans. Second, only the use of model-matched stimuli allowed us to characterize thoroughly their differences. Thus, cross-species comparisons is a necessary step in understanding brain processes. Third, the use of fUS imaging enabled the exploration of the organization of the brain of small species in unprecedented details, and well beyond its well-known frequency tuning.

chapter 4: Space encoding in auditory cortex. The two precedent chapters established that fUS imaging could be used to assess spatially organized activity patterns. In this final set of results, I exposed how more complex, distributed codes can also be accessed. I provided some novel clues about the encoding of spatial azimuth in the auditory cortex, across primary and secondary areas. FUS imaging revealed that single hemispheres could contain a full representation of space, that was lying along an at least two-dimensional space. These results recapitulated many of the results observed with single electrodes, and contributed to identifying the local structure of azimuth sensitive clusters throughout auditory cortex. These clusters were present in both primary and non-primary areas, and in both dorsal and ventral streams, questioning further the existence of a 'where' pathway in ferrets. Future studies should build upon this paved way in order to better quantify the dimensionality of the encoding and explore the validity of current models.

5.2 Discussion

The specific aspects for each subject were treated and discussed separately within each chapter. In this general discussion, I will provide a more general view of the points shared across subjects and chapters, and expose the future directions of this whole project.

Filling the gap of mesoscale imaging techniques

In the Introduction, I have argued that coarse resolution yet high-sampling techniques can provide valuable perspectives on brain organization and computation, such as with the example of the characterization of a face area in the macaque using fMRI (Tsao et al., 2003). Throughout this manuscript, we have attempted to demonstrate that fUS imaging can be used as a new prism to look at brain activity, and reveal mesoscale, 3-dimensional organization that would be hard to access otherwise. Thus, fUS imaging fills up the so far quasi empty technical window of 3D, mesoscale techniques that can be used in small mammals. In all of the studies that we performed and that I presented throughout this thesis, we revealed previously uncharacterized mesoscale organization, among which: the top-down projections from frontal cortex to the fundus of the pseudo-sylvian sulcus in chapter 2; the spatially organized components of responses to natural sounds in chapter 3; the azimuth-sensitive clusters in chapter 4. Moreover, because we could map the cortical organization in each individual ferret, further studies could aim at designing an individual approach combining large-scale fUS mapping and targeted single electrode investigations, as in Macé et al. (2018).

An opportunity for more diversity in neuroscience

Cross-species comparisons have proven very fruitful in probing neural mechanisms and processing pathways underlying fundamental cognitive capacities (Rauschecker and Scott, 2009; Grothe et al., 2010; Walker et al., 2019). As an example, we demonstrated in this thesis that ferret brains could actually present speech selectivity to an unexpected level (the speech-'selective' component being one of the most robust, across animals, and experiments), and only a cross-species comparison with model-matched stimuli as controls enabled us to identify the true specificities of the human brain (i.e., high-order feature selectivity). Another example is the absence of columnar organization for orientation selectivity in the visual cortex of rodents, in contrast to what is systematically observed in cats or macaques, which questioned the actual computational significance of orientation-selective columns (Kaschube, 2014).

Neuroscience is abundant in remarkable studies using remarkable animal models, some being well known (such as the aplysia (Castellucci et al., 1970)), others much less (such as the cuttlefish (Reiter et al., 2018)). Amongst mammals, the complexity of their nervous system and of its organization can be an obstacle to its exploration. We have argued that fUS imaging could provide rapid yet precise mapping of sensory systems across brain structures, depth, and stimulus type. Just like the studies developed in this thesis with ferrets, we are hopeful that fUS imaging could help increase the diversity of species (especially small mammals) used, and thus provide a broader view of the brain's organization and evolutionary trajectory.

Cortical fields kept their mysteries

The large size of the performed craniotomies enabled us to explore the relative organization over different cortical fields, such as primary auditory cortex (A1 and AAF), secondary auditory cortex (PEG and AEG), as well as tertiary regions in at least one animal (VP). Fundamental differences exist at the single-cell level between these (Guo et al., 2012; Elgueda et al., 2019), but their global organization has remained poorly understood beyond frequency tuning. Despite the large number of stimuli types used in our studies, spanning a fair number of acoustic dimensions (frequency, spectro-temporal modulation, high-order features, azimuth), we did not observe systematic, robust differences across cortical fields. Several reasons could explain such an absence. First, the resolution of the technique, which averages over many cells, limits the detection of small differences in receptive field properties across cortical regions (Guo et al., 2012; Elgueda et al., 2019). Second, these differences could become striking only during behavior (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al., 2019; Ahveninen et al., 2006). Third, the specificity of some of these fields could actually come from the way they combine all those properties. Only few studies investigated the encoding across numerous dimensions in ferret auditory cortex (Bizley et al., 2009), and how the encoding of these acoustic properties actually interact thus remains poorly known. One hypothesis, which has been explored further in the visual cortex (DiCarlo and Cox, 2007), states that the cortical hierarchy performs object recognition ('what' pathway). Higher cortical areas, such as IT, thus possess 'untangled' object category representations (though non-categorical properties are still being encoded (Hong et al., 2016)). In our experiments, most sound categories were largely defined already by their low level acoustic properties, on which animals actually seemed to rely for their behaviorally relevant categories. Indeed, ferrets' movement was larger on both model-matched and original vocalizations than on any other sound, suggesting that sounds with the low-level acoustics

of vocalizations may be perceptually grouped in a category that has a global behavioral meaning. Introducing additional, category-orthogonal dimensions to the sounds could be a necessary step to unveil the specific representation of sound categories (as such) within auditory cortical fields. In the visual domain, object position or orientation are often used as category-orthogonal dimensions. Further studies could for example use sounds of varying nature and spatial locations. Other types of feature interactions, such as audiovisual combinations (Bizley and King, 2008), could also be further explored and help characterize the specific role, if any, of each field.

A single voxel characterization?

In that direction, one of the underlying questions raised in the Introduction was the following: how are all layers of complexity combined one with another within auditory cortex? Our ability to explore this question will directly depend on how precise the patterns of vascular responses to these different dimensions can be. Throughout the three chapters, I provided a global vision of the spatial organization of vascular responses to several sound dimensions. The clusters of functionally homogeneous voxels were often spanning quite large areas ($>$600 μm), with sometimes sharp functional boundaries between domains (see Mapping the auditory hierarchy, figure 2.2, or Natural sounds processing, figures 3.3 and 3.9). This observation was consistent with the existence of broadly tuned vascular domains (Harrison et al., 2002; O'Herron et al., 2016). However, in O'Herron et al. (2016), even close by vessels could display significantly different (yet broad) tunings, raising the question of the experimental and theoretical limits on the discriminability between adjacent voxels in the auditory cortex. In chapter 2, we proposed a minimal value by focusing on sharp functional transition, that depended on the sounds and the number of trials used (300 μm). Yet, the variation in the voxels' functional tuning *within* vascular domain remains poorly explored. Our approach, using a large variety of sounds, could help explore this fundamental limitation of blood flow based imaging techniques.

Understanding connectivity patterns and top-down modulation

The organization of a brain area generally makes sense if the inputs and even more the outputs of this region are known, as well as the global computational goal of the circuit. The latter, behavior, will be discussed in the next section. The former resides in the connectivity scheme that exists between brain regions. As an example, we demonstrated in chapter 2 that frontal cortex connects only a subpart of the auditory cortex, namely the fundus of the pseudo-sylvian sulcus (PSS). The

part of the frontal cortex investigated here (dorso-lateral frontal cortex, dlPFC) has been shown to extract task-relevant stimuli (Fritz et al., 2010). Thus, this top-down pathway is a good candidate for the adaptive modulation of auditory processing in auditory cortex during behavior (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al., 2019). Three main questions remain open.

First, are there any other parts of frontal cortex that connect auditory cortex ? Our investigation has been limited to a proof of concept, but further studies should explore this question more thoroughly. The combination of fUS and electrodes allows one to explore within FC which area connects auditory cortex, as well as the functional properties of the source FC area itself. Thus, one can explore at a global scale the type of information sent to auditory cortex.

Another question is 'to which circuit is it sent?'. Understanding the computations taking place in the region that receives these top-down inputs, i.e., PSS cortex, as well as its connectivity scheme with the rest of the auditory cortex, might provide fundamental clues to understand the top-down control of auditory processing. However, we overall failed to find functional responses unique to PSS cortex (having explored responses to artificial and natural sounds, vocalizations, as well as visual stimuli). Further studies should thus explore more systematically its functional characteristics, as well as its connections with secondary and primary auditory regions.

Finally, the feedforward connectivity between auditory cortex and more frontal areas is still to be elucidated – yet it is fundamental so as to understand to what end are local computations performed. Recent studies in the monkey have shown a complex pattern of connectivity: while early auditory fields show connectivity with frontal areas, high-order, vocalization-sensitive areas were engaging local networks of processing (Petkov et al., 2015). Further experiments, based on the work presented here, could aim at testing this hypothesis in the ferret.

Technically, the advances of targeted optogenetics will help deciphering more precisely (e.g., in a cell type-specific manner) these connectivity patterns (Lee et al., 2010). Important controls must be set up to take into account the effect of light on vascular dilatation (Rungta et al., 2017).

Towards behavioral paradigms

Any brain computation is to be interpreted through the prism of behavior (Krakauer et al., 2017). Furthermore, behavior itself modifies brain activity to a large and yet mostly unknown extent (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al., 2019; Ahveninen et al., 2006; Stringer et al., 2019c; Musall et al., 2018). Thus, incorporating behavorial paradigms in our experiments could particularly enrich our understanding of auditory processing. Several studies have now used fUS during behavior

in rats (Sieu et al., 2015; Urban et al., 2015), mice (Macé et al., 2018), and monkeys (Dizeux et al., 2019). Yet, in our hands, any movement of the animal provoked substantial artefacts, or variations in global blood perfusion that seemed unrelated to the underlying neural activity. The possible origins for such differences have been discussed in chapter 3. In order to tackle this issue, we designed a head-fixed fUS probe holder, which was tested on frontal areas recordings. The overall quantity of observed movement-related artefacts in the signal was not significantly reduced, suggesting that the design should still benefit from improvements. Designing better head-fixed fUS probes for ferrets would be an interesting path to follow.

Overcoming this technical limitation would then open up a certain number of promising research perspectives. Some have already been hinted at in specific chapters (e.g., exploring the modulation of 'what' vs. 'where' pathways in chapter 4, or the representation of high-order features within auditory cortex in animals trained to discriminate real speech in chapter 3). From another perspective, the role of frontal cortex in sensory processing has remained elusive. Dorso-lateral frontal cortex has been shown to present target-specific responses when animals are engaged in a task, and has thus been hypothesized to top the auditory hierarchy before motor command (Fritz et al., 2010). However, other parts of the frontal cortex, such as orbito-frontal or medial frontal cortex seem to display such properties (personal recordings at University of Maryland). This has not been systematically explored, due to the difficulty to extensively record in and characterize such a large area. FUS imaging could thus provide an interesting window on that question.

5.3 Future directions, in brief

Finally, I propose here a condensed overview of the potential prospects hinted at in this thesis.

Related to chapter 2, Mapping the auditory hierarchy:

- Exploring the global connectivity patterns between frontal and auditory areas to understand top-down control
- Characterizing the functional organization of both sending and receiving areas
- Characterizing the connectivity between PSS cortex and the rest of the auditory cortex to understand how top-down effects spread across cortical fields

Related to chapter 3, Natural sounds processing:

- Exploring the underlying mechanism for original vs. model-matched vocalization perceptual discrimination by:
  - Refining the movement cancellation procedure in order to extract purely functional domains
  - Exploring other brain areas (the most rostral part of ventral posterior auditory field, or frontal areas)
- Characterizing the formation of high-order auditory feature selective domains by:
  - Training adult ferrets to discriminate original vs. model-matched speech, and record potential changes across auditory cortex (in the line of e.g., Polley et al. (2006) that used a simple acoustic feature)
  - Creating specific representation of speech in juvenile ferrets by, e.g., associating real speech with rewards (in the line of the FFA formation in monkeys), and observing whether speech-specific domains or patterns appear in auditory cortex

Related to chapter 4, Space encoding in auditory cortex:

- Further testing the multi-channel hypothesis by improving statistical power and analysis
- Deciphering the effects of behavioral engagement in the representation of space

Related to all:

- Exploring the delimitations of vascular domains and their local functional heterogeneity
- Characterizing the representation of each acoustic dimension throughout auditory cortex, and the possible interactions among them

Chapter 6

General conclusion

This thesis aimed at exploring how sounds are topographically encoded within the brain. Brain computations are in part apparent through the way functional responses are organized within brain structures. The thesis combined new computational and experimental tools to expose the various spatially organized modules of processing that overlap within single brain areas. This new approach will hopefully help us understand how the auditory system combines acoustic features at different levels to create navigable representations of the world. Specifically, our results provide new clues on the evolution of speech processing, and on how the brain of different species might apprehend various levels of complexity in the natural world.

Bibliography

Adams, D. L., Piserchia, V., Economides, J. R., and Horton, J. C. (2015). Vascular supply of the cerebral cortex is specialized for cell layers but not columns. *Cerebral Cortex*, 25(10):3673–3681.

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.-H. H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., Belliveau, J. W., Jääskeläinen, I. P., Hämäläinen, M., and Levänen, S. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39):14608–14613.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., and Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, 20(10):1404–1412.

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. a., and Fritz, J. B. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron*, 82(2):486–499.

Bandyopadhyay, S., Shamma, S. A., and Kanold, P. O. (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nature Neuroscience*, 13(3):361–368.

Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nature Neuroscience*, 14(4):423–425.

Baumann, S., Joly, O., Rees, A., Petkov, C. I., Sun, L., Thiele, A., and Griffiths, T. D. (2015). The topography of frequency and time representation in primate auditory cortices. *eLife*, 2015(4):1–15.

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

Bendor, D. and Wang, X. (2008). Neural response properties of primary, rostral, and rostrotemporal

core fields in the auditory cortex of marmoset monkeys. *Journal of Neurophysiology*, 100(2):888–906.

Benvenuti, G., Chen, Y., Ramakrishnan, C., Deisseroth, K., Geisler, W. S., and Seidemann, E. (2018). Scale-Invariant Visual Capabilities Explained by Topographic Representations of Luminance and Texture in Primate V1. *Neuron*, 100(6):1504–1512.e4.

Berwick, J., Johnston, D., Jones, M., Martindale, J., Martin, C., Kennerley, A. J., Redgrave, P., and Mayhew, J. E. (2008). Fine detail of neurovascular coupling revealed by spatiotemporal analysis of the hemodynamic response to single whisker stimulation in rat barrel cortex. *Journal of Neurophysiology*, 99(2):787–798.

Bimbard, C., Demené, C., Girard, C., Radtke-Schuller, S., Shamma, S., Tanter, M., and Boubenec, Y. (2018). Multi-scale mapping along the auditory hierarchy using high-resolution functional UltraSound in the awake ferret. *eLife*, 7:1–14.

Binns, K. E., Grant, S., Withington, D. J., and Keating, M. J. (1992). A topographic representation of auditory space in the external nucleus of the inferior colliculus of the guinea-pig. *Brain Research*, 589(2):231–242.

Bizley, J. K. and King, A. J. (2008). Visual – auditory spatial processing in auditory cortical neurons. *Brain Research*, 1242:24–36.

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., and King, A. J. (2006). Physiological and Anatomical Evidence for Multisensory Interactions in Auditory Cortex. *Cerebral Cortex*, 17(9):2172–2189.

Bizley, J. K., Nodal, F. R., Nelken, I., and King, A. J. (2005). Functional organization of ferret auditory cortex. *Cerebral Cortex*, 15(10):1637–1653.

Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., and Schnupp, J. W. H. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(7):2064–75.

Blakemore, C. and Cooper, G. (1970). Development of the Brain depends on the Visual Environment. *Nature*, 228(1969):1969–1970.

Boido, D., Rungta, R. L., Osmanski, B. F., Roche, M., Tsurugizawa, T., Le Bihan, D., Ciobanu, L., and Charpak, S. (2019). Mesoscopic and microscopic imaging of sensory responses in the same animal. *Nature Communications*, 10(1):1–13.

Borchers, S., Himmelbach, M., Logothetis, N., and Karnath, H.-O. (2012). Direct electrical stimulation of human cortex - the gold standard for mapping brain functions? *Nature reviews. Neuroscience*, 13(1):63–70.

Boyce, S. W., Zingg, B. M., and Lightfoot, T. L. (2001). Behavior of Mustela Putorius Furo (The Domestic Ferret). *Veterinary Clinics of North America: Exotic Animal Practice*, 4(3):697–712.

Brewer, A. A. and Barton, B. (2016). Maps of the Auditory Cortex. *Annual Review of Neuroscience*, 39(1):385–407.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole) [Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech)]. *Bulletin de la Société Anatomique*, 6:330–357.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues von Dr. K. Brodmann, ...* J.A. Barth.

Buonomano, D. V. and Merzenich, M. M. (1998). Cortical plasticity: From Synapses to Maps. *Annual Review of Neuroscience*, 21(1):149–186.

Buzsáki, G., Geisler, C., Henze, D. A., and Wang, X. J. (2004). Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends in Neurosciences*, 27(4):186–193.

Castellucci, V., Pinsker, H., Kupfermann, I., and Kandel, E. R. (1970). Neuronal Mechanisms of Habituation and Dishabituation of the Gill-Withdrawal Reflex in Aplysia. *Science*, 167(3926):1745–1748.

Chen, G., Wang, F., Gore, J. C., and Roe, A. W. (2013). Layer-specific BOLD activation in awake monkey V1 revealed by ultra-high spatial resolution functional magnetic resonance imaging. *NeuroImage*, 64(1):147–155.

Chen, X., Leischner, U., Rochefort, N. L., Nelken, I., and Konnerth, A. (2011). Functional mapping of single spines in cortical neurons in vivo. *Nature*, 475(7357):501–505.

Cherniak, C. (1990). The Bounded Brain: Toward Quantitative Neuroanatomy. *Journal of Cognitive Neuroscience*, 2(1):58–68.

Cheung, M. M., Lau, C., Zhou, I. Y., Chan, K. C., Zhang, J. W., Fan, S. J., and Wu, E. X. (2012). High fidelity tonotopic mapping using swept source functional magnetic resonance imaging. *NeuroImage*, 61(4):978–986.

Chi, T., Ru, P., and Shamma, S. a. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, 118(2):887–906.

Chklovskii, D. B. and Koulakov, A. A. (2004). MAPS IN THE BRAIN: What Can We Learn from Them? *Annual Review of Neuroscience*, 27(1):369–392.

Cushing, H. (1909). A note upon the faradic stimulation of the postcentral gyrus in conscious patients. *Brain*, 32(1):44–53.

Dax, G. (1863). Observations tendant à prouver la coïncidence constante des dérangements de la parole avec une lésion de l'hémisphère gauche du cerveau. *Comptes rendus de l'Académie des Sciences*, 56:536.

de Cheveigné, A., Di Liberto, G. M., Arzounian, D., Wong, D. D., Hjortkjær, J., Fuglsang, S., and Parra, L. C. (2019). Multiway canonical correlation analysis of brain data. *NeuroImage*, 186(October 2018):728–740.

de Cheveigne, A. and Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, 171(2):331–339.

De Felipe, J., Alonso-Nanclares, L., and Arellano, J. (2002). Microstructure of the neocortex: Comparative aspects. *Journal of Neurocytology*, 31(2002):299–316.

De Martino, F., Moerel, M., Ugurbil, K., Goebel, R., Yacoub, E., and Formisano, E. (2015). Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences*, 112(52):16036–16041.

Dear, S. P., Fritz, J., Haresign, T., Ferragamo, M., and Simmons, J. A. (1993). Tonotopic and functional organization in the auditory cortex of the big brown bat, Eptesicus fuscus. *Journal of Neurophysiology*, 70(5):1988–2009.

Deffieux, T., Demene, C., Pernot, M., and Tanter, M. (2018). Functional ultrasound neuroimaging: a review of the preclinical and clinical state of the art. *Current Opinion in Neurobiology*, 50:128–135.

Dehaene, S. and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6):254–262.

Demene, C., Baranger, J., Bernal, M., Delanoe, C., Auvin, S., Biran, V., Alison, M., Mairesse, J., Harribaud, E., Pernot, M., Tanter, M., and Baud, O. (2017). Functional ultrasound imaging of brain activity in human newborns. *Science Translational Medicine*, 9(411).

Demené, C., Deffieux, T., Pernot, M., Osmanski, B. F., Biran, V., Gennisson, J. L., Sieu, L. A., Bergel, A., Franqui, S., Correas, J. M., Cohen, I., Baud, O., and Tanter, M. (2015). Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity. *IEEE Transactions on Medical Imaging*, 34(11):2271–2285.

Demené, C., Tiran, E., Sieu, L. A., Bergel, A., Gennisson, J. L., Pernot, M., Deffieux, T., Cohen, I., and Tanter, M. (2016). 4D microvascular imaging based on ultrafast Doppler tomography. *NeuroImage*, 127:472–483.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8):333–341.

Dizeux, A., Gesnik, M., Ahnine, H., Blaize, K., Arcizet, F., Picaud, S., Sahel, J. A., Deffieux, T., Pouget, P., and Tanter, M. (2019). Functional ultrasound imaging of the brain reveals propagation of task-related brain activity in behaving primates. *Nature Communications*, 10(1):1–9.

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

Erb, J., Armendariz, M., De Martino, F., Goebel, R., Vanduffel, W., and Formisano, E. (2018). Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex. *Cerebral Cortex*, 29(9):3636–3650.

Errico, C., Pierre, J., Pezet, S., Desailly, Y., Lenkei, Z., Couture, O., and Tanter, M. (2015). Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature*, 527(7579):499–502.

Fahey, P. G., Muhammad, T., Smith, C., Froudarakis, E., Cobos, E., Fu, J., Edgar, Y., Yatsenko, D., Sinz, F. H., Reimer, J., and Tolias, A. S. (2019). A global map of orientation tuning in mouse visual cortex. *BioRxiv*.

Fitch, W. T. (2011). Speech perception: A language-trained chimpanzee weighs in. *Current Biology*, 21(14):R543–R546.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience*, 6(11):1216–1223.

Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., and Shamma, S. a. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *i*, 13(8):1011–1019.

Froemke, R. C., Carcea, I., Barker, A. J., Yuan, K., Seybold, B. a., Martins, A. R. O., Zaika, N., Bernstein, H., Wachs, M., Levis, P. a., Polley, D. B., Merzenich, M. M., and Schreiner, C. E. (2013). Long-term modification of cortical synapses improves sensory perception. *Nature neuroscience*, 16(1):79–88.

Froemke, R. C., Merzenich, M. M., and Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450(7168):425–429.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6):568–573.

Gesnik, M., Blaize, K., Deffieux, T., Gennisson, J. L., Sahel, J. A., Fink, M., Picaud, S., and Tanter, M. (2017). 3D functional ultrasound imaging of the cerebral visual system in rodents. *NeuroImage*, 149(January):267–274.

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, 65(6):852–858.

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):983–1012.

Guo, W., Chambers, a. R., Darrow, K. N., Hancock, K. E., Shinn-Cunningham, B. G., and Polley, D. B. (2012). Robustness of Cortical Topography across Fields, Laminae, Anesthetic States, and Neurophysiological Signal Types. *Journal of Neuroscience*, 32(27):9159–9172.

Gămănuț, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., and Burkhalter, A. (2018). The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. *Neuron*, 97(3):698–715.e10.

Hackett, T. A., Rinaldi Barkat, T., O'Brien, B. M. J., Hensch, T. K., and Polley, D. B. (2011). Linking Topography to Tonotopy in the Mouse Auditory Thalamocortical Circuit. *Journal of Neuroscience*, 31(8):2983–2995.

Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 394(4):475–495.

Harel, N., Lin, J., Moeller, S., Ugurbil, K., and Yacoub, E. (2006). Combined imaging-histological study of cortical laminar specificity of fMRI signals. *NeuroImage*, 29(3):879–887.

Harrington, I. A., Stecker, G. C., Macpherson, E. A., and Middlebrooks, J. C. (2008). Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. *Hearing Research*, 240(1-2):22–41.

Harrison, R. V., Harel, N., Panesar, J., and Mount, R. J. (2002). Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 12(3):225–233.

Head, H. and Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, 34(2-3):102–254.

Heffner, H. E. and Heffner, R. S. (1984). Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science*, 226(4670):75–76.

Heimbauer, L. A., Beran, M. J., and Owren, M. J. (2011). A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Current Biology*, 21(14):1210–1214.

Higgins, N. C., McLaughlin, S. A., Rinne, T., and Stecker, G. C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. *Proceedings of the National Academy of Sciences of the United States of America*, 114(36):E7602–E7611.

Hong, H., Yamins, D. L. K., Majaj, N. J., and DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, 19(4):613–622.

Horton, J. G. and Adams, D. L. (2005). The cortical column: A structure without a function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):837–862.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.

Imbault, M., Chauvet, D., Gennisson, J. L., Capelle, L., and Tanter, M. (2017). Intraoperative Functional Ultrasound Imaging of Human Brain Activity. *Scientific Reports*, 7(1):1–7.

Inouye, T. (1909). Die sehstorungen bei schussverletzungen der kortikalen sehsphare. *Nach Beobachtungen an Verwundeten der letszten japanischen Kriege*.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1):35–39.

Jenkins, W. M. and Merzenich, M. M. (1984). Role of cat primary auditory cortex for sound-localization behavior. *Journal of Neurophysiology*, 52(5):819–847.

Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T., and Guíc-Robles, E. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of neurophysiology*, 63(1):82–104.

Jiang, X., Chevillet, M. A., Rauschecker, J. P., and Riesenhuber, M. (2018). Training Humans to Categorize Monkey Calls: Auditory Feature- and Category-Selective Neural Tuning Changes. *Neuron*, 98(2):405–416.e4.

Kaas, J. H. (1989). Why Does the Brain Have So Many Visual Areas? *Journal of Cognitive Neuroscience*, 1(2):121–135.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, 44(2):107–112.

Kaas, J. H. (2015). *Topographic Maps*, volume 24. Elsevier, second edi edition.

Kalatsky, V. A., Polley, D. B., Merzenich, M. M., Schreiner, C. E., and Stryker, M. P. (2005). Fine functional organization of auditory cortex revealed by Fourier optical imaging. *Proceedings of the National Academy of Sciences*, 102(37):13325–13330.

Kanold, P. O., Nelken, I., and Polley, D. B. (2014). Local versus global scales of organization in auditory cortex. *Trends in Neurosciences*, 37(9):502–510.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11):4302–4311.

Kaschube, M. (2014). Neural maps versus salt-and-pepper organization in visual cortex. *Current Opinion in Neurobiology*, 24(1):95–102.

Katz, L. C. and Crowley, J. C. (2002). Development of cortical circuits: Lessons from ocular dominance columns. *Nature Reviews Neuroscience*, 3(1):34–42.

Kilgard, M. P. and Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science (New York, N.Y.)*, 279(5357):1714–1718.

King, A. J. and Hutchings, M. E. (1987). Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: A map of auditory space. *Journal of Neurophysiology*, 57(2):596–624.

Klink, P. C., Dagnino, B., Gariel-Mathis, M. A., and Roelfsema, P. R. (2017). Distinct Feedforward and Feedback Effects of Microstimulation in Visual Cortex Reveal Neural Mechanisms of Texture Segregation. *Neuron*, 95(1):209–220.e3.

Knudsen, E., du Lac, S., and Esterly, S. D. (1987). Computational Maps In The Brain. *Annual Review of Neuroscience*, 10(1):41–65.

Knudsen, I. (1982). Auditory and Visual Maps of Space in the Optic Tectum of the Owl. *Journal of Neuroscience*, 2(9):1177–1194.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., Maciver, M. A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, 93(3):480–490.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4.

Kriengwatana, B., Escudero, P., and ten Cate, C. (2015). Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Frontiers in Psychology*, 5(OCT):1–13.

Kuchibhotla, K. V., Hindmarsh Sten, T., Papadoyannis, E. S., Elnozahy, S., Fogelson, K. A., Kumar, R., Boubenec, Y., Holland, P. C., Ostojic, S., and Froemke, R. C. (2019). Dissociating task

acquisition from expression during learning reveals latent knowledge. *Nature Communications*, 10(1).

Kuhl;, P. K. and Miller, J. D. (1975). Speech Perception by the Chinchilla : Voiced-Voiceless Distinction in Alveolar Plosive Consonants. *Science*, 621(October):69–72.

Langner, G., Dinse, H. R., and Godde, B. (2009). A map of periodicity orthogonal to frequency representation in the cat auditory cortex. *Frontiers in Integrative Neuroscience*, 3(October):1–7.

Larionow, W. (1899). Ueber die musikalischen Centren des Gehirns. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 76(11):608–625.

Leaver, A. M. and Rauschecker, J. P. (2016). Functional topography of human auditory cortex. *Journal of Neuroscience*, 36(4):1416–1428.

Lee, C. C. and Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neuroscience*, 14(1):108–116.

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., Fenno, L. E., Ramakrishnan, C., and Deisseroth, K. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, 465(7299):788–792.

Liu, R. C., Linden, J. F., and Schreiner, C. E. (2006). Improved cortical entrainment to infant communication calls in mothers compared with virgin mice. *European Journal of Neuroscience*, 23(11):3087–3097.

Livingstone, M. S., Vincent, J. L., Arcaro, M. J., Srihasam, K., Schade, P. F., and Savage, T. (2017). Development of the macaque face-patch system. *Nature communications*, 8:14897.

Logothetis, N. K., Augath, M., Murayama, Y., Rauch, A., Sultan, F., Goense, J., Oeltermann, A., and Merkle, H. (2010). The effects of electrical microstimulation on cortical signal propagation. *Nature neuroscience*, 13(10):1283–91.

Lokmane, L., Proville, R., Narboux-Nême, N., Györy, I., Keita, M., Mailhes, C., Léna, C., Gaspar, P., Grosschedl, R., and Garel, S. (2013). Sensory map transfer to the neocortex relies on pretarget ordering of thalamic axons. *Current Biology*, 23(9):810–816.

Lomber, S. G. and Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature neuroscience*, 11(5):609–616.

Macé, E., Montaldo, G., Cohen, I., Baulac, M., Fink, M., and Tanter, M. (2011). Functional ultrasound imaging of the brain. *Nature Methods*, 8(8):662–664.

Macé, É., Montaldo, G., Trenholm, S., Cowan, C., Brignall, A., Urban, A., and Roska, B. (2018). Whole-Brain Functional Ultrasound Imaging Reveals Brain Modules for Visuomotor Integration. *Neuron*, 100(5):1241–1251.e7.

Macé, E., Montaldo, G., Osmanski, B.-F., Cohen, I., Fink, M., and Tanter, M. (2013). Functional ultrasound imaging of the brain: theory and basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 60(3):492–506.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY, USA.

Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., Shane, J. C., McKnight, D. J., Yoshizawa, S., Kato, H. E., Ganguli, S., and Deisseroth, K. (2019). Cortical layer–specific critical dynamics triggering perception. *Science*, 5202:eaaw5202.

Mazengenya, P. and Bhika, R. (2017). The structure and function of the central nervous system and sense organs in the canon of medicine by Avicenna. *Archives of Iranian Medicine*, 20(1):67–70.

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7):293–299.

McKone, E., Crookes, K., Jeffery, L., and Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive Neuropsychology*, 29(1-2):174–212.

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, 123(2):899–909.

Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G., and Somers, D. C. (2015). Short-Term Memory for Space and Time Flexibly Recruit Complementary Sensory-Biased Frontal Lobe Attention Networks. *Neuron*, 87(4):882–892.

Michel, M. M., Chen, Y., Geisler, W. S., and Seidemann, E. (2013). An illusion predicted by V1 population activity implicates cortical topography in shape perception. *Nature Neuroscience*, 16(10):1477–1483.

Minderer, M., Brown, K. D., and Harvey, C. D. (2019). The Spatial Structure of Neural Encoding in Mouse Posterior Cortex during Navigation. *Neuron*, 102(1):232–248.e11.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6(C):414–417.

Moore, D. R., Semple, M. N., and Addison, P. D. (1983). Some acoustic properties of neurones in the ferret inferior colliculus. *Brain Research*, 269(1):69–82.

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4):408–434.

Mrsic-Flogel, T. D., Versnel, H., and King, A. J. (2006). Development of contralateral and ipsilateral frequency representations in ferret primary auditory cortex. *European Journal of Neuroscience*, 23(3):780–792.

Musall, S., Kaufman, M. T., Gluf, S., and Churchland, A. K. (2018). Movement-related activity dominates cortex during sensory-guided decision making. *BioRxiv*.

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. *Journal of neurophysiology*, 99(4):1928–1941.

Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., Schnupp, J. W. H., and King, A. J. (2004). Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. *Journal of neurophysiology*, 92(4):2574–88.

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, 88(6):1281–1296.

Norman-Haignere, S. V. and McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLOS Biology*, 16(12):1–46.

O'Herron, P., Chhatbar, P. Y., Levy, M., Shen, Z., Schramm, A. E., Lu, Z., and Kara, P. (2016). Neural correlates of single-vessel haemodynamic responses in vivo. *Nature*, pages 1–17.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., and Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *Journal of Neuroscience*, 26(50):13025–13036.

Op De Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G., and DiCarlo, J. J. (2008). A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cerebral Cortex*, 18(7):1676–1694.

Ortiz-Rios, M., Azevedo, F. A., Kuśmierek, P., Balla, D. Z., Munk, M. H., Keliris, G. A., Logothetis, N. K., and Rauschecker, J. P. (2017). Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. *Neuron*, 93(4):971–983.e4.

Osmanski, B.-F., Martin, C., Montaldo, G., Lanièce, P., Pain, F., Tanter, M., and Gurden, H. (2014a). Functional ultrasound imaging reveals different odor-evoked patterns of vascular activity in the main olfactory bulb and the anterior piriform cortex. *NeuroImage*, 95:176–184.

Osmanski, B.-F., Pezet, S., Ricobaraza, A., Lenkei, Z., and Tanter, M. (2014b). Functional ultrasound imaging of intrinsic connectivity in the living rat brain with high spatiotemporal resolution. *Nature Communications*, 5:5023.

Pallas, S. L., Roe, A. W., and Sur, M. (1990). Visual projections induced into the auditory pathway of ferrets. I. Novel inputs to primary auditory cortex (AI) from the LP/pulvinar comples and the topography of the MGN-AI projection. *Journal of Comparative Neurology*, 298(1):50–68.

Palmer, A. R. and King, A. J. (1982). The representation of auditory space in the mammalian superior colliculus. *Nature*, 299.

Panniello, M., King, A. J., Dahmen, J. C., and Walker, K. M. (2018). Local and global spatial organization of interaural level difference and frequency preferences in auditory cortex. *Cerebral Cortex*, 28(1):350–369.

Patel, G. H., Kaplan, D. M., and Snyder, L. H. (2014). Topographic organization in the brain: Searching for general principles. *Trends in Cognitive Sciences*, 18(7):351–363.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, 11(3):367–374.

Petkov, C. I., Kikuchi, Y., Milne, A. E., Mishkin, M., Rauschecker, J. P., and Logothetis, N. K. (2015). Different forms of effective connectivity in primate frontotemporal pathways. *Nat Commun*, 6(May 2014):6000.

Poeppel, D., Idsardi, W. J., and Van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1071–1086.

Polley, D. B., Steinberg, E., and Merzenich, M. M. (2006). Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences. *Journal of Neuroscience*, 26(18):4970–4982.

Provost, J., Papadacci, C., Demené, C., Gennisson, J. L., Tanter, M., and Pernot, M. (2015). 3-D ultrafast doppler imaging applied to the noninvasive mapping of blood vessels in Vivo. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 62(8):1467–1472.

Radtke-Schuller, S. (2018). Cyto- and Myeloarchitectural Brain Atlas of the Ferret (Mustela putorius) in MRI Aided Stereotaxic Coordinates. *Springer*.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.

Rauschecker, J. P. and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22):11800–11806.

Razak, K. A. (2011). Systematic representation of sound locations in the primary auditory cortex. *Journal of Neuroscience*, 31(39):13848–13859.

Read, H. L., Winer, J. A., and Schreiner, C. E. (2001). Modular organization of intrinsic connections associated with spectral tuning in cat auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 98(14):8042–8047.

Recanzone, G. H. (2008). Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *Journal of Neuroscience*, 28(49):13184–13193.

Reed, A., Riley, J., Carraway, R., Carrasco, A., Perez, C., Jakkamsetti, V., and Kilgard, M. P. (2011). Cortical Map Plasticity Improves Learning but Is Not Necessary for Improved Performance. *Neuron*, 70(1):121–131.

Reiter, S., Hülsdunk, P., Woo, T., Lauterbach, M. A., Eberle, J. S., Akay, L. A., Longo, A., Meier-Credo, J., Kretschmer, F., Langer, J. D., Kaschube, M., and Laurent, G. (2018). Elucidating the control and development of skin patterning in cuttlefish. *Nature*, 562(7727):361–366.

Rideau Batista Novais, A., Pham, H., Van de Looij, Y., Bernal, M., Mairesse, J., Zana-Taieb, E., Colella, M., Jarreau, P. H., Pansiot, J., Dumont, F., Sizonenko, S., Gressens, P., Charriaut-Marlangue, C., Tanter, M., Demené, C., Vaiman, D., and Baud, O. (2016). Transcriptomic regulations in oligodendroglial and microglial cells related to brain damage following fetal growth restriction. *Glia*, 64(12):2306–2320.

Rocca, J. (2003). *Galen on the brain: Anatomical knowledge and physiological speculation in the second century AD*. Brill.

Roe, A. W., Pallas, S. L., Hahm, J. O., and Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, 250(4982):818–820.

Romanski, L. M. and Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, 5(1):15–16.

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. *Nature neuroscience*, 2(12):1131–1136.

Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature Neuroscience*, 13(3):353–360.

Rubin, J. M., Bude, R. O., Carson, P. L., Bree, R. L., and Adler, R. S. (1994). Power Doppler US: a potentially useful alternative to mean frequency-based color Doppler US. *Radiology*, 190(3):853–856.

Rungta, R. L., Osmanski, B. F., Boido, D., Tanter, M., and Charpak, S. (2017). Light controls cerebral blood flow in naive animals. *Nature Communications*, 8(May 2016):1–9.

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., and Formisano, E. (2014). Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS Computational Biology*, 10(1).

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., and Formisano, E. (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences*, 114(18):4799–4804.

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

Sergent, J., Ohta, S., and MacDonald, B. (1992). Functional Neuroanatomy of Face and Object Processing. *Brain*, 115(1):15–36.

Sharma, J., Angelucci, A., and Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780):841–847.

Shepard, K. N., Chong, K. K., and Liu, R. C. (2016). Contrast Enhancement without Transient Map Expansion for Species-Specific Vocalizations in Core Auditory Cortex during Learning. *eNeuro*, 3(6).

Sieu, L.-A., Bergel, A., Tiran, E., Deffieux, T., Pernot, M., Gennisson, J.-L., Tanter, M., and Cohen, I. (2015). EEG and functional ultrasound imaging in mobile rats. *Nature Methods*, 12(9):831–834.

Silva, A. C. and Koretsky, A. P. (2002). Laminar specificity of functional MRI onset times during somatosensory stimulation in rat. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15182–7.

Singh, N. C. and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6):3394–3411.

Sperry, R. W. (1963). Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns and Connections. *Proceedings of the National Academy of Sciences*, 50(4):703–710.

Srihasam, K., Mandeville, J. B., Morocz, I. A., Sullivan, K. J., and Livingstone, M. S. (2012). Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron*, 73(3):608–619.

Srihasam, K., Vincent, J. L., and Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature neuroscience*, 17(12):1776–1783.

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

Stecker, G. C. and Middlebrooks, J. C. (2003). Distributed coding of sound locations in the auditory cortex. *Biological Cybernetics*, 89(5):341–349.

Stringer, C., Michaelos, M., and Pachitariu, M. (2019a). High precision coding in mouse visual cortex. *bioRxiv*, page 679324.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019b). High-dimensional geometry of population responses in visual cortex. *Nature*.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019c). Spontaneous behaviors drive multidimensional, brainwide activity. *Science (New York, N.Y.)*, 364(6437):255.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences*, 105(1):394–398.

Sur, M., Garuaghry, P. E., and Roe, A. W. (1988). Experimentally Induced Visual Projections. *Science*.

Tanter, M. and Fink, M. (2014). Ultrafast imaging in biomedical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(1):102–119.

Theunissen, F. E. and Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, 15(6):355–366.

Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience*, 20(6):2315–2331.

Thivierge, J. P. and Marcus, G. F. (2007). The topographic brain: from neural connectivity to cognition. *Trends in Neurosciences*, 30(6):251–259.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 292(5515):290–293.

Tiran, E., Ferrier, J., Deffieux, T., Gennisson, J. L., Pezet, S., Lenkei, Z., and Tanter, M. (2017). Transcranial Functional Ultrasound Imaging in Freely Moving Awake Mice and Anesthetized Young Rats without Contrast Agent. *Ultrasound in Medicine and Biology*, 43(8):1679–1689.

Tolias, A. S., Sultan, F., Augath, M., Oeltermann, A., Tehovnik, E. J., Schiller, P. H., and Logothetis, N. K. (2005). Mapping cortical activity elicited with electrical microstimulation using fMRI in the macaque. *Neuron*, 48(6):901–911.

Tootell, R., Switkes, E., Silverman, M., and Hamilton, S. (1988). Functional Anatomy Organization of Macaque Striate Cortex . *The Journal of Neuroscience*, 8(5):1531–1568.

Trout, J. D. (2003). Biological Specializations for Speech: What Can the Animals Tell Us? *Current Directions in Psychological Science*, 12(5):155–159.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9):989–995.

Urban, A., Dussaux, C., Martel, G., Brunner, C., Macé, E., and Montaldo, G. (2015). Real-time imaging of brain activity in freely moving rats using functional ultrasound. *Nature Methods*, 12(9):873–878.

Urban, A., Macé, E., Brunner, C., Heidmann, M., Rossier, J., and Montaldo, G. (2014). Chronic assessment of cerebral hemodynamics during rat forepaw electrical stimulation using functional ultrasound imaging. *NeuroImage*, 101:138–149.

Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, 360(6388):537–542.

Versnel, H., Mossop, J. E., Mrsic-Flogel, T. D., Ahmed, B., and Moore, D. R. (2017). Optical Imaging of Intrinsic Signals in Ferret Auditory Cortex: Responses to Narrowband Sound Stimuli. *Journal of Neurophysiology*, 88(3):1545–1558.

Walker, K. M., Gonzalez, R., Kang, J. Z., McDermott, J. H., and King, A. J. (2019). Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLife*, 8:1–20.

Watanabe, T. and Sasaki, Y. (2015). Perceptual Learning: Toward a Comprehensive Theory. *Annual Review of Psychology*, 66(1):197–221.

Weinberger, N. (1995). Dynamic Regulation of Receptive Fields and Maps in the Adult Sensory Cortex. *Annual Review of Neuroscience*, 18(1):129–158.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn.

Wilson, F. A., Scalaidhe, S. P., and Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260(5116):1955–1958.

Winkowski, D. E., Bandyopadhyay, S., Shamma, S. a., and Kanold, P. O. (2013). Frontal Cortex Activation Causes Rapid Plasticity of Auditory Cortical Processing. *Journal of Neuroscience*, 33(46):18134–18148.

Winkowski, D. E. and Kanold, P. O. (2013). Laminar Transformation of Frequency Organization in Auditory Cortex. *Journal of Neuroscience*, 33(4):1498–1508.

Wood, K. C., Town, S. M., Atilgan, H., Jones, G. P., and Bizley, J. K. (2017). Acute inactivation of primary auditory cortex causes a sound localisation deficit in ferrets. *PLoS ONE*, 12(1):1–26.

Wood, K. C., Town, S. M., and Bizley, J. K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. *Nature Communications*, 10(1):1–15.

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. *Journal of Neurophysiology*, 96(6):3323–3337.

Yao, J. D., Bremen, P., and Middlebrooks, J. C. (2013). Rat primary auditory cortex is tuned exclusively to the contralateral hemifield. *Journal of Neurophysiology*, 110(9):2140–2151.

York, G. K. and Steinberg, D. A. (2011). Hughlings Jackson's neurological ideas. *Brain*, 134(10):3106–3113.

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1):37–46.

Zhang, L. I., Bao, S., and Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience*, 4:1123–1130.

Zhuang, J., Ng, L., Williams, D., Valley, M., Li, Y., Garrett, M., and Waters, J. (2017). An extended retinotopic map of mouse cortex. *eLife*, 6:1–29.

RÉSUMÉ

Le monde extérieur regorge de sons complexes, que chaque animal doit interpréter afin de survivre. Pour ce faire, leur cerveau se doit de représenter toute la richesse de la structure acoustique de ces sons, jusque dans leurs propriétés les plus complexes. Dans cette thèse, cette question est explorée à travers un nouveau prisme, l'imagerie fonctionnelle ultrasonore (fUS). Dans un premier temps, l'imagerie fUS est utilisée pour étudier avec une haute fidélité l'organisation topographique du système auditif, ainsi que ses connexions avec d'autres aires cérébrales. Dans un deuxième temps, elle permet d'explorer des aspects fondamentaux de la façon dont le cortex auditif encode les sons naturels, ainsi que les spécificités humaines pour le traitement du langage. Enfin, elle révèle des formes topographiques mais non continues d'encodage, avec l'exemple de la localisation spatiale des sons. À travers ces trois aspects sont révélés les différents modules de traitement de l'information auditive, spatialement organisés, qui se superposent au sein d'une aire cérébrale unique.

MOTS CLÉS

Imagerie fonctionnelle UltraSonore; Traitement de l'information auditive; Topographie; Sons naturels; Furet

ABSTRACT

The world teems with complex sounds that animals have to interpret in order to survive. To do so, their brain must represent the richness of the sounds' acoustic structure, from simple to high-order features. Understanding how it does it, however, remains filled with challenges. In this thesis, these questions were explored through a new technical prism, namely functional UltraSound imaging (fUSi). First, fUSi was used to investigate with a high fidelity the topographical organization of the auditory system, as well as its connectivity with other brain areas. Second, it provided fundamental clues for our understanding of how natural sounds are encoded in the auditory cortex, and hints at the human particularities for speech processing. Last, it gave us access to non-continuous topographical encoding, with the example of spatial localization. Through these three aspects, we exposed the different spatially organized modules of processing that overlap within a single brain area.

KEYWORDS

Functional UltraSound imaging; Auditory processing; Topography; Natural sounds; Ferret