The computational architecture of the human auditory cortex

Chapter 5

Summary and Conclusions

Summary of the main findings

The present thesis investigated the sensory representation of natural sounds in the human auditory cortex. Specifically, the studies presented in the previous chapters were motivated by two general research questions. First, which acoustic features define the representational space of natural sounds in the human auditory cortex? Second, how do representations change across auditory cortical regions?

The study presented in Chapter 2 compared competing computational models of sound representation. The tested models described the processing of three relevant properties of natural sounds: frequency, temporal modulations and spectral modulations. Results showed that a model that represents spectral and temporal modulations *jointly* and in a *frequency-specific* fashion provides the best account of fMRI responses to natural sounds. Besides providing insights into the representation of natural sounds in the human auditory cortex, this finding implies that caution should be taken when interpreting results based on stimuli that vary along one dimension alone, as an interaction occurs among frequency, spectral and temporal modulations. Based on feature preference maps, Chapter 2 suggested the presence of a large-scale spatial topography of spectral and temporal modulation tuning and that the functional specialization of auditory cortical fields can be partially accounted for by their modulation tuning.

The spatial variation of spectral and temporal response properties throughout the auditory cortex was further investigated in Chapter 3. The main goal of this chapter was to characterize the contribution of distinct macro-anatomical auditory cortical regions to the encoding of natural sounds. We addressed this question by assessing which acoustic features could be accurately reconstructed from patterns of fMRI responses to natural sounds. Results showed that auditory regions on the superior temporal plane encode a broader range of acoustic features compared to auditory regions along the superior temporal gyrus, where high frequencies and fast temporal modulations are less faithfully represented. The analysis of the acoustic discriminability of incoming sounds revealed that already in primary regions - and progressively more in surrounding non-primary regions – the observed spectral and temporal tuning increases the acoustic distance among speech/voice sounds, but not among exemplars of other categories of natural sounds. These results support the hypothesis that auditory cortical representational mechanisms – including those at an early hierarchical level - are optimized for a subset of natural sounds that are behaviorally most relevant, namely speech/voice for human listeners.

Chapter 4 tackled the challenge – so far unexplored – of reconstructing arbitrary natural sounds based on patterns of fMRI responses. Results showed that a time-resolved computational model that represents sounds according to their time-varying spectro-temporal modulation content enables the detailed reconstruction of natural sound spectrograms (and waveforms) from fMRI cortical response patterns. These findings provide a direct proof of validity of the encoding hypothesis postulated in Chapter 2. Additionally, they open up the possibility to study covert auditory processing such as selective attention and imagery, which is pivotal to the development of fMRI-based communication systems and brain computer interfaces.

In conclusions, the main findings of this thesis provide insights on how natural sounds are encoded in human auditory cortex and the proposed methodological approach constitutes an advance in the way this question can be addressed in future studies.

Model-based analysis of fMRI responses: Encoding vs Decoding

In the present thesis we applied both encoding (Chapter 2) and decoding (Chapters 3 and 4) techniques. A discussion on general merits of decoding/encoding approaches can be found in Naselaris et al. (2010). Below we discuss their differences in relation to the specific research questions of the thesis.

Both encoding and decoding are well-suited for testing hypotheses of stimulus representation (Naselaris et al., 2010). Whereas encoding can be used to highlight tuning properties of individual voxels, decoding might be more appropriate when the goal is to characterize the representational content of a population of voxels or a cortical region (Mesgarani et al., 2009). In the encoding approach, a (multidimensional) tuning function is estimated for each voxel. This function represents the voxel's input-output relationship, i.e. the relative contribution of each stimulus feature to the voxel response. Due to its univariate (i.e. voxel-wise) nature, results from individual voxels must be integrated in order to obtain an input-output relationship for an entire cortical region. One approach is to generate cortical maps (i.e. spatial distributions) of the feature that elicits the highest response at each voxel (see Chapter 2). Conclusions based on this approach are inferred based on the assumption that stimulus features encoded with greater fidelity are those that maximally contribute to individual voxels responses and that are "preferred" by a larger number of voxels. However, the interpretation of tuning functions is still issue of debate - even at neuron's level - and higher responses might not necessarily mean better encoding (Butts and Goldman, 2006). In summary, possible limitations of the encoding approach are the following: 1) tuning functions from individual voxels must be integrated into a population input-output relationship; 2) results are provided in terms of average activation levels (tuning functions), while conclusions must be drawn in the stimulus domain (i.e. which features are best encoded); therefore, knowledge of the link between activation levels and amount of stimulus information that is encoded is necessary.

These issues may be tackled with a decoding approach, which reconstructs stimulus features based on patterns of voxels' activity (Chapter 3). First, within the decoding framework, data from individual voxels are jointly modeled. This allows a multivariate characterization of a cortical region, without the need of integrating results derived from the modeling of individual voxels. Furthermore, the combined analysis of signals from multiple voxels may increase the sensitivity for stimulus information that may be represented in patterns of activity, rather than in individual voxels. Second, the amount of information that a given cortical region conveys about a set of stimulus features is made explicit in the accuracy with which those features can be reconstructed. Thus, as opposed to the encoding approach, there is no need to translate results obtained in the activation (or tuning function) space into conclusions drawn in the stimulus space.

Future directions

The work presented in this thesis provides the basis for future research aiming at unraveling the representation of natural sounds in the human auditory system. The present thesis focused on a sound representation that captures cortical tuning only along three stimulus dimensions (frequency, spectral modulations and temporal modulations). A challenge for future studies will be to extend the methodological framework presented here to explore increasingly complex models of sound representation.

A natural future direction for the work presented in this thesis is to investigate the temporal dynamics of spectral and temporal tuning properties, as well as their flexibility with respect to contextual changes, training, or task demands. The study presented in Chapter 3 revealed a cortical filtering mechanism that is designed to facilitate the acoustic discriminability of speech samples, arguably the natural sound with the highest behavioral relevance for human listeners. How are these mechanisms shaped by experience? Does the cortical tuning change as a function of the behavioral relevance that particular sound classes hold for specialized listeners (e.g. musicians)? Such long-term cortical reorganization mechanisms are only one aspect of functional plasticity. Electrophysiological studies in animals have shown that tasks' properties can influence neuronal tuning over short temporal scales (e.g. Fritz et al., 2003; David et al., 2012; Yin et al., 2014). Similarly, neuroimaging studies in humans have revealed that task-specific demands lead to different processing of the same physical stimulus (e.g. Alain et al., 2001; Bonte et al., 2014). However, previous reports have described this differential processing in terms of changes in localized activation strength (e.g. Alain et al., 2001) or distributed activation patterns (e.g. Bonte et al., 2014), leaving unanswered the question of which tuning mechanisms underlie these observations. The methodological framework presented in this thesis may help bridge this gap by estimating the feature tuning of auditory cortical regions under different behavioral conditions. This will provide a "task's fingerprint", that will allow us to hear sounds the way they are "heard" by the auditory cortex during the execution of a given task. From a technological perspective, the outcome of these studies may inform the design of engineering devices aiming at reproducing human auditory perceptual abilities.

Recent advances in MRI technology allow recording fMRI signals with unprecedented sub-millimeter resolution, opening the possibility to zoom into the laminar structure of the human auditory cortex (Koopmans et al., 2010, 2011; Polimeni et al., 2010). In combination with the methods described in the present thesis, these technological developments will allow studying the sensory representation of natural sounds across cortical layers.

References

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) "What" and "where" in the human auditory system. Proc Natl Acad Sci U S A 98:12301–12306.

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014) Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J Neurosci 34:4548–4557.

Butts DA, Goldman MS (2006) Tuning Curves, Neuronal Variability, and Sensory Coding. PLoS Biol 4:e92.

David S V, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci U S A 109:2144–2149.

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

Koopmans PJ, Barth M, Norris DG (2010) Layer-specific BOLD activation in human V1. Hum Brain Mapp 31:1297–1304.

Koopmans PJ, Barth M, Orzada S, Norris DG (2011) Multi-echo fMRI of the cortical laminae in humans at 7 T. Neuroimage 56:1276–1285.

Mesgarani N, David S V, Fritz JB, Shamma S a (2009) Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J Neurophysiol 102:3329–3339.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2010) Encoding and decoding in fMRI. Neuroimage:1–11.

Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. Neuroimage 52:1334–1346.

Yin P, Fritz JB, Shamma SA (2014) Rapid spectrotemporal plasticity in primary auditory cortex during behavior. J Neurosci 34:4396–4408.