ABSTRACT

When we hear a sound, we perceive not the spectro-temporal patterns of the auditory signal but a piano playing, a name spoken, approaching steps. The ambiguous time-varying vibrations of the tympanic membrane reflect the entangled emissions of all concurrently active sources. Yet the resulting perception is of an orderly 'auditory scene' that is organized into ethologically-relevant 'auditory objects', separate entities related to things and events in the environment. The introspection is that we can perceive the existence of multiple auditory objects, as well as selectively attend to one object, mostly without any explicit effort. How the brain performs this task is only partially understood. That it can be done also without direct access to information regarding the active sources in the environment, is indication that auditory objects can be defined based on the auditory data stream.

In this work I suggest to bridge the gap between the sound signal and its perceptual organization by identifying auditory objects with predictive models. The basic idea is that natural sounds have distinctive underlying regular patterns that govern their evolution in time. The underlying process in auditory scene analysis is thus identified with regularity extraction from sensory input. By detecting regularities in the sensory data stream the brain can extrapolate from current sensory evidence into the future. Predictive models inferred from the statistics of the auditory input, tested against it and adjusted accordingly. They thus constitute the basis for an intrinsic definition of auditory objects, one based solely on the physical properties of the sound without additional perceptual criteria. Key to this process is the role of prediction errors. A mismatch between the predictions of a predictive model and actual auditory input can trigger an update of the perceptual organization, possibly inducing a new auditory object in the scene. The predictive formulation of auditory objects thus supports a

dynamic representation in a changing acoustic environment and is universal enough to be applied in varied acoustic scenarios.

Whereas the general mechanisms of regularity extraction are probably ubiquitous in the auditory system, I suggest the predictive formulation is predominantly relevant to primary auditory cortex (A1). In the first study I analyzed single unit responses in human cortex to an artificial sound with random statistics and to a natural sound stream, which included speech and music. Whereas in the artificial context the neurons responded as particularly narrow spectral filters, cortical responses to the natural sound scenes were non-trivially related to the physical dimensions of the sound. The results suggested non-linear mechanisms participated in shaping the neuronal responses to real world sound scenes, possibly reflecting an additional functional role of A1 in such auditory context. The second study suggested a specific link between cortical responses and object identification. In the analysis of MEG extracranial responses to rudimentary auditory scenes that comprised noises and pure tones I identified the 'object potential'- a response not correlated with change detection, but elicited only when a tone was heard as a new separate object in the presence of noise. The neural response was thus understood by considering the composition of complex sounds in terms of objects.

Primary auditory cortex (A1) has been suggested to play a central role in auditory scene analysis, and has also been implicated in the encoding of regularity violations in sounds. The predictive framework, suggested here as the appropriate tool to investigate responses of cortical neurons, is a unified description of these two functional roles. Specifically, I proposed the prediction error marks the outlines of auditory objects by significantly modulating A1 responses. To test this suggestion, I constructed a general procedure based on the Information Bottleneck method to generate predictive models of natural sounds directly from the statistics of the waveforms. The prediction errors of the models were then tested against experimental results obtained with the same sounds.

First I implemented the procedure with a set of sounds used in the lab of D. Pressnitzer to study timbre recognition, specifically comparing performance with human voices and musical instruments. The psychoacoustic results suggested a special behavioral status of the voices, which were recognized faster and more accurately even when sound segments were very short (~10 ms). Prediction errors of models fitted to the entire ensemble were likewise sensitive to the timbre categories - errors were consistently larger for voices, with differences also found for very short sound samples. Importantly, the calculation of the prediction error was shown to be equivalent to spectral analysis of the sound. This suggests the predictive framework as a principled method for highlighting spectral differences between sound sets. Alternatively, the spectral representation of sounds, ubiquitous in the auditory system, can be considered as a way to estimate the average prediction error. Prediction error thus corresponded in this study to the behavioral results and the physical properties of the sounds, as well as pointed to a possible link between the two.

.

Building upon the feasibility of the predictive procedure in an experimental scenario, the last study directly tested the hypothesis that A1 responses are affected by the extent to which a current sound fulfills or violates the regularities detected in earlier acoustic context. Local field potentials (LFP) and multi-unit activity (MUA) from rat auditory cortex were recorded during the presentation of Ligeti's *Musica Ricercata II* - a musical piece for piano with an apparent pitch structure and simple statistics. As could be expected, note level had a prominent effect on the neuronal response - the loud notes in the music elicited most responses. Additionally, a prediction error calculated for each piano note (in two different manners) was positively correlated with neural responses. Notably, combining the level of the note with the prediction error in a single measure yielded maximal positive correlation with the neuronal response, more than any of the measures separately. This finding provided strong support for the claim that A1 neurons encode prediction error in addition to the encoding of the spectro-temporal features of the sound. Under the predictive formulation suggested here this result can be interpreted in the wider functional context of object-based encoding of sounds in A1.

This dissertation used a data-driven approach to the question of predictive auditory objects. Results highlight the prediction error as a possible link between the theoretical formulation and experimental results, including neuronal responses and even behavior. However, the experiments did not explicitly test the hypothesis that the core of ASA is the processing of prediction errors. The findings justify an effort to experimentally test for a more concrete link between auditory scene analysis and predictive modeling.

CONTENTS