Information Theoretic Approach to the Study of Auditory Coding

Gal Chechik

Under the supervision of Prof. Naftali Tishby and Dr. Israel Nelken

This dissertation develops information theoretic tools to study properties of the neural code used by the auditory system, and applies them to electro-physiological recordings in three auditory processing stations: auditory cortex (AI), thalamus (MGB) and inferior colliculus (IC). It focuses on several aspects of the neural code: First, robust estimation of the information carried by spike trains is developed, using a variety of dimensionality reduction techniques Secondly, measures of informational redundancy in small groups of neurons are developed. These are applied to neural activity in a series of brain regions, demonstrating a process of redundancy reduction in the ascending processing pathway. Finally, a method to identify relevant features, by filtering out the effects of lower processing stations is developed. This approach is shown to have numerous applications in domains extending far beyond neural coding. These three components are summarized below.

The problem of the *neural code* of sensory systems combines two interdependent tasks. The First is identifying the *code words*: i.e., the components of neural activities from which a model of the outside world can be inferred. Common suggestions for these components are the firing rates of single neurons, correlated and synchronized firing in groups of neurons, or specific temporal firing patterns across groups of neurons. The second task is to identify the components of the sensory inputs about which neurons in various brain regions carry information. In the auditory realm, these can be "physical" properties of acoustic stimuli, such as frequency content or intensity level, or more abstract properties such as pitch or even semantic content of spoken words.

We first address the first task, that of identifying code words that are informative about a predefined set of natural and complex acoustic stimuli. The difficulty of this problem is due to the huge dimension of both neural responses and the acoustic stimuli, which forces us to develop techniques to reduce the dimensionality of spike trains, while losing as little information as possible. Great care is taken to develop and use unbiased and robust estimators of the mutual information. Six different dimensionality reduction approaches are tested, and the level of information they achieve is compared. The findings show that the maximal information can almost always be extracted by considering the distribution of temporal patterns of spikes. Surprisingly, the first spike latency carries almost the same level of information. In contrast, spike counts convey only half of the maximal information level. In all of these methods, IC neurons conveyed about twice more information about the identity of the presented stimulus than AI and MGB neurons.

To study neural coding strategies we further examine how small groups of neurons interact to code auditory stimuli. For this purpose we develop measures of informational redundancy in groups of cells, and describe their properties. These measures can be

reliably estimated in practice from empirical data using stimulus conditioned independence approximation. Since redundancy is biased by the baseline single-unit information level, we study this effect and show how it can be reduced with proper normalization. Finally, redundancy biases due to ceiling effect on maximal information are discussed.

The developed measures of redundancy are then applied to quantify redundancy in processing stations of the auditory pathway. Pairs and triplets of neurons in the lower processing station, the IC, are found to be considerably more redundant than those in MGB and AI. This demonstrates a process of redundancy reduction along the ascending auditory pathway, and puts forward *redundancy reduction as a potential generic organization principle for sensory systems*. Such a process was hypothesized 40 years ago by Barlow based on a computational motivation and is experimentally demonstrated here for the first time.

We further show that the redundancies in IC are correlated with the frequency characterization of the cells; namely, redundant pairs tend to share a similar best-frequency. This effect is much weaker in MGB and AI, suggesting that even the low redundancy in these stations is not due to similar frequency sensitivity. This result has great significance for the study of auditory coding since it cannot be explained by the standard model for cortical responses, the spectro-temporal receptive field (STRF). Finally, we measured informational redundancy in the information that single spikes convey about the spectro-temporal structure. Redundancy reduction is observed here as well. Moreover, IC cells convey an order of magnitude more information about these spectro-temporal structures than MGB and AI neurons. Since AI neurons convey half the information that IC neurons do about stimulus identity we conclude that cortical neurons code the identity of the stimuli well without characterizing their "physical" aspects. This observation hints that the cortex is sensitive to complex structures in our stimulus set, which cannot be identified with the common parametric stimuli.

In the last part of the work, we address the second task in neural coding identification. Here the goal is to develop methods to which characterize the stimulus features that cortical neurons are sensitive. One difficulty is that many features that cortical neurons are informative about result from processing at lower brain regions. For example, AI neurons are sensitive to frequency content and level, but these properties are already computed at the cochlea. This is in fact a special case of a fundamental problem in unsupervised learning. The task of identifying relevant structures in data in an unsupervised manner is ill defined since real world data often contain conflicting structures. Which of them is relevant depends on the task. For example, documents can be clustered according to content or style; speech can be classified according to the speaker or the meaning.

We provide a formal solution to this problem in the framework of the *information bottleneck* (IB) method. In IB, a source variable is compressed while preserving information about another relevance variable. We extend this approach to use side information in the form of additional data, and the task is now to compress the source (e.g. the stimuli) while preserving information about the relevance variable (e.g. cortical responses), but removing information about the side variable (e.g. 8th nerve responses). The irrelevant structures are therefore implicitly and automatically learned from the side data. We present a formal definition of the problem, as well as its analytic and algorithmic solutions. We show how this approach can be used in a variety of domains in addition to auditory processing.