

Research paper

Information theory in auditory research

Israel Nelken^{a,*}, Gal Chechik^b

^a Department of Neurobiology and The Interdisciplinary Center for Neural Computation, Silberman Institute of Life Sciences, Safra Campus, Givat Ram, Hebrew University, Jerusalem 91904, Israel

^b Computer Science Department, Stanford University, Stanford CA 94305, USA

Received 10 September 2006; received in revised form 22 November 2006; accepted 3 January 2007

Available online 16 January 2007

Abstract

Mutual information (MI) is in increasing use as a way of quantifying neural responses. However, it is still considered with some doubts by many researchers, because it is not always clear what MI really measures, and because MI is hard to calculate in practice. This paper aims to clarify these issues. First, it provides an interpretation of mutual information as variability decomposition, similar to standard variance decomposition routinely used in statistical evaluations of neural data, except that the measure of variability is entropy rather than variance. Second, it discusses those aspects of the MI that makes its calculation difficult. The goal of this paper is to clarify when and how information theory can be used informatively and reliably in auditory neuroscience.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Auditory system; Information theory; Entropy; Mutual information; Variance decomposition; Neural code

1. Introduction

In recent years, information-theoretic measures are increasingly used in neuroscience in general, and in auditory research in particular, as tools for studying and quantifying neural activity. Measures such as entropy and mutual information (MI) can be used to gain deep insight into neural coding, but can also be badly abused. This paper is an attempt to present those theoretical and practical issues that we found particularly pertinent when using information-theoretic measures in analyzing neural data.

The experimental context for this paper is that of measuring a stimulus–response relationship. In a typical experiment, a relatively small number of stimuli ($\sim < 100$) are presented repeatedly, typically 1–100 repeats for each stimulus. The main experimental question is whether the neuronal activity was different in response to the different stimuli. If so, it is concluded that the signal whose activity is monitored (single-neuron responses, evoked potentials, optical

signals, and so on) was selective to the parameter manipulated in the experiment.

The MI is a measure of the strength of association between two random variables. The MI, $I(S; R)$, between the stimuli S and the neural responses R is defined in terms of their joint distribution $p(S, R)$. When this distribution is known exactly, the MI can be calculated as

$$I(S; R) = \sum_{s \in S, r \in R} p(s, r) \log_2 \left(\frac{p(s, r)}{p(s)p(r)} \right)$$

where $p(s) = \sum_{r \in R} p(s, r)$ and $p(r) = \sum_{s \in S} p(s, r)$ are the marginal distributions over the stimuli and responses, respectively.

The easy way to use the MI is to test for significant association between the two variables. Here the null hypothesis is that the two variables are independent. The distribution of the MI under the null hypothesis is (with appropriate scaling) that of a χ^2 variable, leading to a significance test for the presence of association (e.g. Sokal and Rohlf, 1981; where it is called the G -statistic). Using the MI in this way, only its size relative to the critical value of the test is of importance.

* Corresponding author. Tel.: +972 2 6584229; fax: +972 2 6586077.
E-mail address: israel@cc.huji.ac.il (I. Nelken).

A more complicated way of using the MI is to try to estimate its actual value, in which case it is possible to make substantially deeper inferences regarding the relationships between the two variables. This estimation is substantially more difficult than performing the significance test. The reasons to undertake this hard estimation problem, and the associated difficulties, are the main subject of this paper.

2. Why mutual information?

2.1. The Mutual Information as a measure of stimulus effect

Neuronal responses are high-dimensional: to fully characterize in detail any single spiking response to a stimulus presentation, it is necessary to specify many values, such as the number of spikes that occurred during the relevant response window and their precise times. Similarly, membrane potential fluctuates at >1000 Hz, and therefore more than 200 measurements are required to fully specify a 100 ms response. We usually believe that most of the details in such representations are unimportant, and instead of specifying all of these values, typically a single value is used to summarize single responses – for example, the total spike count during the response window, or first spike latency, or other such simple measures, that will be called later ‘reduced measures’ of the actual response.

Having reduced the representation of the responses to a single value, it is now possible to test whether the stimuli had an effect on the responses. Usually, the effect that is tested is a dependence of the firing rate of the neuron on stimulus parameters. For example, to demonstrate frequency selectivity, we will look for changes in firing rates of a neuron as a function of tone frequency.

To understand what information-theoretic measures tell us about neuronal responses, let us consider the standard methods for performing such tests in detail. A test for a significant difference between means is really about comparing variances (Fig. 1): the variation between response means has to be large enough with respect to variation between responses to repeated presentations of the same stimulus.

Initially, all the responses to all stimuli are pooled together, and the overall variability is estimated by the variance of this set of values around its grand mean. Fig. 1 shows the analysis of artificial data that represents 20 repeats of each of two stimuli (these are actually samples of two Poisson distributions with expected values of 5 and 10). In Fig. 1a, the overall distribution of all responses (both of stimulus 1 and of stimulus 2) is presented. The total variance is 10.9 (there are no units, since these are spike counts), corresponding to a standard deviation of about 3 spikes.

Part of the overall variation occurs because responses to repeated presentation of the same stimulus are noisy – this is called *within-stimulus* variability. Another part of this overall variation is due to the fact that different stimuli cause different responses. A stimulus effect is significant if

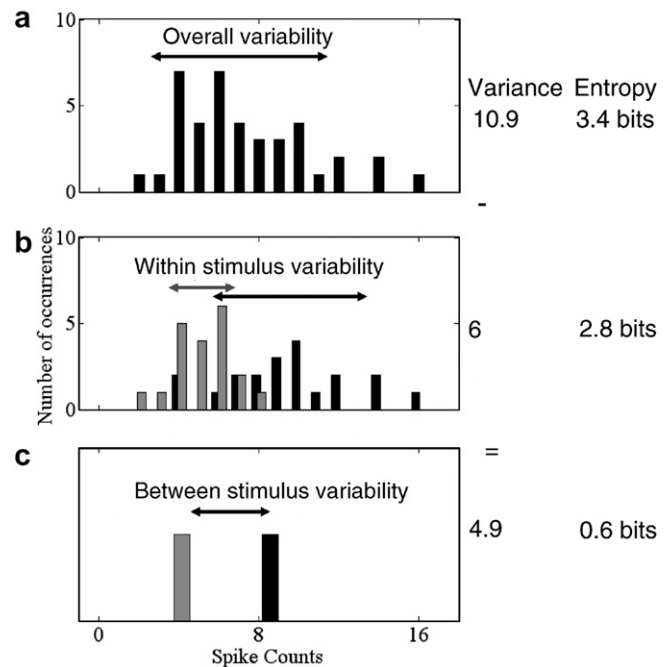


Fig. 1. Variability analysis using variances and entropies. (a) Overall distribution of 40 measurements (2 stimuli, 20 repeats of each stimulus). (b) Distribution of the responses to the two stimuli. (c) Samples means.

the second variability source, *between-stimulus* variation, is large enough relative to the first variability source, the within-stimulus variability. Conceptually, the next step is to compute within-stimulus variability. To do that, the variance of all responses to each stimulus, around their own mean, is computed (Fig. 1b). The two histograms represent the responses to stimulus 1 (black) and stimulus 2 (gray), with variances of 10.1 and 2 (standard deviations of about 3 and 1.5 spikes). This set of variances is then averaged across stimuli, and used as an estimate of the within-stimulus variability – for the data in Fig. 1, the average within-stimulus variance is about 6.

It can be shown mathematically that within-stimulus variability will always be smaller than the overall variance, and the difference between them is the variability between the means of the responses to the different stimuli (Fig. 1c). Thus, the goal of dividing variance into two sources, the within-stimulus variance and the across-stimulus variance, is achieved.

This decomposition has good statistical properties, in the sense that the two variability sources are uncorrelated. Statistical theory can now be used to determine when the ratio between the two variability sources should be considered as larger than expected under the assumption of no stimulus effect (Sokal and Rohlf, 1981), leading to specific statistical tests (e.g. the *F*-test of the 1-way ANOVA) in Fig. 1 the *F*-test (or the equivalent *t*-test for equality of means, which is essentially the same thing here) comes out highly significant.

The recipe given above is extremely powerful, and therefore unsurprisingly is extensively used. However, it has

Download English Version:

<https://daneshyari.com/en/article/4356265>

Download Persian Version:

<https://daneshyari.com/article/4356265>

[Daneshyari.com](https://daneshyari.com)