Group Redundancy Measures Reveal Redundancy Reduction in the Auditory Pathway

Gal Chechik Amir Globerson Naftali Tishby
Institute of Computer Science and Engineering
and The Interdisciplinary Center for Neural Computation
Hebrew University of Jerusalem, Israel
ggal@cs.huji.ac.il

Michael J. Anderson Eric D. Young
Department of Biomedical Engineering
Johns Hopkins University, Baltimore, MD, USA

Israel Nelken
Department of Physiology, Hadassah Medical School
and The Interdisciplinary Center for Neural Computation
Hebrew University of Jerusalem, Israel

Abstract

The way groups of auditory neurons interact to code acoustic information is investigated using an information theoretic approach. Identifying the case of stimulus-conditioned independent neurons, we develop redundancy measures that allow enhanced information estimation for groups of neurons. These measures are then applied to study the collaborative coding efficiency in two processing stations in the auditory pathway: the inferior colliculus (IC) and the primary auditory cortex (A1). Under two different coding paradigms we show differences in both information content and group redundancies between IC and cortical auditory neurons. These results provide for the first time a direct evidence for redundancy reduction along the ascending auditory pathway, as has been hypothesized by Barlow (1959). The redundancy effects under the single-spikes coding paradigm are significant only for groups larger than ten cells, and cannot be revealed with the standard redundancy measures that use only pairs of cells. Our results suggest that redundancy reduction transformations are not limited to low level sensory processing (aimed to reduce redundancy in input statistics) but are applied even at cortical sensory stations.

1 Introduction

How do groups of sensory neurons interact to code information and how do these interactions change along the ascending sensory pathways ? One view is that sensory systems are composed of a series of processing stations, representing more and more complex aspects of sensory inputs. Among the computational principles proposed to govern the mapping between these stations are the transformation to sparse and distributed representations, information maximization and redundancy reduction [1, 7]. These ideas provide a computational framework to investigate changes in the neural code along the sensory hierarchy. In order to investigate such changes in practice, one has to develop methods to quantify interactions among groups of neurons, and compare these measures in various processing stations of a sensory system.

Interactions and high order correlations between neurons were mostly investigated within single brain areas on the level of pairs of cells, showing both synergistic and redundant interactions [5, 6, 4]. The current study focuses on developing redundancy measures for larger groups of neurons and comparing these measures in different processing stations.

To this end we use an information theoretic approach and identify a case for which redundancy may be measured more reliably even for groups of neurons. We then apply our measures to electro-physiological recordings from two auditory stations: the inferior colliculus and the primary auditory cortex, and use complex acoustic stimuli that are critical for the investigation of auditory cortical coding [8].

2 Redundancy measures for groups of neurons

To investigate high order correlations and interactions within groups of neurons we start by defining information measures for groups of cells and then develop information redundancy measures for such groups. The properties of these measures are then further discussed for the specific case of stimulus-conditioned independence.

Formally, high order correlations between two variables X and Y are quantified in terms of their *mutual information* (MI) [11, 3]. This well known quantity, now widely used in analysis of neural data, is defined by

$$I(X;Y) = D_{KL}[P(X,Y)||P(X)P(Y)] = \sum_{x,y} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{1}$$

and measures how close the joint distribution $P(X,Y)$ is to the factorization by the marginal distributions $P(X)P(Y)$ (D_{KL} is the Kullback Leiber divergence [3]).

For larger groups of cells, an important generalized measure quantifies the information that several variables provide about each other. This *multi information* measure is defined as

$$\begin{aligned} I(X_1,...,X_n) &= D_{KL}[P(X_1,...,X_n)||P(X_1)...P(X_n)] = \\ &= \sum_{x_1,...,x_n} p(x_1,...,x_n) log\left(\frac{p(x_1,...,x_n)}{p(x_1)...p(x_n)}\right) \quad . \end{aligned} \tag{2}$$

Similar to the *mutual information* case, the *multi information* measures how close the joint distribution is to the factorization by the marginals, and is always positive.

We now turn to develop measures for group redundancies. Consider first the simple case of a pair of neurons (X_1, X_2) conveying information about the stimulus S, for which the redundancy-synergy index ([2, 4]) is defined by

$$RS_{pairs}(X_1, X_2, S) = I(X_1, X_2; S) - [I(X_1; S) + I(X_2; S)] \tag{3}$$

Intuitively, RS_{pairs} measures the amount of information on the stimulus S gained by observing the joint distribution of both X_1 and X_2, as compared with observing the two cells independently. In the extreme case where $X_1 = X_2$, the two cells are completely redundant and provide the same information about the stimulus, yielding $RS_{pairs} = I(X_1, X_2; S) - I(X_1; S) - I(X_2; S) = -I(X_1; S)$, which is always non-positive. On the other hand, positive RS_{pairs} values testify for synergistic interaction between X_1 and X_2 ([2, 4]).

For larger groups of neurons, two different measures of redundancy-synergy should be considered. The first measure quantifies the residual information obtained from a group of N neurons compared to all its $N-1$ subgroups. As with inclusion-exclusion calculations this measure takes the form of a telescopic sum

$$RS_{N|N-1} = I(X^N; S) - \sum_{\{X^{N-1}\}} I(X^{N-1}; S) + ... + (-1)^{N-1} \sum_{\{X_i\}} I(X_i; S) \tag{4}$$

where $\{X^k\}$ are all the subgroups of size k out of the N available neurons . This measure may obtain positive (synergistic), vanishing (independent), or negative (redundant) values. Unfortunately, it involves 2^N information terms, making its calculation unfeasible even for moderate N values [1].

A different RS measure quantifies the information embodied in the joint distribution of N neurons compared to that provided by N single independent neurons, and is defined by

$$RS_{N|1} = I(X_1, ..., X_N; S) - \sum_{i=1}^{N} I(X_i; S) \tag{5}$$

Interestingly, this synergy-redundancy measure may be rewritten as the difference between two multi-information terms

$$\begin{aligned} RS_{N|1} &= I(X_1, ..., X_N; S) - \sum_{i=1}^{N} I(X_i; S) = \\ &= H(X_1, ..., X_N) - H(X_1, ..., X_N|S) - \sum_{i=1}^{N} H(X_i) - H(X_i|S) = \\ &= I(X_1; ...; X_N|S) - I(X_1; ...; X_N) \end{aligned} \tag{6}$$

where $H(X) = -\sum_x p(x) log(p(x))$ is the entropy of X [2]. We conclude that the index $RS_{N|1}$ can be separated into two terms: one that is always non-negative, and measures the coding synergy, and the second which is always non-positive and quantifies the redundancy among the group of neurons. The current work focuses on the latter redundancy term $I(X_1; ...; X_N)$.

[1] Our results below suggest that some redundancy effects become significant only for groups larger than 10-15 cells.

[2] When comparing redundancy in different processing stations, one must consider the effects of the baseline information conveyed by each set of neurons. We thus use the normalized redundancy (compare with [10] p.315 and [2]) defined by $RS_{N|1} = \left(I(X_1, ..., X_N; S) - \sum_{i=1}^{N} I(X_i; S)\right) / I(X_1; ...; X_N; S)$.

The formulation of $RS_{N|1}$ in equation 6 proves highly useful in the case where neural activities are independent given the stimulus $P(\vec{X}|S) = \Pi_{i=1}^{N} P(X_i|S)$. In such scenario, the first (synergy) term vanishes, thus limiting neural interactions to the redundant regime. More importantly, under the independence assumption we only have to estimate the marginal distributions $P(X_i|S = s)$ for each stimulus s instead of the full distribution $P(\vec{X}|S = s)$. It thus allows to estimate an exponentially smaller number of parameters, which in our case of small sample sizes, provides more accurate information estimates. This approximation hence allows to investigate redundancy among considerably larger groups of neurons than the two-three neurons considered in the literature.

How reasonable is the conditional-independence approximation ? It is fully accurate in the case of non-simultaneous recordings which is indeed the case in our data. The approximation is also reasonable whenever neuronal activity is mostly determined by the presented stimulus and to a lesser extent by interactions with nearby neurons, but the experimental evidence in this regard is mixed (see e.g.[6]). To summarize, the stimulus-conditioned independence assumption limits us to interactions in the redundant regime, but allows us to compare the extent of redundancy among large groups of cells in different brain areas.

3 Experimental Methods

To investigate redundancy in the auditory pathway, we analyze extracellular recordings from two brain areas of Halothane anesthetized cats: 16 cells from the *Inferior Colliculus* (IC) - the third processing station of the ascending auditory pathway - and 19 cells from the *Primary Auditory Cortex* (A1) - the fifth station. Neural activity was recorded non-simultaneously from a total of 6 different animals responding to a fixed set of stimuli. Because cortical auditory neurons respond considerably different to simplified and complex stimuli [8], we refrain from using artificial over-simplified acoustic stimuli but instead use a set of stimuli based on bird vocalizations which contains complex 'real-life' acoustic features. A representative example is shown in figure 1.

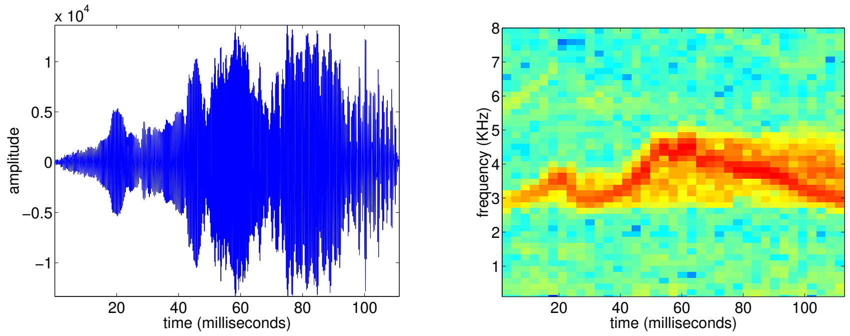


Figure 1: A representative stimulus containing a short bird vocalization recorded in a natural environment. The set of stimuli consisted of similar natural and modified recordings. **A.** Signal in time domain **B.** Signal in frequency domain.

4 Experimental Results

In order to estimate in practice the information content of neural activity, one must assume some representation of neural activity and sensory inputs. In this paper we consider two extreme cases: coding acoustics with single spikes and coding the stimulus identity with spike counts.

4.1 Coding acoustics with single spikes

The current section focuses on the relation between single spikes and short windows of the acoustic stimuli shortly preceding them (which we denote as *frames*). As the set of possible frames is very large and no frame actually repeats itself, we must first pre-process the stimuli to reduce frames dimensionality.

To this end, we first transform the stimuli into the frequency domain (roughly approximating the cochlear transformation) and then extract overlapping windows of 50 millisecond length, with 1 millisecond spacing. This set is clustered into 32 representatives, which capture different acoustic features in our stimulus set. This representation allows us to estimate the joint distribution (under the stimulus-conditioned independence assumption) of cells activity and stimuli, for groups of cells of different sizes. Figure 2 shows the incremental mutual information as a function of number of cells for both A1 and IC neurons, compared with the average information conveyed by single cells. The difference between these two lines measures the redundancy $RS_{N|1}$ of equation 6. The information conveyed by IC neurons saturates already with 15 cells, exhibiting significant redundancy for groups larger than 10 cells. More importantly, single A1 neurons provide less information (note the y-axis scale) but their information sums almost linearly.

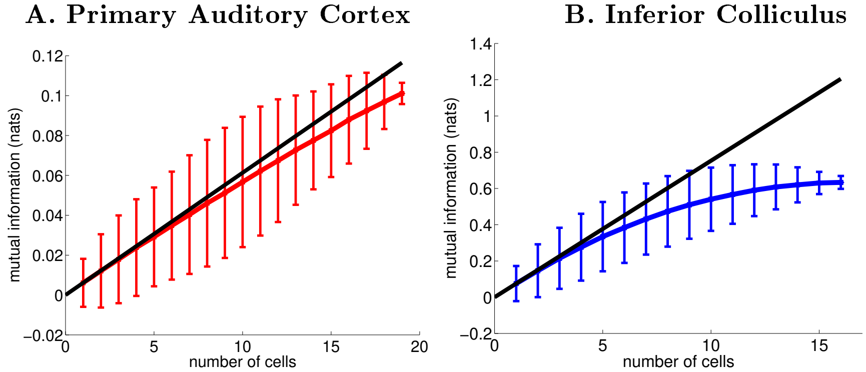


Figure 2: Information about stimulus frames as a function of group size. Information calculation was repeated for several subgroups of each size, and with several random seed initializations. The dark linear curve depicts the average information provided by neurons treated independently $\frac{k}{N}\sum_i I(X_i)$ while the curved line depicts average information from joint distribution of sets of neurons $Mean[I(X_1, ...X_k; S)]$. All information estimations were corrected for small-samples bias by shuffling methods [9].

4.2 Coding stimuli by spike counts

We now turn to investigate a second coding paradigm, and calculate the information conveyed by A1 and IC **spike counts** about the **identity of the presented stimulus**. To this end, we calculate a histogram of spike counts and estimate the counts' distribution as obtained from repeated presentations of the stimuli.

Figure 3A depicts MI distribution obtained from single cells and figure 3B presents the distribution of redundancy values among all cell-pairs (equation 3) in IC and A1. As in the case of coding with single spikes, single A1 cells convey on average less information about the stimulus. However, they are also more independent, thus allowing to gain more information from groups of neurons. IC neurons on the other hand, provide more information when considered separately but are more redundant.

To illustrate the high information provided by both sets, we trained a neural network classifier that predicts the identity of the presented stimulus according to spike counts of a limited set of neurons. Figure 4 shows that both sets of neurons achieve considerable prediction accuracy, but IC neurons obtain average accuracy of more than 90 percent already with five cells, while the average prediction accuracy using cortical neurons rises continuously [3].

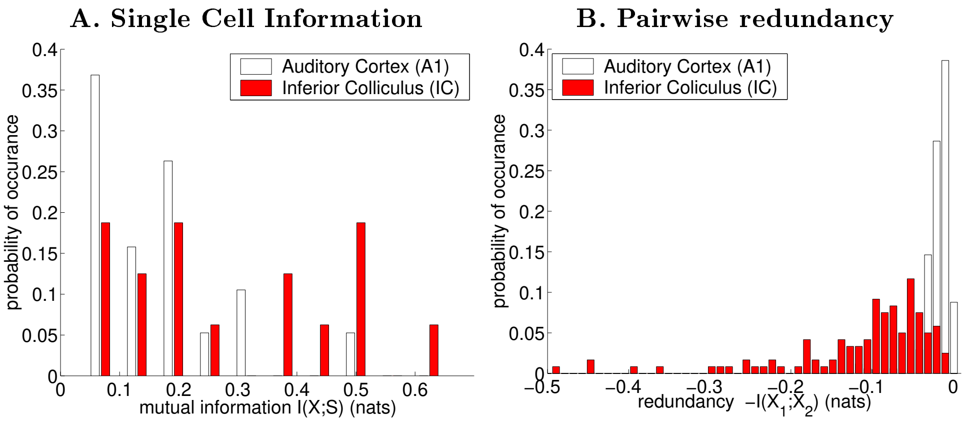
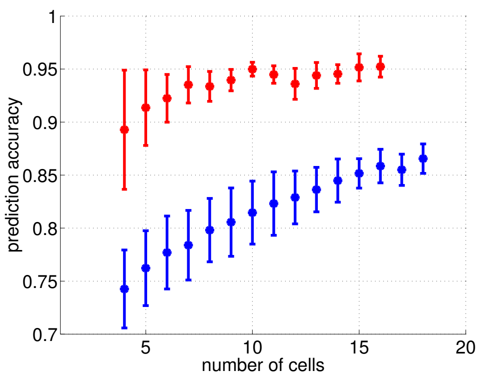


Figure 3: **A.** Distribution of MI across cells. A1 cells (light bars) convey slightly less information on average than IC neurons. **B.** Distribution of the normalized redundancy . A1 pairs (light bars) are concentrated near zero, while IC pairs have significantly higher redundancy values. Spike counts were collected over a window that maximizes overall MI. Number of bins in counts-histogram was optimized separately for each stimulus and every cell. Information estimations were corrected for small-samples bias by shuffling methods [9].

Figure 4. Prediction accuracy of stimulus identity as a function of number of cells used by the classifier. Error bars denote standard deviation across several subgroups of the same size. For each subgroup, a one-hidden layer neural network was trained separately for each stimulus using some stimulus presentations as a training set and the rest for testing. Performance reported is for the testing set.



5 Discussion

We have developed information theoretic measures of redundancy among groups of neurons and applied them to investigate the collaborative coding efficiency in the auditory modality. Under two different coding paradigms, we show differences in both information content and group redundancies between IC and cortical auditory

[3]The probability of accurate prediction is exponentially related to the input-output mutual information, via the relation $P_{correct} = exp(-\text{missing nats})$ yielding $MI_{nats} = ln(\text{no. of stimuli}) + ln(P_{correct})$. Classification thus provides lower bounds on information content.

neurons. Single IC neurons carry more information about the presented stimulus, but are also more redundant. On the other hand, auditory cortical neurons carry less information but are more independent, thus allowing information to be summed almost linearly when considering groups of few tens of neurons. The results provide for the first time direct evidence for redundancy reduction along the ascending auditory pathway, as has been hypothesized by Barlow [1]. The redundancy effects under the single-spikes coding paradigm are significant only for groups larger than ten cells, and cannot be revealed with the standard redundancy measures that use only pairs of cells.

Our results suggest that redundancy reduction transformations are not limited to low level sensory processing (aimed to reduce redundancy in input statistics) but are applied even at cortical sensory stations. We suggest that an essential experimental prerequisite to reveal these effects is the use of complex acoustic stimuli whose processing occurs at the cortical stations.

The above findings are in agreement with the view that along the ascending sensory pathways, the number of neurons increase, their firing rates decrease, and neurons become tuned to more complex and independent features. Together, these suggest that the neural representation is mapped into a representation with higher effective dimensionality. Interestingly, recent advances in kernel-methods learning have shown that nonlinear mapping into higher dimension and over-complete representations may be useful for learning of complex classifications. It is therefor possible that such mappings provide easier readout and more efficient learning in the brain.

References

[1] H.B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In *Mechanisation of thought processes*, pages 535–539. Her Majesty's stationary office, London, 1959.

[2] N. Brenner, S.P. Strong, R. Koberle, R. de Ruyter van Steveninck, and W. Bialek. Synergy in a neural code. *Neural Computation*, 13(7):1531, 2000.

[3] T.M. Cover and J.A. Thomas. *The elements of information theory.* Plenum Press, New York, 1991.

[4] I. Gat and N. Tishby. Synergy and redundancy among brain cells of behaving monkeys. In M.S. Kearns, S.A. Solla, and D.A.Cohn, editors, *Advances in Neural Information Proceedings systems*, volume 11, Cambridge, MA, 1999. MIT Press.

[5] T.J. Gawne and B.J. Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons ? *J. Neuroscience*, 13(7):2758–2771, 1993.

[6] P.M. Gochin, M. Colombo, G. A. Dorfman, G.L. Gerstein, and C.G. Gross. Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.*, 71:2325–2337, 1994.

[7] J.P. Nadal, N. Brunel, and N. Parga. Nonlinear feedforward networks with stochastic outputs: infomax implies redundancy reduction. *Network: Computation in neural systems*, 9:207–217, 1998.

[8] I. Nelken, Y. Rotman, and O. Bar-Yosef. Specialization of the auditory system for the analysis of natural sounds. In J. Brugge and P.F. Poon, editors, *Central Auditory Processing and Neural Modeling.* Plenum, New York, 1997.

[9] LM. Optican, T.J. Gawne, B.J. Richmond, and P.J. Joseph. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol. Cyber*, 65:305–310, 1991.

[10] E. T. Rolls and A. Treves. *Neural Networks and Brain Function.* Oxford Univ. Press, 1998.

[11] C.E. Shanon. A mathematical theory of communication. *The Bell systems technical journal*, 27:379–423,623–656, 1948.