[P2-32]

Modelling disordered A1 map and ordered V1 map

Hiroki Terashima (PY)[1,2], and Masato Okada[1,3]

[1] Department of Complexity Science and Engineering, The University of Tokyo
[2] JSPS Research Fellow DC1 [3] RIKEN Brain Science Institute
E-mail: hterashima@mns.k.u-tokyo.ac.jp

Abstract— Unlike the smooth retinotopic map of the primary visual cortex, the tonotopic map of the primary auditory cortex is disordered, which seemingly contradicts the uniformity of the cortex. We give a unified description of the two maps through a common adaptive process, suggesting the discrepancy of order may result from different natural stimlus statistics.

Keywords— Primary auditory cortex (A1), tonotopy, natural stimulus statistics, retinotopy, topographic independent component analysis (TICA)

1 Introduction

While the sensory cortices are known to exhibit common topographic map structures, their uniformity has been challenged by recent findings on the primary visual cortex (V1) and the primary auditory cortex (A1). Two-photon calcium imaging has shown that V1 has smooth, precisely arranged orientation and retinotopic maps [1]. In contrast, arrangement of best frequencies or the tonotopy of A1 neurons has been revealed to be highly disordered, although it shows a weak global trend [2]. The discrepancy of map smoothness apparently suggests the two areas process their inputs differently.

We propose a hypothesis stating that the seemingly different maps of V1 and A1 still result from a common adaptive process: this discrepancy does not reflect their different ways of information processing but different natural stimulus statistics. Our previous study [3] has shown that a model for the spatially localized receptive fields of V1 neurons, sparse coding, can also explain those of A1 neurons that also seem to differ from V1 because they are often non-localized in the frequency domain. This suggests that a difference emerges from different stimulus statistics. Whereas natural stimuli of vision correlate only locally, those of audition often do with distant frequencies. By using this point of view, we give an integrated interpretation for the seemingly dissimilar topography of A1 and V1.

2 Methods

2.1 Model

Since we hypothesize that A1 and V1 adopt a common adaptive process, we discuss a model previously proposed for the V1 maps, called topographic independent component analysis (TICA) [4], as a model for the A1 map. It extends the independent component analysis by simultaneously making neighbour units active, eventually reproducing V1-like smooth retinotopy and orientation map.

Topographic independent component analysis is a two-layer neural network, in which the first-layer unit activity s_i is defined as the inner product of an input I and weights (filter) $\boldsymbol{w}_i$. The neighbourhood function $h(i, j)$ defines connections between units of the first and second layers, with 1 showing that i and j are neighbours, and 0 otherwise. Activity of a second layer unit c_{it} is the sum of the "local energy" of its neighbourhoods in the first layer. The filters $\boldsymbol{w}_i$ are learned by maximizing the likelihood L.

$$c_{it} = \sum_j h(i,j) \langle \boldsymbol{w}_j, I_t \rangle^2 \tag{1}$$

$$\log L(\{I_t\}; \{\boldsymbol{w}_i\}) = \sum_i \sum_t G(c_{it}) \tag{2}$$

where $G(c_{it}) = -\sqrt{0.005 + c_{it}} = \log p_i(c_{it})$, imposing a sparse prior on the second layer activity.

2.2 Artificial inputs

To systematically evaluate the effects of input "auditoriness", one-dimensional inputs $I(x)$ are generated with various degrees of correlation between distant coordinates. Its domain is $x = 0, 1, \ldots, 15$, and is a torus for excluding the boundary effect.

Each sample is initialized by the standard normal distribution. To produce vision-like local correlation, a constant value 4 is added at k points, where k is from a uniform distribution over $\{3, 4, 5, 6\}$. The points' coordinates are from a normal distribution with a random centre and $\sigma = 2$. When adding the constant value at x, we also add another at $x_{dist} = x+5$ with probability p_a or a parameter defining its "auditoriness", which in turn regulates the degree of distant correlation.

2.3 Natural inputs

To model and compare the maps of A1 and V1, biologically plausible stimuli of vision and audition are used as inputs. The visual ones are 25×25 pixel patches extracted from natural images attached to a demo program provided by Hyvärinen and Hoyer [4].

The auditory inputs are 25×25 patches extracted from spectrograms of human vocalizations [1]. Original spectrograms are generated using NSL toolbox [5], whose frequency domain consists of 128 points from 90 to 3623 Hz in log scale. Patches of 200 ms are extracted from the spectrograms and scaled down in the frequency domain.

The patches ($N = 50,000$ in both cases) are low-passed by using 400 main principal components, with the four highest excluded.

2.4 Degree of map disorder D

For quantitative comparison of map disorder, the degree of disorder D regarding a specific feature, such

[1] Narrative recordings from Linguistics Handbook of IPA

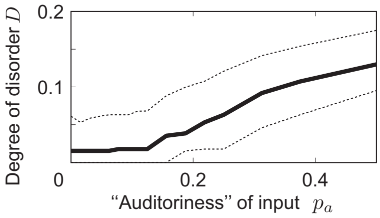

Figure 1: Map disorder correlates with auditoriness of artificial inputs.

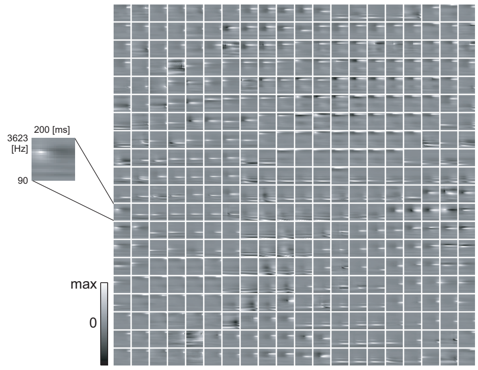

Figure 2: Map of spectro-temporal filters that adapt to natural auditory inputs with disordered tonotopy

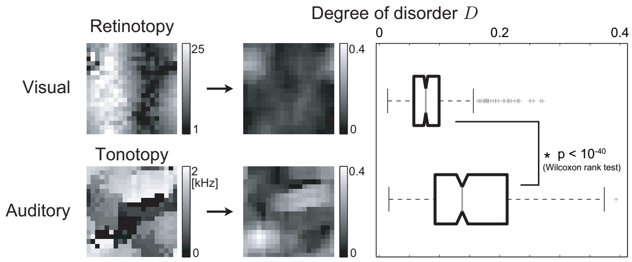

Figure 3: More disordered tonotopy than retinotopy

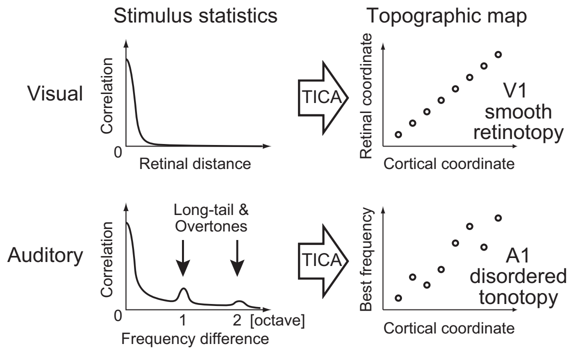

Figure 4: Suggested source of map differentiation

as the best frequency, is defined as below at each map coordinate. Suppose a one-dimensional map coordinate x for simplicity. First, the feature $f(x)$ is scaled so that its value range is $[0, 1]$. Next, for each unit at x, its neighbourhood is defined as $NB(x) = \{x-2, x-1, \ldots, x+2\}$, and the residual error of linear fitting within the neighbour is defined as r. The degree $D(x)$ is defined as

$$D(x) = \sqrt{\frac{\sum_{x_i \in NB(x)} r(x_i)^2}{N_{NB}}}. \tag{3}$$

The degree of disorder for a two-dimensional map is defined likewise with a 5 × 5 neighbourhood. When the input domain is a torus, another D is computed using modified f values that are added 1 if within $[0, \frac{1}{2})$, and the smaller D is taken.

3 Results

3.1 Maps adapting to artificial inputs

To test the hypothesis, Ds of maps adapting to different inputs were compared. First, to evaluate the quantitative effects of the input structure on map disorder, artificial inputs were generated with a specific “auditoriness” p_a ($N = 20,000$), for each session. The TICA model adapted to the inputs, resulting in a one-dimensional map composed of 16 filters $\boldsymbol{w}_i$. For each filter, D was computed, where the feature $f(x)$ is the coordinate with the maximum absolute value.

Figure 1 shows that the degree of disorder D correlates with the input “auditoriness” p_a. The three lines show quartiles (25, 50 (bold), 75 %) in 100 sessions. While vision-like inputs ($p_a \sim 0$) induce smooth maps ($D \sim 0$), audition-like inputs induce disordered ones.

3.2 Maps adapting to natural inputs

Next, for modelling the A1 map, a 20 × 20 torus TICA model adapted to human vocalizations. Figure 2 shows the map of learned filters, which resemble receptive fields of A1 neurons. The feature f of each filter is the frequency coordinate with the highest absolute value. The map of f is disordered, that is, best frequencies of neighbour units are not necessarily be similar to each other.

Its Ds were compared with those of another map adapting to natural visual inputs; the visual map consists of filters resembling V1 simple cells with smoothly varying retinotopy and orientation as reported in [4]. Each filter was fitted using a Gabor function, whose centre coordinate is its f. Figure 3 shows that the visual map is less disordered than the auditory one.

4 Summary

We provided a unified interpretation for the seemingly different maps of A1 and V1. The discrepancy does not reflect their different ways of information processing but the different natural stimulus statistics, as summarized in Figure 4. The disordered tonotopy may emerge from the unique statistics of auditory stimuli, that is, the correlation between distant frequencies.

References

[1] Smith S. L. & Häusser M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nature Neuroscience*, 13, 1144–1149.

[2] Rothschild G., Nelken I., & Mizrahi A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature Neuroscience*, 13, 353–360.

[3] Terashima H. & Hosoya H. (2009). Sparse codes of harmonic natural sounds and their modulatory interactions. *Network: Computation in Neural Systems*, 20, 253–267.

[4] Hyvärinen A. & Hoyer P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41, 2413–2423.

[5] http://www.isr.umd.edu/CAAR/pubs.html