Data-driven approaches for unveiling the neurophysiological functions of the auditory system

Shigeto Furukawa1,∗, Hiroki Terashima1, Takuya Koumura1 and Hiroaki Tsukano2
1NTT Communication Science Laboratories, NTT Corporation, 3–1 Morinosato-Wakamiya, Atsugi, 243–0198 Japan
2Brain Research Institute, Niigata University, 1–757 Asahimachi-dori, Chuo-ku, Niigata, 951–8585 Japan

Abstract: Typical neurophysiological experiments employ “hypothesis-driven” approaches: Researchers set a specific hypothesis, based on which stimuli and their parameters are chosen. However, there is always a concern that the hypothesis or stimulus parameter could be irrelevant to the essence of the brain function. The present paper review the authors’ recent studies that have applied some “data-driven” approaches as relatively hypothesis-free methodologies to traditional questions in auditory neurophysiology, such as neural frequency tuning and cortical topography. The results provide some new insights into the functional organization of the cortex and the optimality of the brain structure for auditory processing.

Keywords: Neurophysiology, Machine learning, Auditory system, Neural network

PACS number: 43.64.Bt, 43.64.Fy, 43.64.Qh [doi:10.1250/ast.41.63]

1. INTRODUCTION

Typically, setting a specific hypothesis is the first step for researchers in neurophysiology when experiments or model-based studies are designed (Fig. 1 left). This hypothesis is the basis of selecting appropriate stimuli and analyses. When the experiment fails to clarify the validity of the hypothesis, the researcher revises his or her hypothesis and conducts another experiments with a different set of stimuli. Even when the experimental results are consistent with the hypothesis, however, there is also a concern that the hypothesis or stimulus parameter could be irrelevant to the essence of the brain function. The neural system is vastly complex, but researchers often have neither enough time nor resources to fully explore alternative hypotheses and stimulus parameters.

Emerging data-driven approaches can complement hypothesis-driven ones to reveal the “true” figure of the brain function (Fig. 1 right). Here, a range of naturalistically complex sounds are often chosen as stimuli without the basis of a precisely defined hypothesis, and machine-learning techniques are used to discover systematic patterns or structures embedded in the apparently complex neural responses observed. The present paper will review the authors’ recent studies that have applied such relatively hypothesis-free methodologies to traditional questions in auditory neurophysiology, such as neural frequency tuning and brain structure.

2. FUNCTIONAL ORGANIZATION OF THE MOUSE AUDITORY CORTEX

Topographical organization of the auditory cortex has the fundamental importance in the auditory neuroscience. Despite efforts of many neuroscientists and development of measurement techniques, researchers often reach somewhat different conclusions on the functional organization of the auditory cortex. This problem applies also to the mouse, an animal species that is attracting researcher’s attentions rapidly in recent years [1].

At least two types of limitations inherent to the existing experimental paradigm make it difficult to solve the discrepancies among studies. One is that a small (and perhaps biased) repertoire of synthesized stimuli have been used, such as bursts of pure tone and random noise. The other limitation is that the researchers of the existing studies have to select acoustical features or parameters (e.g., tone frequency) to be examined prior to the experiment. However, natural auditory scenes that the auditory system encounters are far more complex than those experimental stimuli. Also, it is possible that researcher may overlook acoustical features that are essentially important for the auditory system.

∗e-mail: furukawa-shig@umin.org
The present study attempted to avoid these limitations by using a machine-learning technique, which would allow us to analyze complex cortical responses to complex stimuli without setting a particular hypothesis. Stimuli were generated by pitch-shifting 165 natural sounds in the sound dataset originally for human experiments [2]. The pitch-shifting was conducted to fit the frequency components to mouse’s audible frequency range. We believe that the stimuli were as complex as natural sounds, although they were not truly natural because of the pitch-shifting.

We made fluorescent imaging of the auditory cortex from anaesthetized transgenic mice, in which the calcium-sensitive protein GCaMP8 was expressed. The experimental protocols were approved by the Committee for Animal Care at Niigata University. The cortical activities were recorded as 21,504 (= 128 × 168) pixels of the image data. Thus, cortical responses to the complex stimuli can be expressed as a high-dimensional matrix, $D$ (165 sounds × 21,504 pixels). Given that the cortical surface can be divided into a few subdivisions, it is expected that the matrix $D$ can be expressed as combinations of a small number of components, i.e., $D \sim RW$, where $R$ and $W$ have $165 \times N$ and $N \times 21,504$ pixels, respectively, and $N$ is the number of components. Each row of $W$ should represent spatial components corresponding to cortical surface. The concept and method of this matrix decomposition technique was proposed by an earlier study that examined human fMRI data [2].

We derived five spatial components from the responses to the 165 stimuli. Each component was characterized by spatial patterns of excitatory and inhibitory pixel clusters. Comparing with the imaging data recorded for tonal stimuli, we found that three of the five components correspond well to tonotopic regions and one component to non-tonotopic regions, which had been proposed by the earlier studies (e.g., [1]). Figure 2 (bottom) summarizes the characteristics of the above components and illustrates cortical subdivisions. The results generally indicates that the traditional (hypothesis-driven) and the present (data-driven) methods mutually support their validities (compare top and bottom panels in Fig. 2). Interestingly, the

---

![Hypothesis-driven approach vs Data-driven approach](image)

**Fig. 1** Typical study flows in hypothesis-driven and data-driven approaches.

---

![Figure 2](image)

**Fig. 2** Top: Divisions of the mouse auditory cortex proposed by an earlier study using tonal stimuli (extracted from Fig. 1 of Ref. [1]). Bottom: Divisions derived by the present study.
remaining one component does not appear to fit any subdivisions proposed earlier. This component may indicate the existence of a new region that have not as yet been identified (marked as “?” in Fig. 2).

3. HIERARCHICAL PROCESSING OF AMPLITUDE MODULATION [3]

This section reviews another type of data-driven approach, being explored recently in our research group. It is a model-based study for understanding the significance of the hierarchical structure of the auditory brainstem.

Temporal variation of amplitude envelope, called amplitude modulation (AM), carries essential information for auditory perception of natural sounds. Models that incorporate an array of modulation filters can account well for various aspects of human auditory perception [4,5]. A vast amount of physiological studies have been conducted to examine neural sensitivities to AM at the various stages along the auditory pathway. There is a broad agreement that (1) the distribution of preferred AM frequencies of the neurons tends to shift from higher to lower ones along the ascending pathway, and that (2) the neural representation of AM is transformed from temporal-code (via synchronized firing to the AM waveform) to the rate-code (i.e., spike rate varies with AM frequency, with or without synchrony) [6]. Those physiological data provide critical information for understanding how the auditory system is structure.

The present study was conducted, instead, to approach the question of why, i.e., what is the significance of the auditory system having such a structure. We modelled the auditory system with a deep neural network (DNN). The DNN is a version of dilated causal convolutional neural network [7], with 13 layers and 128 units per layer (Fig. 3 left; see [3] for details). The DNN was trained to classify natural sounds based on sound waveforms, which is a natural task for the auditory system. It should be noted that the DNN was not designed to simulate particular neural structure, except that it consists of multiple layers of multiple units or “neurons.”

After optimizing the DNN for the recognition task, we applied typical experimental procedures to characterizing the neurons’ AM sensitivities. That is, we presented amplitude-modulated white-noise with various modulation frequency to the optimized network, and recorded the responses of individual units to the stimuli. Neural preference to modulation frequency was evaluated in terms of response magnitude (comparable to neural spike rate) and of synchrony to the AM waveform (comparable to neural synchronous firing).

We found that neural selectivity (or tuning) to AM frequency had emerged in the DNN. There was a systematic trend in neural distribution along ascending hierarchy of layers that preferred AM frequency (expressed as the best or upper-cutoff AM frequency responses) decreased (Fig. 3 right). Also, the majority of units exhibited AM

![Fig. 3](image-url)  
**Fig. 3** Left: Illustration of the structure of the DNN used in the present study. Right: Distributions of the units’ best and upper-cutoff AM frequencies. Each panel represents one nucleus (left) of the auditory system or one layer of the DNN.
tuning only in temporal synchrony (temporal code) but not in response magnitude (rate code) in relatively lower-level layers, whereas a proportion of the rate-code units started appearing in higher-level layers. This trend was similar to that of the actual auditory system as described earlier [6].

We quantified the similarity between the DNN and actual nervous system by computing the distance between the distributions of the units’ best and upper-cutoff AM frequencies for the DNN layers and for the brainstem nuclei (as shown in Fig. 3 right). We found that the similarity between the DNN and the brainstem increased with the progress of DNN optimization. That is, better-performing DNNs tended to resemble more to the nervous system.

Various control examinations suggest that the naturalness of the stimuli and the task objectives, rather than the DNN properties, are essential factors for the similarity. The results suggest that the AM representation in the auditory system may be a consequence of optimization (through the process of evolution or development) of the nervous system for natural sound recognition.

We should point out that we optimized the DNN for generic tasks without explicit implementation of AM sensitive mechanisms. Thus, the current approach can be readily expanded to a variety of stimulus dimensions other than AM and can serve as a potential tool to explore yet-unrecognized functions of the auditory system in an unbiased way. This makes a significant contrast with prevailing physiological models, which are hypothesis-driven and aim to simulate the behaviors of real neurons or to implement particular functions.

4. CONCLUDING REMARKS

The two studies reviewed in this article, although targeting different aspects of auditory function and adopting different methodologies, commonly demonstrate advantages of data-driven approaches; i.e. they do not require a predefined, strong hypothesis. The results provide unbiased and independent views on the functions of the auditory system, which might be overlooked by prevailing hypothesis-driven approaches. This does not mean, however, that the data-driven approaches will replace the standard hypothesis-driven ones. We hope that the new techniques would help to accelerate the auditory science by allowing researchers to evaluate a validity of hypotheses or by providing a new hypothesis that would be worth tested.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP15H05915 (Grant-in-Aid for Scientific Research on Innovative Areas “Innovative SHITSUKSAN Science and Technology”).

REFERENCES

[1] H. Tsukano, M. Horie, T. Bo, A. Uchimura, R. Hishida, M. Kudoh, K. Takahashi, H. Takebayashi and K. Shibuki, “Delineation of a frequency-organized region isolated from the mouse primary auditory cortex,” J. Neurophysiol., 113, 2900–2920 (2015).
[2] S. Norman-Haignere, N. G. Kanwisher and J. H. McDermott, “Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition,” Neuron, 88, 1281–1296 (2015).
[3] T. Koumura, H. Terashima and S. Furukawa, “Cascaded processing of amplitude modulation for natural sound recognition,” J. Neurosci., 39, 5517–5533 (2019).
[4] T. Dau, B. Kollmeier and A. Kohlrausch, “Modeling auditory processing of amplitude modulation I. Detection and masking with narrow-band carriers,” J. Acoust. Soc. Am., 102, 2892–2905 (1997).
[5] R. McWalter and T. Dau, “Cascaded amplitude modulations in sound texture perception,” Front. Neurosci., 11, 485 (2017).
[6] P. X. Joris, C. E. Schreiner and A. Rees, “Neural processing of amplitude-modulated sounds,” Physiol. Rev., 84, 541–577 (2004).
[7] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior and K. Kavukcuoglu, “WaveNet: A generative model for raw audio,” arXiv: 1609.03499 (2016).