Beyond AI Project

Expanding Brain Function Using AI
(AI-assisted Expansion of Perception, Sensibility, and Cognitive Performance)

Project Leader: Yuji Ikegaya, PhD
Professor, Grad Sch Pharmaceut Sci, Univ Tokyo

For the first three years, we tried to establish an experimental system that would powerfully motivate animals to perform specific behavioral tasks and allow us to control various behaviors and functions of the animals. We explored the possibility of rewarding animals by electrically stimulating the medial forebrain bundle (one of the reward systems). We have shown that the experimental system actually works and have established a stable experimental system that allows us to target the medial forebrain bundle. Using this technique, we demonstrated that electrical stimulation of the medial forebrain bundle induced sniffing activity and gamma oscillations in the primary motor cortex prior to locomotion (Yoshimoto et al., 2022). In addition, delayed reinforcement using medial forebrain bundle stimulation inhibited subsequent extinction learning (Shibata et al., 2022). Using these techniques, we approached the brain-AI fusion project as follows.

Neurofeedback, which visualizes and self-regulates brain activity, is effective in expanding visual and auditory perceptual abilities (Chang et al., PLoS One, 2017). However, there are some neural activity patterns that cannot be acquired by self-regulatory neurofeedback (Sadtler et al., Nature, 2014), which is a limitation in this project to expand brain function. To overcome this limitation, we hypothesized that brain function could be maximized by extracting latent information from the brain using machine learning and feeding it back directly to the brain as electrical stimulation. To test this hypothesis in an animal model, we created a task in which rats are congenitally unable to discriminate. The task was designed to discriminate between English and Spanish.

We prepared 50 sentences, each 2.5 to 6 seconds long, spoken in English and Spanish by the same speaker using a speech synthesis model trained on bilingual speech. Rats were placed in a box with two nosepoke holes on each side, and English or Spanish phrases were presented randomly. The rats continued to perform 500 trials per day for one week, but the discrimination rate was still 50% of chance. We also recorded local field potentials (LFPs) during language presentation and decoded which language was presented based on the LFPs. We used a convolutional neural network (CNN) with kernels in each of the temporal and interelectrode directions, as we thought that deep learning would be useful for exhaustive information extraction. The classification accuracy of the LFPs corresponding to the phrases included in the training of the model was about 70%, and the classification accuracy of the LFPs corresponding to the phrases not included in the training was about

60%, both of which exceeded the chance level of 50%. In addition, when the features of the final layer used by the CNN to make decisions were dimensionally reduced by Uniform Manifold Approximation and Projection (UMAP), there was an overall tendency for the clusters to separate into clusters for each language. These results suggest that although language is indistinguishable at the behavioral level, language features are generalized and latent at the level of neural activity.

We then used electrical stimulation of the left or right somatosensory cortex (S1) and the reward system, the medial forebrain bundle (MFB), to provide feedback to the brain. To see if the rats could discriminate between these left and right stimuli, we performed a left-right discrimination task using electrical stimulation of the S1. Rats were rewarded by stimulation of the MFB for poking their noses into the left hole when the right S1 was stimulated and into the right hole when the left S1 was stimulated. As a result, we confirmed that about 90% of the correct nose pokes were made in the correct hole.

We are currently investigating whether languages can be discriminated at the behavioral level when we provide latent feedback information at the neural activity level, which is not expressed at the behavioral level but can be extracted by a CNN, by electrically stimulating the S1. We performed a language discrimination task in which the animal's right S1 was stimulated after the CNN detected English and the left S1 was stimulated after the CNN detected Spanish. As a result, the CNN was able to discriminate correctly above chance level. To confirm that this learning was maintained when the feedback stimulus was turned off, we conducted a trial in which no stimulus was given for 10% of the trials. As a result, the percentage of correct responses remained above chance even in the absence of stimulation. As a control experiment, we are also working on a paradigm in which the right S1 is stimulated when English is presented and the left S1 is stimulated when Spanish is presented, independent of LFPs. So far, trials with stimulation have improved to about 90% correct, while trials without stimulation were at chance level. We believe that this result was observed because we could not induce plastic changes, as we could always obtain correct information without plastic changes in our own neural activity. In summary, we have shown that intelligence augmentation is possible through co-learning between the brain and artificial intelligence. If this technology is extended to humans, it will open up new possibilities as a treatment method for improving cognitive dysfunction. In addition, this technology is expected to be applied to healthy people as well; this technology will enable them to acquire new concepts and values of things they have never noticed before, and contribute to the improvement of human well-being as a whole.

It is well known that it is difficult to guide neurofeedback in the direction of increasing dimensions that explain neural activity patterns during spontaneous activity (Sadtler et al., Nature, 2014). In our study, however, we believe that neural activity patterns representing new English or Spanish phrases can be induced because the neural circuits of S1 are directly activated during co-learning with artificial intelligence. If this leads to a change that increases the dimension of neural activity, it would be a

groundbreaking discovery that would change conventional wisdom.

Selected Publications

1. Norimoto, H., Makino, K., Gao, M., Shikano, Y., Okamoto, K., Ishikawa, T., Sasaki, T., Hioki, H., Fujisawa, S., and Ikegaya, Y. Hippocampal ripples down-regulate synapses. Science, 359:1524-1527, 2018.
2. Norimoto, H., Ikegaya, Y. Visual cortical prosthesis with a geomagnetic compass restores spatial navigation in blind rats. Curr. Biol., 21:1091-1095, 2015.
3. Ishikawa, D., Matsumoto, N., Sakaguchi, T., Matsuki, N., Ikegaya, Y. Operant conditioning of synaptic and spiking activity patterns in single hippocampal neurons. J. Neurosci., 34:5044-5053, 2014