Authors:
Osamu Hoshino, Masayuki Miyamoto, Tsuyoshi Ono and Kazuharu Kuroiwa
hoshino@cc.oita-u.ac.jp
masayuki@cc.oita-u.ac.jp
tsuyoshi@hwe.oita-u.ac.jp
kuroiwa@cc.oita-u.ac.jp

Corresponding author:
Tsuyoshi Ono
Hoshino Lab.
Department of Human Welfare Engineering
Oita University
700 Dannoharu, Oita 870-1192, Japan
tel:+81-97-554-7301
fax:+81-97-554-7301
e-mail: tsuyoshi@hwe.oita-u.ac.jp

Title:

A NEURAL NETWORK MODEL FOR ENCODING AND PERCEPTION OF VOWEL SOUNDS

Authours:
Osamu Hoshino, Masayuki Miyamoto, Tsuyoshi Ono and Kazuharu Kuroiwa
Oita University at Oita
hoshino@cc.oita-u.ac.jp
masayuki@cc.oita-u.ac.jp
tsuyoshi@hwe.oita-u.ac.jp
kuroiwa@cc.oita-u.ac.jp

Abstract:

We propose a neural network model for the encoding and perception of vowel sounds. Pairs of formant frequencies (F1, F2) of five Japanese vowels were separately applied to the network, during which synaptic connections were modified according to the Hebbian learning rule. After the leaning, five point attractors corresponding these vowels had been created in the network dynamics. When the network was stimulated with formant pairs that belong to the same vowel but differ from each other, one point attractor corresponding to the vowel was always induced. That is, the network can invariantly perceive the same vowel sound as identical one. We suggest that the encoding of vowel sounds by distinct attractors may be an effective strategy for the brain to encode and perceive vowel sounds, by which the sounds spoken by different people, including ones whom we have met for the first time, can be perceived correctly.

Preference for presentation: poster

Preference for presentation: poster

Summary:
Regardless of acoustic differences among individuals in vocalization, one can invariantly perceive speeches independent of speakers. We hear to someone speaking and understand what he or she is saying. We can also understand speeches that are spoken by people whom we have met for the first time. It seems that at birth we do not understand speeches due to the lack of vocabulary. The ability of speech perception seems to arise from experiences in hearing and learning of vocabularies (and their combinations) through our daily lives. It is interesting how humans establish in the brain such an auditory function of invariant perception of speeches.

The vowel is one of the fundamental elements for human speech sounds and is characterized by the so-called formant frequencies, which are the peak frequencies in the spectrum of vocalized vowel sounds. Based on a psychological experiment, Peterson and Berney (1952) made two-dimensional coordinates (F1-F2) for the first (F1) and second (F2) formant frequencies, and drew ten elliptical enclosures relevant to the English vowels. If, in the F1-F2 coordinates, different formant pairs (F1, F2) of vowels point at loci that are within the same enclosure, vowels of these formant pairs are classified as the same vowel. Different formant pairs belonging to the same vowel imply that people speak the same vowel in different acoustic ways. Suga (1988) made similar two-dimensional coordinates for Japanese vowels, on which five elliptical enclosures relevant to vowels, /a/, /i/, /u/, /e/ and /o/, were drawn.

In a nonprimary auditory cortical area, Langer et al. (1981) found neurons that were specifically responsive to spoken vowels of German. Using synthetic vowels generated by formant pairs, they demonstrated that these neurons responded to a specific combination of formant frequencies (F1, F2). Based on these findings, they proposed a two-dimensional neuronal model in which neurons are tonotopically arranged along each axis. The neuron located at a locus in the F1-F2 coordinates simultaneously receives F1 and F2 formant frequencies and fires responding to the combinatory stimulus

(F1, F2). These results suggest that the combination-sensitive neurons may contribute to processing the information about vowels. One question, then raised there, is how auditory systems place a variety of formant pairs (F1, F2) that belong to the same vowel into one relevant vowel category and in what manner auditory systems use it for the perception of vowel sounds.

The purpose of the present study is to propose a neural network model for the encoding and perception of vowel sounds. The model was constructed based on a two-dimensional model proposed by Langer et al. (1981). Different formant pairs (F1, F2) belonging to vowel /x/ (x = a, i, u, e, o) were randomly applied to the network, during which synaptic connections were modified according to the Hebbian learning rule. This process continued until the point attractor corresponding to /x/ was created in the dynamic system of the network. We let the network learn these five vowels. After the leaning process, the state of the network randomly itinerates among these point attractors. Within the basin of each point attractor, a group of neighboring neurons fires in a coherent manner. In the two-dimensional map, the boundaries of the groups form enclosures corresponding to these vowels. By simulating the model, we investigated neuronal bases of invariant perception of vowel sounds.

Different formant pairs belonging to the same vowel /x/, which means that the vowel is spoken by different people, were separately applied to the network. The stimulation with the pairs always induced one identical point attractor corresponding to the vowel. This implies that the network classified these sounds as one identical vowel sound. If the model was stimulated with formant pairs that did not belong to any of the vowels, the state of the network tended to itinerate randomly among the point attractors and did not converge at any point attractor. That is to say, the sound has not been classified as a vowel.

When the network was stimulated with pairs that had not been presented in the learning process but belong to the vowel /x/, the identical point attractor corresponding to /x/ was induced. Similar results were obtained for perceptual processes of all the vowels. These results indicate that the network recognizes vowel sounds

spoken by different people whom one has met for the first time. Humans, surely, can have such capability.

We consider that the encoding of vowel sounds by distinct attractors may be an effective strategy for the brain to place a variety of formant pairs of vowels into distinct vowel categories and may be a plausible mechanism for the invariant perception of vowel sounds.

If the model was stimulated with formant pairs that belong to multiple vowels, the state of the network tended to itinerate randomly among the point attractors corresponding to the vowels. This means that the network has succeeded in rough categorization of the applied sound but failed to identify it.

This problem, in most part, could be overcome in a three-dimensional auditory map, in which additional formant information such as the third formant frequency of vowels is placed on the third tonotopic axis. Although the real auditory system may have a much higher and more complex dimensional map in the brain, the present simple two-dimensional map has gained good insights into the basic neuronal mechanisms for the encoding and perception of vowel sounds.

References

G.E. Peterson, and H.L. Barney, Control methods used in a study of the vowels, J. Acoust. Soc. Am., Vol. 24, pp. 175-184, 1952.

N. Suga, Auditory function: Neurobiological bases of hearing, New York, Wiley, 1988.

G. Langer, D. Bonke, and H.Scheich, Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds, Exp. Brain Res., Vol. 43, pp. 11-24, 1981.