How does the infant brain learn to segment speech, invariant of speaker?

Department: Artificial Intelligence. Size: 45 EC.

Supervision: Aditya Gilra, CWI (Research Institute for Mathematics and Computer Science), Amsterdam.
(aditya.gilra@cwi.nl)

Background

Infants learn to represent phonetic features of speech, invariant of speaker, even at 3 months after birth [1]. Over the first year, via vocal babbling and imitation, the infant refines its representations of speech [2]. It is unclear how much the neural representations for speech are learned or refined via the interplay of perception and production [3]. In this project, you will build a neural network model of learning to segment and cluster speech units irrespective of speaker identity and speaking rate.

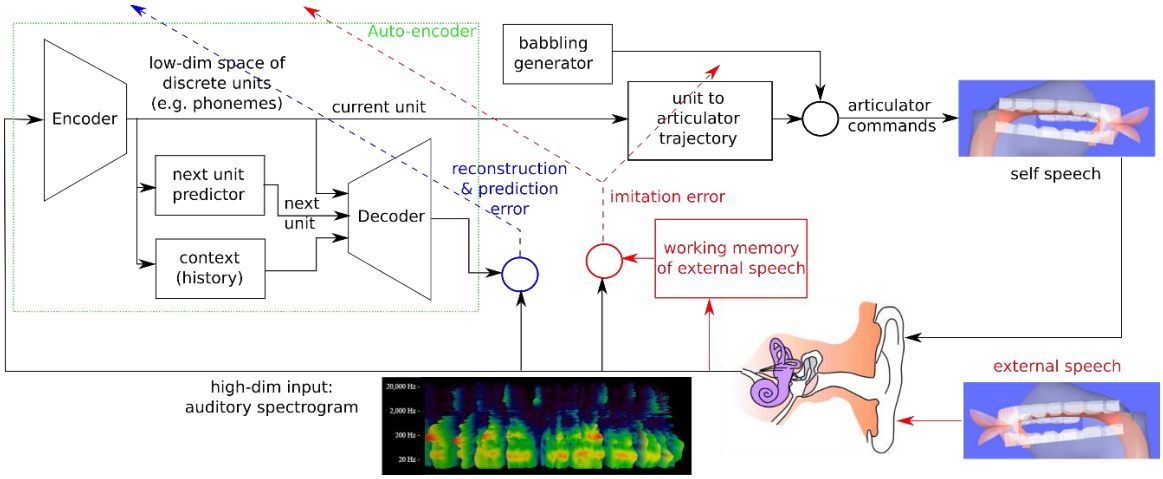

Figure 1: A block diagram for learning to imitate external speech stored in phonological working memory (in red), using a predictive autoencoder architecture (in green box) that converts high-dimensional continuous representations of speech input (spectrogram obtained from cochlea) into low-dimensional discrete units of speech (e.g. phonemes or syllables). (Ear subfigure from Wikimedia, spectrogram from Chang lab NeuroSpeech player, and vocal articulator from VocalTractLab.)

Goals and scope

The student will apply machine learning techniques like autoencoders and contrastive learning, in a bio-plausible manner, to enable speech segmentation in closed loop using acoustic filters similar to human ear and the VocalTractLab model (https://www.vocaltractlab.de/). The project consists of three tasks:

1. Learn a speaker-invariant auto-encoder to represent speech units that are useful in both perceiving and producing speech in closed loop as shown in Figure 1.
2. Analyze the role of speech production in segmentation by trading off loss terms or varying self-vocal parameters.
3. Compare the speech units and representations that are learned by your encoder with those observed in neuroscience experiments. Amend your auto-encoder to match neuroscience representations.

Student profile

Background in machine learning, preferably with exposure to neuroscience or dynamical systems.

References

[1] Gennari, G., Marti, S., Palu, M., Fló, A., & Dehaene-Lambertz, G. (2021). Orthogonal neural codes for speech in the infant brain. Proceedings of the National Academy of Sciences, 118(31), e2020410118.

[2] Liberto, G. M. D., Attaheri, A., Cantisani, G., Reilly, R. B., Choisdealbha, Á. N., Rocha, S., Brusini, P., & Goswami, U. (2022). Emergence of the cortical encoding of phonetic features in the first year of life (p. 2022.10.11.511716). bioRxiv.

[3] Choi, D., Yeung, H. H., & Werker, J. F. (2023). Sensorimotor foundations of speech perception in infancy. Trends in Cognitive Sciences, 27(8), 773–784.