Modelling predictive learning of the speech signal

Jessie S. Nixon[1] & Fabian Tomaschek[2]
[1]Universität Oldenburg, [2]Universität Tübingen
jessie.nixon@uni-tuebingen.de, fabian.tomaschek@uni-tuebingen.de,

Recent years have seen huge growth in interest regarding the role of prediction in language. So far, the role of prediction of the speech stream as a means by which infants learn the sound system of their language has received little attention. Most statistical learning models of phonetic learning have used clustering methods. But recent work suggests that in a *second language*, error-driven learning contributes to learning speech cues (Nixon, 2020). We modelled early infant speech learning as error driven, using incoming speech signal to predict upcoming speech signal.

We trained a simple two-layer cue-outcome network using error-driven learning (an implementation of Rescorla & Wagner, 1972) on sound files from a child-directed speech corpus. Because the model focuses on young infants, no lexical items or a priori sound units (phonemes, phonetic features) were used in training. Instead, infant learning was simulated with a 25 ms moving window, which moved in 10 ms steps. Intensity (log power values) at different frequency bands (104 equal mel steps) was used as *cues* to predict intensity of upcoming spectral components at each frequency band (*outcomes*). The output of the training was a network of cue-outcome connection weights.

The model was tested for discrimination of vowel and fricative continua based on data from the literature: infant discrimination data for vowels [i - I] (Swoboda et al., 1976) and fricatives [s – ʃ] (Eilers & Minifie, 1975). The model predicted linear perception of the vowel continuum, as found by Swoboda et al. and discrimination of the endpoints of the fricatives, as found by Eilers and Minifie. Interestingly, the model also predicted nonlinear perception of the fricatives – this has not yet been tested for infants, but has been found in adults. Linear vowel perception and nonlinear fricative perception is an interesting prediction of the model, as this is known in speech perception, but not yet well understood. Further model evaluation was carried out and will also be discussed.

In summary, the model – without a priori speech units, trained to use incoming speech to predict upcoming speech – learned to weight cues in such a way as to discriminate vowel and fricative pairs. Thus, predictive, error-driven learning of the acoustic signal may play a role early infant speech learning.

References: • Eilers, R. & Minifie, F. (1975). Fricative discrimination in early infancy. JSHR, 18, 158–167. • Nixon, J. S. 2020. Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. Cognition, 197, 104081. • Rescorla, R. & Wagner, A. 1972. A theory of Pavlovian conditioning. Ap.-Cent-Crofts, NY. 64– 99. • Swoboda, P. et al. (1976). Continuous vowel discrimination in normal

& at risk infants. Child Development, 459–465.