ERROR-DRIVEN LEARNING OF THE CONTINUOUS ACOUSTIC SPEECH SIGNAL

Jessie S. Nixon and Fabian Tomaschek (University of Tübingen)
jessie.nixon@uni-tuebingen.de

Infants begin honing perception to the surrounding language in the first few months of life. It has been argued that perceptual learning of native speech sounds begins too early to result from minimal pairs (Maye et al., 2002). Based on this observation it has been proposed that infants learn in an unsupervised way, through statistical clustering mechanisms. However, a number of computational models have demonstrated that an unsupervised, purely statistical approach may not be sufficient to model acquisition of speech sounds (e.g. Feldman, Griffiths, Goldwater, and Morgan, 2013; McMurray & Hollich, 2009). Error-driven learning models (e.g. Rescorla & Wagner, 1972) propose instead that learning occurs through prediction and prediction error. Perceptual cues are used to predict important events. Based on feedback from the predictions, expectations about future events are adjusted. Previous research has shown that speech sound acquisition in a *second language* appears to be error-driven (Nixon, 2018, 2020; see also Nixon & Tomaschek, 2023). The present study investigates speech sound acquisition in the first language.

In this study, we investigate whether early infant acquisition of speech cues could occur through error-driven, discriminative learning of the acoustic speech signal. Cue weights develop as a result of the cues' informativity for predicting upcoming signal. We use a simple two-layer cue-outcome Rescorla-Wagner network (Rescorla & Wagner, 1972) trained on a corpus of spontaneous conversational speech in German. Because the model focuses on the first few months of life, no lexical items and no a priori sound units, such as phonemes or phonetic features, are used as either inputs or outputs of the model. Instead, 25 ms by 85 Hz spectral-temporal slices of running speech are used as both input cues and outcomes. The model output is a matrix of cue-outcome connection weights.

To gauge model performance, consonant and vowel continua were created and the model was tested against infant fricative and vowel perception data from the literature. Generalised additive mixed models (GAMMs) showed that differences in connection weights to target and competitor sounds occurred in the expected spectral frequency ranges for the different contrasts. Moreover, the model predicted a particular pattern that is known, but remains somewhat of a mystery in speech perception research. Namely, the model predicted a *linear* classification pattern for the vowel pairs and a *non-linear* classification curve for the fricative pairs.

In summary, using unstructured acoustic input cues to predict upcoming signal in running speech, the model learned to weight cues in such a way as to discriminate pairs of vowels and consonants – the standard measure of speech perception ability. The results suggest that error-driven learning of the acoustic signal may be a feasible alternative to statistical clustering models for infant acquisition of speech cues.

References

Best, C. T., McRoberts, G. W., Sithole, N. M. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of experimental psychology: human perception and performance 14 (3), 345

Feldman, N. H., Griffiths, T. L., Goldwater, S., Morgan, J. L., 2013. A role for the developing lexicon in phonetic category acquisition. Psychological review 120 (4), 751.

Maye, J., Werker, J. F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82 (3).

McMurray, B., & Hollich, G. 2009. Core computational principles of language acquisition: can statistical learning do the job? Introduction to special section. Developmental Science.

Nixon, J. S. 2018. Effective acoustic cue learning is not just statistical, it is discriminative. In Proceedings of the 19th Annual Conference of the International Speech Communication Association, September 2-6, Hyderabad.

Nixon, J. S. 2020. Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. Cognition, 197, 104081.

Nixon, J. S., and Tomaschek, F. 2023. Emergence of speech and language from prediction error: error-driven language models. LCN, 38:4, 411-418. DOI: 10.1080/23273798.2023.2197650

Rescorla, R. and Wagner, A. 1972. A theory of Pavlovian conditioning. Black, A. H., Prokasy, W. F. (Eds.), Classical conditioning II: Current research theory. Ap.-Cent-Crofts, New-York. 64– 99.