A gradually developing lexicon leads to robust emergence of phonological features in a neural network

Klaas Seinhorst, Paul Boersma and Silke Hamann — University of Amsterdam
{seinhorst, paul.boersma, silke.hamann}@uva.nl

We present a neural network in which phonological features emerge as a result of auditory and lexical acquisition. Our model requires three levels of representation: an auditory form ([AudF]), a phonological surface form (/SF/), and a lexical form (“meaning”). These levels are implemented as separate layers in the neural network. Earlier work within this model showed that phonological categories emerge at SF after distributional learning of auditory tokens alone (Boersma, Benders & Seinhorst 2020). In Chládková’s (2014: ch. 5) simulations, there was a lexicon in place from the onset of learning; in Seinhorst, Boersma and Hamann (2019) the lexicon was created halfway through the learning process.

We use this last model to explore phonological category emergence when the number of lexical categories and the number of peaks in the auditory distribution are not identical. We assume one auditory continuum, for instance spectral centre of gravity (COG) in sibilants; the auditory tokens are drawn from Gaussian probability distributions. We create a simple language with only two lexical categories, defined on the COG continuum. One of the categories is unimodal, i.e. its probability distribution has one peak: this category is implemented with a central COG (perhaps [ɕ]). The other category is bimodal: this category is implemented with either a low or a high COG (perhaps [ʂ] and [s], respectively). Fig. 1a shows a network at the onset of learning, Fig. 1b shows the same network after 8,000 auditory input tokens (i.e. fixed activations at AudF). At this point, nodes 1 and 5 have specialised in the middle part of the auditory continuum: if we activate a node in this part of AudF and spread the activity to SF, as in Fig. 1b, only these two nodes become active. SF nodes 3 and 6 have specialised in the rightmost part, and nodes 2 and 4 in the leftmost part. From this situation we can conclude that the network has induced three phonological categories at SF, that is, there are now three possible activation patterns, one associated with each peak in the AudF distribution. These categories are directly observable in our model, while they had to be inferred in earlier computational models of auditory distributional learning (Feldman et al. 2009; McMurray et al. 2009).

After these 8,000 learning steps, a lexicon is added to the model, and the network is fed 8,000 sound–meaning pairs (i.e. simultaneous fixed activations at AudF and in the lexicon). The lexicon will tell the learner that there are only two lexical categories, and this new information comes to be reflected in the phonology. In Fig. 1c, nodes 2, 3, 4 and 6 now form a single category, and these nodes are all connected to high as well as low COGs: the former audition-based ternary contrast was updated to a lexically-based binary contrast.

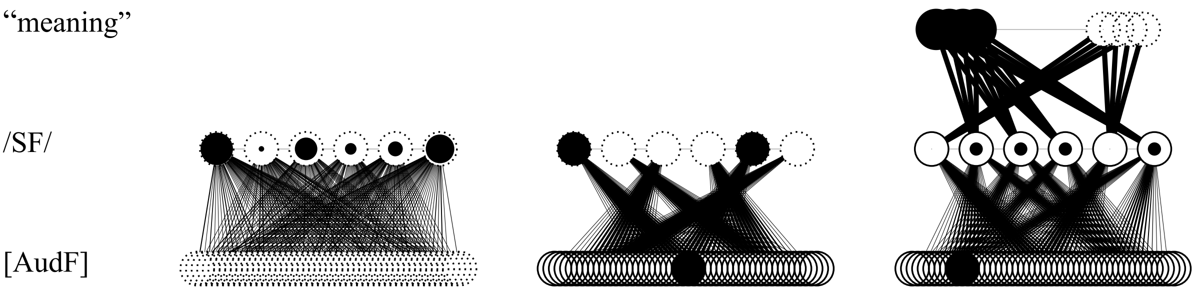

Fig. 1(a). *A network before learning begins: only AudF and SF exist.* **(b).** *The same network after 8,000 learning steps (sounds only).* **(c).** *The same neural network after 8,000 more steps (sound-meaning pairs).*

However, neither the presence of a lexicon from the onset of learning (as in Chládková 2014), nor its sudden creation halfway through the learning process (as in Seinhorst, Boersma and Hamann 2019), is very realistic, as perceptual learning and lexical acquisition in children appear to proceed in tandem (a.o. Swingley 2008). For this reason, we also created a neural network model in which lexical information becomes available to the learner gradually: in this model, the network is fed pairs of sounds and meanings from the onset of learning, but the activations of the nodes in the lexicon layer are only slightly larger than 0 initially, growing logistically towards 1 throughout the learning process. We subject this model to various tests, such as manipulations of the steepness and midpoint of the logistic activation curve in the lexicon, and various kinds of errors in word learning (e.g. inhibition in the lexicon; sound–meaning associations that do not exist in the input language). The model also has an additional auditory continuum, say periodicity, for increased ecological validity. Fig. 2a shows a network that was trained on a language with four lexical categories (perhaps "s ʃ z ʒ") with lexical inhibition, after 12,000 learning steps. Most connections between the lexicon and the phonology resist activation spreading, and the SF–AudF interface looks less organised than the one in Fig. 1b: not all SF nodes have been recruited by a category. However, after 4,000 more steps, all lexical connections are excitatory, and all SF nodes now participate in a category.

"meaning"

/SF/

[AudF]

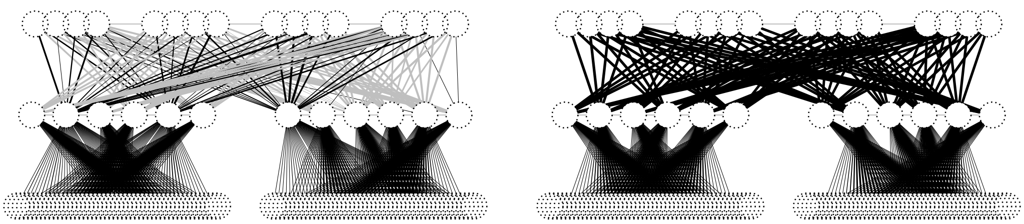

Fig. 2(a). *A network with lexical inhibition after 12,000 steps…* **(b).** *and after 16,000 steps.*

These manipulations do not affect the robustness of phonological feature emergence: in all these scenarios, learners end up displaying the same categorical behaviour found by Chládková (2014), Seinhorst, Boersma and Hamann (2019), and Boersma, Benders and Seinhorst (2020), assuming that in those cases where the word learning process is hampered, the probability that an error occurs decreases over time.

References:

Boersma, Paul, Titia Benders & Klaas Seinhorst (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling* 8 (1): 103–177.

Chládková, Kateřina (2014). *Finding phonological features in perception*. Doctoral dissertation, University of Amsterdam.

Feldman, Naomi, Thomas Griffiths & James Morgan (2009). Learning phonetic categories by learning a lexicon. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*: 2208–2213.

McMurray, Bob, Richard Aslin & Joseph Toscano (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science* 12: 369–378.

Seinhorst, Klaas, Paul Boersma & Silke Hamann (2019). Iterated distributional and lexicon-driven learning in a symmetric neural network explains the emergence of features and dispersion. In *Proceedings of the 19th International Congress of Phonetic Sciences*.

Swingley, Daniel (2008). The roots of the early vocabulary in infants' learning from speech. *Current Directions in Psychological Science* 17: 308–312.