How phonological elements can be both auditory-based and substance-free

Paul Boersma, University of Amsterdam
paul.boersma@uva.nl

The organizers of the GLOW 37 Phonology Workshop ask the following question about "the relationship between phonological primitives and phonetics":

> "As far as we can see, there are roughly three options: one can either assume that the primitives represent elements of articulation (as in most feature theories or in Articulatory Phonology); or elements of acoustics (as in Element Theory). Or is the mainstream view incorrect, in that phonological primitives bear no direct relationship to phonetics at all (as in Substance-Free Phonology)?"

We address this question from the viewpoint of a recurrent Artificial Neural Network. The result of our computer learning simulations is that in the first step, when the infant has access only to sound, phonological categories appear on the basis of auditory information alone; the relationship between these phonological proto-features and audition is arbitrary from a typological standpoint, because it is based on the distributions and correlations of auditory cues in the language environment, but the proto-features can still be described as what the organizers call "acoustics-based". In the next step, the infant has access not just to sound, but to meaning as well; this leads her to develop an indirect relationship between audition and meaning, mediated through an intermediate phonological representation; the learning process sharpens the relationship between audition and phonological features, with the possibility that some of these features merge or split. In the third step, the infant has access to compositional meaning and to morphological alternation; as a result, the computer-learned phonological features are restructured along the lines of natural classes of the undergoers and triggers of phonological alternation. In the final state, the phonological features have become substance-free to a large extent, reflecting the distributions and correlations between the uses of the earlier lower-level features in phonological alternation in the language environment at hand.

Our answer to the organizers' question would thus be that phonological features start out being audition-based, and that they shift towards being substance-free to the extent that such substance freedom is supported by the specific target language.

The ingredients of our recurrent Artificial Neural Network model have been designed to be as general as possible. The network consists of two layers of nodes (in step 1) or of three layers of nodes (in steps 2 and 3). The two-layer version (Boersma, Benders & Seinhorst "Neural network models for phonology and phonetics") has an input layer that represents auditory activity distributions (the most typical example is a tonotopic map of the basilar membrane), and an output layer that starts out being inactive but gradually acquires the ability to represent phonological proto-features. The three-layer version is constructed from the two-layer version by adding a new output layer that can represent meaning; the phonological layer thereby changes its status from being an output layer to being an intermediate layer. Each consecutive pair of layers is exhaustively linked by bidirectional connections, which can just as easily carry information from a lower level to a higher level as from a higher level from a lower level. The phonological layer is connected to the auditory and semantic layers, but the auditory layer is not connected directly to the semantic layer. The bidirectionality of the connections ensures that the network can comprehend as well as produce: listening proceeds by spreading the activation from the auditory layer to the phonological layer and perhaps on to the semantic layer (once it exists), and speaking proceeds by spreading the activation from the semantic layer through the phonological layer to the auditory layer (and from there to an articulatory layer, which we model in the simplest way possible). Stability of the network is ensured by the fact that the connection *between* layers are excitatory and the connections *within* layers are inhibitory. Learning proceeds by a bidirectional update rule for connection weights, the *inoutstar* learning rule, which is a symmetrized version of Grossberg's instar and outstar rules.

The model makes predictions that can be tested. The predicted shift from auditory-based features to (partially) substance-free features can be tested with infants in the lab. The predicted dependence of the final degree of substance freedom on the ubiquity and nature of the phonological alternations in the language at hand can be tested by correlating this diachronically with symmetries and asymmetries in phonological and phonetic changes. Finally, the fact that the model predicts that adult phonological features reflect both their auditory origins and their participation in alternations can guide theoretical phonologists towards considering the possibility that the auditory-based viewpoint and the substance-free viewpoint are not incompatible after all.