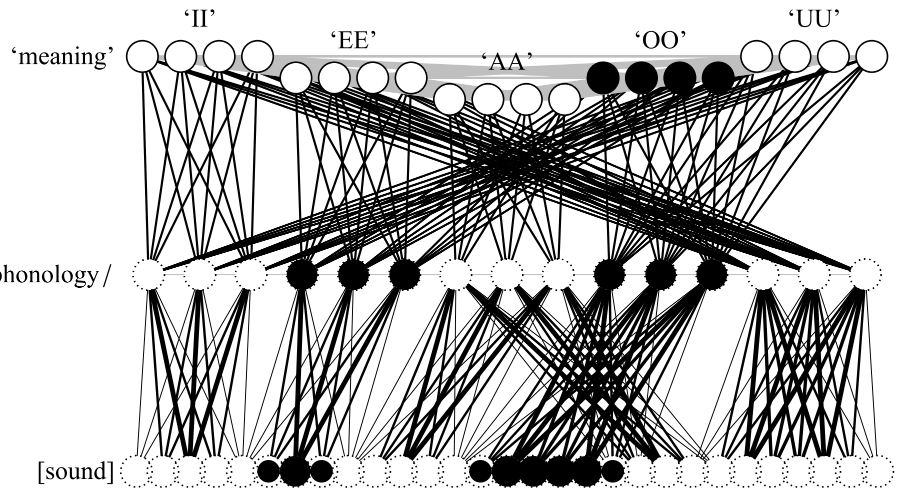
How to learn features from phonetic distributions and phonological alternations

Paul Boersma & Kateřina Chládková, University of Amsterdam

We study artificial neural networks in which phonological elements and constraints emerge from a combination of phonetic and phonological input data and innate restrictions on neural structure. Here we try to answer one question: in a toy vowel-only language, will the emergent phonological elements be **phonemes** or **features**?

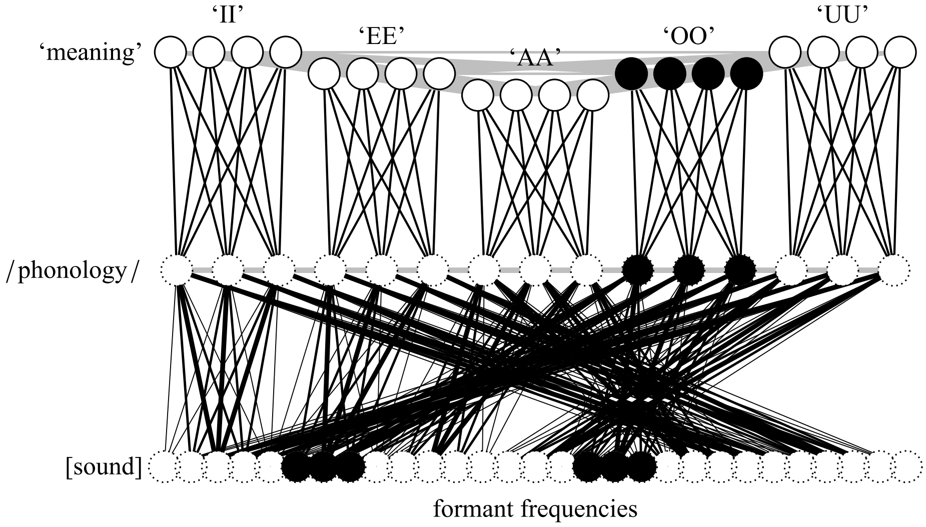
Imagine a language containing only five possible sentences, each consisting of only one word with an indivisible meaning. The (abstract) meaning 'II' is pronounced as [i], the meaning 'EE' as [e], 'AA' as [a], 'OO' as [o], and 'UU' as [u]. The following neural network can produce and perceive all these meanings and sounds:



The figure exemplifies the production of the meaning 'OO'. In the top layer, only this meaning is activated, and all other meanings are inactive. Activity can subsequently spread from the top layer through the middle (phonological) layer to the bottom (auditory-phonetic) layer, through the connections between the nodes. During this activity spreading, the activities of the meaning nodes are kept constant, whereas the activities of the phonological and phonetic nodes are free to be influenced by any nodes with which they are connected, both top-down and bottom-up. After the activities have settled into an equilibrium, the auditory spectrum at the bottom, where low frequencies are on the left and high frequencies on the right, ends up with a mid F1 value (say, 500 Hz) and a low range of F2 values (around 1000 Hz), as is appropriate for the sound [o]. Besides production, the network also performs comprehension, when the sound level is given an auditory input and the upper two levels are free to change their activities. Crucially, the network contains inhibitory connections (the gray lines) within each level.

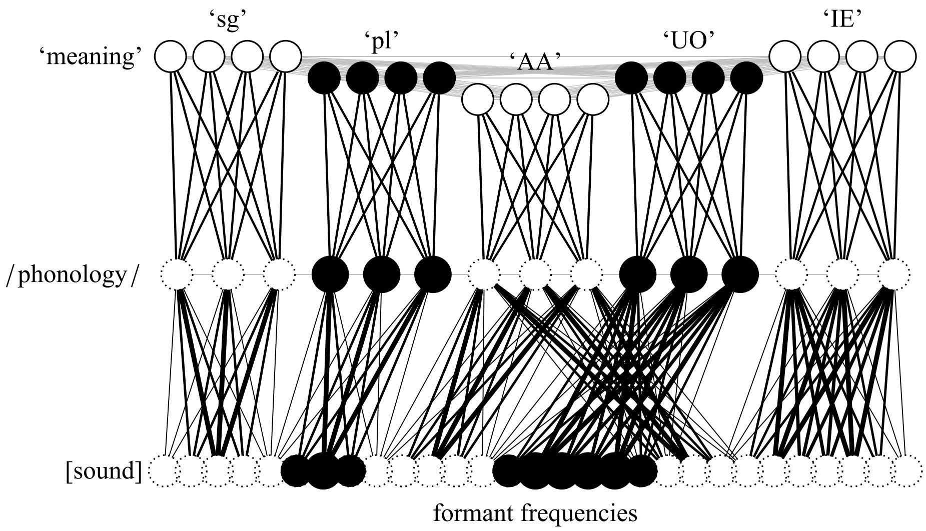
The phonological level in the figure exhibits featural behavior. When studying the connections, you see that the middle level consists of five three-node groups, which correspond to the features *high* (a low F1, shared by [i] and [u]), *mid* (a mid F1, shared by [e] and [o]), *low & central* (a high F1 and a mid F2, i.e. [a]), *back* (a low range of F2 values, shared by [o] and [u]), and *front* (a high range of F2 values, shared by [e] and [i]).

The network can therefore **represent** features, but the question is: can the network also **learn** features? To find this out, we trained the network above in the following way. Initially, all connections were weak and random. Then, we applied 20,000 sound–meaning pairs to the network. At each incoming sound–meaning pair, the activities in the meaning and sound levels were set accordingly and kept constant, and the activities in the phonological level were allowed to be influenced by those above and below, through the connections. After the activities settle into an equilibrium, the connection strengths are modified according to a symmetrized Hebbian learning algorithm. The resulting trained network is not the one above, but a network in which each of the five vowel phonemes has its own representation:



In this network, the five triplets at the middle level represent the phonemes /i/, /e/, /a/, /o/ and /u/, rather than the five features of the previous figure.

We have to conclude that the network above cannot learn features just from the phonetic similarities between some of the vowels, despite the fact that e.g. [i] and [u] share their F1 values, and [u] and [o] have overlapping F2 values. **But will the network perhaps learn features from natural classes, i.e. if there are phonological alternations?** To find this out, we simulated a language with five possible sentences, four of which have a meaning composed of one of the words ‘UO’ or ‘IE’ and one of the number morphemes ‘sg’ (singular) or ‘pl’ (plural). The words ‘UO’ and ‘IE’ are pronounced as [u] and [i], respectively, in the singular, and as [o] and [e], respectively, in the plural. Together, the data point towards the four natural classes {/u/,/o/}, {/i/,/e/}, {/u/,/i/} and {/o/,/e/}. A possible network that results is the following:



The five triplets at the middle level represent the features *high*, *mid*, *low & central*, *back*, and *front*, respectively. We conclude that with this kind of networks, the acquisition of phonological features cannot rely on phonetic similarity alone but has to be helped by the existence of phonological alternations. The empirical prediction for phonology is that the representation of vowels will depend on how strongly they alternate.