Learning from multiple acoustic cues for phoneme acquisition: Dutch infants’ perception, language input, and neural networks

Titia Benders & Paul Boersma
University of Amsterdam, Amsterdam Center of Language and Communication

**Issue:** We investigate how Dutch infants acquire the /ɑ/-/aː/ contrast, which is signaled by vowel duration and spectral quality, to learn whether and how infants combine cues during phoneme acquisition.

**Perception:** 42 11–15-month-old Dutch infants were tested on their discrimination on the full phoneme contrast ([ɑ]–[aː]), the contrast involving only duration (e.g. [ɑ]–[ɑː]), and the contrast involving only quality (e.g. [ɑ]–[a]). As infants discriminate the full vowel contrast better than either single cue contrast (Figure 1), Dutch infants know by 11 months of age that both duration and quality add to the contrast between /ɑ/ and /aː/.

**Input:** Eighteen Dutch mothers were recorded when playing with their infant, and the duration and F2 of /ɑ/ and /aː/ were measured. A mixture of Gaussians model with one cluster was the best fit to the data when only one acoustic cue was given. With both duration and F2 available, a model with more than one cluster proved the best fit, and the two (largest) clusters matched the actual /ɑ/ and /aː/ (Figure 2). Attending to two cues simultaneously appears necessary for Dutch infants to discover they learn a language with two low vowels.

**Simulation:** Preliminary results show that also a neural network is only able to learn the two categories from the input data if both acoustic cues are provided.

**Conclusion:** Infants most likely acquire phoneme categories by attending to all available acoustic cues simultaneously. This suggests that infants acquire features only while learning more complex units.
Figure 1. The infants’ ratio scores on trials with the full-vowel contrast, duration-only contrast, and quality-only contrast. In the experiment, every second trial presented a contrast, while on the other trials a repetition of the same vowel was presented. A ratio score is the infants’ looking time to the trial on which the contrast is presented, divided by the average looking time to the two surrounding repetition trials. A ratio score over 1 indicates that infants react to the contrast.

Figure 2. Top: The vowel duration and second formant of /a/ (red) and /aː/ (blue) in infant-directed speech in Dutch. Bottom: The two clusters from a mixture of Gaussians fitted onto the same data. Note how well the actual categories (top) and estimated clusters (bottom) are in accordance.