Which speech sound categories matter most for word segmentation?

Key Words: *word segmentation, early acquisition, speech perception, computational modeling*

Between 6 and 9 months, infants begin to segment speech (Jusczyk et al., 1999). The distribution of phonemes provides cues for word boundaries, but what if infants have not yet learned all the speech sound categories of their language? How damaged is their ability to learn word boundaries? I use a supervised, probabilistic model of word segmentation to answer this question, DiBS (Daland & Pierrehumbert, 2011). If even the optimal learner is greatly weakened in its ability to learn word boundaries, an unsupervised model, the standard for word segmentation models, would do even worse. I neutralized place of articulation, voice, nasal, and continuant contrasts in the Dutch National Corpus, one contrast for each condition. I trained the model on the corpus, then tested which types of neutralization resulted in the worst damage to the learner, as compared with a control where all contrasts were present (ie, the corpus was unaltered).

Precision and d' for each condition tested are given in Table 1. Results show that indeed some contrasts are more important than others: namely, the neutralization of nasal and continuant contrasts are more damaging to d' and precision than are place or voice. Indeed, precision actually *increases* slightly when place or voice are neutralized.

Most probabilistic, segment-based computational models of speech segmentation assume that infants already know all relevant speech sound categories by the time they begin to segment speech, an assumption I challenge in my model of the input. Even if infants can learn speech sound categories in a laboratory setting, it is likely that they have not yet learned *all* the categories of their language. Further, it is likely that in noisier, real-life listening conditions, they misperceive contrasts that they have learned or partially learned. A large body of literature suggests that some sound categories are more acoustically salient than others (Miller & Nicely, 1955; Smits et al., 2003; Lin, 2005), and also that the more acoustically distinguishable contrasts may indeed be learned earlier (Narayan et al., 2010).

From my results, I make two conclusions. Firstly and unsurprisingly, neutralizing categories is detrimental to model performance; the quality of the input - what an infant might realistically be hearing and retaining - should not be overlooked when evaluating model performance. More interestingly, the sound categories whose neutralization least affected model performance are those that have fewer acoustic cues: place and voice. That is, these results show a correlation between the perceptual salience of a contrast and its usefulness in word segmentation - and by extension, its usefulness to the language learner. This means that even if the learner does not have access to all the existing sound categories of her language, the language is organized such that the learner is less impacted by the lack of sound categories that are less distinctive, and hence more difficult to learn. This is evidence for both learnability and perceptual salience influencing and shaping phonological organization.

Word count: 486

Condition	Control	Place	Voice	Nasal	Continuant
False Positives	123501	102553	101920	114475	123025
True Positives	355613	312260	306327	327946	331804
False Negatives	226741	270094	276027	254408	250550
True Negatives	1468824	1489772	1490405	1477850	1469300
Hit Rate	0.6106	0.5362	0.5260	0.5631	0.5698
False Alarm Rate	0.0776	0.0644	0.0640	0.0719	0.0773
Precision	0.7422	0.7528	0.7503	0.7413	0.7295
d'	1.6840	1.6550	1.6300	1.6270	1.5810

Table 1: Results

References

Daland, R., & Pierrehumbert, J. (2011). Learning diphone-based segmentation. *Cognitive Science*, *35*(1). 2010.

Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, *39*(3-4), 159–207.

Lin, Y. (2005). *Learning Features and Segments from Waveforms: A Statistical Model of Early Phonological Acquisition*. Ph.D. thesis, University of California Los Angeles.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(B101-B111).

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some english consonants. *The Journal of the Acoustical Society of America*, *27*, 338.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420.

Oostdijk, N. (1999). Building a corpus of spoken dutch.

Smits, R., Warner, N., McQueen, J., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of dutch diphone perception. *The Journal of the Acoustical Society of America*, *113*, 563.