The Identifiability of Consonants and of Syllable Boundaries in Infant-Directed English

Corpus-based models of infant phonetic category learning usually assume that infants induce categories from experienced instances, clustering segments into categories defined by statistical distributions. Corpus-based models of word segmentation, in turn, usually assume that infants can categorize each phonetic segment, and sometimes assume that syllable boundaries are given in the signal. Both of these starting assumptions apparently conflict with the well-known result that even whole words extracted from conversation are frequently unidentifiable by adult native speakers (e.g. Pollack & Pickett, 1963) even in child-directed speech (e.g. Bard & Anderson, 1983). Existing learning models are only tenable if segments and boundaries are sometimes locally identifiable. Here, we asked: when are consonants of infant-directed speech recognizable, and to what extent are onset and coda consonants identifiable as such? Setting quantitative bounds on identifiability helps evaluate the plausibility of learning models.

An hourlong session from each of two American English mothers speaking to their 10-month-olds (Brent & Siskind, 2001) was orthographically transcribed, hand-aligned at the word and phone levels, and phonetically transcribed (e.g., Adriaans & Swingley, 2017). The words in virtually all sentences were readily interpretable in context. All vowel-consonant-vowel sequences where the consonant was at a word boundary (either as coda or onset) were extracted into v.cv (n=1008) or vc.v (n=407) 3-segment audiofiles consisting of the consonant and the entirety of the surrounding vowel segments. These files were divided into sets, and presented online to 51 trained native-English adults who judged, for each vcv clip, the identity of the consonant, and the consonant's word position {coda, onset}. Each token was judged by at least 6 listeners.

Analysis of these responses showed that many instances of maternal-speech consonants from these sessions were unintelligible, and many were impossible to assign to a syllable. Considering all consonant categories, the median identification proportion for nominal word onsets was 55.6% (25^{th} %ile, 52.2%; 75^{th}, 59.4%); and for codas, 26.7% (25^{th}, 9.2%; 75^{th}, 35.2%). Figure 1 shows the confusion matrices for onsets and codas, organized by manner of articulation. In most cases, particularly for onsets, the modal response was the correct one. However, many errors remained, and for some sounds at onset (and nearly all of the codas), most sounds were not correctly identified. Perhaps surprisingly, in most cases this was not because sounds competed with phonologically similar competitors, like 1-feature mismatches. There were some such cases, such as voicing errors in fricatives and stops, but for the most part, it seems that sounds were either correctly identifiable, or unintelligible, leading to guessing.

Participants were also quite poor at telling whether a consonant was an onset or coda. Over items, the median success proportion was 60% (25^{th} %ile, 44; 75^{th} %ile, 75). Though significantly above 50%, these proportions also reflected a bias toward responding "onset", which matched the stimuli (71% onsets, following expected distributions from English). Only a third of participants showed a significant contingency (by chi-square test) between their responses and the true syllable position. Could it be that syllable positions were more discernable for the consonants that were easier to identify? Some, but not much. As Figure 2 shows, for only some sounds (mainly codas), consonants' positions tracked identifiability (numbers are *r*s). Put another way, even the easiest-to-identify consonants' syllable affiliations were often a mystery to adult listeners.

Although these results only concern vcv sequences from two mothers speaking to their 10-month-olds, they suggest several conclusions relevant to modeling of early word processing in infants. First, models should not assume that all tokens are good instances for training phonetic categories. More likely, some instances are superior training tokens; the question is whether infants can identify them as such. Second, models that take syllables as inputs to "statistical learning" should not presuppose that syllable boundaries are given in the signal (see Jusczyk et al., 1999). Third, models assuming the emergence of protolexical islands of familiarity seem more plausible than full segmentation models (Goodsitt et al., 1993). Ongoing work assesses the generality of these effects and seeks correlates of identifiability.

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *JASA*, 141, 3070-3078.

Bard, E.G., & Anderson, A.H. (1983). The unintelligibility of speech to children. *JCL, 10*, 265-292.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33-B44.

Goodsitt, J.V., Morgan, J.L., & Kuhl, P.K. (1993). Perceptual strategies in prelingual speech segmentation. *JCL, 20*, 229-252.

Pollack, I., & Pickett, J.M. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, 6, 165-171.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61*, 1465-1476.

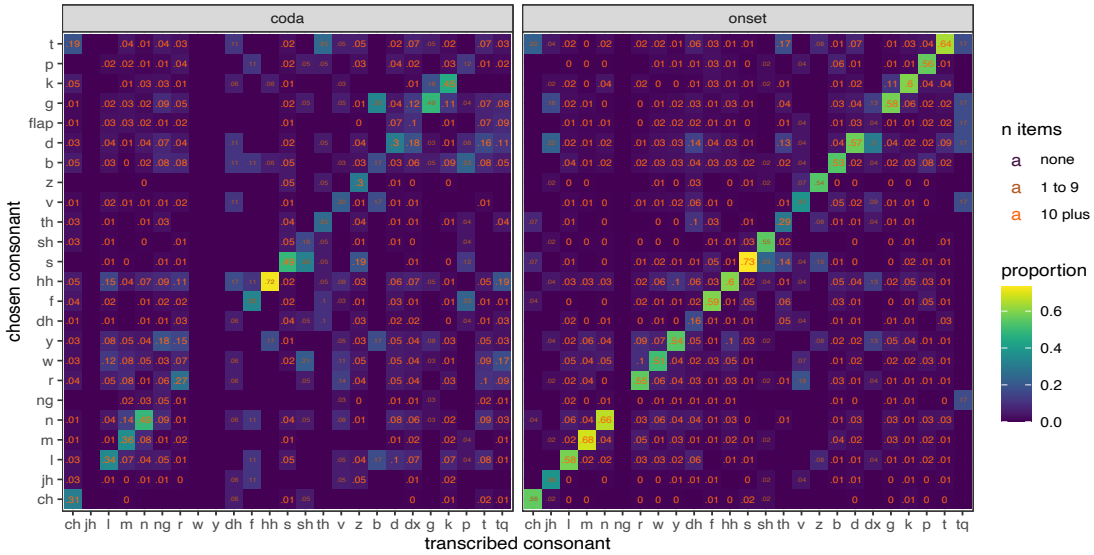

Figure 1. Confusion matrices for codas (left) and onsets (right). Proportion is shown in each cell. Numbers based on <10 items are shown in a smaller, darker font. Warmer colors show greater convergence. Onsets were ID'd correctly about 56% of the time; codas, about 27% of the time.

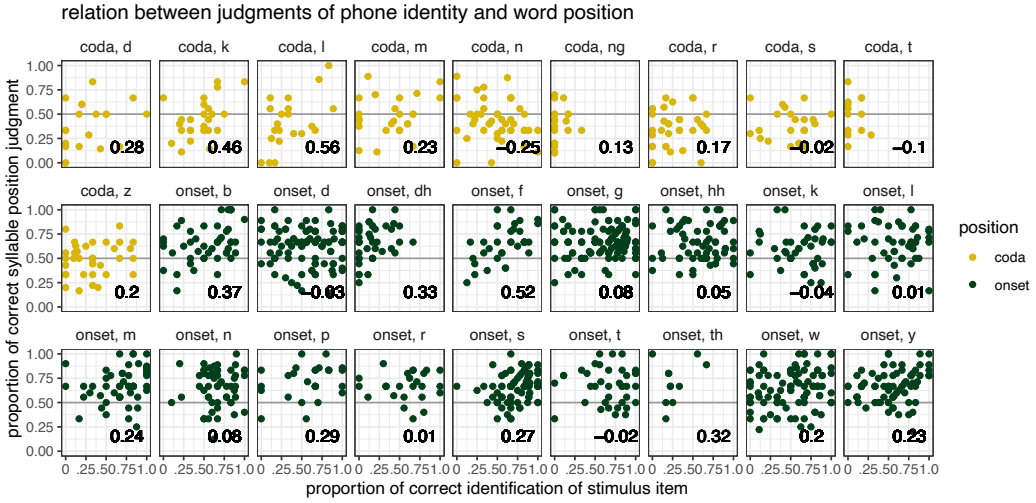

Figure 2. For each item, how often its syllable position was correctly judged (y axis) plotted against how often its identity (which consonant) was correctly judged (x axis), considering consonants tested with more than 10 stimulus items. For most sounds, identifiability of the consonant did not make judgments of syllable position more accurate, even for stop consonants.