Fragile phonetic contrasts in longitudinal infant-directed speech:
Implications for infant speech perception

C.R. Narayan (*York Univ.*), V. Woldenga-Racine (*York Univ.*), A. Peters (*Univ. of Toronto*), K.B. Gorman (*Google Research*)

Abstract
Synchronic typologies and historical sound change suggest phonetic contrasts lie on a cline of robustness, with certain typologically rare contrasts being prone to change over time. The /nV/-/ŋV/ contrast is acoustically fragile, exhibiting considerable overlap between tokens along perceptually relevant acoustic dimensions relative to the acoustically robust /nV/-/mV/ contrast. Infants' perception of this fragile nasal contrast in Filipino (Austronesian) differs from their perception of the robust contrast, which is discriminated early in infancy. We examined the naturalistic infant-directed and adult-directed Filipino of five mothers recorded approximately once a month over the course of one year. Analysis of the nasal-vowel onsets in the infant-directed corpus revealed striking patterns, with 54% of nasal onsets being /n-/, 44% /m-/, and only 2% /ŋ-/. This asymmetric pattern was found within individual mothers' speech as well as at every stage in development over the course of one year. Murmur duration, f0, F2 and F3 (at the NV juncture) of /Na-/ tokens were used to classify tokens in both a supervised and unsupervised fashion. LDAs (supervised) showed very poor separation of velar nasal tokens. Gaussian mixture models (unsupervised) showed broad distribution of velar nasals across five components, while /m/ and /n/ were generally restricted to two different components. Classification of using LDA or GMMs did not improve within speaker over the course of the infant's first year. Implications for speech perception are discussed.

Introduction
- Certain linguistically relevant phonetic contrasts are well separated in multidimensional acoustic space (e.g., English [p]-[k]), while others remain more elusive (English [f]-[θ])[1]
- Consequently the perception of phonetic contrasts lie on cline of more/less discriminable, owing in part to the raw similarity/difference of tokens in acoustic space[2]
- The acoustic salience of phonetic contrasts have been shown to affect their discrimination in development—that is, it is not always clear that contrasts follow one particular developmental trajectory
- While some contrasts (e.g., Hindi [t]-[ʈ])[3] may be discriminable in early infancy, others (e.g., Filipino [n-]-[ŋ-][4]; Eng. [f]-[θ][5]; Mandarin T1-T4[6]) remain more elusive
    - Filipino infants discriminate bilabial/alveolar nasal onsets more consistently than they do velar/alveolar nasal onsets[4]
- Many studies of IDS show acoustic features are *enhanced* relative to ADS[7] (though not for all speech sounds[1])
- We examine the acoustics of nasal-onset syllables in Filipino IDS in comparison with ADS as well as longitudinally in order to determine whether the acoustic nature of velar nasals in mothers' speech might explain speech perception patterns

Methods
Cross-linguistic corpus of infant-directed speech (CCIDS)[7]
- High-quality naturalistic audio recordings of 16 mothers [5 Sri Lankan Tamil, 5 Filipino, and 6 Korean] interacting with their infants (IDS) and adults (ADS) in their own homes over the course of the first year of the infant's development
- The five Filipino participants in the corpus used the standard variety of Tagalog spoken in Manila, Philippines
- The CCIDS project set out to record participants once a month for 12–14 months beginning when the infant was 4 months old
- Over 10hrs of recordings per mother were transcribed by native Filipino speakers

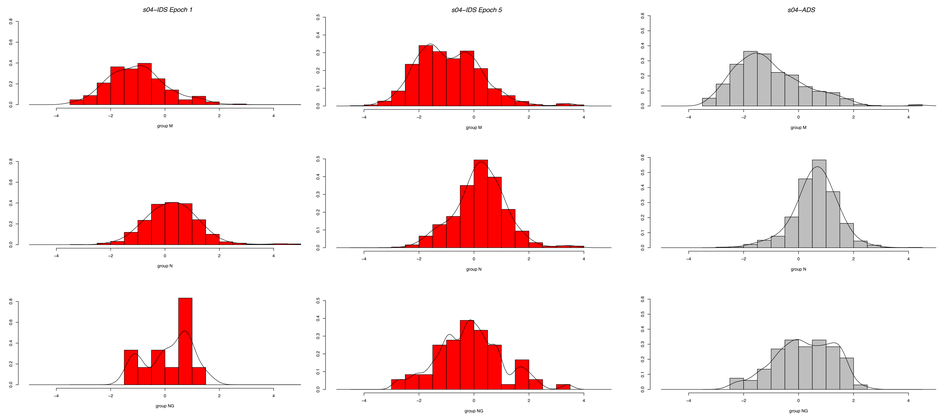
Automatic phoneme alignment
- A Filipino spelling dictionary (in Arpabet) was developed from an existing dictionary to account for all of the words (including English words used in code-switching contexts) in the corpus
- Acoustic models of the data were developed and evaluated (each speaker and register) for forced phoneme alignment using the ProsodyLab aligner
- Two rounds of alignment were checked by phoneticians to correct misalignments. Updated alignments were then used to train subsequent models

Measurements
- Acoustic measurements of NV word onsets were collected automatically and included the following: Nasal murmur duration, Average f_0 of nasal murmur (ms), Average frequency of the nasal resonance/formant (bark), static F1, F2, F3 (bark) taken at the NV juncture
- These measurements had been shown to classify Filipino nasals in carefully articulated lab speech,[2] with additional dynamic measures only marginally improving classification

Results
Across all speakers: 54% /n/, /44% /m/, 2% /ŋ/; /ŋ/ murmur is longer than /m/ or /n/; /ŋ/ has lower F3 and higher F2 than /n/; /m/ lower F2 than /n/ and /ŋ/. Type/token frequency for /ŋ-/ words lower than other nasals and includes very frequent adverbial words, e.g., "nga" (En. *indeed*), "ngayon" (En.*now/today*)

Modeling
Only nasals with following low-back vowels were modeled as they share similar spectral properties at the NV juncture
- *Linear Discriminant Analysis* (LDA)—A supervised method for reducing multiple dimensions (*c*) to at least *c*-1 dimensions
- *Gaussian Mixture Models* (GMM)—An unsupervised probabilistic model where data are generated from a number of Gaussian distributions in multidimensional space
    - Bayesian variational inference was used to select the minimal number of clusters justified by the data and under reasonable settings

Results
LDA
Histograms of first LDA dimension, IDS (epochs 1 and 4) and ADS for speaker 04



- LDA for both IDS and ADS corpora never achieved more than 25% correct classification of velar onsets, while 95% alveolars tokens and 70% of bilabials were correctly classified
- None of the five speakers in the corpus showed an effect of enhancement of velar nasal acoustics as the infant got older
- Acoustics of IDS nasals did not differ from those of ADS nasals in the corpus

GMMs
- Data were modeled using varying number of Gaussian components. Using 2 or 3 components achieve separation between /m/ and /n, ŋ/
- The addition of more components does not clearly segregate /ŋ/ from the other two nasals

Percent of nasal tokens assigned to a given component, three speakers IDS

	s03	/m/	/n/	/ŋ/
Component	0	0.07	0.06	0.23
	1	0.20	0.26	0.23
	2	0.05	0.40	0.07
	3	0.53	0.15	0.31
	4	0.16	0.12	0.15

B-cubed *F1* =0.40

	s04	/m/	/n/	/ŋ/
	0	0.12	0.25	0.10
	1	0.21	0.11	0.19
	2	0.30	0.35	0.44
	3	0.06	0.06	0.11
	4	0.32	0.23	0.17

B-cubed *F1* =0.36

	s07	/m/	/n/	/ŋ/
	0	0.03	0.23	0.12
	1	0.27	0.40	0.29
	2	0.31	0.16	0.22
	3	0.35	0.17	0.23
	4	0.04	0.04	0.15

B-cubed *F1* =0.37

Conclusions and Implications
- Both the supervised LDA and unsupervised GMMs do a decent job of separating /m-/ and /n-/ tokens but a relatively poor job segregating /ŋ-/ from the other places of articulation
- There is neither a register effect (Filipino IDS nasals are *not* more clearly articulated than ADS) nor a longitudinal effect (nasal place acoustics remain relatively similar across the infant's first year)
- Velar nasals are acoustically inseparable from other nasals in IDS and could explain early speech perception results
- How do infants eventually get the three-way nasal contrast? → Feedback from lexical knowledge[8]? → majority of /ŋ-/ tokens in corpus belong to a few types
- How might we better define the acoustics of velar nasals to achieve separation from other classes?

For refs see handout