# AN ACOUSTIC ANALYSIS OF NUNGON VOWELS IN CHILD- VERSUS ADULT-DIRECTED SPEECH

Hannah Sarvasy[1,2*], Jaydene Elvin[2,3], Weicong Li[1,2], Paola Escudero[1,2]

[1]MARCS Institute, Western Sydney University [2]ARC Centre of Excellence for the Dynamics of Language
[3]Department of Linguistics, California State University, Fresno
h.sarvasy@westernsydney.edu.au

## ABSTRACT

Vowels in child-directed speech (CDS) can differ from those in adult-directed speech (ADS) in the same language. Some speech communities, including in Papua New Guinea (PNG), are said to lack special CDS styles. The present study compares vowel acoustics and duration in CDS and ADS for the Papuan language Nungon of PNG. Vowels in conversations between two mothers and their children (2;2 and 3;2) are compared with vowels in monologues by two other women. Multi-point phonetic analyses were applied to evaluate formant trajectories for all six phonemic vowels in CDS and ADS. Preliminary analyses suggest that formant trajectories in CDS are similar to those in ADS, but F0 range variation differs significantly, and the ADS duration difference between short and long vowels is reduced in CDS. These results suggest that vowels in Nungon CDS follow an established cross-linguistic pattern in the ways they differ from ADS vowels.

**Keywords**: Nungon, vowel, acoustics, CDS, ADS, Papuan

## 1. INTRODUCTION

It is well known that adults show different patterns of speech when communicating with infants and children than they do with other adults. The special speech style adults use when addressing children is known as infant- or child-directed speech (IDS/CDS). IDS and CDS typically differ from adult-directed speech (ADS) in several ways. Among these are use of higher pitch and exaggerated pitch contours [3, 4, 6, 9, 19]. In some cases [e.g., 6], there is evidence of increased within-category variation (or formant movement) or hyper-articulation of vowels [3].

Studies have shown that not only do infants prefer IDS over ADS [e.g., 23] as it attracts their attention, but that it can facilitate word recognition [18] and language development in general (see [21] for a review). Finally, when comparing prosodic features in ADS vs. IDS across different emotions, [19] found few differences between the two registers in terms of their acoustic features (e.g., pitch, pitch contour, rhythm, etc.) but robust differences in these features across the different emotions in the IDS and ADS samples.

In this study, we compare vowel production in ADS and CDS in Nungon, a Papuan (non-Austronesian) language spoken by 1,000 people in remote hamlets of the Saruwaged Mountains, northeastern PNG. Nungon belongs to the Finisterre-Huon language group [12] and is actually an umbrella term for the four southern dialects of a dialect chain, grouped together based on morphosyntax and lexicon. Of these, the dialect that has been best-described is that of Towet village [14]; this is the target dialect here.

While many studies show that IDS and CDS is used cross-culturally, one other speech community in PNG (Bosavi) has been said to lack a special CDS style, for ideological reasons. That is, Schieffelin [17] has suggested in the only fully-analysed longitudinal study of L1 acquisition in PNG to date that CDS in the Bosavi speech community lacks at least some of the linguistic modifications attested for English CDS. Schieffelin attributes this to Bosavi language ideology, with parents maintaining that children must hear and repeat 'hard language,' or adult-like speech [17].

Unlike the claims for Bosavi in [17], Towet Nungon is known to have a special CDS style, including modification of certain consonants [15]. For this reason, in this study we investigate whether Nungon CDS also features vowel modification, measured acoustically. Specifically, we target vowel formants, measured across the length of the vowel, pitch range, and duration.

According to Sarvasy [14], Nungon has six phonemic vowels, /i e a ɔ o u/, with contrastive vowel length. The first acoustic-based vowel plot for Nungon was in [14]; more extensive acoustically measured vowel differentiation and duration contrasts are established in [16].

## 2. METHODS

The current study compares data from two different cohorts of two speakers each. Work in progress will compare ADS and CDS for the same speakers, but at this point the two cohorts do not overlap.

## 2.1 SPEAKERS

The ADS cohort includes two women aged ~26 and ~45. Both grew up within the Towet village community and are fluent in the Towet Nungon dialect. The CDS speakers are two women aged ~22 and ~32. Both grew up in Towet village and are fluent in Towet Nungon.

## 2.2 ADS RECORDINGS AND TRANSCRIPTIONS

Vowel acoustics for ADS were analysed based on tokens in narratives recorded between September 2011 and March 2013. The speakers delivered monologues on themes of their own choice such as personal anecdotes, and were recorded in close range using the built-in twin microphones of a Zoom H4N Handy audio recorder with no external microphone, at a 44.1 kHz sampling rate, in WAV format. One narrative of 1:45 duration (for the younger woman) and two narratives of 1:46 and 0:44, as well as a brief introductory statement of 0:21 (for the older woman), served as source material. These were digitally transcribed together with native Nungon speakers in the field in 2011–2013.

## 2.3 CDS RECORDINGS AND TRANSCRIPTION

The two women were recorded monthly for an hour at a time interacting with their children over two years between 2015 and 2017 as part of a longitudinal study of five children acquiring Towet Nungon. Of these monthly recording sessions, two were used for analysis here, one approximately hour-long recording for each speaker. The children were aged 2;2 (female) and 3;2 (male). The women sat in a natural indoor or outdoor setting with the target child, a local research assistant, and one or more others. Interaction was videotaped using a Canon IXUS 190 digital camera held by a research assistant or mounted on a tripod. Interaction was audio-recorded with a Zoom H5 recorder mounted on a tripod and pointed toward the target child. Recordings were done at a 44.1 kHz sampling rate, in WAV format. Recordings were transcribed in Mid-CHAT format [11] by native Towet Nungon speakers on Lenovo laptops in the village.

## 2.5 ACOUSTIC ANALYSIS

The ADS and CDS transcriptions were searched for words that included the six phonemic Nungon vowels in word-initial, word-medial and word-final environments, and not adjacent to nasals. While the ideal was to find the same word, e.g. *agep* ‘firm,’ in all transcriptions, the uncontrolled nature of subject matters made this difficult. If tokens of a word exceeded 20 in a single transcription, only the first 20 tokens were used. The corresponding audio was hand-screened for obviously poor recording quality and obscuring background noise. 223 vowel tokens of eligible words were selected for ADS and 464 for CDS.

The transcriptions were then manually aligned at utterance level before the use of Munich Automatic Segmentation System (MAUS) for forced alignment at the word and phonetic level. WebMAUS [8] performed this task in two steps: first, it conducted grapheme-to-phoneme conversion; then it produced alignment in the ‘language independent’ mode that does not require language training in advance, pragmatic for under-described languages [7, 10]. Later, the segmentation by MAUS at phonetic level was manually checked and adjusted, as there were a large number of cases in which misalignment occurred.

The vowels’ duration, fundamental frequency and formant values were extracted using the analysis techniques of previous studies [e.g., 5]: 30 evenly distributed points starting from the 20% point to the 80% point of the vowel duration in Praat [2]. The obtained series of formant values for each vowel were smoothed using discrete cosine transforms (DCTs). Then the vowel duration and formants values were averaged over the speakers and across the different positions.

# 3. RESULTS

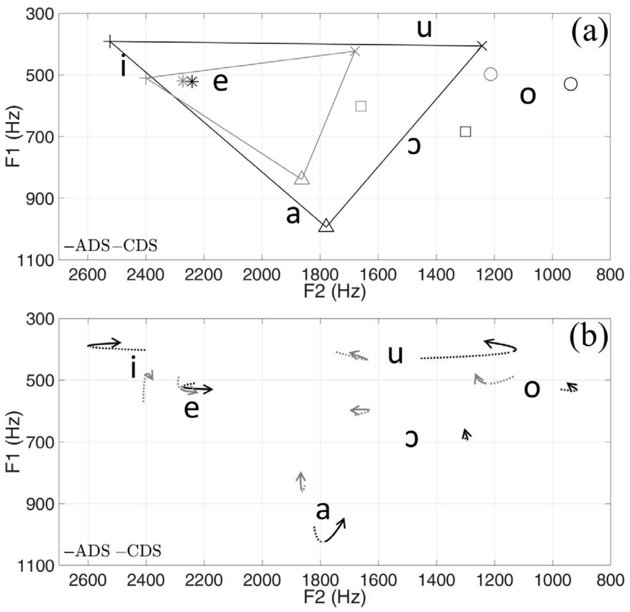
## 3.1. VOWEL PLOTS

Our approach to measuring vowel formants in both ADS and CDS was two-fold: we established the means across speakers of formant values to be able to compare vowel triangles between ADS and CDS, and we also calculated formant trajectories for both registers. Figure 1 shows the results of these measurements.

First, comparison of the vowel triangles for ADS and CDS in Figure 1a reveals a difference in size: while the ADS vowel triangle size is 38039 $Hz^2$, the CDS vowel size is only 14196 $Hz^2$. As in Dutch IDS [1], the Nungon vowel space in CDS is reduced through a decreased contrast in F1 between high and low vowels (increased F1 value for /i/, decreased F1 value for /a/), and increased F2 for the back vowels. [1] showed that the primary factor in Dutch IDS vowel triangle reduction was affect by investigating fricative consonants in IDS and ADS. As we currently lack this data for Nungon, we cannot make this claim definitively.

A full description of the formant trajectories for ADS can be found in [16]. Visual comparison of the trajectories of the two registers in Figure 1b suggests an overall greater degree of formant movement in the ADS corpus than in the CDS corpus for the cardinal vowels /i/, /a/ and /u/. On the other hand, it seems that /ɔ/ and /o/ are more dynamic in CDS than in ADS, with an increase in F2 in each. Interestingly, while /i/ in ADS is primarily characterized by initial fronting and then backing (in F2), the production of /i/ in CDS is characterized by an increase in F1.

**Figure 1**: formants of the six vowels in ADS (black) and CDS (grey): (a) average formant values and vowel triangles; (b) formant trajectories.




### 3.2. FUNDAMENTAL FREQUENCY VARIATION

Following studies like [20], we evaluate F0 range within individual vowel tokens. Table 1 shows the average, mean maximum and mean minimum F0 values for the six Nungon vowels in both ADS and CDS. The vowels in CDS exhibit greater F0 ranges than those in ADS.

We performed a two-sample t-test to evaluate the degree of difference between the F0 ranges (max−min) of all the vowels for ADS (M = 11.54 Hz, SD = 0.85 Hz) and CDS (M = 19.80 Hz, SD = 1.22 Hz). The results were: $t(614) = -4.71$, $p < 0.001$, $d = 0.40$. This shows that F0 ranges differ significantly between the two registers.

**Table 1**: F0 mean, max and min of the six vowels in ADS and CDS.

| vowel | register | F0 mean (Hz) | F0 min (Hz) | F0 max (Hz) |
|---|---|---|---|---|
| /i/ | CDS | 200 | 188 | 211 |
| | ADS | 200 | 195 | 205 |
| /e/ | CDS | 260 | 243 | 271 |
| | ADS | 227 | 218 | 233 |
| /a/ | CDS | 214 | 203 | 224 |
| | ADS | 208 | 203 | 214 |
| /o/ | CDS | 226 | 221 | 231 |
| | ADS | 215 | 207 | 222 |
| /ɔ/ | CDS | 209 | 203 | 216 |
| | ADS | 210 | 204 | 214 |
| /u/ | CDS | 219 | 212 | 229 |
| | ADS | 233 | 228 | 238 |

### 3.3. VOWEL DURATION

We then investigated vowel duration, to see whether vowel duration contrasts established for Nungon ADS in [16] patterned similarly in Nungon CDS.

We performed a two-sample t-test for each phonologically short and long vowel pair in each corpus, comparing, for instance, /a:/ and /a/. In ADS, the two-sample t-test on the duration of long vowels (M = 166 ms, SD = 30 ms) versus short vowels (M = 92 ms, SD = 46 ms) yielded the results: $t(221) = 7.40$, $p < 0.001$, $d = 1.67$. This shows a significant difference in duration between long and short vowels, consistent with [16]. But in CDS, a two-sample t-test on the duration of long vowels (M = 118 ms, SD = 49 ms) and short vowels (M = 105 ms, SD = 82 ms) does not show such a distinction: $t(221) = 1.17$, $p = 0.244$, d = 0.16.

The present study points to a reduction in duration distinctions between phonologically long and short vowels in Nungon CDS, compared with Nungon ADS.

## 4. DISCUSSION

While the current study uses two separate speaker sets, with CDS and ADS produced by different speakers, the results here point to intriguing possibilities to be explored in the future with more complete, intra-speaker analyses.

First, the vowel triangle formed by the three cardinal vowels is significantly smaller in Nungon CDS than in Nungon ADS (Figure 1a). This echoes results for Dutch in [1], among others. It is as yet unclear whether affect (more smiling in CDS than in ADS) is the major factor in this reduction in vowel triangle area for Nungon, as in Dutch. It is possible, however, that a factor in the recording setup for Nungon ADS here made for hyper-articulated vowels in that corpus, instead. In the ADS recordings, the primary interlocutor was a non-native Nungon speaker (Sarvasy), and such situations are known to produce hyper-articulation of vowels [21].

For vowel trajectories, Figure 1b shows no clear distinction between the shapes of vowel trajectories in Nungon CDS and ADS.

All vowels in CDS exhibit greater pitch ranges than those of ADS. Not only is there on average a greater F0 range within each CDS vowel token than within those of ADS, CDS mean pitch values also have greater standard deviation than those in ADS, indicating more variation. A two-sample t-test shows there is no significant difference in F0 between CDS and ADS, but a Levene's test shows that ADS and CDS have different variance levels in F0 ($F[1, 614] = 68.5$, $p < .001$).

For duration, two-sample t-test results indicate that the vowel duration contrasts of ADS are neutralized in CDS. However, duration results are limited by the fact that some of the data for long vowels in CDS (/i/, /ɔ/ and /u/) are missing from the current dataset.

## 5. CONCLUSION

This study represents a first foray into comparison of CDS and ADS for Nungon, an under-described language of PNG. Results point to a reduced vowel triangle in CDS compared with ADS, greater pitch range within individual vowels in CDS than in ADS, and less duration contrast between phonologically long and short vowels in CDS than in ADS. We did not find, however, a clear pattern of differentiation between vowel trajectories in Nungon CDS and ADS.

## ACKNOWLEDGMENTS

This study was supported by a Language Documentation grant to Sarvasy through the ARC Centre of Excellence for the Dynamics of Language, CE140100041, and MARCS Institute funds for Li. We would also like to thank Jason Peed for help with the manual segmentation and adjustment of the data.

## REFERENCES

[1] Benders, T. 2013. Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, 36, 847–862.

[2] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

[3] Burnham, D., Kitamura, C., Vollmer-Conna, U. 2002. What's new, pussycat? On talking to babies and animals. *Science*, 296(5572), 1435–1435.

[4] Escudero, P., Mulak, K. E., Elvin, J., Traynor, N. M. 2017. "Mummy, keep it steady": Phonetic variation shapes word learning at 15 and 17 months. *Developmental Science*, 21(5), e12640.

[5] Kashima, E., Williams, D., Mark Ellison, T., Schokkin, D., Escudero, P. 2016. Uncovering the acoustic vowel space of a previously undescribed language: The vowels of Nambo. *J. Acoustic. Soc. Am.*, *139*(6), EL252–EL256.

[6] Katz, G. S., Cohn, J. F., Moore, C. A. 1996. A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, 67(1), 205–217.

[7] Kisler, T., Schiel, F., Sloetjes, H. 2012. Signal processing via web services: The use case WebMAUS, in *Digital Humanities Conference* Hamburg.

[8] Kisler, T., Reichel, U.D., Schiel, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347.

[9] Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Lacerda, F. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684–686.

[10] Jones, C., Demuth, K., Li, W., Almeida, A. 2017. Vowels in the Barunga variety of North Australian Kriol. *Proc. Interspeech 2017* Stockholm, 219–223.

[11] MacWhinney & Bates CHAT manual

[12] McElhanon, K. 1973. *Toward a Typology of the Finisterre-Huon Languages.* Canberra: Pacific Linguistics.

[13] Sarvasy, H. 2013. Across the great divide: How birth-order terms scaled the Saruwaged Mountains, Papua New Guinea. *Anthropological Linguistics* 55:3, 234–255.

[14] Sarvasy, H. 2017. *A Grammar of Nungon: A Papuan Language of Northeast New Guinea.* Leiden: Brill.

[15] Sarvasy, H. Forthcoming. The root nominal stage: a case study of early Nungon verbs. *Journal of Child Language.*

[16] Sarvasy, H., Elvin, J., Li, W., Escudero, P. Forthcoming. Vowel acoustics of Nungon, Papua New Guinea. *ICPhS 2019*.

[17] Schieffelin, B. B. 1990. *The Give and Take of Everyday Life: Language Socialization of Kaluli Children*. Cambridge: CUP.

[18] Song, J. Y., Demuth, K., Morgan, J. 2010. Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, 128(1), 389–400.

[19] Trainor, L. J., Austin, C. M., Desjardins, R. N. 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11(3), 188–195.

[20] Trainor, L. J., Desjardins, R. N. 2002. Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9(2), 335–340.

[21] Uther, M., Knoll, M. A., Burnham, D. 2007. Do you speak E-NG-LI-SH? A comparison of foreigner- and

infant-directed speech. *Speech Communication*, 49(1), 2–7.
[22] Wegmann, U. 1994. Orthography paper: Yau. Unpublished ms. Ukarumpa: Summer Institute of Linguistics.
[23] Werker, J. F., Pegg, J. E., McLeod, P. J. 1994. A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development*, 17(3), 323–333.