

# Production of Long and Short Vowels in Malayalam

Anugil Velayudhan<sup>1</sup>, Sandra Surendran<sup>2</sup>, Rahul Krishna H S<sup>3</sup>

<sup>1</sup>Department of Language and Linguistic Science, University of York, UK.

<sup>2</sup>School of English, Communication and Philosophy, Cardiff University, UK.

<sup>3</sup>Department of Linguistics, University of Kerala, India.

av906@york.ac.uk, SurendranS@cardiff.ac.uk, rahulkrishna@keralauniversity.ac.in

## 1. Introduction

Malayalam—a South Dravidian language—retains all Proto-Dravidian vowels (/a, i, u, e, o/) with phonemic length contrast [1]. Although prior studies (e.g., [2], [3]) have addressed Malayalam vowel acoustics, none have analysed their dynamic temporal patterns. In this study, we employed a three-point model that captures formant values at 20%, 50%, and 80% of vowel duration, a method shown to enhance vowel identification and classification accuracy [4]. Accordingly, the objectives of this study are: (1) to examine durational and acoustic differences across gender (Male/Female) and speaking conditions (IS/CW); and (2) to evaluate vowel classification performance using three-point formant measurements with (F+D) and without duration (F+NoD).

## 2. Methodology

Twelve native Malayalam speakers (6 male, 6 female), aged 20–30, were recorded in a soundproof studio at 44.1 kHz and 16-bit resolution. Each speaker produced 10 vowels in isolation (IS) and in a /V<sub>1</sub>.C<sub>1</sub>V<sub>2</sub>/ word context (CW), with five repetitions per condition. Of the 1,200 tokens, 10 were excluded due to mispronunciations or noise. Vowel duration and F1–F3 values were extracted using a Praat script using *To Formant (Burg)* method, with maximum formant frequencies of 5 kHz (male) and 5.5 kHz (female), 5 formants, a 0.025s window length, and a 0.01s time step. A linear mixed-effects model (LMM) with a random intercept for speaker was fitted using the *lme4* package to test the effects of gender (M/F) and condition (IS/CW) on vowel duration and formants (F1–F3) at intervals (20%, 50%, and 80%). Statistical significance was assessed using likelihood ratio tests (ANOVA). Linear Discriminant Analysis (LDA) was used to evaluate vowel classification, with accuracy for each vowel assessed against a 10% chance level.

## 3. Results

<sup>1</sup> In general, female speakers exhibit consistently higher F1 and F2 values across both long and short vowels compared to male speakers. Notably, the short vowel /a/ demonstrates greater variability among speakers of both genders (Figure 1).

**Main Effects:** Gender and condition significantly affect F1 and F3 (Gender– F1:  $\chi^2(1) = 25.40$ , F3:  $\chi^2(1) = 13.26$ ; Condition– F1:  $\chi^2(1) = 31.25$ , F3:  $\chi^2(1) = 434.12$ , all  $p < 0.001$ ). F2 is unaffected by gender and condition. Intervals (formant dynamics) show marginal effects on F1 and F2, but a strong effect on F3 ( $\chi^2(2) = 25.05$ ,  $p < 0.001$ ).

**Durational Analysis:** Gender did not significantly affect vowel duration ( $\chi^2(1) = 1.93$ ,  $p = .165$ ), whereas condition

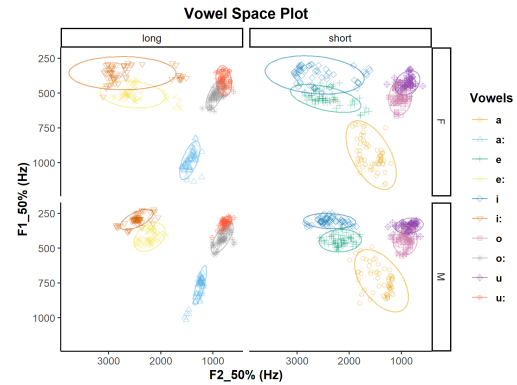


Figure 1: Raw F1 and F2 values for long and short Malayalam vowels measured at 50% of vowel duration, from 6 male and 6 female speakers.

had a significant effect ( $\chi^2(1) = 1030.7$ ,  $p < .001$ ). Long vowels had a significantly greater mean duration ( $M = 364$  ms) compared to short vowels ( $M = 188$  ms), with an estimated difference of 176 ms ( $t(1180) = 56.38$ ,  $p < .0001$ ), indicating an approximate 2:1 ratio [3].

**LDA Scores:** Including duration along with formants improves classification accuracy for all vowels. Short vowels show greater gains, with /a/ increasing from 54.17% to 97.50%. The highest accuracy without duration is for /a:/ (86.21%), and the lowest is for /i/ (50.00%). With duration, /a/ achieves the highest score (97.50%), while /o:/ has the lowest (84.03%). The overall classification accuracy of short and long vowels increases from 64.45% (without duration) to 89.68% (with duration). These results confirm that vowel duration is a key feature for distinguishing short and long vowels in Malayalam.

## 4. References

- [1] B. Krishnamurti, *The Dravidian Languages*, Cambridge University Press, pp. 48–77, 2003.
- [2] S. Namboodiripad and M. Garellek, “Malayalam (Namboodiri Dialect),” *Journal of the International Phonetic Association*, vol. 47, no. 2, pp. 109–118, 2017.
- [3] S. Velayudhan and J. M. Howie, “Acoustical measurements of distinctive vowel quantity in Malayalam,” *Language and Speech*, vol. 17, no. 1, pp. 95–102, 1974.
- [4] W. Almurashi, J. Al-Tamimi, and G. Khatlab, “Static and dynamic cues in vowel production in Hijazi Arabic,” *The Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. 2917–2927, 2020.

<sup>1</sup>We did not include any normalised data results due to the space limitation.