

Machine Learning Analysis of Sinhala Vowel Length Perception

Authors:

Kumudri Goonetilleke, University of Aizu

Alexander Kilpatrick, University of Aizu

Rikke Bundgaard-Nielsen, University of Melbourne

Abstract ID: 322

Event: Phonetics 2026 Hong Kong

Subject: 3. Speech perception

Presenter Name: Kumudri Goonetilleke

Abstract

Japanese and Sinhala share strikingly similar vowel inventories. Japanese employs five vowels (a, i, u, e, o), while Sinhala contains these same five plus the additional vowel /æ/, with both languages exhibiting robust phonemic contrasts between short and long versions of each vowel and comparable duration ratios of approximately 2.6:1. This structural overlap makes Japanese listeners a particularly suitable population for investigating cross-linguistic vowel length perception. While both systems encode duration phonemically, they differ in vowel quality distributions, providing an opportunity to examine how phonetic similarity, durational contrast, and unfamiliar vowel categories interact in perception. The central question is whether Japanese listeners treat Sinhala vowel length contrasts as phonological categories independent of vowel quality, or whether unfamiliar vowel quality disrupts the perception of an otherwise familiar durational contrast.

The present study investigated how native Japanese listeners perceive vowel duration contrasts in Sinhala using controlled /hVba/ sequences, where V = /a, i, u, e, o, æ/ (each realized as short or long). This design allows us to test how well Japanese listeners generalize duration-based phonemic contrasts to a non-native but structurally similar system, and whether unfamiliar vowels disrupt category formation. The experiment comprised two parts. In Part 1 (Identification and Goodness Rating), participants heard naturally produced /hVba/ tokens and classified each as short or long, assigning a 1–7 goodness-of-fit score. In Part 2 (AXB Discrimination), participants performed a forced-choice task deciding whether B is more like A or X with respect to vowel length. Accuracy and reaction times were recorded to assess perceptual sensitivity, decision confidence, and cue weighting.

We expected Japanese listeners to show high overall accuracy and clear categorical clustering for vowels overlapping with their native system. However, systematic deviations were predicted for /æ/, which is absent from the Japanese inventory. According to the Perceptual Assimilation Model (PAM), /æ/ is likely assimilated to /a/, leading to category ambiguity, lower accuracy, and longer reaction times.

To extend the analysis beyond descriptive statistics, we apply an XGBoost classifier to predict trial-level classification accuracy from features including vowel quality, duration category, listener group, and reaction time. To the best of our knowledge, this is the first application of XGBoost to this type of task. Feature importance scores will quantify the relative contribution of vowel quality versus duration to perceptual accuracy across vowel types, complementing traditional statistics with a data-driven account of cue weighting. It also offers potential for informing automated pronunciation assessment tools handling cross-linguistic variation in perceptual categories. Existing results show that duration-based contrasts generalize successfully for vowels shared between inventories. These findings contribute to models of cross-language speech perception and have implications for AI-powered pronunciation training systems accommodating learners with typologically similar but non-identical native phonologies. The planned XGBoost analysis will further quantify cue weighting patterns and mitigate collinearity in the dataset. This is expected to corroborate PAM-based predictions at the trial level.