

L2 PROSODIC FEATURES IN AUTOMATIC ACCENT CLASSIFICATION

Leônidas Silva-Jr

Abstract

This study aims to determine the influence of L2 prosodic features on automatic accent classification. Specifically, we investigated which features—durational, melodic, intensive, or voice quality—are most effective at classifying accents as either 'Native' (L1-English) or 'Foreign' (L2-English and L1-Brazilian Portuguese). Our hypothesis is that a multidimensional matrix of prosodic features is necessary to finely distinguish L1-L2 linguistic differences and enhance automatic accent classification. For Methodology, this research integrates principles from phonetics, L2 prosody, and Artificial Intelligence (AI). The dataset comprised 160 read-speech samples from three groups: 80 L1-English (L1E) American speakers, 40 L2-English (L2E) proficient Brazilian speakers, and 40 L1-Brazilian Portuguese (L1BP) speakers. Samples were based on a phonetically balanced text (an Aesop's fable). Acoustic processing involved forced-alignment via Montreal Forced Aligner (MFA), manual correction, re-alignment into prosodic-level units, and automatic feature extraction using a Praat script. Statistical analysis included a Kruskal-Wallis test followed by a Dunn test for pairwise comparisons (L1E-L2E, L1E-L1BP, L2E-L1BP). Finally, L1E was categorized as 'Native' and L2E/L1BP as 'Foreign' targets for the Automatic Speech Recognition (ASR) system. Preliminary results indicate that long-term spectral (and other intensive) features of voice quality, followed by durational features, were consistent in differentiating the language groups globally and in pairwise comparisons. These features also showed a moderate-to-high influence on the accent classification performance. The Machine Learning algorithms achieved classification accuracy levels ranging from 71% to 100%. Variables related to duration and intensity were found to be significant predictors in the accent classification models. The major conclusion is that prosodic acoustic features, particularly those related to intensity and duration, are highly influential in the automatic classification of foreign accent. The significance of this study extends to L2 pedagogy and to the L2 Forensic field. For the former, the predictive power of prosodic features suggests that current practices in pronunciation classes that prioritize a segment-narrow-focus approach should be revisited and re-prioritized to align with suprasegmental instruction (e.g., teaching stress, rhythm, intonation, and voice modulation). For the latter, the identification of reliable, automatically-extracted prosodic features provides new, measurable acoustic parameters that can be applied to speaker profiling and (foreign) accent characterization in unknown or disguised speech samples. This alignment enhances both the technical performance of ASR models and the potential for improving L2 communication effectiveness and forensic speaker analysis.

Keywords: L2 Prosody. ASR. Acoustic Phonetics. L2 Pronunciation Pedagogy. L2 Forensic Field.