Articulation rate as a metric in spoken language assessment

Calbert Graham, Andrew Caines, Paula Buttery, Francis Nolan
University of Cambridge
crg29, apc38, pjb48, fjn1 (all @cam.ac.uk)

The speech patterns of non-native speakers often differ from those of native speakers in complex ways. Research over the past few decades has documented many segmental differences between native and non-native speech. More recently, there has been a growing body of research focusing on the non-segmental (prosodic) aspects of L1 and L2 speech production. Both segmental and prosodic differences between L1 and L2 speech have been described in terms of both learners' realisation of acoustic phonetic properties and their perceptual ability compared to native speakers. The observed tendency is for L2 learners to exhibit a foreign accent in their L2 speech, the degree of which depends on a number of factors including their L2 proficiency, native language and age of onset of acquisition (cf. Piske et al. 2001 for a review).

Foreign-accentedness ratings are often used in these studies as a measure of the intelligibility and comprehensibility of L2 speech. Automated pronunciation scoring provides a consistent and more cost-efficient alternative to human marking. In automating the process of pronunciation assessment there is considerable interest in deriving metrics that are based on the perceptual cues humans use to judge pronunciation. Previous research reported on the use of phonetic features such as vowel characteristics in automated spoken language assessment (e.g. Graham et al. 2016). The present study extends this work on the significance of phonetic features in automated evaluation of L2 speech by investigating the use of articulation rate as a predictor of proficiency score. The dataset analysed in this study is taken from Cambridge English BULATS test of business English comprising elicited spontaneous speech (in the form of a short bio and a monologue testing the business knowledge of the candidate). There were 240 speakers (approx. 1000 recordings). Speakers are of various source language backgrounds (i.e. Gujarati, Polish, Arabic, Dutch). The data were orthographically transcribed using multiple crowd-sourcers and a speech recogniser according to the procedure described in van Dalen et al. 2015. The transcribed data were then automatically segmented and aligned using a HTK MLP-based algorithm to determine word and phone boundaries.

We used predictive modelling techniques to evaluate the relationship between various articulation rate metrics and the proficiency score on non-native speakers. The metrics are derived from articulated segments (phones) as the unit of measurement, taking stress and the differences in the articulation of different segment types (e.g. fricatives vs. stops) into account.

We found that the optimal predictive model was one in which the phonetic details of phoneme articulation were taken into account in the analysis of articulation rate. We also found that model performance varied according to the L1 background of speakers. The implications of these results are discussed.

References

Graham, C., Buttery, P., & Nolan, F. Vowel characteristics in the assessment of L2 English pronunciation. Interspeech 2016, pp. 1417-1422, Causal Productions.

Piseke, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. Journal of Phonetices, 29, 191-215.

van Dalen, R., Knill, K., Psiakoulis, P., & Gales, M. (2015). Improving Multiple-Crowd-Sourced Transcriptions Using a Speech Recogniser. ICASSP.