

## Abstract

Forced alignment software is now widely used in contemporary sociolinguistics, and is quickly becoming a crucial methodological tool as an increasing number of studies begin to utilise ‘big data’. This study investigates the possibility of taking forced alignment one step further towards the goal of complete automation; specifically, it expands the functionality of FAVE-align to fully automate the coding of three sociolinguistic variables in British English: (th)-fronting, (td)-deletion, and (h)-dropping.

This involved the expansion of pronouncing dictionaries to reflect the surface output of these variable rules; FAVE then compares the fit of competing acoustic models with the speech signal to determine the surface variant. It does so with an impressive degree of accuracy, largely comparable to inter-transcriber agreement for all variables; however, the pattern of its mistakes, which are largely false positives, suggests a difficulty in identifying the voiceless segments of (td) and (th). Although it is reassuring that inter-transcriber agreement was also lowest for these tokens, it should be noted that FAVE’s accuracy decreases in faster speech rates while no comparable effect is found for agreement among human transcribers.