Free Classification: Clustering Project

Clustering

## 0.1 Data

Eighty-four participants (one participant was not used due to some weird coding issue) took part in a free speech classification task. The speech samples were selected from The Speech Accent Archive. The talkers included three American English regional dialects, three international English dialects, and nine nonnative accents. The nonnative accents were split into three accents from East Asia, three accents from South Asia, and three accents from Southeast Asia. The American English dialects included the New England dialect, the Southern dialect, and the Midland dialect. The international English dialects included British English, Australian English, and Africaans. The native languages of the nonnative-accented talkers were Mandarin, Korean, and Japanese from East Asia, Bengali, Gujarati, and Urdu from South Asia, and Indonesian, Tagalog, and Thai from Southeast Asia. We have data in wide format. Each row is a talker type and each col is a participant.

# 1 Full Code

The full script of executive code contained in this document is reproduced here.

# 2 References