Design of Varshney-Chung English Accent Corpus
Lav Varshney

The United States have been divided into accent regions by various dialectology researchers studying American accents of English. Central New York has been placed in the Inland North [1], General American [2], and North-Central [3] regions by different linguists. One distinctive accent feature that is found in this accent is the use of the vowel /æ/ in a word like half, and in fact forms part of the British stereotype of an American accent [4]. In addition, this accent is rhotic, meaning that speakers pronounce written r in all positions [4].

A large majority of Singapore inhabitants are ethnically Chinese. Consequently, many of the features of Singaporean English are attributed to interference from the Chinese phonological core, for example limitations on the range of possible syllable-final consonants. One of the most prominent features of Singaporean English is the use of syllable-timed rhythm, as opposed to the stress-timed rhythm of other accents. This is something that applies to all that is said. There are also occasional unusual placements of word stress, for example college. The vowel system is of the British type. The voicing distinction is often neutralized in the word-final position, all obstruents being made voiceless, so that believe and belief become homophones in [-f] [4].

The accent of educated Indians for whom English is a second language, the first language being Hindi or some other indigenous language of the sub-continent is referred to as General Indian English. This accent is very distinct from the Anglo-Indian accent associated with the community of the same name, who speak English as their first language. In General Indian English, the vowel system is much like the British system. It is generally rhotic. One feature of Indian English is the absence of /v/ vs. /w/ opposition. The voiceless plosives, /p,t,k/, are typically unaspirated in all positions. Nearly all Indic languages exploit the presence vs. absence of aspiration phonologically, like ह in vs. एr opposition, but the English phonemes have come to be equated with the unaspirated rather than the aspirated phonemes. In words spelt gh-, spelling pronunciation leads northern Indian speakers of English to use /gθ/- [gh], thus [ghost] ghost, as compared with British /gθəst/. Some people pronounce [θ] or [pʰ] in place of [f], for example [pʰɪT] fit. Some speakers distinguish words spelt with wh from those spelt with plain w by pronouncing the former with [uθ] or [uθ], thus which [uθɪT] ~ wɪθɪT] vs. witch [wɪT] ~ wɪT]. In terms of word stress, one major variant is the pronunciation and stress on the word atmosphere [ˈætmosfɪr] [4].

Sentences in Corpus
1. I believe my colleague liked the private atmosphere in Kenya.
2. Which is bigger, a ghost or a witch?
3. The dwarf’s throne was in a huge castle encircled by a deep moat
4. Aluminum silverware can often be flimsy
We are designing an accent classification system to classify the English accents of North India, Singapore, and Central New York; the corpus was designed with this in mind. The first two sentences in the corpus are designed to highlight the differences in these accents and facilitate classification, whereas the third and fourth sentences are general phonemically rich sentences. Sentence 1 and Sentence 2 were designed by Lav Varshney to incorporate known accent variants in North Indian, Singaporean, and Central New Yorker accents of English, as described above. Sentence 3 was taken from the Australian National Database of Spoken Language, and was selected somewhat randomly. Sentence 4 was taken from the TIMIT (Texas Instruments – Massachusetts Institute of Technology) Corpus, and was also selected somewhat randomly.

References