

Abstract

As globalization increases, learning a second language became necessity. It provides us to interact with other communities, expands social networks, helps in migration and education. Second language (L2) acquisition refers to the process through which learner learn a language beyond their first (L1) or native language. The challenge in learning a second language is the learner try to use their L1 phonology to L2 language while speaking. English is the worlds most spoken language after including native and non-native speakers. Local languages and dialects blended with native English and gave rise to a wide variety of world English. In India, ‘English’ used as a language of greater communication among the people. Bengali is the seventh most spoken language in the world. It is the official state language of the Eastern Indian state of West Bengal and the national language of Bangladesh. Most of the Bengali speakers wants to speak English fluently. But, due to phonemic (segmental) and prosodic (suprasegmental) variation between Bengali and English, Bengali learner find it difficult to learn English. Also, classroom based learning is limited and getting individual feedback about their pronunciation by the teachers is difficult.

In last few decades, technology was emerging and technology based learning became popular. Computer assisted pronunciation training (CAPT) tool helps in pronunciation learning by giving proper feedback automatically after detecting and diagnosis of pronunciation errors in learner speech using mispronunciation detection and diagnosis (MDD) module. But for non-native (L2) speech scenarios, getting accurate feedback is challenging due to language transfer which occurs at both segmental and suprasegmental levels. While most existing MDD systems focus primarily on segmental errors, suprasegmental aspects are often underexplored, especially in the context of children’s L2 speech. This gap becomes more critical for low resource language pairs, such as Bengali accented English, where annotated corpus and robust modeling approaches are scarce. For automatically detecting pronunciation errors in L2 speech, it is necessary to develop a robust methods that can model variations of language transfer effectively. For this purpose, this thesis focuses to enhance phonetic

and prosodic MDD for low resource Bengali accented English speech produced by children (L1 as Standard Colloquial Bengali).

This thesis introduces a Bengali accented English (BengLish) children speech corpus for studying pronunciation errors in non-native (L2) child speech. The corpus is designed to capture both segmental and suprasegmental variations influenced by L1 transfer and includes detailed multi tier annotations at word, phoneme, syllable and stress levels. The corpus analysis reveals systematic L1 transfer effects including vowel uncertainty, consonant substitutions and epenthetic insertions. Significant suprasegmental challenges, particularly in lexical stress and rhythm, were also observed due to Bengali influence. This newly created corpus support the development, training and evaluation of MDD systems.

For automatically detecting segmental errors, a framework is proposed which utilized Branchformer encoder based attention-connectionist temporal classification (CTC) model that capture both global and local features which improve mispronunciation detection performance on the newly created corpus. The results show enhanced accuracy, higher F1 score and reduced phoneme error rate compared to baseline methods. After analysis, it is reveal that proposed architecture able to capture L1 influence with frequent substitution for non-native phonemes. Results also show that the phoneme distribution affects performance of models which suggests that increasing low frequency phonemes can further enhance the model performance.

Finally, a lexical stress detection system is presented for suprasegmental error analysis for non-native speech. The architecture integrates forced alignment with multidimensional acoustic features such as pitch, duration, energy and MFCCs capturing prosodic and spectral cues at the syllable level. It also utilized contextual features by including features from preceding and succeeding syllables. Using speaker independent evaluation, the model demonstrates strong generalization. Ensemble methods, particularly Gradient Boosting, achieve superior performance, emphasizing the key role of prosodic cues in detecting lexical stress.

The BengLish corpus, combined with the proposed modeling approaches, offers a strong foundation for developing robust, interpretable and linguistically informed MDD systems which helps to design personalized and age appropriate pronunciation feedback system for improving English pronunciation of Bengali speaking children.

Keywords: Segmental features, Suprasegmental features, Bengali accented English, Second language acquisition, Language transfer, Mispronunciation detection and diagnosis (MDD), Computer assisted language learning (CALL), Computer assisted pronunciation training (CAPT), Deep learning, Machine learning, Lexical stress, Non-native speech.