Abstract

The individuals with cleft lip and palate (CLP) suffer from speech disorders due to articulatory impairments. As a result, measurable reductions are observed in the speech intelligibility and quality. Despite the advances in surgical management, the problems of articulation and resonance still remain to some degree in the individuals with CLP. For repaired CLP speakers, another direction of improvement in the speech intelligibility could be achieved through exploration of signal processing algorithms. Among all the speech disorders demonstrated by individuals with CLP, certain disorders relatively have a more severe impact on speech. In CLP speech, it is reported that the speech intelligibility is affected by two factors mainly: hypernasality and articulation errors. These two speech disorders are addressed in this thesis. In this thesis, we consider enhancement of CLP speech intelligibility by modifying frequently observed phoneme-specific distortions (fricative misarticulation, stop misarticulation, and vowel nasalization). The CLP speech modification is performed with an assumption that the ground truth of the phoneme distortions is available.

The first work in the thesis addresses the modification of fricative /s/ misarticulation. Based on the deviant characteristics, the misarticulated fricatives are segmented automatically followed by categorization of the type of distortion into palatalization, phoneme specific nasal air emission, and glottal stop. The fricative distortions that involve change in place of articulation are corrected using spectral compression and spectral tilt modification. While the insertion method is used when both change in the place and manner of articulation are observed.

In the next work, misarticulated stops are modified. As the stop consonants play an important role in speech perceptivity, it is important to address them as well. The work focuses on three unvoiced stop consonants /k/, /t/, and /T/. Three type of misarticulations are studied: glottal, palatal, and velar stop substitutions. An event-based modification approach is used to correct the misarticulated stops, where at first, automatic detection of burst onset and vowel onset events is carried out. The misarticulated stops are modified using spectral conversion method.

For an entire word enhancement, apart from stops, the vowel distortion needs to be addressed. The nasalized vowels (hypernasality) are observed to influence the speech intelligibility and quality both. Hence, vowel modification is performed as the third work. The issues related to nasalization is studied for vowels /a/, /i/, and /u/. In this work, the CLP distortions are analyzed in children speech and they exhibit high-pitch speech. Additionally, hypernasality introduces nasal resonances in the oral sounds, with consistent nasal resonances in the low-frequency region. Therefore, an accurate representation of the spectral envelope is necessary, for which extended weighted linear prediction (XLP) method is used. The transformation is achieved using spectral conversion method. The deviated spectral characteristics results in interference/additional signal components in the residual signal. Therefore, a weighting function is used for de-emphasizing the interfering signal components in the XLP residual signal.

Finally the word-level intelligibility is attempted by combining the specific phoneme modification techniques discussed in the earlier works. Several issues exist in performing the entire word-level intelligibility because many times, both articulation error and hypernasality are observed in the same word. It is challenging to detect such misarticulations in an unsupervised method. Hence, with certain assumptions and prior knowledge, the enhancement task is carried out.

To transform the CLP speech, different attempts are made in the thesis. The notable contributions of the thesis are listed below:

- A database is developed for analysis and enhancing the CLP speech. Database comprised of nonsensical words, vowel phonations, meaningful words, and short phrases. However, only some nonsensical words and vowel phonations are used in this thesis.
- Misarticulated fricative /s/ is first studied because it is observed as one of the frequently occurring speech distortions in the database and findings of various studies also support the same.
- Misarticulated unvoiced stops /k/, /t/, and /T/ are analyzed and modified.
- Nasalized vowels /a/, /i/, and /u/ are modified using temporal as well as spectral processing.
- Finally, phoneme specific modification techniques are combined to achieve entire word-level intelligibility.