Analysis and Automatic Detection of Aspirated Fricative and Aspirated Nasals

Saswati Rabha

Abstract

Unlike aspiration in stops, aspiration in non-stop consonants is quite rare. Most of the languages that have aspirated non-stop consonants are low-resource languages. Hence, data-driven, quantitative, and statistical analysis of their aspiration phenomena is fairly limited. From the literature review, it has been observed that there is still a need to explore a novel framework that will utilize the advantages of both linguistic and signal-processing knowledge-based approaches for acoustic-phonetic analysis and automatic classification of aspirated fricatives and aspirated nasals. To address these issues, we study two phonemes,/s/ and /s^h/ in a North-eastern language of India of Tibeto-Burman origin - Rabha, where contrast exists between aspirated and unaspirated counterparts. Also, for the study of aspirated nasals (/m/- /m^h/, /n/-/n^h/ and /ŋ/-/$ŋ^h$), we choose the Angami language, another North-eastern language of India of Tibeto-Burman origin. Both languages are low-resourced and are characterized by these unique sounds.

Initially, a detailed discussion about the development of a speech database for the thesis work and the procedure of database preparation is provided. In order to study the aspirated and unaspirated fricatives and nasals, the database was created for the Rabha and Angami as there is no standard built-in database available in the public domain for those languages. Data was collected from the Korean speakers as no open-source Korean database was available that was relevant to our area of interest. While the reviewed literature presents Korean fricatives as two-way contrast fricatives: aspirated and unaspirated, and the presence of the two-way contrast of the Angami nasals has been reported, no study has been attempted for the phonological fricative contrast in Rabha. Hence, this perception experiment would give an insight into the existence or non-existence of phonemic pairs of Rabha fricatives, where aspiration may serve as a cue. The discrimination test and identification test of the perception experiment

confirm that the Rabha fricatives are phonemically distinguished in terms of aspiration. Further, production analysis on the recorded speech stimuli of the fricatives and nasals was conducted.

The subsequent work unearths the details of aspiration characteristics from linguistic studies and further utilizes acoustic-phonetic information to extract source features using signal processing methods. A set of acoustic features is proposed to detect aspiration in fricatives and nasals automatically. Acoustic features, such as vocal tract constriction (VTC), normalized autocorrelation peak strength (NAPS), the strength of excitation (SoE), and variance of successive epoch intervals (VSEI) are used to detect aspiration in fricatives and nasals. Results show that VTC, NAPS, and SoE can detect aspiration in the nasals, whereas SoE and VSEI can detect aspiration in fricatives.

Further, to capture the distinctive aspiration associated with the fricatives, the source features, along-with durational features, were used to model a two-class classification system. The extracted features are combined to improve the classification of fricatives based on aspiration. The acoustic parameters are fed to support vector machine (SVM), Gaussian mixture model (GMM), and Random Forest (RF) classifiers. The integrated features fed to the SVM-GMM-RF hybrid model outperform baseline MFCCs for classifying fricatives in Rabha and Korean, irrespective of the language.

Another work investigates the nature of peak ERB_N trajectories to analyze the spectral dynamics across the time course of fricatives and nasals. The work is motivated by the fact that differences in the articulation of the two-way contrast non-stop consonants lead to differences in the contours of peak ERB_N in the aspirated and unaspirated variants. Such differences help to propose an efficient approach for analyzing and automatically classifying the aspirated fricatives and aspirated nasals from their unaspirated counterparts. The experimental investigation has shown that combining the static and dynamic acoustic parameters improved the classification accuracy compared to the baseline MFCC features.

Finally, we propose a framework that can utilize different acoustic-phonetic knowledge at

various stages to classify aspirated and unaspirated non-stop consonants automatically. Also, possible applications of the proposed methods are discussed by demonstrating the details of an aspiration detection-based phone recognition system on the collected corpus. The proposed method improves the performance of an automatic phoneme recognizer by reducing the confusion between aspirated and unaspirated counterparts.

Keywords: Fricatives, Nasals, Aspiration, under-resourced language, speech recognition, speech analysis, perception, low-resourced languages, Rabha, Angami, Korean.