

A Study: Can MFCC be accepted as a Better Feature for Speech Emotion Recognition

Akalpita Das¹, Purnendu Acharjee², Pranhari Talukdar³
 Assistant Professor¹, Assistant Professor², Professor³
 dasakalpita@gmail.com¹, pbacharyaa@gmail.com², phtassam@gmail.com³

Abstract: Speech synthesis is the machine generated acoustic signal which resembles human speech sounds. Formants are the basic and most important features of human speech to analyze. That’s why Formants frequency is considered as the most distinguishing frequency components of human speech. It generally refers to the specific resonance frequencies of vocal tract which have maximum energy concentration during the utterances of vowel sounds.

KeyWord: MFCC, Cepstral, voiced, unvoiced, plosive.

I. INTRODUCTION

In the present study it is observed in Assamese linguistic structure 1) **high** or **plane** and 2) **low** or **tense**. It is also observed that all such tones are lexically assigned, tightly bound and monosyllabic at morpheme level or root level and stress is trochaic/initial/falling. Each word in Assamese is minimally having one foot. The present study may help in designing a diaphone based speech synthesis, speech recognition [67], also purpose in developing speech-to-speech translation [55]. As per table, the speech energy level comparisons between Male and Female in Assamese language. For such spectral analysis, each speech spectra corresponding to each of vowels and consonants are divided into 12 unique frames (**1 frame =250 samples**) and for each frame calculating value of 12 Cepstral coefficients. For carrying dialect of Assamese, data are taken from three districts like Kamrup, Barpeta and Goalpara. Each vowel and consonant spectra are sampled with sampling frequency rate **16000 samples** per second and **12 frames** per sample and **12 Cepstral** coefficients from each frame have been calculated. The equivalent graph using MFCC representations for the words of type CV, VC, CVC and VCV with three emotions are shown to compare Male in three emotions, Female in three emotions.

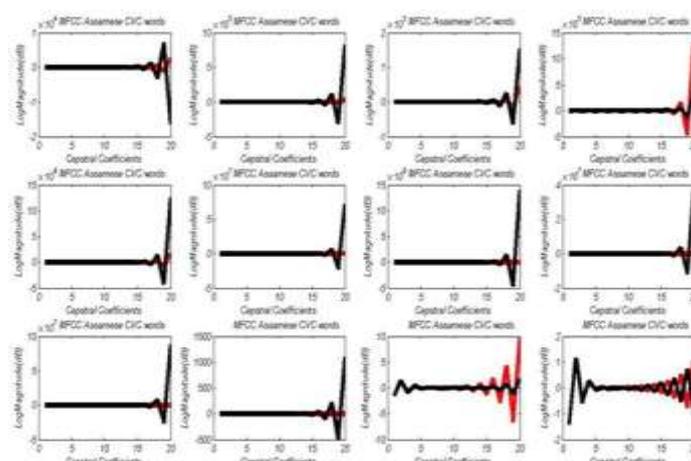


Figure 1.1 : MFCC comparison for Assamese CVC male and female word /jam/ (যাম)

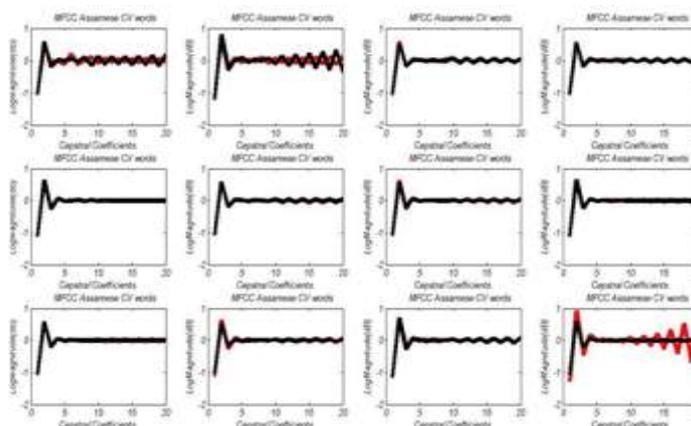


Figure 1.2: MFCC comparison for Assamese CV male and female word /ma/ (মা)

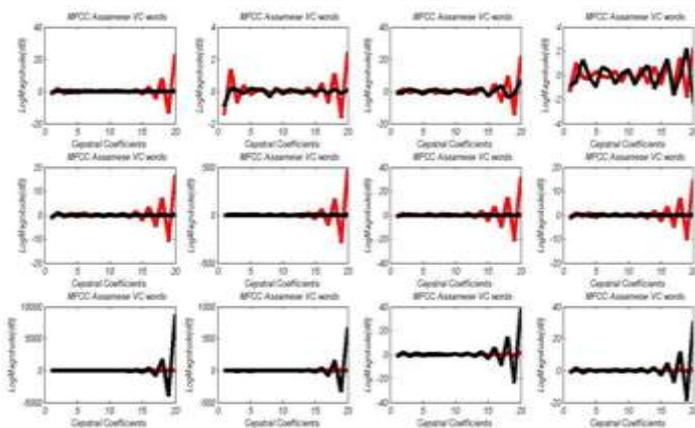


Figure 1.3: MFCC comparison for Assamese VC male and female /ek/ (এক)

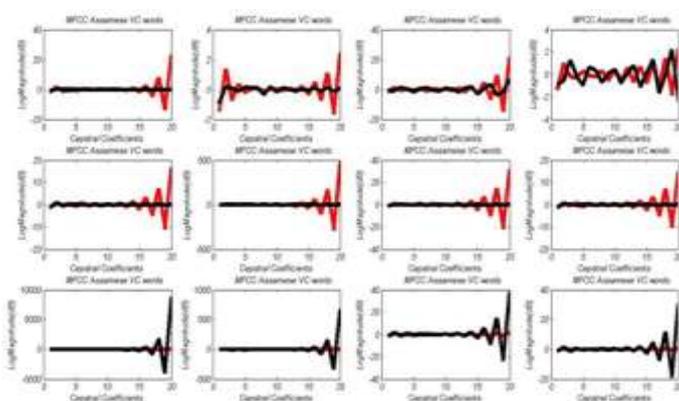


Figure 1.4: MFCC comparison for Assamese VCV male and female /asa/ (আসা)

II. OBSERVATIONS

From the study it is observed that all vowels and consonants in Assamese, as shown in tables, include plosives, nasals and unvoiced. The Voice sounds in Assamese are produced by forcing the air through the glottis where by making some proper adjustments of the tension of the vocal cords results in opening and closing of the cords and a almost periodic pulses of air thus produced and this pulse excites the vocal tract. Unvoiced speech sections are generally produced by forcing air through a constriction formed at a point in the vocal tract (toward the mouth end) and thus produces turbulence of air to get unvoiced speech.

1. In case MFCC study for Assamese different words in three emotions (type CV, VC, CVC and VCV) it is observed as:
 - (i) In case of CVC words of Assamese Male, considerable changes are found in frame 8, 9 and 12. In case of CVC words of Assamese Female, considerable changes are found in frame 8, 9, 10 and 12.
 - (ii) In case of VC words of Assamese Male, considerable changes are found in frame 9,

10, 11 and 12. In case of VC words of Assamese Female, considerable changes are found in frame 8, 9, 10, 11 and 12.

- (iii) In case of CV words of Assamese Male, considerable changes are found in frame 9, 10, 11 and 12. In case of CV words of Assamese Female, considerable changes are found in frame 10, 11 and 12.

- (iv) In case of VCV words of Assamese Male, considerable changes are found in frame 8, 9, 10 and 12. In case of VCV words of Assamese Female, considerable changes are found in frame 9, 11 and 12.

III. CONCLUSION

Based on excitation, the speech sounds can be classified as: (1)voiced, (2)unvoiced and (3)plosive. The Voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they used to vibrate in a relaxation oscillation. The Unvoiced sounds are produced by forming constrictions at some point of the vocal tract, and forcing air through the constriction at a very high velocity enough to produce turbulence.

1. These unique features Assamese may help us to developed and enhance a better Prosody based Text to Speech Tool and speech recognition tool as well. Tone has been identified as a critical marker of the identity of speech. Therefore, Chronological architecture of Consonant changes with the change of Sex.
2. Further we can say that these unique features of Assamese language may help us to developed and enhance better online teaching or explanation tool cum document reader with proper emotion.