.Abstract

Nepali community lacks the technology to be able to convert Nepali Speech to written text which has made us backward in adapting newer way of interacting with computer systems. While the services like Siri, Cortana and Google Now are on rise, we are still not able to use them in our local language – which is a problem, because not all of us are good English speakers and listeners. Lack of Nepali Speech Recognition Engine has also restricted the developers to use their mind in more innovative ways. Thus, to solve this problem, we are trying to create an automated speech recognition library to provide an interface to developer. Commonly used applications use Mel Frequency Cepstral Coefficients to decrease the dimensionality of the speech signal and later use Hidden Markov Model as an acoustic model to determine the sequence of phonemes which can generate a specific word. We use similar methods in order to create a Nepali Acoustic Model and use it for our Automated Speech Recognition. Our Initial Focus will be to recognize isolated words which we plan on extending to recognizing connected words.

Keywords: *Automated Speech Recognition, Hidden Markov Model, Mel Frequency Cepstral Coefficients, Acoustic Model*

Table of Contents

Abbreviations

ASR: Automated Speech Recognition

ANN: Artificial Neural Network

DCT: Discrete Cosine Transform

DFT: Discrete Fourier Transform

FFT: Fast Fourier Transform

HMM: Hidden Markov Model

Hz: Hertz

KHz: Kilo Hertz

MFCC: Mel Frequency Cepstral Coefficient

NMF: Non-negative Matrix Factorization

PC: Personal Computer

TTS: Text to Speech

WT: Wavelet Transformation

List of Figures