.**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 Hidden Markov Model or Time-Delayed Neural Networks for phoneme classification to design Automated Speech Recognition. We are trying to adopt either of those models in our Nepali Speech Recognition System. This can be done by applying the Nepali Grammar and Linguistics to a similar system. The basic flow can be summarized as: Getting User Speech Data, Extracting Phonemes, Analyzing Phoneme sequence and using probabilistic models to deduce whole words and sentences.

**Keywords:** *Automated Speech Recognition, Hidden Markov Model, Time-Delayed Neural Networks*

**Table of Contents**

|  |  |
| --- | --- |
| Abstract | i |
| Table of Contents | ii |
| Abbreviations | iii |
| List of Figures | iv |
| Introduction | 1 |
| Objectives | 1 |
| Scope | 2 |
| Literature Review | 2 |
| Methodology | 3 |
| Project Breakdown | 5 |
| References | 6 |

**Abbreviations**

ASR: Automated Speech Recognition

HMM: Hidden Markov Model

Hz: Hertz

KHz: Kilo Hertz

MLP: Multi-Layered Perceptron

PC: Personal Computer

TTS: Text to Speech

TDNN: Time Delayed Neural Network

**List of Figures**

Figure 5.1 Overview of Automated Speech Recognition Engine 4

Figure 5.2.1 Signal Preprocessing and Framing 5