International Journal of Engineering Sciences & Emerging Technologies, Oct. 2023.
ISSN: 22316604 Volume 11, Issue 2, pp: 108-114 ©IJESET

AN AUTOMATIC SPEECH RECOGNITION APPROACH USING MODIFIED VOICE ACTIVITY DETECTION MECHANISM

Neha Verma[1], Ajeet Singh[1], Atul Pratap Singh[2]

[1]Department of CSE, Moradabad Institute of Technology, Moradabad, UP, India
nehav0624@gmail.com
ajeetsingh252@gmail.com
[2]Department of CSE, Swami Vivekanand Subharti University Meerut, UP, India
atulpratap26385@gmail.com

ABSTRACT

The effectiveness of Automatic Speech Recognition (ASR), which can be employed in loud circumstances, is being researched. The effectiveness of common parameterization methods was evaluated in relation to lustiness and compared to the background signal. By merging the essential components of PLP and MFCC, a hybrid feature extractor is created for Mel frequency cepstral coefficients (MFCC), Perceptual linear predictive (PLP) coefficients, and their modified forms. Only the ASR method's training phase was used to apply the VAD-based frame dropping calculation. This technique has the benefit of eliminating pauses and speech segments that may be considerably distorted, It helps with more accurate phone modelling. The examination and contribution of the improved vocal activity detection approach are the main topics of the second section.

KEYWORDS: *Automatic Speech Recognition, Short-Time Fourier Transform, Perceptual Linear Prediction, Mel Cepstrum Frequency Coefficients, and Speech Presence Probability*

1. INTRODUCTION

Automatic speech recognition (ASR) is the process of using machines to modify a human speaker's string of words. It is preferable for an ASR system to be resilient to unpleasant fluctuation because the goal of ASR is to have speech as a suboptimal form of contact between a machine and a person [5]. It is frequently difficult for speech recognition systems to identify endpoints when background noise and non-speech events are present [1]. When ambient acoustic noise is present, speech recognition systems that were trained in quiet contexts sometimes perform worse. Dilapidation typically occurs as a result of the disparity between accurate acoustic models and noisy speech data. To reduce this mismatch and improve recognition accuracy in noisy environments, a lot of effort has been done [2]. It is possible to approach the subject of noise resilience in automated speech recognition (ASR) in a number of fundamentally different ways. One method is to just expose the machine to the type of noise it hears during the recognition step. Only for that particular form of noise, this type of system, known as a "matched system," is probably superior than many other noise-compensation techniques. To respond to these new forms of noises, the system must be retrained over an enormously long period of time and a huge library of new noise types. Multi condition training, which teaches the system on noisy speech heard under the loudest noise circumstances and eliminates the need to retrain the system every time the background noise changes [5], is a more practical alternative to matched training.

2. RELATED WORK

2.1 Review of existing VAD techniques and their limitations: