Speech Recognition using Machine Learning

Dr.M.Rajaiah, Dean Academics & HOD, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr.B.Hari Babu, Assistant Professor ,Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr.N.Sateesh kumar reddy, UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr.K.Madan, UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr.P.Hema krishna reddy, UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr.P.Gowtham, UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

ABSTRACT :

Speech recognition is one of the fastest-growing engineering technologies. It has several applications in different areas, and provides many potential benefits. A lot of people are unable to communicate due to language barriers. We aim to reduce this barrier via our project, which was designed and developed to achieve systems in particular cases to provide significant help so people can share information by operating a computer using voice input. This project keeps that factor in mind, and an effort is made to ensure our project is able to recognize speech and convert input audio into text; it also enables a user to perform file operations like Save, Open, or Exit from voice-only input. We design a system that can recognize the human voice as well as audio clips, and translate between English and Hindi. The output is in text form, and we provide options to convert audio from one language to the other. Going forward, we expect to add functionality that provides dictionary meanings for Hindi and English words. Neural machine translation is the primary algorithm used in the industry to perform machine translation.

Keywords:
Speech recognition, Speech emotion recognition, Statistical classifiers, Dimensionality reduction techniques, Emotional speech databases, Vision processing, Computational intelligence,

1.INTRODUCTION

Speech is the most common and primary mode of communication among humans. The communication between humans and machines is referred to as the human-computer interface (Gaikwad et al., 2010). The method of translating Voice recognition is the process of converting a speech signal into a series of words using an Algorithm implemented as a computer program (SR). One of the most exciting fields of signal processing in speech processing. The SR field aims to build techniques and systems for using speech as a computer input (Gaikwad et al., 2010). Increasingly time, sophisticated skills for recognizing patterns such as voice, handwriting, facial features, and so on have been developed. The search for computer programs that allow machines to acquire the aforementioned skills from past experience gave rise to machine learning (Mitchell., 1997). of the sense of machine learning, it is claimed that "A computer program is said to learn from experience E concerning some class of tasks in T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." While speech processing systems have been used in a variety of business applications, the problem for speech processing remains largely unresolved (Padmanabhan and Johnson., 2015). More intelligent contact between humans and machines was desire as technology progresses in the latest years. Because of that, the ML and SP communities are being increasingly intertwined. In preliminary work, SP researchers find formal theoretical findings and mathematical guarantees from ML to be minus useful. Also as result, they pay little attention to these findings than they should, potentially losing out on feedback and advice offered by ML hypotheses and formal structures, even though dynamic SP activities were often outside this existing state-ofthe-art in ML. An analysis of the various machine learning (ML) strategies for speech processing (SP) was discuss in this review.

2.OBJECTIVE OF THE SYSTEM:

Different types of information may be gleaned from speech signals. Examples of this kind of knowledge are: Speech recognition, detection speech signals (Hinton et al., 2012), Speaker recognition, speaker identity (Chung et al., 2018), Speech Emotion recognition (Ayadi et al.,2011),

Health recognition, patient's health status (Johnson et al.,2014), Language recognition spoken language knowledge (Muthusamy et al., 1994), Accent recognition, which products information about the speaker accent (Biadsy., 2011),Age recognition, which products information about the speaker age (Li et al., 2013) and Gender recognition, which Find gender speaker (Li et al., 2013).In the digital era, speech as an essential part of human being's daily communication is a challenging area for many signal processing experts and researchers .Speech recognition technologies have the possibility to analysing individual speech boundaries in Human languages and recognizing them. The concept of audio processing studying techniques are to be developed and models for detecting and analysing the input speech into machines (Anasuya and Katti, 2009). Nowadays, Automatic signal speech processing systems discover widespread applications in tasks that need a human machine interface (HMI) (Gaikwad et al., 2010). Recently, automatic speech processing systems have found wide applications in many domains that require a human-machine interface. As a computer program, an algorithm is deployed to convert audio signals into a sequenced manner that would go through some serious process of analysing and detecting (Gaikwa et al., 2010). Speech recognition devices convert the spotted acoustic sound signals to the related written portrayal of the words spoken, for example, converting voice messages to texts. Speech recognition

computer or program's ability to understand and execute speech event which can simplify voice translation tasks (Senhildevi and Chandra, 2012). Audio recognition is used to mention recognition model to training for a specific speaker. The physical appearance of the vocal tract of a person and an individual's behavior features influenced the overall recognition process. Text-to-speech is measured as a portion of audio or audio processing was used to translate the contents of a signal aloud viewing an automatic screening of a blind user (Senhildevi and Chandra, 2012). Audio syntheses are text-to-speech (TTS) applications used through computers to easily convert text messages into speech. The synthesizer may be as a peripheral a card inserted into the system, a box connected to the processer via a wire or computer applications (Gaikwad et al., 2010).

3.PROPOSED SYSTEM:

In this research, the work is based on the flowchart below. According to working model of speech. The models illustrated previously are made up of millions of parameters, from which the instruction corpus needs to be learned. We make use of additional information where appropriate, such as text that is closely linked to the speech we are about to translate . It is possible to write this text in the source language, the target language, or both. Future development will reach billions of smart phone users for the most complex intelligent systems focused on deep learning. There is a lengthy list of vision and voice technologies that can increasingly simplify and assist the visual and auditory processing of humans to a greater scale and consistency, from sensation and emotion detection to the development of self-driving autonomous transport systems. This paper serves scholars, clinicians, technology creators, and consumers as an exemplary analysis of emerging technologies in many fields, such as behavioral science, psychology, transportation, and medicine.

4.LITERARTURE SURVEY:

Mehmet Berkehan Akçay et al. [1] explained that neural networks are mainly limited to industrial control and robotics applications. However, recent advances in neural networks through the introduction of intelligent travel, intelligent diagnosis and health monitoring for precision medicine, robotics and home appliance automation, virtual online support, emarketing, weather forecasting and natural disaster management, among others, have contributed to successful IS implementations in almost every aspect of human life. G. Tsontzos et al. [2] clarified how feelings allow us to better understand each other, and a natural consequence is to expand this understanding to computers. Thanks to smart mobile devices capable of accepting and responding to voice commands with synthesized speech, speech recognition is now part of our daily lives. To allow devices to detect our emotions, speech emotion recognition (SER) could be used. T. Taleb et al. [7] said they were motivated by understanding that these standards place higher boundaries on the improvement that can be achieved when using HMMs in speech recognition. In an attempt to improve robustness, particularly under noisy conditions, new modeling schemes that can explicitly model time are being explored, and this work was partially funded by the EUIST FP6 HIWIRE research project. Spatial similarities, including dynamic linear models (LDM), were initially proposed for use in speech recognition. Vinícius Maran et al. [6] explained that learning speech is a dynamic mechanism in which the processing of phonemes is marked by continuities and discontinuities in the path of the infant towards the advanced production of ambient language segments and structures. Y. Wu et al. [3] noted that discriminative testing has been used for speech recognition for many years now. The few organizations that have had the resources to implement discriminatory instructions for large-scale speech recognition assignments have mostly used the full shared information system in recent years (MMI). Instead, in the extension of the studies first presented, we reflect on the minimum classification error (MCE) paradigm for discriminatory instruction. Peng et al. [4] stated that identification of speakers refers to identifying people by their voice. This technology is increasingly adopted and used as a kind of biometrics for its ease of use and non-interactivity, and soon became a research hotspot in the field of biometrics.

Block Diagram

The below Figure shows the Block diagram of the proposed system which has been implemented:

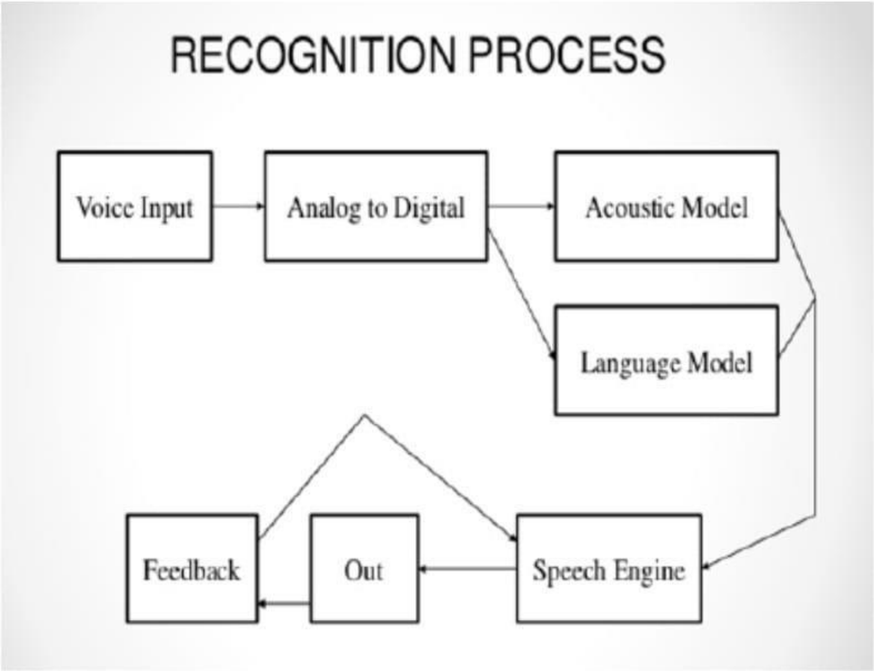

Fig. 3. Working Model of Speech Recognition.

Connected Speech: Linked words or connected speech are identical to independent speech, and except for brief delays between them, they make separate utterances. ·Continuous Speech: Continuous speech allows the user to speak almost naturally; it is also called computer dictation. ·Spontaneous Speech: At a simple level, this can be viewed as speech that is naturalsounding and not rehearsed. An ASR device with random speech abilities should be able to accommodate a variety of normal speech features, such as sentences that run together, and that include "ums" and "ahs" and even slight stutters. Machine Translation: Machine translation usually models whole sentences with the use of an artificial neural network to predict a sequence of terms. Typically, it models entire sentences in a single integrated model through the use of an artificial neural network to predict the sequence of words. Initially, word sequence modeling is usually carried out using a recurrent neural network (RNN). Unlike the traditional phrase-based translation method that consists of many small subcomponents that are tuned separately, neural machine translation is used to build and train a single, broad neural network that reads a phrase and outputs the correct translation. Neural machine translation by end-to-end systems is said to be a neural machine translation system because only one model is needed for translation. The transfer of scientific, metaphysical, literary, commercial, political, and artistic knowledge through linguistic barriers is an integral and essential component of human endeavor.

5.RESULT :

Voice detection with real-time predictive voice translation device optimization using multimodal vector SoftMax (score, axis = 1). It is implemented on the last axis by default, but we want to implement it on the first axis here, as the score form is as follows: batch size, max length, secret size. The length of our input is Max length. Since we are attempting to assign a weight to each input, it is important to add SoftMax on that axis. Context vector = sum (weights of focus * EO, axis = 1). The same explanation as above applies for an axis selection of 1.

Embedding output = The input is transferred through an embedding layer to the decoder. Integrated vector = concept (embedding output, context vector).

6.CONCLUSION :

In the past few years, the complexity and precision of speech recognition applications have evolved exponentially. This paper extensively explores the recent advancements in intelligent vision and speech algorithms, their applications on the most popular smart phones and embedded platforms, and their application limitations. In spite of immense advances in success and efficacy from deep learning algorithms, training the machine with other knowledge sources, which are the framework, also contributes significantly to the class subject.

REFERENCES :

Mehmet Berkehan Akçay, Kaya Oğuz, Speech emotion recognition: Emotional models, databases,features, preprocessing methods, supporting modalities, and classifiers, Speech Communication, Volume116, 2020, Pages 56-76, ISSN 0167-6393.

G. Tsontzos, V. Diakoloukas, C. Koniaris and V. Digalakis, "Estimation of General Identifiable Linear Dynamic Models with an sApplication in Speech characteristics vectors, Computer Standards & Interfaces, Volume 35, Issue 5, 2013, Pages 490-506, ISSN 0920-5489.

Y. Wu et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," arXiv preprint arXiv:1609.08144, pp. 1-23, 2016.

Shuping Peng, Tao Lv, Xiyu Han, Shisong Wu, ChunhuiYan, Heyong Zhang,Remote speaker recognition based on the enhanced LDV-captured speech, Applied Acoustics,Volume 143, 2019, Pages 165-170, ISSN 0003-682X.

International Conference on SoftComputing and Networks Security (ICSNS), Coimbatore, 2015, pp. 1-5 A. A. Varghese, J. P. Cherian and J. J. Kizhakkethottam, "Overview on emotion recognition system," 2015

Author Profiles

Dr.M.Rajaiah, Currently working as an Dean Academics & HOD in the department of CSE at
ASCET (Autonomous), Gudur, Tirupathi(DT).He has published more than 35 papers in, Web of Science, Scopus Indexing.

Mr.B.Hari Babu, currently working as an Assistant professer in the department of CSE at ASCET autonomous, gudur, Tirupati(DT).

N.Sateesh Kumar Reddy, B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.

K.Madan, B.Tech student in the department of CSE at AudiSankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.

P.Hema Krishna Reddy, B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.

P.Gowtham B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.