1.Automatic Speech Recognition: Systematic Literature Review

ABSTRACT A huge amount of research has been done in the field of speech signal processing in recent years. In particular, there has been increasing interest in the automatic speech recognition (ASR) technology field. ASR began with simple systems that responded to a limited number of sounds and has evolved into sophisticated systems that respond fluently to natural language. This systematic review of automatic speech recognition is provided to help other researchers with the most significant topics published in the last six years. This research will also help in identifying recent major ASR challenges in real-world environments. In addition, it discusses current research gaps in ASR. This review covers articles available in five research databases that were completed according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol. The search strategy yielded 82 conferences and articles related to the study’s scope for the period 2015–2020. The results presented in this review shed light on research trends in the area of ASR and also suggest new research directions.

CONCLUSION The most important way for humans to communicate with each other and acquire information is through speech. This paper provides a systematic literature review of automatic speech recognition with the most significant topics published in the last six years. A total of 82 conferences and articles studies were reviewed from five research databases: IEEE Xplore Digital Library, ACM Digital Library, Scopus, the Web of Science, and Science. First, a brief introduction to ASR was provided. The methodology of this research, including research questions, search strategy, and quality assessment process, was then described. After that, a review of the selected studies published from 2015 to 2020 in ASR based on their characteristics was organized and presented. The publication trends in speech recognition were then detailed. In addition, the major challenges and current research gaps in ASR were reviewed. Future directions for research in ASR were investigated. It is expected that this examination will help other researchers, as it provides a review of ASR studies published in recent years. Finally, statistics of mostly utilized datasets in reviewed papers were provided.

2.A Comprehensive Review of Speech Emotion Recognition Systems

ABSTRACT During the last decade, Speech Emotion Recognition (SER) has emerged as an integral component within Human-computer Interaction (HCI) and other high-end speech processing systems. Generally, an SER system targets the speaker’s existence of varied emotions by extracting and classifying the prominent features from a preprocessed speech signal. However, the way humans and machines recognize and correlate emotional aspects of speech signals are quite contrasting quantitatively and qualitatively, which present enormous difficulties in blending knowledge from interdisciplinary fields, particularly speech emotion recognition, applied psychology, and human-computer interface. The paper carefully identifies and synthesizes recent relevant literature related to the SER systems’ varied design components/methodologies, thereby providing readers with a state-of-the-art understanding of the hot research topic. Furthermore, while scrutinizing the current state of understanding on SER systems, the research gap’s prominence has been sketched out for consideration and analysis by other related researchers, institutions, and regulatory bodies.

CONCLUSION The capability to drive speech communication using programmable devices is currently in research progress, even if human beings could systematically achieve this errand. The focus of SER research is to design proficient and robust methods to recognize emotions. In this paper, we have offered a precise analysis of SER systems. It makes use of speech databases that provide the data for the training process. Feature extraction is done after the speech signal has undergone preprocessing. The SER system commonly utilizes prosodic and spectral acoustic features such as formant frequencies, spectral energy of speech, speech rate and fundamental frequencies, and some feature extraction techniques like MFCC, LPCC, and TEO features. Two classification algorithms are used to recognize emotions, traditional classifiers, and deep learning classifiers, after the extraction of features. Even if there is much work done using traditional techniques, the turning point in SER is deep learning techniques. Although SER has come far ahead than it was a decade ago, there are still several challenges to work on. Some of them are highlighted in this paper. The system needs more robust algorithms to improve the performance so that the accuracy rates increase and thrive on finding an appropriate set of features and efficient classification techniques to enhance the HCI to a greater extend.

3. Multilingual Speech and Text Recognition and Translation using Image

Abstract - The aim of our project to automate the application to overcome from the language barrier among countries and also states within the country, the above mentioned application will perform the various features in the application. The application recognizes speech (human matter) in one language to another user defined language to communicate in expressive manner. It includes 4 modules voice recognition, translation and speech synthesis and image translation and gives audio of the translated language. Also the application accepts text written and converts it into the language needed. Application is able to recognize the text present in the image which stored in system or captured using camera and translate the text into the language needed and display the translation result back on to the screen of system.

CONCLUSION In this proposed system, we implemented the system for user who phasing problems of language barrier and also it user interface is also user friendly so that user can easily interact with this system .so because of this system don’t have to use dictionary for understanding the meaning of word, so it automatically reduce the user task for understanding the languages for communication.

4. Speech to text conversion and summarization for effective understanding and documentation

Abstract- Speech, is the most powerful way of communication with which human

beings express their thoughts and feelings through different languages.

The features of speech differs with each language. However, even while

communicating in the same language, the pace and the dialect varies with

each person. This creates difficulty in understanding the conveyed message

for some people. Sometimes lengthy speeches are also quite difficult to

follow due to reasons such as different pronunciation, pace and so on. Speech

recognition which is an inter disciplinary field of computational linguistics

aids in developing technologies that empowers the recognition and

translation of speech into text. Text summarization extracts the utmost

important information from a source which is a text and provides

the adequate summary of the same. The research work presented in this paper

describes an easy and effective method for speech recognition. The speech is

converted to the corresponding text and produces summarized text. This has

various applications like lecture notes creation, summarizing catalogues for

lengthy documents and so on. Extensive experimentation is performed to

validate the efficiency of the proposed method

CONCLUSION

Speech recognition and text summarization are two vast areas to be explored. The proposed research

work aims to reduce the time and effort of manual documentation of lengthy speeches in an event.

Speech recognition and text summarization can ease the work of documentation. Even for the verification of

the summarized content, the system can be automated to read out the summarised content with the help of

text to speech conversion. As of now, speech summarization for sentences terminating with a full stop or

containing a small pause shown by comma is experimented. The future work is to include all punctuation

marks in the recognized speech which helps in improving the text summarization performance. This model

can be used where ever there is a requirement of summarising lengthy lectures into precise documents as the

automated system will convert the speech to text and also summarise the content. It can be of great help for

students to archive lecture notes from classes, conferences or seminars.

5. Speech Recognition using Machine Learning

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 voiceonly 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. Two recurrent neural networks used in tandem to construct an encoder–decoder structure are the architecture behind neural machine translation. This work on speech recognition starts with an introduction to the technology and the applications used in different sectors. Part of the report is based on software developments in speech recognition.

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.

6. Text to Speech Conversion

Abstract The present paper has introduced an innovative, efficient and real-time cost beneficial technique that enables user to hear the contents of text images instead of reading through them. It combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS) in Raspberry pi. This kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction from color images is a challenging task in computer vision. Text-to-Speech conversion is a method that scans and reads English alphabets and numbers that are in the image using OCR technique and changing it to voices. This paper describes the design, implementation and experimental results of the device. This device consists of two modules, image processing module and voice processing module. The device was developed based on Raspberry Pi v2 with 900 MHz processor speed.

Conclusion Text-to-Speech device can change the text image input into sound with a performance that is high enough and a readability tolerance of less than 2%, with the average time processing less than three minutes for A4 paper size. This portable device, does not require internet connection, and can be used independently by people. Through this method, we can make editing process of books or web pages easier.

7. A critical review and analysis on techniques of speech recognition: The road ahead

Abstract. Recognition of human speech has long been an intriguing issue among artificial intelligence and processing researchers. Speech is the most crucial and essential method of communication among the human beings. Several research efforts have been prepared in the field of speech recognition in the previous decades. Accordingly, a survey of speech recognition strategies suitable for human identification is discussed in this study. The main motivation of this survey is to explore the existing speech recognition strategies so that the researchers can include all the necessary metrics in their works in this domain and the limitations in the existing ones can be overcome. In this review, diverse issues included in speech recognition methodologies is distinguished and distinctive speech recognition procedures were studied to discover which qualities is tended to in a given system and which is disregarded. Hence, we offer a detailed survey of 50 methods from standard publishers from the year of 2000 to 2015. Here, we categorize the research based on three dissimilar perspectives, like techniques utilized, applications and parameter measures. In addition, this study gives an elaborate idea about speech recognition techniques.

Conclusion Speech recognition is a standout amongst the most facilitating zones of machine information since individuals do an everyday movement of speech recognition. In this survey, we have reviewed various speech recognition techniques and tabulated different applications and parameters under speech signal. 50 articles were selected from the year 2000 to 2015, associated with speech recognition. Accordingly, these 50 articles are categorized into three ways and also, we have studied their limitations and time complexity. Importantly, three different factors such as, techniques, application and parameter measure were considered for comparing and reviewing the existing works. The detailed review performed in this paper will give the achievement happened in the speech recognition to further formulate the research ideas to overcome the current benchmark results for the researchers. At last, some of the research issues are also addressed to lead the further research in the same direction. Furthermore, the latest intelligent and algorithms were also discussed to incorporate those techniques for speech recognition in future.

8. Implementation of Text to Speech Conversion

Abstract- Text-To-Speech (TTS) conversion is a computerbased system that can be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. While in text to speech, there are many systems which convert normal language text in to speech. The main aims of this paper are to study on Optical Character Recognition with speech synthesis technology and to develop a cost effective user friendly image to speech conversion system using MATLAB. In this work, the OCR system is implemented for the recognition of capital English character A to Z and number 0 to 9. Each character is recognized at once. The recognized character is saved as text in notepad file. In this work a text-to-speech conversion system that can get the text through image and directly input in the computer then speech through that text using MATLAB.

CONCLUSION In this work, image into text and then that text into speech is converted by MATLAB. E-text into speech is also converted successfully. By this approach text from a word document, Web page or e-Book can be read and can generate synthesized speech through a computer's speakers. For image to text conversion, firstly image is converted into gray image. Gray image is converted into binary image by thresholding and then it is converted into text by MATLAB. Microsoft Win 32 SAPI library has been used to build speech enabled applications, which retrieve the voice and audio output information available for computer. In this work, one character can be converted into text at once. As a further extension, OCR system can be developed for converting words or sentences image into text.

9. Speech Emotion Recognition Using Convolutional Neural Network (CNN)

Abstract--The Automated Speech Emotion Recognition is a tough process because of the gap among acoustic characteristics and human emotions, which depends strongly on the discriminative acoustic characteristics extracted for a provided recognition task. Different persons have different emotions and altogether a different way to express it. Speech emotion do have different energies, pitch variations are emphasized if considering different subjects. Therefore, the speech emotion detection is a demanding task in computing vision. Here, the speech emotion recognition is based on the Convolutional Neural Network (CNN) algorithm which uses different modules for the emotion recognition and the classifiers are used to differentiate emotions such as happiness, surprise, anger, neutral state, sadness, etc. The dataset for the speech emotion recognition system is the speech samples and the characteristics are extracted from these speech samples using LIBROSA package. The classification performance is based on extracted characteristics. Finally we can determine the emotion of speech signal.

CONCLUSION After constructing various models, we got the better CNN model for the emotion distinction task. We reached 71%accuracy from the previously available model. Our model would’ve performed better with more data. Also our model performed very well when distinguishing among a masculine and feminine voice. Our project can be extended to integrate with the robot to help it to have a better understanding of the mood the corresponding human is in, which will help it to have a better conversation as well as it can be integrated with various music applications to recommend songs to its users according to his/her emotions, it can also be used in various online shopping applications such as Amazon to improve the product recommendation for its users. Moreover, in the upcoming years we can construct a sequence to sequence model to create voice having different emotions. E.g. asad voice, an excited one etc.

10. VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT

Abstract: VOICE RECOGNITION SYSTEM:SPEECH-TO-TEXT is a software that lets the user control computer functions and dictates text by voice. The system consists of two components , first component is for processing acoustic signal which is captured by a microphone and second component is to interpret the processed signal, then mapping of the signal to words. Model for each letter will be built using Hidden Markov Model(HMM). Feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). Feature training of the dataset will be done using vector quantization and Feature testing of the dataset will be done using viterbi algorithm. Home automation will be completely based on voice recognition system.

Conclusion: In this paper the fundamentals are discussed and its recent progress is investigated. The various approaches available for developing a Voice Recognition System based on adapted feature extraction technique and the speech recognition approach for the particular language are compared in this paper. The main aim of our project is to develop a system that will allow the computer to translate voice request and dictation into text using MFCC and VQ techniques. Feature extraction and feature matching will be done using Mel Frequency Cepstral Coefficients and Vector Quantization technique. The extracted feature will be stored in .mat file. A distortion measure which is based on minimizing the Euclidean distance will be used while matching the unknown speech signal with the database of the speech signal. In near future, home automation will be completely based on Voice Recognition System