IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 3, March 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

Implementation of Voice Recognition for Student Group Discussion

Tejas Dinesh Bhat

Department of Information Technology, Vishwakarma Institute of Technology, Savitribai Phule Pune University, India

ABSTRACT: Development of voice recognition for group discussion system is beneficial in many ways. It helps the HR or manager in administrative to track the voice with efficiency. It also promotes your confidence level while *re-listening* the recorded voice. It is an effective tool in problem solving, decision making and personality assessment. This paper presents neural network-based approach to recognize the user from group. For the same the features have been extracted from the voice samples and then by using CNN the samples are trained and tested. The results with proposed system are illustrated at the end.

KEYWORDS: Voice recognition, CNN, group discussions recordings.

I. INTRODUCTION

Speech recognition tool greatly enhance a student skill in language both when reading and writing. with this technology, student can obtain information on important elements of phonemic awareness such as the corresponding between sounds and symbol as students speak, they see their words on the screen Each person has their unique characteristic in speech and voice that can be captured and analyze to make this new class attendance more efficient and effective. Voice recognition can be divided into two, which are speech recognition and speaker recognition. Both are using voice biometric differently. In brief, speech recognition covers the ability to match a voice pattern against an acquired or provided vocabulary. Normally, the vocabulary given is small and the user needs to record a new word to expand the vocabulary. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech signals. It can be divided into two tasks, which are identification and verification. Speaker identification is added to decide which unknown voice belongs to from amongst a set of known speakers. Speaker verification accepts or rejects the identity claim of a speaker. For this project speaker identification in speaker recognition is used. The unknown speaker is later labeled as test data and the

possible because different speakers have different spectra for similar sound. Spectra are the location and magnitude of peaks in spectrum. The phrase "voice recognition" refers to acknowledging the speaker rather than the content of their speech. Recognizing the speaker can help systems that have been trained on a single person's voice translates speech more quickly, or it can be used to authenticate or verify the speaker's identity as part of a security procedure.

Speech recognition has a lengthy history in terms of technology, with multiple waves of key advancements. Advances in deep learning and big data have lately improved the sector. The advancements are proven not only by the increasing number of academic articles published in the subject, but also by the widespread industrial use of a range of deep learning approaches in the design and deployment of voice recognition systems throughout the world.

Group discussion is an important activity in academic, business and administrative spheres. It is a systematic and purposeful interactive oral process. Here the exchange of ideas, thoughts and feelings take place through oral communication. The exchange of ideas takes place in a systematic and structured way. The participants sit facing each other almost in a semi-circle and express their views on the given topic/issue/problem. This is a challenge is faced by many participants. In a GD, some participants might be shy or introvert and do not feel much confident while speaking in a group even if they know the topic. Low pitch of voice is also considered as a sign of low confidence in the GD. The objective of a group discussion is to mainly check your team playing skills because as a manager, you will be working in organizations with people. Employers are looking for candidates who have potential to be executives and to lead teams of people.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

Why GDs:-

The reason why institutes put you through a Group discussion and an interview, after testing your technical and conceptual skills in an exam, is to get to know you as a person and gauge how well you will fit in their institute. The Group discussion tests how you function as a part of a team. As a manager, you will always be working in teams, as a member or as a leader. Therefore how you interact in a team becomes an important criterion for your selection. Managers have to work in a team and get best results out of teamwork. That is the reason why management institutes include GD as a component of the selection procedure.

II. LITERATURE SURVEY

Glenn Arwin M. Bristol et.al [1] Voice recognition will help them to access email. This study also reduces the cognitive load taken by visually impaired users to remember and type characters using a keyboard. If this system is implemented, self-esteem and social and emotional well-being of the visually impaired users will be lifted up for they will now feel they are being valued in society and fair treatment and access to technology main function of this study is to use a keyboard of the user that will respond through voice. The purpose of this study is to help a visually impaired person to use the modern application to interact with voice recognition systems with the use of email into different types of modern gadgets, Line computers, or mobile phones.

O S Stefanenko et.al[2] This paper describes a method for recognizing voice command based on a fuzzy logic system capable of perceiving fuzzy commands, i.e. commands containing fuzzy terms, for example, „close", „closer", „close to", „closer than", „further" and „very far". The developed approach has the ability to be trained for a specific user. The developed fuzzy logic system is used to recognize linguistically inaccurate commands in order to increase the expressiveness of the language for control of a moving robot.

Zuoming Wang et.al [3] this study investigates the measurement of social identification, interpersonal attraction, and cohesiveness in virtual groups. Different theoretical claims about relationships in computer-mediated groups rely on measurement strategies that are shown to reflect dramatically inconsistent semantic and administration features. A review of conceptual approaches and definitions for these constructs is presented. Data were collected from groups working asynchronously via the Internet under different geographic distributions, whose members completed a variety of measures related to these constructs. Analyses generated three likely dimensions of attraction. The research highlights the need for greater specificity in reports of the actual measures used in group research, and additional conceptual concerns regarding the contested relationships among these constructs.

Hao Shao et.al [4] First the sounds from different musical instruments are used as a standard sound source to show the high accuracy of the nanofiber device to distinguish different sounds, e.g., from both the same and different instruments, and then people's voices are tested. Experimental results show that the device can distinguish people's voices with high resolution and the influence of background noise on speech recognition is very small. The acoustic sensor is stable and can be used in sound detection for a long time. Electrospun PAN Membranes may be useful for the development of voice recognition systems for security, environmental protection, scientific research, and other high-tech applications.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

Huaixiang Hu et.al [5].FPGA has the characteristics of high parallelism, low power consumption, and high flexibility. It can support high-performance computing acceleration for speech recognition applications. In order to implement speech recognition system on an specified computing platform, a speech recognition system architecture is proposed, which is based on software and hardware cooperative computing technology that focus on FPGA. First, a system architecture based on gated convolutional neural network is designed, with the simplification of the traditional speech recognition system architecture. Then a modular design idea is used to map the gated convolutional neural network operator to the FPGA hardware logic. Further through the state machine to schedule and control the data stream and operators, to achieve the flexible acceleration of the voice recognition network model.

Naomi Winstone et.al [6] there is growing recognition that socio-constructivist representations of feedback processes, where students build their own understanding through engaging with and discussing feedback information, are more appropriate than cognitivist transmission-oriented models. In parallel, practice has developed away from hard-copy handwritten or typed feedback comments, towards the provision of e-feedback in learning management systems (LMS). Through thematic analysis of activity-oriented focus groups with 33 undergraduate students, the present study aimed to explore 1) students' experience of engaging with feedback in the LMS; 2) barriers to students' engagement; and 3) students' perceptions of the potential for technology to ameliorate these barriers. The data reveal particular barriers to engagement created by the LMS environment; grades and feedback are commonly separated spatially, limiting attention to the latter.

Joan Palmiter Bajorek et.al [7] Speech recognition has significant race and gender biases. As with facial recognition, web searches, and even soap dispensers, speech recognition is another form of AI that performs worse for women and non-white people. To be clear, I do not believe that the creators of these systems set out to build racist or sexist products. It's doubtful these biases are intentional, but they are still problematic.
Eduardo Rodríguez-Orozco et.al [8] a new embedded chaotic cryptosystem is introduced herein with the aim to encrypt digital images and performing speech recognition as an external access key. The proposed cryptosystem consists of three technologies: (i) a Spartan 3E-1600 FPGA from Xilinx; (ii) a 64-bit Raspberry Pi 3 single board computer; and (iii) a voice recognition chip manufactured by Surplus.

Abhinav Dhall et.al [9] this paper details the sixth Emotion Recognition in the Wild (EmotiW) challenge. EmotiW 2018 is a grand challenge in the ACM International Conference on Multimodal Interaction 2018, Colarado, USA. The challenge aims at providing a common platform to researchers working in the affective computing community to benchmark their algorithms on 'in the wild' data. This year EmotiW contains three sub-challenges: a) Audio-video based emotion recognition; b) Student engagement prediction; and c) Group-level emotion recognition. The databases, protocols and baselines are discussed in detail.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

Shelli Wynants et.al [10] given the variety of learning and engagement needs of the increasingly diverse student population in higher education, flexible approaches to teaching are critical for improving student success. Professional development that provides faculty exposure to effective, evidence-based instructional strategies in an online context may enhance their teaching practices. This study explored the advantages and disadvantages of the online context from the perspectives of ten faculties who completed an online disability awareness program, designed using two promising models:

1. Ability to work in a team
2. Communication skills
3. Reasoning ability
4. Leadership skills
5. Initiative
6. Assertiveness
7. Flexibility
8. Creativity
9. Ability to think on ones feet

From the literature it is clear that there is a need of neural network-based approach to get the maximum accuracy. So, the proposed system integrates techniques such as Principal Component Analysis Companies conduct group discussion after the written test so as to check on your interactive skills and how good you are at communicating with other people. The GD is to check how you behave, participate and contribute in a group, how much importance do you give to the group objective as well as your own, how well do you listen to viewpoints of others and how open-minded are you in accepting views contrary to your own. The aspects which make up a GD are verbal communication, non-verbal behavior, and conformation to norms, decision-making ability and cooperation. You should try to be as true as possible to these aspects.

III. SYSTEM ARCHITECTURE

System introduced in this paper i.e. “Development of Voice Recognition for Student Group discussion”, is a real-time data processing system. Laptop takes input as speech from mic. Then Laptop compares speech from user with speech from dataset and accordingly classifies the signal. Classification here means system classify speech signal into two parts. If speech matches with speech from dataset, system allows particular students voice can we listen properly. And while discussion in students the manager will may be destroy the concentration it will be hard find the who student give the best performance. So these systems identify particular student voice and regains listen the speech and identify student performance who is caple for our organization.

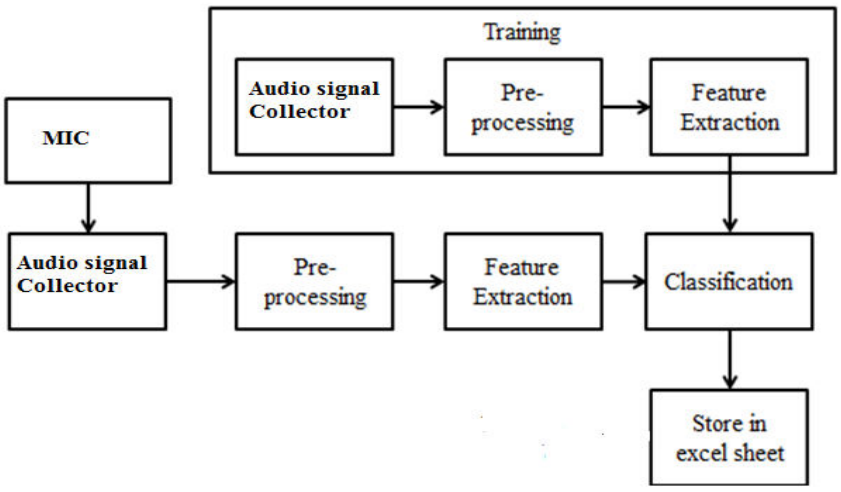

Figure 1: System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

Mic: Mic gives the audio signal and is feed as input to the audio signal collector.

Audio signal collector: Its collect the audio and input to the python for further processing.

Pre-processing: Pre-processing of image commonly involves removing low-frequency background noise, normalizing the intensity of the individual particles images, removing reflections, and masking portions of images. Image pre-processing is the technique of enhancing data images prior to computational processing.

Feature extraction: Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved.

Classification: Image classification refers to the task of extracting information classes from two or many class of image. Voice is detected and recognized in this phase. When voice is recognized, a tick mark is created against name of student whose voice is recognized. Data is stored in excel sheet at the end of the month.

Algorithm -Convolutional Neural Network

Convolutional Neural Network can do a lot of good things if they are fed with a bunch of signals for instance to learn some basic signals such as frequency changes, amplitude changes. Since, they are multi neural networks; the first layer is fed with this information. The second layer is fed with some recognizable features. To illustrate this, a signal of two-dimensional array of pixels is considered. It is a check board with each square on the board is either light or dark color. By observing the pattern CNN decides whether it is a signal with frequency change or amplitude change.
The convolutional neural network match the parts of the signal instead of considering the whole signal of pixels as it becomes difficult for a computer to identify the signal when the whole set of pixels are considered [8][9]. The mathematics behind matching these is filtering. The way this is done is by considering the feature that is lined up with this patch signal and then one by one pixel are compared and multiplied by each other and then adds it up and divide it with the total number of pixels. This step is repeated for all the pixels that is considered. The act of convolving signals with a bunch of filters, a bunch of features which creates a stack of filtered images is called as convolutional layer. It is a layer because it is operating based on stack that is in convolution one signal becomes a stack of filtered signals. We get a lot of filtered signals because of the presence of the filters. Convolution layer is one part.

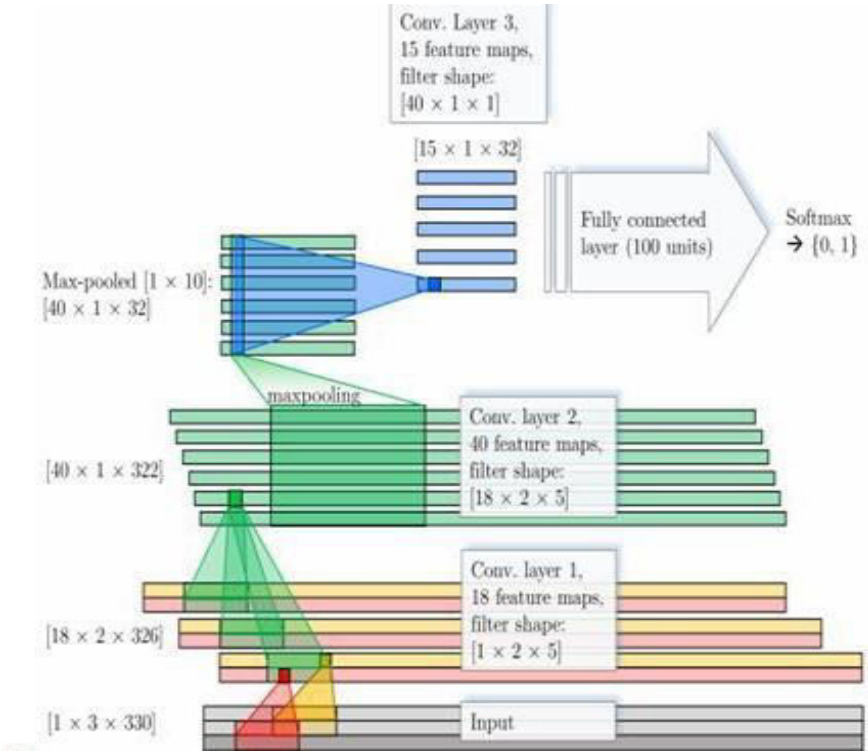

Figure 2: CNN structure

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

The next big part is called as pooling that is how a signal stack can be compressed. This is done by considering a small window pixel which might be a 2 by 2 window pixel or 3 by 3. On considering a 2 by 2 window pixel and pass it in strides across the filtered signals, from each window the maximum value is considered. This passed through the whole signal. At the end it is found that by considering only the maximum values the size of the filtered signal is reduced [10]. The third part is normalization; in this if a pixel value is negative then the negative values are replaced with zeros. This is done to all the filtered signals. This becomes another type of layer which is known as a rectified linear unit, a stack of signals which becomes a stack of signals with no negative values. Now the three layers are stacked up so that one output will become the input for the next. The final layer is the fully connected layer.

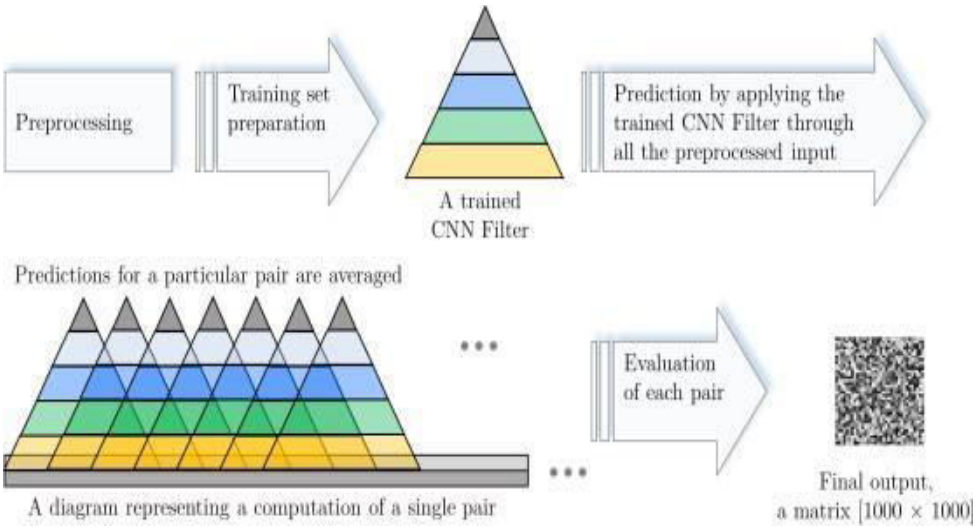

Figure 3: CNN simplified workflow model.

The standard feed-forward fully connected Neural network (NN) is a computational model composed of several layers. An input to a particular unit is outputs of all the units in the previous layer (or input data for the first layer). The unit output is a single linear regression, to which output value a specific activation function is applied. Convolutional neural network (CNN) is a type of NN where the input variables are related spatially to each other. To take into account very important spatial positions, CNNs were developed. Not only they are able to detect general spatial dependencies, but also are capable of specific patterns recognition. Shared weights, representing different patterns, improve the convergence by reducing significantly the number of parameters. CNN recognize small patterns at each layer, generalizing them (detecting higher order, more complex patterns) in subsequent layers. This allows detection of various patterns and keeps the number of weights to be learnt very low.

This CNN module is applied on the feature of the voice samples. The MFCC features have been calculated from the voice samples. The MFCC feature extraction technique basically includes windowing the signal, applying the DFT, taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by applying the inverse DCT. The training of these features gives trained neural network file which then can be easily used for the testing.

IV. RESULTS

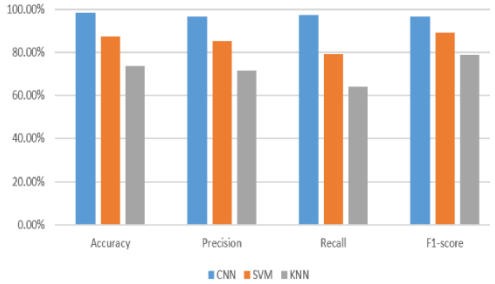

Figure 4: Comparison of performance parameters of CNN with SVM and KNN

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

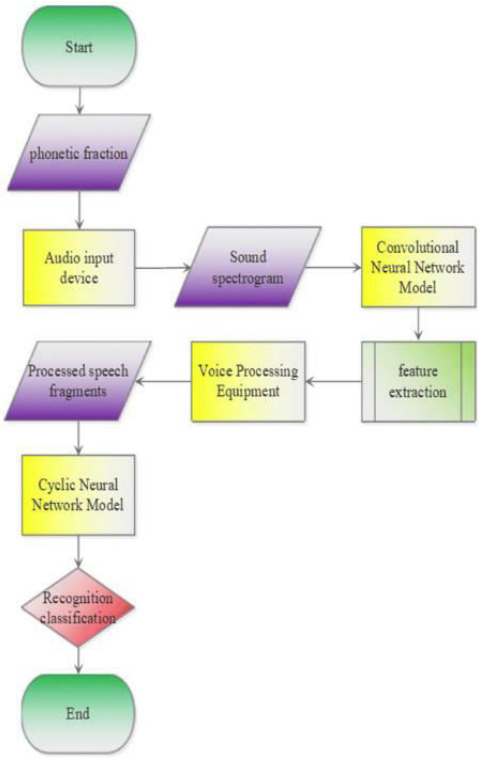

Figure 5: Flow chart of the proposed system architecture

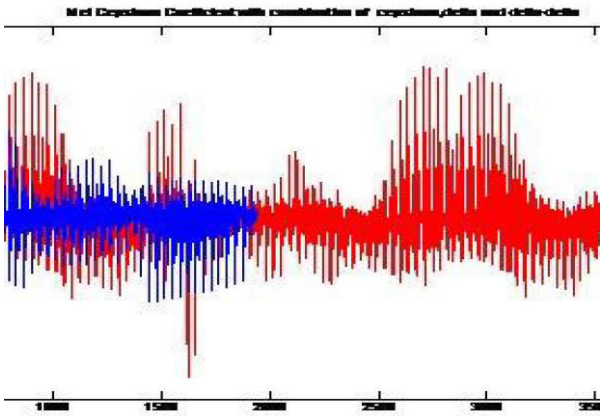

Figure 6: Frequency domain analysis of voice sample

In above result it will shows that how voice recognition will do, and gives better performance than other related system in this system we have to find that better performance student to give prospected result and HR or manager to find them student ability or performance it is capable for over organization and solving them group discussions other issues.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003052 |

V. CONCLUSION

In this system, voice recognition of group discussion gives the more benefits to the manger or HR as well as student to find them student ability, performance communication skills and other skills. The main moto of the system is to design recognize the audio of the any particular user. The performance parameters of the CNN are much better than the SVM and KNN which indicates that the voice recognition with MFCC features along with CNN is the best combination to get more accurate results. Proposed system architecture gives accuracy, precision, recall and F1-score as 99.74%, 98.34%, 99.14% and 98.28% respectively.

REFERENCES

1. Bristol, G. A. M. (2022). Integrating of Voice Recognition Email Application System for Visually Impaired Person using Linear Regression Algorithm. Proceedings of The International Halal Science and Technology Conference, 14(1), 56–66
2. S Stefanenko, L V Lipinskiy, A S Polyakova, J A Khudonogova and E S Semenkin." An intelligent voice recognition system based on fuzzy logic and the bag-of-words technique" IOP Conf. Series: Materials Science and Engineering 1230 (2022) 012020 IOP Publishing doi:10.1088/1757-899X/1230/1/012020
3. Wang, Z.M. (2022).The Assessment of Interpersonal Attraction and Group Identification in Virtual Group Network. Social Networking, 11, 1-14.
4. Hao Shao, Hongxia Wang, Yuying Cao, Xiang Ding, Jian Fang, Wenyu Wang, Xin Jin,Lu Peng, Daquan Zhang, and Tong Lin." High-Performance Voice Recognition Based on Piezoelectric Polyacrylonitrile Nanofibers"IEEE 2021.
5. Huaixiang Hu[1], Jiatong Li[1], Chunchun Wu[1], Xueyang Li[1] and Yuping Chen[1]"Design and Implementation of Intelligent Speech Recognition System Based on FPGA"Published under licence by IOP Publishing Ltd Journal of Physics: Conference Series, Volume 2171, International Conference on Computer, Big Data and Artificial Intelligence (ICCBDAI 2021) 12/11/2021 - 14/11/2021
6. T1 - "Check the grade, log out": students' engagement with feedback in learning management systems" Winstone, Naomi,Bourne, Jessica, Medland, Emma, Niculescu, Irina, Rees, Roger. 2021.DA - 2021/05/19.N1 - doi: 10.1080/02602938.2020.1787331
7. Joan Palmiter Bajorek." Voice Recognition Still Has Significant Race and Gender Biases" MAY 10, 2019
8. Rodríguez-Orozco, Eduardo, Enrique E. García-Guerrero, Everardo Inzunza-Gonzalez, Oscar R. López-Bonilla, Abraham Flores-Vergara, Jose R. Cárdenas-Valdez, and Esteban Tlelo-Cuautle. 2018. "FPGA-based Chaotic Cryptosystem by Using Voice Recognition as Access Key" Electronics 7, no. 12: 414.
9. Abhinav Dhall, Amanjot Kaur." EmotiW 2018: Audio-Video, Student Engagement and Group-Level Affect Prediction" ICMI'18, October 16-20, 2018, Boulder, CO, USA
10. Shelli Wynants, California State University, Fullerton Jessica Dennis, California State University, Los Angeles." PROFESSIONAL DEVELOPMENT IN AN ONLINE CONTEXT: OPPORTUNITIES AND CHALLENGES FROM THE VOICES OF COLLEGE FACULTY" IEEE 2018

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details