iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3002

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Age and Gender Identification by Indian Multilingual Speech Sample

Barkha Shrivastava[1], Vinay Jain[2]
[1, 2]*Electronics & Telecom (Communication System), Shri Shankaracharya Technical Campus (SSGI), Bhilai (CG)*

Abstract: ***The human voice is contained sound made by a person utilizing the vocal cord for talking, singing, snickering, crying and yelling. It is especially a bit of human sound creation in which the vocal string is the fundamental sound source, which assumes a vital job in the discussion. The uses of discourse or voice preparing innovation assume a pivotal job in human PC communication. The framework enhances gender orientation ID, age amass characterization, age and feeling acknowledgment execution. The examination work utilizes new and effective techniques for highlight extraction of discourse or voice and grouping of standard strategy on the different sound datasets. Mel Frequency Cepstral Coefficients highlight extraction and determination is performed to locate an increasingly reasonable list of capabilities for building speaker models. The proposed framework utilizes Gaussian Mixture Model is a super vector for framework include choice and highlight displaying. Bolster Vector Machine arrangement and highlight coordinating procedure is utilized to order the component for various age bunches like youngster, adolescent, youthful, grown-up and higher ranking than increment the resultant execution and precision. The database is made utilizing the sound records for each age gathering of speaker and for every feeling as an info, performs highlight extraction and distinguishes the gender orientation, arrange age gathering, and perceive age and feeling.***

Keywords: Mel Frequency Cepstral Coefficient (MFCC), Gaussian Mixture Model (GMM), Expectation-Maximization (EM), Maximum a Posteriori (MAP), Hidden Markov Models (HMMs), Suprasegmental Hidden Markov Models (SPHMMs), Interactive Voice Response System (IVRs).

I. INTRODUCTION

Human cooperation with PCs in done from various perspectives and the interface among human and the PC is urgent to encourage this communication. Most extreme work area applications, web utilizing programs like Firefox, chrome and web wayfarer. The PCs make utilization of the common Graphical User Interfaces (GUI). Voice User Interfaces (VUI) is utilized for discourse acknowledgment and blending frameworks. Human Computer Interaction (HCI) plans to enhance the interface among clients and PCs by making PCs progressively usable and open to clients require. There are numerous speaker qualities that have valuable applications. The most prevalent incorporate gender orientation, age, wellbeing, dialect, tongue, emphasizes communist, idiolect, passionate state and consideration state.

These attributes have numerous applications in exchange frameworks, discourse blend, legal, call steering, discourse interpretation, dialect learning, appraisal frameworks, speaker acknowledgment, meeting program, law implementation, human robot association and brilliant workspaces. For instance, the verbally expressed exchange framework gives benefits in the areas of fund, travel, planning, mentoring. The frameworks need to accumulate data from the client naturally so as to give opportune and important administrations. Most phones based administrations today utilize spoken discourse frameworks to either course calls to the proper operator or even handle. The total administration is given by a programmed framework. For instance, shopping frameworks can prescribe reasonable merchandise fitting to the age and gender orientation of the customer. The speaker explicit attributes of the flag can be misused by audience members and mechanical applications to depict and arrange speakers, in view of age, gender orientation, complement, dialect, feeling or wellbeing critical normal for human discourse or voice based interfaces is the constancy of the phonetic, syntactic and lexical properties of the expression or word verbally expressed by the client. Human voice based gender orientation; age gathering and exact age estimation are troublesome. To start with, ordinarily there is a contrast between the ages of a speaker as saw the recognized age and their genuine age is evaluated age. The law requirement has been worried about various biometric highlights to recognize the every human uniqueness. Distinctive biometric highlights can be utilized for exceptional human distinguishing proof, for example, fingerprints, facial, hand geometry design, signature elements and voice designs. In some criminal cases, the accessible proof is as recorded discussions or phonetic voice. The discourse examples can incorporate interesting and critical data to law authorization work force.

A. Voice Features

The long discourse highlights separated from the more extended sections of discourse flag, for example, whole sentences, words, syllables are known as supra segmental or prosodic highlights. They typically speak to the discourse properties like cadence, stretch, inflection, clamor and term. Acoustic corresponds of prosodic highlights are pitch, vitality, term and their subordinates. The feeling explicit data about shapes and sizes of the vocal tract, in charge of creating diverse sound units and the related development of articulators are caught utilizing phantom highlights. The attributes of glottal movement, explicit to the feelings are assessed utilizing excitation source highlights.

The talk data on feeling acknowledgment has been joined with acoustic associates to enhance the general execution of feeling order, reiteration or revision data was utilized for the talk data, likewise embraced redundancy as their talk data.

II. LITERATURE REVIEW

In [1] presents a measurement decrease procedure which plans to enhance more prominent productivity and the exactness of speaker's age gathering and exact age estimation frameworks dependent on the human voice flag. Two distinct gender orientations based age estimation approaches considered, the first is the age gathering (senior, grown-up, and youthful) arrangement and the second is an exact age estimation utilizing relapse strategy. These two methodologies utilize the GMM super vectors as highlights for a classifier show. Age gather characterization doles out an age gathering to the speaker and age relapse appraises the speaker's exact age in years.

In paper [2] presents gender recognition is an incredibly helpful errand for a broad assortment of voice or discourse based applications. In the expressed dialect frameworks INESC ID, the gender orientation distinguishing proof part is starting and the fundamental segment of our voice preparing framework, where it is used before speaker bunching, so as to abstain from blending speakers among male and female gender in a similar group. Gender orientation data (male or female) is likewise used to make gender subordinate acoustic module for discourse acknowledgment.

In [3] present new gender recognition and an age estimation approach are proposed. To build up this technique, in the wake of choosing an acoustic highlights show for all speakers of the example database, Gaussian blend loads are removed and associated with construct a super vector for every speaker. At that point, half and half engineering of General Regression Neural Network (GRNN) and Weighted Supervised Non Negative Matrix Factorization (WSNMF) are produced utilizing the made super vectors of the preparation informational collection. The half breed technique is utilized to distinguish the gender orientation speaker while testing and to gauge their age. Distinctive biometric highlights can be utilized for criminological distinguishing proof. Picking a strategy relies upon its utilization and proficient unwavering quality of a specific application and the accessible information type. In some wrongdoing cases, the accessible proof or evidence may be as recorded voice. Discourse examples can incorporate novel and essential data for law authorization faculty.

In [4] basically centered around improving feeling acknowledgment and distinguishing proof execution dependent on a two phases that is mix of gender orientation recognizer and feeling recognizer. The framework work is a gender orientation subordinate, content free and speaker autonomous feeling recognizer. Both Hidden Markov Model (HMM) and Supra segmental Hidden Markov Model (SPHMM) have utilized as classifiers in the two phase design. This design has been assessed on two unique and separate discourse databases. The two databases are passionate prosody discourse and transcripts database and human voice gathered database.

In [5] investigates the location of explicit sort feelings utilizing talk data and dialect in blend with acoustic flag highlights of feeling in discourse signals. The primary spotlight is on an identifying sort of feelings utilizing spoken dialect information got from a call focus application. Most past work in sort feeling acknowledgment has utilized just the acoustic highlights data contained in the discourse. The framework contains three wellsprings of data, lexical, acoustic and talk is utilized for speaker's feeling acknowledgment.

In [6] create models for distinguishing different qualities of a speaker dependent on spoken the content alone. These qualities or characteristics incorporate whether the speaker is talking local dialect, the speakers age and gender orientation, the provincial data detailed by the speakers. The exploration investigates different lexical highlights data just as highlights motivated by phonetic (a dialect related) data and various word and lexicon of effect in dialect. This framework proposes that when sound or voice information isn't accessible, by investigating successful sound capabilities just from articulated content and framework mixes of various order calculations, specialist assemble measurable models to distinguish these qualities of speakers, proportional to systems that can investigate the sound data.

In [7] present speaker trademark acknowledgment and recognizable proof field has made broad utilization of speaker MAP adjustment methods. The adjustment permits speaker demonstrate include parameters to be assessed utilizing less discourse

information than required for Maximum Likelihood (ML) preparing technique. The Maximum Likelihood Linear Regression (MLLR) and Maximum a Posteriori (MAP) procedures have commonly been utilized for speaker show adjustment. As of late, these adjustment procedures have been joined into the element extraction phase of the SVM classifier based speaker distinguishing proof and acknowledgment frameworks. In [15] people, enthusiastic discourse acknowledgment contributes a lot to make amicable human to machine connection, furthermore with numerous potential applications. Three ways to deal with enlarge parallel classifier are looked at for perceiving feelings from a discourse by the discourse database. Classifier connected on prosody, otherworldly, MFCC and other normal highlights. One is standard order plans (one versus one) and two strategies are coordinated a cyclic Graph (DAG) and Unbalanced Decision Tree (UDT) that can frame a paired choice tree classifier. The various leveled arrangement strategy of highlight driven progressive SVMs classifiers is planned, it utilizes distinctive component parameters to drive each layer and the feeling can be sub isolated layer by layer. At last, examination of the characterization rate of those three expands twofold grouping, DAG framework plays out the best to test database and standard classifier isn't a long ways behind, the UDT is the poorest due to depending on upper layer arrange preparing.

In [8] the extraction and coordinating procedure is executed after the flag preprocessing is performed. Non parametric strategy for demonstrating is the human voice handling System. The nonlinear arrangement called as Dynamic Time Warping (DTW) utilized as highlights coordinating systems. This paper introduces the method of MFCC include extraction and wrapping system to look at the test designs.

III. METHODOLOGY

The design of this system architecture contains two phases i.e. Training phase and testing phase.

The preparation stage utilized huge sound dataset for preparing the framework utilizing the MFFC highlight extraction system connected for separating the interesting element of sound/voice record and make the component vector. GMM super vector portrayal and measurement decrease for each element type, and so forth. Preparing stage connected to the huge set example informational collection for preparing reason. Train the framework is an over MFCC highlights, removed from discourse expressions of discourse sessions. The discourse sessions used to prepare the framework foundation model ought to be broadened and consistently conveyed over speaker ages and genders. In the testing stage, the discourse session is handled same as preparing stage. A GMM display is prepared, a super vector is framed and the measurement decrease projection framework is connected on it to make a diminished testing highlight vector. SVM order calculation and coordinating strategy are connected to group the outcome and locate the correct outcome for information voice.

A. Feature Extraction

The extraction of the best parametric portrayal of the acoustic signs of the human voice is an essential assignment to deliver a letter acknowledgment execution. The outcome proficiency of highlight extraction stage is vital for the following stage like displaying, arrangement and highlight coordinating since it influences its conduct.

B. Gaussian Mixture Model

A GMM demonstrate is a likelihood thickness work spoken to utilizing a weighted whole of all Gaussian part densities. Demonstrating procedure is regularly utilized parametric model of the likelihood circulation of highlights in a proposed framework, for example, voice tract related unearthly highlights of flag in a speaker acknowledgment framework. The parameters are evaluated from preparing test voice information utilizing the iterative EM calculation or MAP estimation from an all around prepared earlier demonstrating methodology is an outstanding displaying procedure in content autonomous speaker acknowledgment frameworks for edge based highlights.

$$\lambda = \sum_{k=o}^{n} \binom{n}{k} x^k a^{n-k} \qquad \text{------ (1)}$$

The each component density is a D variant Gaussian function of the form, with mean vector u_i and covariance matrix $\sum$ i=1, the complete Gaussian mixture model is parameterized by the covariance matrices, mixture weights and mean vectors from all component densities.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

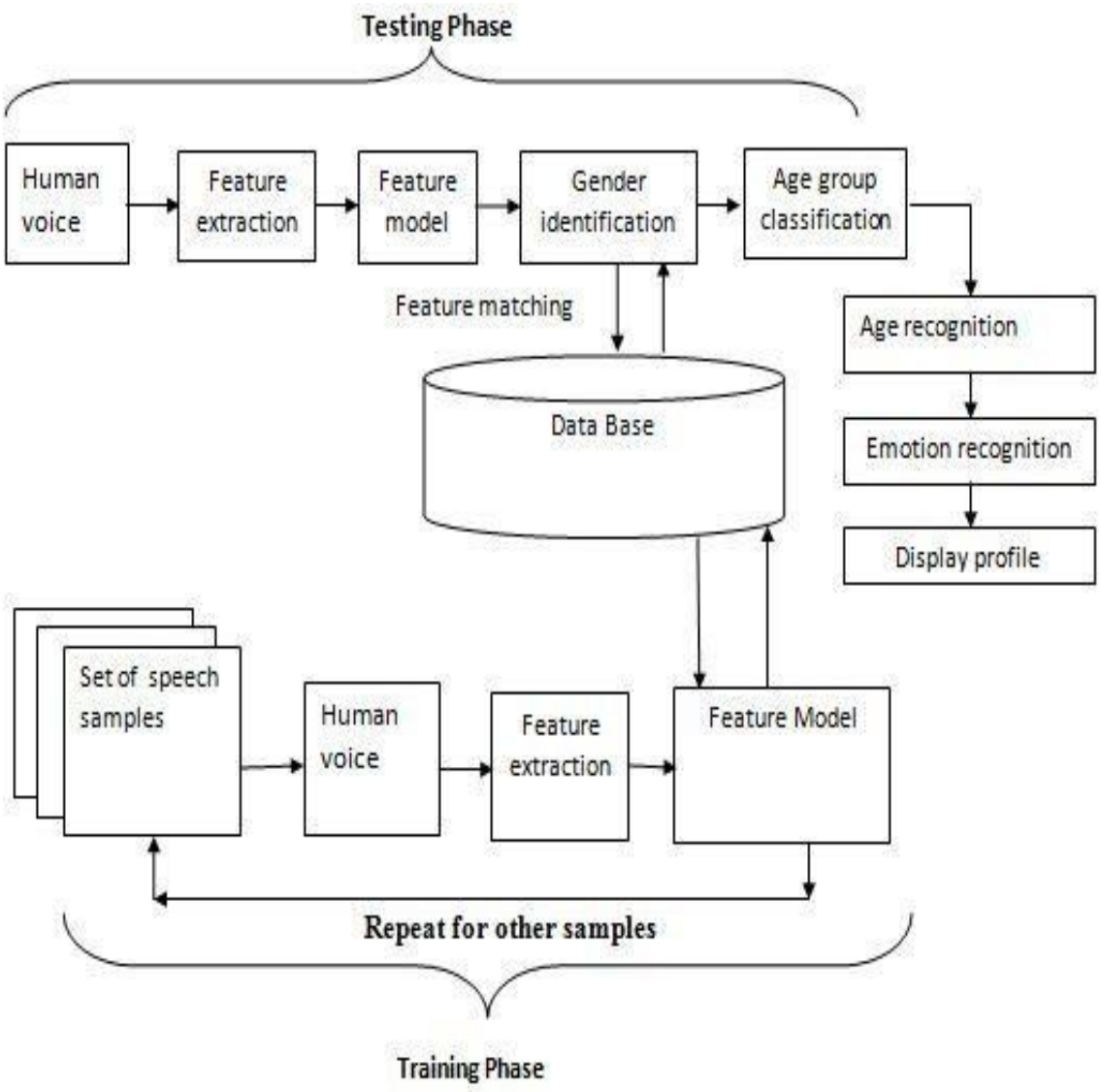


Figure 1: Age and Gender detection block diagram

C. Hidden Markov Model (HMM)

In the Markov demonstrate each state relates to one noticeable occasion. Be that as it may, this model is unreasonably prohibitive, for a substantial number of perceptions the extent of the model detonates, and the situation where the scope of perceptions is persistent isn't shrouded in any way. The Hidden Markov idea broadens the model by decoupling the perception grouping and the state arrangement. For each express a likelihood conveyance is characterized that determines how likely every perception image is to be created in that specific state. As each state can now on a fundamental level produce every perception image it is never again conceivable to see which state succession created a perception arrangement just like the case for Markov models, the states are presently covered up, henceforth the name of the mode.

A Hidden Markov model can be characterized by the accompanying parameters:

The quantity of unmistakable perception images M. An output alphabet = { } 1 2 M V

The number of states is N.

A state space Q = {1, 2,...N}

States will usually be indicated by i, j a state that the model is in' at a particular point in time t will be indicated by qt. Thus, qt = i means that the model is in state i at time t. A probability distribution of transitions between states { }ij A= a , where

$$A_{ij} = P(q_{t+1} = j \mid q_t = i) \quad 1<=i, j<=N$$

IV. RESULT

We have 10 speaker data in three languages and find the mel frequency cepstral coefficients value. Each value is saved in 5x5 matrixes. We have found the minimum and maximum value of 5x5 matrixes. After that we have matched one sample with data base value and determine the Pair wise distance between two sets of observations. The set of Pair wise distance is given in below table.

Speaker	Language	Minimum Value	Maximum Value	Mean Value
1	Hindi	-0.6567	0.5413	0.007
	Marathi	-0.4975	0.4675	0.009
	Rajasthani	-0.4665	0.5424	0.006
2	Hindi	-0.1845	0.1756	0.002
	Marathi	-0.1154	0.1796	0.005
	Rajasthani	-0.2154	0.2457	0.005
3	Hindi	-0.2375	0.3149	0.009
	Marathi	-0.2457	0.2147	0.002
	Rajasthani	-0.3145	0.3475	0.004
4	Hindi	-0.3439	-0.3745	0.008
	Marathi	-0.4986	0.4756	0.01
	Rajasthani	-0.4576	0.4786	0.008
5	Hindi	-0.5124	0.5476	0.005
	Marathi	-0.5486	0.5497	0.009
	Rajasthani	-0.5685	0.5412	0.009
6	Hindi	-0.1549	0.1462	0.001
	Marathi	-0.1469	0.1498	0.002
	Rajasthani	-0.2146	0.2579	0.003
7	Hindi	-0.2685	0.2458	0.003
	Marathi	-0.2568	0.2348	0.001
	Rajasthani	-0.4976	0.4856	0.005
8	Hindi	-0.5689	0.5489	0.007
	Marathi	-0.5698	0.5664	0.006
	Rajasthani	-0.4986	0.4576	0.002
9	Hindi	-0.4587	0.4578	0.003
	Marathi	-0.5124	0.5486	0.004
	Rajasthani	-0.5486	0.5478	0.004
10	Hindi	-0.3654	0.3657	0.006
	Marathi	-0.4975	0.4589	0.006
	Rajasthani	-0.5214	0.5478	0.007

V. ACKNOWLEDGEMENT

I might want to recognize. My appreciation is to various individuals who have helped me in various courses for the fruitful culmination of my theory. I accept this open door to express a profound feeling of appreciation towards my guide, Mr.Vinay Jain Associate Professor (ET&T), Faculty of Engineering and Technology, SSTC-SSGI, Bhilai for giving astounding direction, consolation and motivation all through the task work. Without his priceless direction, this work could never have been a fruitful one. I am appreciative to Mr. Chinmay Chandrakar HOD, ET&T Department, and Dr.P. B. Deshmukh, Director, SSTC-SSGI, Faculty of Engineering and Technology, Bhilai for their thoughtful help and participation. I might likewise want to express gratitude toward Mr. Chandrashekhar Kamargaonkar M.E. In-control, SSTC-SSGI, Mr. Sharad Mohan Shrivastava, Faculty of Engineering and Technology, Bhilai for his caring help and accommodating recommendations. I feel massively moved in communicating my obligation to my folks whose forfeit, direction and endowments helped me to finish my work.

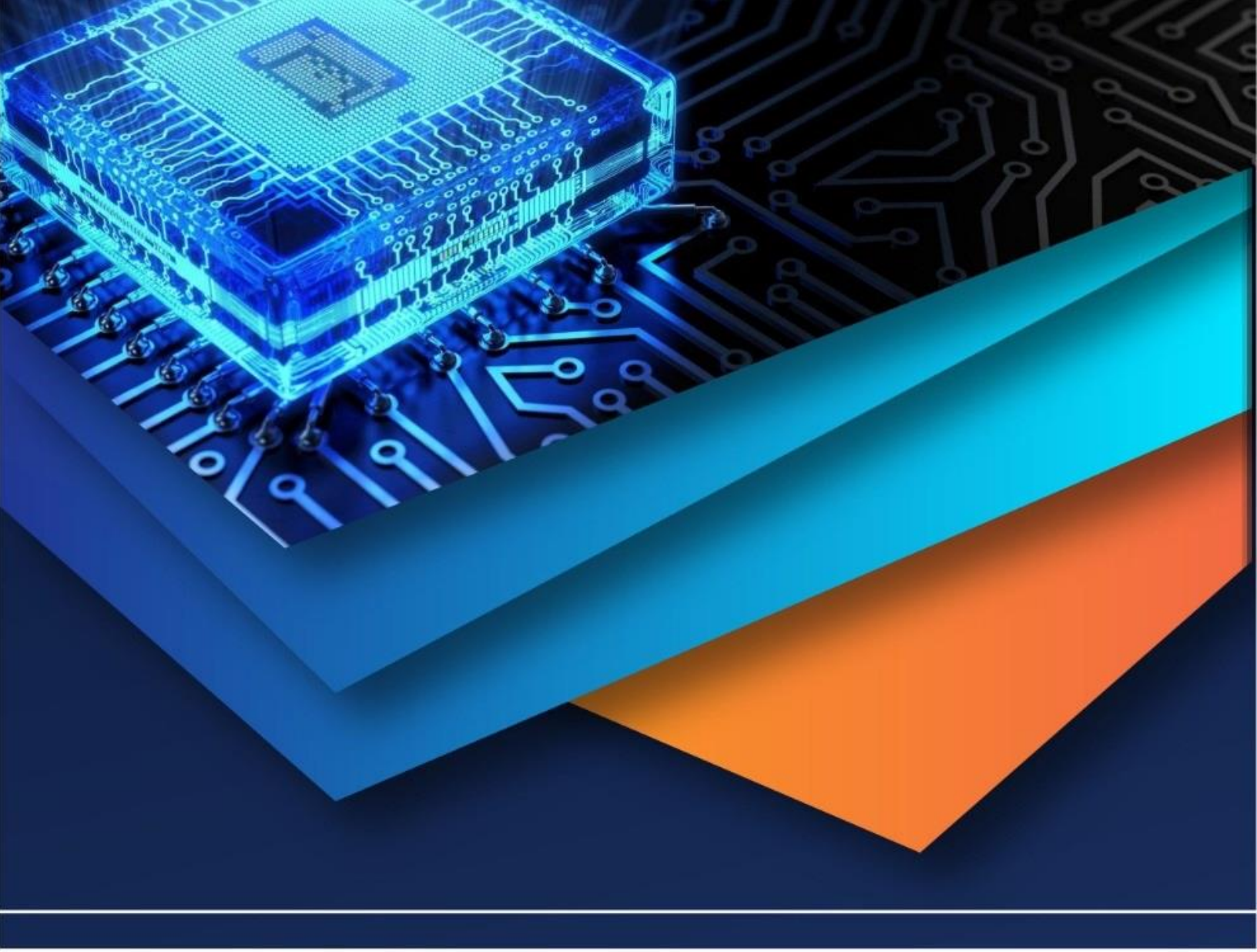
VI. CONCLUSION

Along these lines the proposed framework help to distinguish arrange and perceive correct speaker age with feeling and showing profiles of speaker utilizing the prepared database. The speaker profile is useful in numerous applications like for notice, focusing to specific individuals, naturally distinguishing proof of this component, age, feeling to give office and administration to client in a call focus, in some field speaker's voice can be utilized as the biometric security in light of the fact that every human has a one of a kind

voice example and extraordinary element. The outcome is doable approach to expand the precision and proficiency of framework yield. The future upgrade of the framework can be stretched out to perceive for progressively muddled clamor test (.wav document). The wellbeing state of the speaker can likewise recognize separate the individual speaker characterization and age additionally conceivable to identify for blend mode gender orientation speaker.

REFERENCES

[1] Gil Dobry, Ron M. Hecht, Mireille Avigal and Yaniv Z, SEPTEMBER, 2011. Supervector Dimension Reduction for Efficient Speaker Age Estimation Based on the Acoustic Speech Signal, IEEE transaction V.19, NO. 7.
[2] Hugo Meinedo1 and Isabel Trancoso, 2008Age and Gender Classification using Fusion of Acoustic and Prosodic Features, Spoken Language Systems Lab, INESC-ID Lisboa, Portugal, Instituto Superior Tecnico, Lisboa, Portugal.
[3] Ismail Mohd Adnan Shahin, 2013Gender-dependent emotion recognition based on HMMs and SPHMMs, Int J Speech Technol, Springer 16:133141.
[4] Mohamad Hasan Bahari and Hugo Van h, ITN2008 Speaker Age Estimation and Gender Detection Based on Supervised Non Negative Matrix Factorization, Centre for Processing Speech and Images Belgium.
[5] Shivaji J Chaudhari and Ramesh M Kagalkar, May 2015 Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition, International Journal of Computer Applications (IJCA) (0975 - 8887), Volume 117 No. 17.
[6] Shivaji J. Chaudhari and Ramesh M. Kagalkar, July 2015 A Methodology for Efficient Gender Dependent Speaker Age and Emotion Identification System, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) ISSN 2319-5940,Volume 4, Issue 7.
[7] Chul Min Lee and Shrikanth S. Narayanan, 2005 Toward Detecting Emotions in Spoken Dialogs, IEEE transaction 1063-6676.
[8] [Tetsuya Takiguchi and Yasuo Ariki, 2006 Robust feature extraction using kernel PCA, Department of Computer and System Engg Kobe University, Japan, ICASSP 14244-0469.
[9] Michael Feld, Felix Burkhardt and Christian Muller, 2010 Automatic Speaker Age and Gender Recognition in the Car for Tailoring Dialog and Mobile Services, German Research Center for Artificial Intelligence, INTERSPEECH.
[10] M A. Hossan, Sheeraz Memon and Mark A Gregory, A Novel Approach for MFCC Feature extraction, RMIT university, Melbourne, Australia, IEEE, 2010.
[11] Ruben Solera-Ure, 2008 Real-time Robust Automatic Speech Recognition Using Compact Support Vector Machines, TEC 2008-06382 and TEC 2008-02473.
[12] Wei HAN and Cheong fat CHAN, 2006 An Efficient MFCC Extraction Method in Speech Recognition, Department of Electronic Engineering, The Chinese University of Hong Kong Hong Kong, 78039390-06/IEEE ISCAS.
[13] AU Khan and L. P. Bhaiya, 2008 Text Dependent Method for Person Identification through Voice Segment, ISSN- 2277-1956 IJECSE.
[14] Felix Burkhardt, Martin Eckert, Wiebke Johannsen and Joachim Stegmann, 2010A Database of Age and Gender Annotated Telephone Speech, Deutsche Telekom AG Laboratories, Ernst-Reuter-Platz 7, 10587 Berlin, Germany.
[15] Lingli Yu and Kaijun Zhou, March 2014, A Comparative Study on Support Vector Machines classifiers for Emotional Speech Recognition, Immune Computation (IC) Volume2, Number:1.
[16] Rui Martins, Isabel Trancoso, Alberto Abad and Hugo Meinedo, 2009, Detection of Childrens Voices, Intituto Superior Tecnico, Lisboa, Portugal INESC-ID Lisboa, Portugal.
[17] Chao Gao, Guruprasad Saikumar, Amit Srivastava and Premkumar Natarajan, 2011, Open set Speaker Identification in Broadcast News, IEEE 978-1-45770539.
[18] Shivaji J Chaudhari and RameshMKagalkar, 2014, A Review of Automatic Speaker Age Classification, Recognition and Identifying Speaker Emotion Using Voice Signal, International Journal of Science and Research (IJSR 2014), ISSN(Online): 23197064,Volume 3.
[19] M Ferras, C CLeung, C Barras and Jean Luc Gauvain, 2010, Comparison of Speaker Adaptation Methods as Feature Extraction for SVM-Based Speaker Recognition, IEEE Transaction 1558-7916.
[20] Chao Gao, Guruprasad Saikumar, Amit Srivastava and Premkumar Natarajan, 2011, Open-SetSpeaker Identification in Broadcast News, IEEE 978-1-45770539.
[21] ChaoWang, Ruifei Zhu, Hongguang Jia, QunWei, Huhai Jiang, Tianyi Zhang and LinyaoYu, 2013, Design of Speech Recognition System, IEEE 978-1-4673-27640/13.
[22] Manan Vyas, 2013“Gaussian Mixture Model Based Speech Recognition System Using Matlab”, Signal and Image Proc An International Journal (SIPIJ) Vol.4, No.4.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY