International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 5, October 2020, pp. 4752~4758
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i5.pp4752-4758

Creation of speech corpus for emotion analysis in Gujarati language and its evaluation by various speech parameters

Vishal P. Tank[1], S. K. Hadia[2]
[1]V T Patel Department of Electronics and Communication Engineering, Chandubhai S Patel Institute of Technology (CSPIT), Charotar University of Science and Technology (CHARUSAT), India
[2]Gujarat Technological University, India

Article Info

Article history:

Received Nov 18, 2019
Revised Mar 23, 2020
Accepted Apr 3, 2020

Keywords:

Emotion detection from speech
Energy
Gujarati language
MATLAB software
MFCC
Pitch

ABSTRACT

In the last couple of years emotion recognition has proven its significance in the area of artificial intelligence and man machine communication. Emotion recognition can be done using speech and image (facial expression), this paper deals with SER (speech emotion recognition) only. For emotion recognition emotional speech database is essential. In this paper we have proposed emotional database which is developed in Gujarati language, one of the official's language of India. The proposed speech corpus bifurcate six emotional states as: sadness, surprise, anger, disgust, fear, happiness. To observe effect of different emotions, analysis of proposed Gujarati speech database is carried out using efficient speech parameters like pitch, energy and MFCC using MATLAB Software.

Corresponding Author:

Vishal P. Tank,
V T Patel Department of Electronics and Communication Engineering,
Chandubhai S Patel Institute of Technology (CSPIT),
Charotar University of Science and Technology (CHARUSAT),
Changa-388421, Anand, Gujarat, India.
Email: vishaltank.ec@charusat.ac.in

1. INTRODUCTION

Speech and facial expression mainy two mode by which people interact and communicate to each other, betwixt speech is best mode for information exchange. Speech is a compund signal which contains the sharp details of language, speaker, emotion, and message [1]. It is importance to understand role of different emotions in speech because presecnce of emotions make speech more natural. Word "OKAY" spoken with different emotions have different meanings and inerpretation. Human robot interaction can be possible in better, effective and natural way if valid emotion gets involved in a speech.Finally this helps in to area of artificial intellience.

As mention earlier emotions can be perceived either from speech or facial expression (image processing), but dignosticate from the speech is complicated task. By recognitioning emotions of users add values in day to day life. Emotion recognition task is useful in day to day life in several ways like, lie detection system [2], audio/video retrieval [3, 4], artificial intelligence and robotics, assign priority to customers in various call-centers, improved diagnostic tool, intelligent teaching/tutoring system, language conversion, improved computer games, smart car board system and sorting of voicemail/ messages. Such utilisations make emotion recognition from speech as best research topic in the field of speech processing.

To have a speech database is essential in process of speech emotion recognition as shown in Figure 1 [5]. Researchers and scientists have developed speech corpora invarious languages like English, German, Chinese, Spanish, Japanese, Russian, Swedish, and Italian etc [6]. There are few speech databases available for official Indian languages like Hindi, Telugua andMalyalam [7]. As per author perception,

standard speech courpus is not available in Gujaratil language (official language of India) [8, 9]. In this paper, we are introducing the speech databse recorded in Gujarati language for emotion recognition.

It is very important to know how well the speech corups is prepared and hence its analysis is prime task [10]. To analyse speech features are extracted as described later. Emotion specific information is always present at excitation source, vocal tract, and linguistic levels. Individual emotions have a clear effect on speech spoken by human and it can be observed by evluationg various parameters/features like MFCC [11, 12], pitch, energy etc. The mention features are extracted from prposed Gujarati database to see the effects of various emotions and it's described in this paper. The paper is methodical as follows: 1) introduction, 2) process of speech emotion recognition (SER), 3) Gujarati language and its roots, 4) creation of emotional speech database for Gujarati language, 5) analysis of speech corpus using speech parameters, 6) discussion and concluding remarks, and 7) references.

2. BASIC FRMEWORK OF SPEECH EMOTION RECOGNITION (SER)

The process of speech recognition from speech can be understood as shown in Figure 1. The prerequisite of any SER system is suitable emotional speech database. Couple of researchers have done review for available speech database for SER in various languages like German, Spanish, English, etcand mainly it is categorized in simulated database (actor generated), elicited database and natural database [13].

From this available database one has to extract features from database. Suitable feature selection is an important task because it carries intended information and it decides overall efficiency of system. Generally three kinds of features are extracted from database 1) Excitation Source features like LP residual, glottal excitation signal, 2) Vocal track features like MFCC, LPCC 3) prosodic features like pitch, formants 4) Hybrid features [14, 15].

Various classifiers like GMM (Gaussian mixture model), HMM (hidden Markov model) are trained by extracted futures in it will decide the specific emotion. Normally choosing of a particular classifier is based on experimental results or thumb rule. Generally classifiers are categorized in linear classifiers (Naive Bayes classifier) and nonlinear classifiers (GMM, HMM) [16, 17].

3. GUJARATI LANGUAGE AND ITS ROOTS

Twenty two official languages are reported according to eighth Schedule of Indian constitution and Gujarati is part of it and majority spoken in Gujarat state in India. As per literature, Gujarati is approximately 700 years old and origniates from indo European family befor 1100 to 1500 AD. Gujarati is widely spoken language in India by number of native speakers, spoken by 55.5 million speakers which inherently about 4.5% of the total Indian population with 6th rank. It is the most widely spoken language in the world by number of native speakers as of 2007 with a rank of 26 [18]. Gujarati is the official language in country of India. Gujarati is the 24th ranked language spoken by 56.4 million people in the world and which makes 0.732 % of total world population of world as on March 2019. The location of Gujarat is shown in Figure 2.

According to reference and available literatures, ACS (American Community Survey) data by USA Census Bureau reported that 4.34 lakh of population speak Gujarati language as on 2017. Out side the India, Gujarati is also widely spoken in countries like United States, Canada, British and spoken to a lesser extent in China (particularly Hong Kong), Indonesia, Singapore, Australia.This makes ground truth to carry out research work in Gujarati language [19].

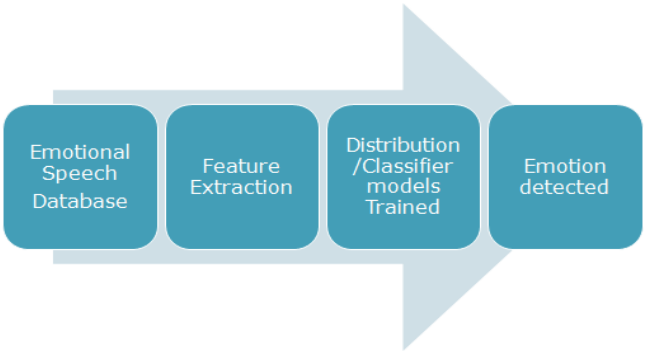

Figure 1. Basic framework for emotion speech recognition from speech

Figure 2. Location of Gujarat State in India available in Wikipedia

4. CREATION OF EMOTIONAL GUJARATI SPEECH CORPUS

Speech corpus can be created in various ways like acted, natural, induced etc [20, 21]. In this paper created Gujarati speech corpus is acted mode [22]. The process is described below.

4.1. Recording method

As shown in Figure 3 recording is performed using mobile phones. The distance between the speaker and the mobile phone is maintained around 20 centimeter. To make a common platform between all recordings only Lenovo (VIBE K5 note) is utilized. For recording, audio recorder application is used which is inbuilt in phone. Recording is done with rate of sampling frequency is 44100Hz and superior quality with wav file (*.wav).

4.2. Composition of proposed Gujarati emotional speech Corpus:

The proposed database is recorded using 9 artists (6 male and 3 Female) who are expertise in DRAMATICS. The recording was done in quite single room at Anand, Gujarat (state), India. All the speakers are in the age group of 20-25 years. For analyzing emotions, 24 different words are recorded with six different emotions as shown Table 1. Each of the speakers/artists has to speak the 24 words in 6 emotions. Speakers were well aware about the training sets and words.

Table 1 contains information about recorded 24 words in Guajarati language and its sequence. Selction of each word is done so that it covers entire range of Gujarati phonemes and its variability. Each emotional file is stored, numbered and labelled in the computer with appropriate emotional state. Table 2 contains six emotional states and its sequence number, like:
02-03-05.wav (02 is speaker-2, 03 is emotion-3, 05 is word-5)
04-05-20.wav (04 is speaker-4, 05 is emotion-5, 20 is word-20)
Total 1296 emotional speech samples (9 users*24 words*6 emotions) are recorded with six different emotional classes.

Table 1. Recorded 24 words in Gujarati language and its English meaning

Serial No.	Gujarati Word	Appropiate English Meaning
1	અભિમાન	Pride
2	કોશિશ	Try
3	પ્રમાણ	Proof
4	બહાર	Out
5	ઈશ્વર	God
6	ઉપકાર	Countenance
7	જેમતેમ	Any How
8	ક્ષત્રિય	Kshtriya
9	ક્રાંતિ	Revolution
10	અહંકાર	Ego
11	હથિયાર	Weapon
12	ઊચનીચ	Unequal
13	ઋણઆદેશ	Garnishee Order
14	એકધારું	Continuous
15	ખોટુ	False
16	ગૌરવ	Glory
17	ઘોષણાપત્ર	Manifesto
18	છેડતી	Molestation
19	ઝટપટ	Fast
20	ડગલું	Step
21	ઢીંગી	Imposter
22	ફરિયાદ	Compaint
23	ઠગાઈ	Cheating
24	તળિયાભાવ	Bottom Price

Table 2. Six emotions in Guajarati and its English meaning

Serial No.	Emotion in Gujarati	Appropiate English Meaning
1	દુઃખી	Sadness
2	આશ્ચર્ય	Surprise
3	ગુસ્સો	Anger
4	અરુચિ	Disgust
5	ડરીગટીલું	Fear
6	ખુશ	Happiness

Figure 3. Recording of speech database using mobile phone

5. EXPERIMENT RESULTS

Speech signal is a denouement of time varying vocal tract system agitated by the time varying excitation source signal. Henceforward speech features are present in both vocal tract system and excitation source characteristics. In this paper classification is observed by three different parameters Energy, MFCC 1 to 13 and Pitch [23]. Each parameter is evaluated and shown Speech parameter Vs Users (Speakers) for different emotion in Graph form and individual speakers and its individual values are presented in table form. Complete evaluation is carried in MATLAB software particularly R2015b version.

Energy: This can be considered as crucial parameter for SER (Speech emotion recognition). Normally energy range or value is low for the sadness (emotion-1), disgust (emotion-4), fear (emotion-5) and high range or value for the joy (emotion-6), anger (emotion-3) and surprise (emotion-2). Energy level evaluation of individual user is shown in graph and values are plotted in normal scale. During analysis frame size is kept with 160 samples and normalized in the range of (+1, -1). Energy of a speech signal can be finding out by using this equation:

$$E = \sum_{\infty}^{-\infty} |x[n]|^2$$

MFCC (Mel frequency cestrum coefficients): MFCC is widely used feature for emotion classification. MFCC purely described the shape of vocal track in form of short power spectrum. Evaluation of MFCC is carried out as follows: 1) divide the speech signal into short frames, 2) for each frame forecast the periodogram and estimate the power spectrum, 3) affix the mel filter bank to the power spectrum, 4) sum the energy in each filter finally take the logarithm of all filter bank energies. take the discrete cosine transform of the log filter bank energies, 5) keep DCT coefficients 2-13, and abdicate rest, and 6) the output of the filters from each frame is used as features and center frequencies of the filters are used in Mel scale by using the equation [24].

$$Mel(F) = 2595 \times \log_{10}(1 + f/700)$$

For the appraisal of the Mel frequency spectrum, 24 triangular filter banks are accumulated. These filters compute the spectrum around each center frequency with increasing bandwidths. In this evaluation of MFCC 1 to 13 features is shown in Tables 3–5. For reference here observation of MFCC-1, MFCC-2, and MFCC-3 are shown otherwise in actual all the parameters MFCC-1 to MFCC-13 are evaluated in same manner.

Pitch: Pitch is also another useful parameter which conveys considerable information for emotion classification. Table 6 and Figure 4 show evaluation of it [25]. In figure individual user wise energy values calculated and plotted against respective emotions. As shown in figure three emotions anger, surprise, happiness are higher energy band emotions and disgust, sadness, fear are falls under lower energy band. Figure clearly shows the separation of emotions using energy parameter of speech. Second important parameter pitch values are evaluated and its values are reported as shown in figure. Here the range of the pitch values are as following.Sandness (214-118 Hz), Surprise (208-152 Hz), Anger (211-165 Hz), Disgust (194-140 Hz), Fear (218-139 Hz), Happiness (194-138 Hz). Mention values clearly distinguish emotions. Our intention of calculating MFCC is emotion classification.

Table 3. Evaluation of MFCC-1 values for Gujarati speech database

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9
Sadness	54.1465	55.6345	51.5519	50.2767	51.8346	55.1357	49.0745	62.8199	56.0789
Surprise	55.5314	59.1399	54.6949	50.9556	53.6297	53.5329	51.4229	62.6117	56.0903
Anger	59.9611	53.9451	53.3597	55.3726	56.6270	49.6293	53.9430	58.2213	57.5959
Disgust	56.2181	52.7455	55.1911	50.3790	51.7035	51.4474	50.6729	59.8750	58.8367
Fear	57.7405	54.6611	51.5343	53.1228	55.7246	49.1865	51.1392	53.0337	58.3565
Happiness	60.1937	54.6131	59.1466	54.3880	54.5966	49.1055	51.6471	52.7353	57.0388

Table 4. Evaluation of MFCC-2 for Gujarati speech database

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9
Sadness	-2.2238	0.9706	-2.9330	-1.8598	-0.5796	-1.6080	-0.5071	-3.9828	-4.8528
Surprise	0.3282	0.9824	-0.9337	-0.3732	-0.9460	-3.2635	-1.0888	-2.1070	-4.2324
Anger	-0.3355	-2.2884	-2.3889	-1.4443	-1.0018	-5.7981	-0.5913	-3.9590	-4.4121
Disgust	0.2999	-1.0024	-2.1522	-1.1472	-0.9003	-2.1104	-0.9030	-2.6625	-5.9572
Fear	1.7066	-3.8935	-3.0708	-2.2713	-0.6393	-3.4203	-2.9169	-4.3413	-5.0345
Happiness	0.4311	0.0741	-2.6731	-2.2302	-1.2104	-4.4427	-2.6107	-4.1124	-5.8747

Table 5. Evaluation of MFCC-3 for Gujarati speech database

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9
Sadness	0.1933	1.0820	-0.5656	-0.2734	-0.1495	0.3566	1.6903	-0.4313	0.6172
Surprise	-0.5114	0.9649	0.0537	-1.4930	0.3122	2.8229	1.6847	-2.2628	0.0849
Anger	-2.2406	2.0109	-0.5874	1.7957	-0.5999	1.1967	2.3585	1.4775	1.6840
Disgust	0.1643	0.6015	0.4887	0.9046	0.4293	3.2405	2.9430	-1.9491	3.0991
Fear	-0.1440	2.8457	-1.1328	2.5336	-0.7280	2.1838	0.9920	-0.6339	1.8572
Happiness	2.6261	0.9625	2.0231	0.0878	2.4471	2.4248	2.5930	1.5067	2.5317

Table 6. Evaluation of pitch values for Gujarati speech database

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9
Sadness	210.0982	214.0308	184.1606	165.6608	173.5516	157.7194	118.7337	135.0328	210.1995
Surprise	183.9260	202.9725	208.1153	199.6098	187.5433	170.0165	152.1450	163.7575	208.4589
Anger	168.5054	204.2806	189.3183	207.5600	204.7712	199.5164	165.5100	193.2691	211.7395
Disgust	188.1440	191.2666	174.0789	190.5393	140.9156	172.7517	129.8313	160.4026	194.7066
Fear	176.5592	168.3506	165.2658	181.8807	192.2072	179.9400	176.5990	139.3690	218.7620
Happiness	167.4267	191.5624	194.9376	191.9864	176.6201	160.8107	138.8605	150.5456	183.5233

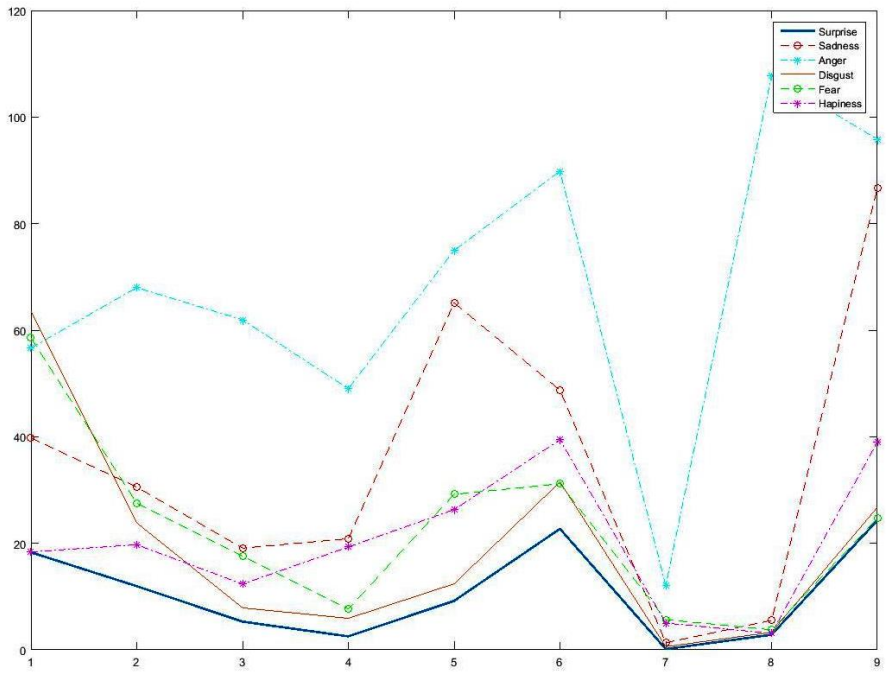

Figure 4. Average energy values for Gujarati speech database for emotions

6. CONCLUSION

In this paper we have contemplated an emotional speech database or speech corpus in Gujarati language. The six basic emotions deliberated for developing database are sadness, surprise, anger, disgust, fear and happiness. Evaluation of speech database is carried by mainly parameters as energy, MFCC 1 to 13, pitch. Results clearly have shown the difference in different emotions. But still database can be further improved and variability in speakers and spoken words makes it most effective. The proposed database is an intermixture of characteristics in terms of different emotions, speakers and words.

Linear model/classifier and Non linear models/classifiers can be explored to further improve the recognition performance. The importance of speech emotion over image is person can change facial expressions easily but hard to change speech. In future, multimodal detection systems can be built up which used image, speech signals and bio-signals all to gather for classification of emotion states of human.

REFERENCES

[1] R. Chakraborty, M. Pandharipande, and S. K. Kopparapu, "Knowledge based framework for intelligent emotion recognition in spontaneous speech," *Procedia Computer Science*, vol 96, pp. 587-596, 2016.

[2] K. S. Rao, Shashidhar G. and Koolagudi, "Emotion Recognition using Speech Features," *Springer*, 2013.

[3] G. Gaurav, *et al*., "Development of application specific continuous speech recognition system in Hindi," *Journal of Signal and Information Processing*, vol. 3, no. 3, 2012.

[4] F. H. Rachman, R. Sarno, C. Fatichah, "Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 3, pp. 1720-1730, 2018.

[5] D. Pravena and D. Govind, "Development of simulated emotion speech database for excitation source analysis," *International Journal of Speech Technology*, vol. 20, pp. 327–338, 2017.

[6] M. Musfafa, M. Yusuf, and M. Malekzadeh, "Speech emotion recognition research: an analysis of research focus," *International Journal of speech Technology (IJST)*, vol. 21, pp. 137-156, 2018.

[7] F. Shah," Study and analysis of speech emotion recognition" *Ph.D. Thesis*, Sodhganga, 2016. [Online]. Available: http://shodhganga.inflibnet.ac.in/handle/10603/122185.

[8] V. Tank and S. K. Hadia, "Development of Emotion Recognition From A Speech In Various Regional Indian Languages: A Review," *IJRECE*, vol. 6, no. 2, pp. 2155-2161, 2018.

[9] R. Kumar, *et al*., "Development of Indian language speech databases for large vocabulary speech recognition systems," in *International Conference on Speech and Computer (SPECOM) Proceedings*, 2005.

[10] M. Swain, A. Routray, P. Kabisatpathy, "Databases, features and classifiers for speech emotion recognition: a review," *International Journal of Speech Technology*, vol. 21, pp. 93-120, 2018.

[11] V. P. Gowda, M. Murugavelu, S. K. Thangamuthu, "Continuous kannada speech segmentation and speech recognition based on threshold using MFCC and VQ," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 4684-4695, 2019.

[12] A. P. Kumar, *et al*., "Continuous Telugu Speech Recognition through Combined Feature Extraction by MFCC and DWPD Using HMM based DNN Techniques," *International Journal of Pure and Applied Mathematics*, vol 114, no. 11, pp. 187–197, 2017.

[13] S. G. Koolagudi and K. Rao, "Emotion recognition from speech: a review," *International journal of speech Technology*, vol. 15, pp. 99-117, 2012.

[14] B. Basharirad and M. Moradhaseli, "Speech emotion recognition methods: A review," in *AIP Conference Proceddings*, pp. 1-7, 2017.

[15] S. K. Koolagudi, *et al*., "Recognition of emotions from speech using excitation source features," in *Elsevier proceedings for international conference on modeling, optimization and computing*, pp. 3409-3417, 2012.

[16] S. Ramakrishnan, "Recognition of Emotion from Speech: A Review," in *Speech Enhancement, Modeling and Recognition – Algorithms and Applications*, InTech, pp. 121-138, 2012.

[17] S. S. Poorna, K. Anuraj., and G. J. Nair, "A Weight Based Approach for Emotion Recognition from Speech: An Analysis Using South Indian Languages. *Communications in Computer and Information Science Soft Computing System*, vol. 837, 2018

[18] "Hindi is the most spoken Indian language in USA, followed by Gujarati and Telugu," *The Sentinel of this land, for its people*, 2018. [Online]. Available: www.sentinelassam.com/news/hindi-is-the-most-spoken-indian-language-in-usa-followed-by-gujarati-and-telegu/

[19] "List of languages by number of native speakers," *Wikipedia, the free encyclopedia*. [Online]. Available: en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers.

[20] S. S. Agrawal, *et al*., "Emotions in Hindi Speech-Analysis, Perception and Recognition," in *2011 International Conference on Speech Database and Assessments (Oriental COCOSDA*, pp. 7-13, 2011.

[21] S. G. Koolagudi, *et al*., "IITKGP-SESC: Speech Database for Emotion Analysis," *Communications in Computer and Information Science*, pp. 485-492, 2009.

[22] S. G. Koolagudi, *et al*., "IITKGP-SEHSC: Hindi Speech Corpus for Emotion Analysis," in *IEEE proceedings on International Conference on Devices and Communications (ICDeCom)*, pp. 1-5, 2011.

[23] A. Jain, N. Prakash, S. S. Agrawal, "Evaluation of MFCC for Emotion Identification in Hindi Speech," *2011 IEEE 3rd International Conference on Communication Software and Networks*, pp. 189-193, 2011.

[24] E. S. Gopi, "Digital speech processing using MATLAB," *Springer*, 2013.

[25] S. S. Poorna, *et al*., "Emotion recognition using multi-parameter speech feature classification," in *IEEE International Conference on Computers, Communications, and Systems*, India, 2015.

BIOGRAPHIES OF AUTHORS

Mr. V P Tank pursed Bachelor of Engineering from Dharmsinh Desai University, Nadiad in 2009 and Master of Engineering from Gujarat Technological University in year 2011. He is currently pursuing Ph.D. and working as Assistant Professor in V T Patel Department of Electronics & Communication Engineering, Charotar University of Science & Technology, since 2012. He is a life time member of IETE since 2014. He has published more than 10 research papers in reputed international journals and conferences and it's also available online. His main research work focuses on Digital Speech processing, Bioelectronics, Digital signal processing. He has 8 years of teaching experience and 1 years of Research Experience.

Dr. S K Hadia is an associate professor at Gujarat Technological University (GTU), Ahemdabad, Gujarat, India. He has completed his Ph.D. from Charotar University of Science and Technology in year 2016. His main research work focuses on Optical communication computer network, Image processing. He has more than 15 years of teaching experience. He has published more than 11 research papers in reputed high impact international journals and conferences.