Speaker Dependent Emotion Recognition System

Priyadarshini N S[#1],Priyanka B M[#2],Priyanka R[#3],Pushpa K[#4],Dr. Jayanth J[*1]

#Under Graduate student, ECE Dept, GSSSETW, Mysuru, India
*Professor in ECE Dept, GSSSETW, Mysuru,570016

Abstract-

Emotion is an integrated feature that creates a void between humans and humanoids. In order to fulfill this void, emotion r ecognition plays an important role. Though there are many ot her methods to recognize emotions, we have chosen speech as a basis for extraction of emotions as it is less effected from env ironmental constrains such as magnetic field, light and other f actor. Emotion recognition has already been implemented in many languages except for Kannada. In this paper, we have cr eated a system where in Kannada speech is an input and emot ion is the output. Praat is used to extract features from the spe ech signal which is given by the speaker as input. A GUI in M ATLAB has been created to interface human speech with the system. The neural network takes the features extracted from praat software and test the data to the trained feed forward n eural network and recognizes the basic emotion of humans suc h as sad, happy, angry .

Keywords-
Emotion recognition, MATLAB, Praat, Speech database neural network training and validation, confusion matrix.

I. ARCHITECTURE

Humanicommunicationidevelopediwithitheioriginiofispeechi whichiapproximatelyidatesibackitoi500,000iBCE.iFromithisispee chihumansicouldiexpressitheiriemotions[4].iTheseiemotionsiofih appy,isad,iangryietciplayianiimportantiroleiiniunderstandingione' sisituationioridesireiiniaibetteriway.Thusiemotionicanibeidefined iasiaistrongifeelingiofione'sicircumstance.iBeforeitheiinventioni ofiemotionirecognitionisystem,itheimachinesicouldinotiefficientl yiidentifyiperson'siemotioniandirespond.iByiintroducingithisisys tem,itheimachinesiareiableitoirespondiefficientlyitoitheiuser'sine edsiandiincreaseitheiutilityiofitheimachinesicorrespondingitoione 'sineeds.iInithisisystemiweihaveicreatediGUIitoirecordispeechifr omitheispeaker.iFromitheirecordedispeechitheifeaturesiareiextrac ted.iTheseiextractedifeaturesiareisentiasitestidataito

recognizeitheibasiciemotionsiofihumansisuchiasihappy,isad,iangr y[3].iThisirecognizediemotioniisidisplayed ioniscreen.
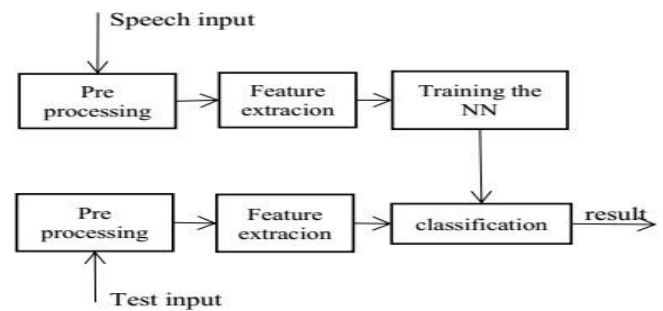
II. METHODOLOGY

Fig.
1. Block diagram

Speechiinput   i:iEmotioniisiidentifiedibased ionitheispeechigivenibyitheispeaker,itheisystemitakesiKannadaispee chiasitheiinputifromitheispeaker.

Pre-processing
i:iTheifeatureiextractioniprocessirequiresispeechisignaliwithoutitheie xternalinoise.iInorderitoiremoveithisinoiseisystemiusesispectralisubt ractioniandiremovesiexternalinon-periodicinoises.

Featureiextraction
i:iInithisiprocess,ifeaturesiareiextractedifromitheioriginalidataset,ibyi decreasingitheiamountiofivariables.iAsitheinumberiofifeaturesiincre ase,itheiaccuracyiincreasesiaccordingly.

Pitch
i:iItirepresentsitheivariationiofiaitoneigivingiprosodiciinformationiof ianiutterance.

Intensity
i:iPowericarriedibyiaisoundiwaveiperiunitiareaiinitheidirectioniperpe ndicularitoithatiarea.

Jitter
i:iDeviationifromitrueiperiodicityiofiaipresumablyiperiodicisignal.

a) *Shimmer*
: It relates to amplitude variation of the sound wave.

1) *Classification*
: The classifier differentiates the emotions of the speech. There are many classifiers such as HMM, GMM, ANN, SVM etc. In this project ANN classifier including feed forward neural network is used.

ANN is popular choice as it can faciliate non linear relationship between features and classes. It has 3 layers (input, output, hidden). In ANN, generally 90% data is used for training and 10% is used for validation. ANN is a highly adaptable learning machine.

2) Output
: The desired emotion to be obtained is represented in the form of emojis hence indicating the recognition of particular emotion[4].

III. DATABASE

Kannada is one of the Southern Dravidian language, and its history divided into three periods: Halegannada from   450–1200 CE, Nadugannada from   1200–1700, and Modern Kannada from 1700 to the present. Kannada is influenced by Sanskrit. Influences of Prakrit and Pali, can also seen in Kannada language. We used read type as our speech corpora. For analyzing the emotion we considered 100 Kannada sentence. The total number of feature in dataset of 600 (100 sentence *3 emotion *2artist) were 9. The proposed emotion in speech corpus are Angry, Sad, Happy.

ThePraatsoftwareisusedtorecordtheKannadasentences. The recording factors considered here are mono channel and sampling frequency of 44.1 kHz. The audio file recorded is saved in WAV file format for further feature extraction to be simple. Multiple sentences in Kannada recited by the speaker are collected through which the features such as pitch parameter (mean pitch, SD pitch, min pitch, max pitch), duration, jitter, shimmer are extracted and tabulated to create the database.

Pitch - degree of highness or lowness of tone. Intensity - degree of loudness.

Jitter - deviation from the tune periodicity.

Shimmer - periodic variation between amplitude peaks.

These features were chosen, as they were giving 90 percent of accuracy. Where as considering other features. By adding other features to these features just increase the accuracy of the emotion recognition system by just 10 percent. Thus limiting our work to these features yields higher efficiency with less number of features and limiting the database. Random audio clips were played and each clips were verified with corresponding emotion with the domain experts.

The data sample from the speaker were artified and was recorded using praat software in noise less environment.

The data collected from speakers here is audio samples. The sentences taken for training the system are roughly 100, however few sentence are given in the TABLE I[7].

TABLE I   English transcription of the Kannada text.

Sent.id	Sentence
1	Shale bahala dooradalede. (School is very far)
2	Nanu oorige hoguthene. (I'm going to town)
3	Swalpa mellage mathadi. (talk in low voice)
4	Nanage sahaya madi. (Please help me)
5	Navu adanu nodidevu. (We saw it)
6	Shalege makkalu baralilla. (Childre n did not come to school)
7	Ninna hesaru yenu. (What is your name)

IV. TRAINING AND VALIDATION

Once the database is created, it is used for training and validation using the MATLAB software, where the MATLAB   has   inbuilt tool for training   neural network namely nntool(neural network toolbox) and for verifying validation and testing namely nftool(neural fitting toolbox). The feed forward back propagation is the network type used for training the system where former update the data in forward direction and back propagation helps to reduce the error by working has feedback system. The TRAINLM is used as training function that change the weight and bias basedon   levenberg-marquardt   optimization   further TRAINLM is the fastest among all the algorithm in MATLAB toolbox.

There are 3 types of layers as usual in feed forward neural network namely input layer, hidden layer and output layer. The input layer consist of   9 neuron ,output consist of 3 neuron single hidden layer is considered with 6 neuron and the system is trained.

70% of the dataset is used for training

15% of dataset is used for validation and

15% of remaining data is used for testing[4].

V. EXECUTION

Graphical User Interface (GUI) created as user-friendly method using MATLAB software. The GUI includes 5 push button and one axes. For selecting the system whether to identify the emotion of male or female two button were included, further after selecting it has the push button for record ,stop and play. The record and stop button perform operation as the name on the button. Play button play the denoised audio of the recorded audio.
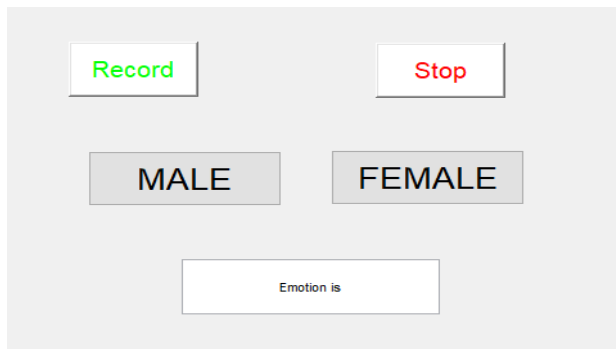
Fig. 2. GUI for emotion detection.

GUI takes the real-

time input from the speaker and is saved in a temporary file for further analysis. Praat software is invoked in GUI through MATLAB to extract the required features from the real-time input. These extracted features are sent as a test data to the already trained neural network. The neural network analyses and compares from already trained dataset and recognizes the emotion. The result is then sent to GUI for display[8].

VI. SOFTWARES

Two softwares were used throughout the project they are MATLAB and PRAAT.
 A. MATLAB: High performance for technical computing. It has data structures, built-

in editing and debugging tools and supports object oriented programming. It generates displays or outputs when commands are executed. It combines calculation and graphic plotting. It is designed for scientific computing.
B. PRAAT
 : It is a computer program with which you can analyse, synthesize, and manipulate speech, and create high-

 quality pictures for your articles and thesis. It is a freeware program for the analysis and reconstruction of acoustic speech.

VII. RESULT

The above project was verified using two methods one with the inbuilt toolbox in MATLAB and the other method by giving dataset manually.

Fig. 3. Confusion matrix for male emotion detection.

The above figure shows that the system detecting male emotion in our experiment is 95.2% accurate.
Similar results were obtained in female speech, emotion
Here in MATLAB the nprtool is used to obtain confusion matrix
The confusion matrix of male and female speech emotion recognition for training, validation, testing Using nprtool is as shown in figures 3 and 4.
The confusion matrix obtained during manual testing method for male emotion detection.

Fig. 4. Confusion matrix for female emotion detection.

TABLE II   system accuracy for male emotion recognition.

	Sad	*Angry*	*Happy*	*overall*
Sad	100	0	0	100
Angry	0	100	0	100
Happy	5	15	80	80
overall				93

The Table II shows the emotion recognition of the male to detect exact emotion in real time by giving random sentence of each emotion, the result obtained were written in percentage. When the happy sentence were given to the system the error was more compared to that of angry and sad sentence. The fact for the reduced accuracy in happy sentence is that in angry sentence pitch will be high and for the sad sentence pitch will be low but in the happy sentence the pitch value will be in between the other two emotion. However, the main feature we considered is pitch and its parameter the accuracy will be dependent on pitch.
The confusion matrix obtained during manual testing for female emotion detection.

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCCDS - 2020 Conference Proceedings

TABLE III   **S**ystem accuracy for female emotion recognition

	Sad	Angry	Happy	overall
Sad	100	0	0	100
Angry	0	100	0	100
Happy	10	20	70	70
overall				90

The Table III shows the emotion recognition of the female at the real time by giving 10 sentence of each emotion, the result obtained were written in percentage. The result obtained during manual testing for male emotion recognition is around 93 percent and for female emotion recognition it was around 90 percent.

The GUI result obtained after angry male sentence is given as input, is as shown in the below figure 5 similarly GUI will display the emotion of different type .

Fig. 5 GUI after the recognition of emotion