International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 32585-32600
© Research India Publications
http://www.ripublication.com

Analysis Towards Optimum Features And Classifiers To Recognize The Emotion Happiness From Speech Signals

A. Milton[a], E.S. Nivea Ghosh[a*], Shilpa S Mohan[a], S. Tamil Selvi[b]

[a]Department of Electronics and Communication Engineering, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil 629 003, Tamil Nadu, India.
E-mail: milton@sxcce.edu.in, nsg2291@gmail.com, smohan.shilpa@gmail.com.
Phone: 9442602309, 9446284771, 9847263069.
[b]National Engineering College, Kovilpatti 628 503, Tamil Nadu, India.
E-mail: stsece@nec.edu.in, Phone: 9443721864.
[*]Corresponding Author

Abstract

In Human computer interaction, speech emotion recognition is a challenging issue. Speech emotion recognition is achieved through several methods, which includes isolation of speech signals and extraction of selected features for classification. The accuracy of an emotion recognition system depends on different factors such as features, functionals used to convert the frame level features to global features and the type of classifier. In speech emotion recognition researches, the emotion happiness is found to be the emotion which is difficult to recognize. In this work, we move towards finding the features and classifiers from a set of features and a set of classifiers that optimize the recognition of the emotion happiness. We evaluate the emotion classification framework on two different emotional databases such as Berlin emotional database and Enterface emotional database and a rich set of features that include pitch, intensity, harmonicity, formant, Mel-frequency cepstral coefficient, power cepstrogram, cochleagram and linear prediction coefficient. The classification techniques used in this paper are; K-nearest neighbour, Naive Bayes and binary decision tree. These three classifiers and the eight features are evaluated to find the optimum features and classifier. In order to search for the best features and classifiers, we have also tested fusion of different combination of features. The experimental results are validated under speaker-independent train-test methodology.

Keywords: Emotion Recognition from Speech Signals, Human Computer Interaction, Happiness Recognition, Fusion of Features.

1 Introduction

Speech emotion recognition (SER) is the identification of the emotional state of a speaker from speech features. SER is gaining momentum in recent years because of its application in artificial intelligence. Still, there is a lack of naturalness in identifying emotion. An exponential growth in available power and significant progress in speech technologies will have a successful application of SER in several domains [1]. An extensive application field of speech emotion recognition are navigation, air travel information system, educational, tutoring, medical emergencies, helping man machine interface for weak and old people [2,3], in call centres to detect satisfied (happy) or unsatisfied (angry) customers and in entertainment electronics to gather emotional user feedbacks [4]. Most of the existing efforts in the field have focused on the recognition of basic emotions from humans such as happiness, sadness, anger, fear, boredom, disgust and neutral [4,5]. One observation in SER researches is particularly difficult to recognize the emotion happiness. A very few number of papers have made attention about this, so in this paper, we focus on improving happiness emotion recognition.

In SER the first step is the design of emotional speech database. Here, we use Berlin [6] and Enterface [7] emotional databases. Second step is to extract speech features with potential value for emotion recognition [5]. Features such as pitch, formant [8,9], linear prediction coefficient (LPC), Mel-frqeuency cepstral coefficient (MFCC) [9,10], intensity, harmonicity [11,1], power cepstrogram and cochleagram are used in this paper. The hunt for optimum feature and classifier is not yet complete in speech emotion recognition due to different research conditions and backgrounds [12]. Third step is to classify emotions by training and testing classifiers with the features. Widely used classifiers are Gaussian mixture model, hidden Markov model, artificial neural network, Support vector machine, binary decision tree, k-nearest neighbour and naive Bayes classifiers [8].

The rest of the paper is organized as follows: Section 2 describes a brief review of the literature in the area of speech emotion recognition. Section 3 presents the extracted features followed by different classifiers used in this paper and the emotion classification analysis methodology. Section 4 explains the results and discuss about the classification performance. Section 6 concludes the paper.

2 Related Works

The importance of the emotion recognition has been proved by an increasing number of related works that can be found in the literature. The basic requirement to implement an emotion recognition system is a valid emotional database. There are different databases created by speech processing community with the help of professional actors which is widely used in research work. The efficiency of the speech emotion recognition system is highly depending upon the naturalness of database used in the system [13]. Databases used by the researchers come under two major categories: the acted database and the natural database. The acted databases reported in literature are recorded by professional or semi-professional actors. The natural database is obtained by recording conversations in real-life situations such as call centres and talk shows [13,14].

When a database is ready, then we can proceed towards extracting useful features from the speech signals. Several commonly used features of speech signal found in the literature are energy, pitch, intensity [12], linear prediction coefficient [14], Mel-frequency cepstrum coefficient, formant [15], harmonicity [1]. Different software available for extracting features and classification are PRAAT [16,17], Open Smile toolkit [18], FEELTRACE [11], SNACK [8] and WEKA [19].

The extracted features are used to classify the emotions with the help of classifiers. K-nearest neighbour [14,8,10,12], artificial neural network [9,20], linear discriminant classifier [21], hidden Markov model [13], Gaussian mixture model [10], support vector machine [13], Bayesian classifier [4], decision tree [23] are the classifiers which are listed in the literature to classify emotions. In all the speech emotion researches, the main focus was improving overall recognition rate and the emotion happiness is found to be the emotion which is difficult to identify by machine. In this paper, we take the first step towards optimizing the recognition of the emotion happiness.

The advantages of this study are i) trying to address the problem of low machine recognition of the emotion happiness, ii) speaker-independent classification and iii) use of a rich set of features including pitch, intensity, harmonicity, formant, MFCC, LPC, power cepstrogram and cochleagram. We utilize K-nearest neighbour (KNN), Naive Bayes (NB) and binary decision tree (BDT) classifiers.

3 Methodology

The proposed analysis is a speaker-independent emotion classification system. In speaker-independent classification, utterances uttered by few speakers are not included in the training patterns. They are unknown for the classifier during the training phase. This means that, we need one set of training patterns to train the classifier and one set of test patterns from unknown speakers to test the generalization capability of the classifier. The block diagram of overall process involved in speech emotion recognition system is shown in Figure1. It shows the procedure of how the input speech signal is processed to recognize the emotional state of the speaker.

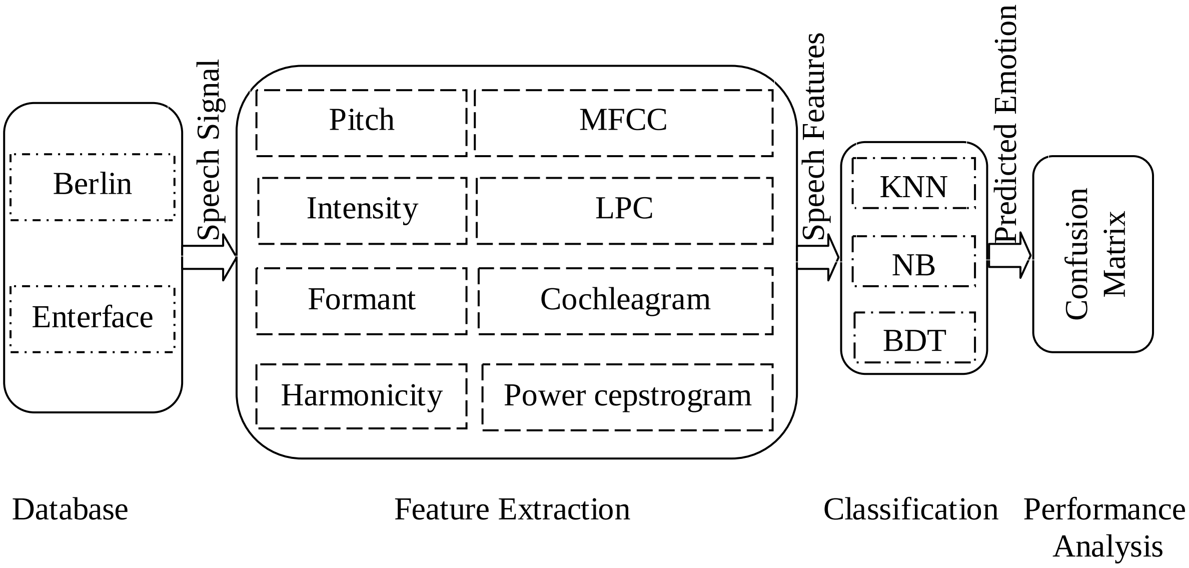

Fig.1 Various stages of the proposed analysis to recognize speech emotions

3.1 Databases

Natural speech databases for emotion recognition are seldom publicly available due to privacy of speakers. In addition, acquisition and labelling of a large size database is very expensive. For these causes, we use two database of acted speech. First one is Berlin emotional database [6]. Ten actors (5 female and 5 male), in the age group of 21-35 years, have uttered 5 short and 5 long sentences. There are 10 texts and 10 speakers in the formation of database. The texts could be used in everyday communication and are interpretable in all applied emotions. The recordings were taken in an anechoic chamber with high-quality recording equipment. In addition to the sound, electro-glottograms were recorded. There are 535 wave files with sampling rate of 16 kHz and 16 bits encoding. The utterances with good voice quality and naturalness have been retained. In case of classification with Berlin database, for speaker-independent classification, utterances uttered by 6 out of 10 speakers are used for training and utterances uttered by the remaining 4 speakers are used for testing purpose.

Second type of acted database is Enterface audio-visual emotional database [7]. The Enterface database containing the six archetypal emotions, 46 subjects were invited to react to six different situations, each of them eliciting one of the following emotions: happiness, sadness, surprise, anger, disgust and fear. The database was recorded by immerging the subject into a situation that evokes a specific emotion, and then let the subject react to the situation in his own language. The database was recorded using a standard mini-DV digital video camera. The recording of the speech signal was realized through the use of a high-quality microphone, specially conceived for speech recordings. The audio sample rate was 48000 Hz, in an uncompressed stereo 16-bit format. Here we discard three subjects whose emotions are not clearly recognized. There are 1287 video samples and 43 subjects used in the experiments.

We have extracted the audio signal in wave format from the audio-video files. In the case of classification with Enterface database, for speaker-independent classification, utterances uttered by 27 out of 43 speakers are used for training purpose and utterances uttered by the remaining 16 speakers are used for testing.

3.2 Features Extraction

Feature extraction module extracts the speech features that provide essential information about the raw speech data. These features represent parameters of different emotional content of the speech samples. Different researchers have used different set of features for their system. We extract selected eight features such as pitch, intensity, formant, harmonicity, MFCC, LPC, cochleagram and power cepstrogram and these are estimated over five functionals such as mean, median, standard deviation, inter-quartile range and mean absolute deviation. We utilize PRAAT software [24] to extract all the features except MFCC and LPC. These two features are extracted using Matlab R2012b.

The speech feature pitch is the vibration frequency of vocal folds. In the time domain, pitch is the reciprocal of the minimum period in voice sound. High pitch is associated with various expressions like happy, graceful, dreamy and exciting. Low pitch is related to dignity, sadness, and boredom [11]. Intensity, the average energy of the speech sequence is widely adopted to represent the loudness of sound and its standard deviation can be used to represent the regularity of the loudness. Formant is the acoustic resonance of the human vocal tract. It represents the spectral maximum. In speech signal, there will be more than one formant. Formants are often measured as amplitude peaks in the frequency spectrum of the speech signal. We have extracted first five formants and the corresponding bandwidths. The speech feature harmonicity represents the degree of acoustic periodicity. It measure the harmonics-to-noise ration i.e. energy distribution between periodic and aperiodic signals.

MFCCs are extracted based on the theory of human auditory perception of sound. The MFCCs representation is also known as a voice quality features. It is the most widely used successful spectral features in automatic speech emotion recognition; it could be because the emotional content of an utterance is strongly related to its voice quality and frequency content. MFCCs are the discrete cosine transform of logarithm of power spectrum in Mel-frequency scale. Here, we have extracted MFCCs with coefficients equal to 24. LPCs represent the spectral envelope of digital signal in speech processing. LPCs are derived by considering a speech sample as the weighted sum of P past speech samples. The weighting coefficients are called LPCs and P is the order of liner prediction. Here, the LPCs are extracted using autocorrelation method with a prediction order of 50.

Cochleagram is a powerful tool to investigate both the temporal and the frequency characteristics of speech signal. Cochleagram, which is the map of auditory nerve firing rate, is the time-frequency representation of sound perceived from cochlea. It is based on the function of basilar membrane vibration and hair cell firing. Cochleagram consists of time and frequency axes; time axis is related to the speech windowing and shifting, and frequency axis is the frequencies that are naturally selected by the cochlea and hair cells of the inner ear. Cochleagram is obtained by filtering the speech signal by an auditory filter bank which represents the activity of basilar membrane followed by a nonlinear rectification which represents the neural firing. We extract cochleagram with 256 central frequencies per frame and the frequency axis is in Bark scale.

Power cepstrogram is the power cepstrum as a function of time. Power cepstrum is the inverse Fourier transform of the logarithm of squared magnitude of Fourier transform of the signal. The two axes of power cepstrogram are amplitude in dB and frequency in seconds. We extract power cepstrogram with 512 point Fourier transform.

3.3 Classification

The features extracted from the feature extraction module are given as input to the classifier. The speech samples considered in this paper belongs to seven emotional classes of Berlin database and six emotional classes of Enterface databases. The classifiers are trained using training set and the classification accuracy is estimated using the unseen test set. Speaker-independent training and test set divisions are done as explained in Section 3.1. The three classifiers used in this paper are explained in the following sub-sections.

3.3.1 K-Nearest Neighbour Classifier

KNN is a non-parametric method for classifying speech samples based on closest training samples in the feature space. KNN classification uses the whole training features in the same form rather than using parameters derived from the features like in NB and BDT classifiers. The similarity of a test feature with all the training features is compared using a distance metric and K nearest training neighbours are identified. The class assigned to the test feature is the class of the maximum number of nearest training features. KNN is a type of instance based learning which is the simplest of all machine learning algorithms. We use KNN with K=6 i.e. 6-nearest neighbours are used to estimate the class of a test feature and the distance metric used to select the nearest neighbours is Euclidean distance.

3.3.2 Binary Decision Tree Classifier

BDT is a classifier in the form of a tree structure. In the tree, decision nodes specify a test on a dimension of features, leaf nodes indicate the class label of the features and arc or edge indicates split of one feature. Path in tree is a disjunction of test to make the final decision. Decision trees classify instances by starting at the root of the tree and moving through it until a leaf node is reached. Decision trees which represent rules are potent and popular algorithms for classification and prediction. Classification by decision tree consists of two phases; tree generation (training) and classification (testing). Decision tree generation consists of two stages that are tree construction and tree pruning. Tree construction is a process in which the training features are at the root and partition features recursively based on selected attributes. Tree pruning is a process of identifying and removing branches that are due to noise or outliers. In classification phase, the generated decision tree is used to predict the class of an unknown test features. Here, we use decision tree with binary splits of each branching node.

3.3.3 Naive Bayes Classifier

NB classification method is based on Bayes rule. The assumption used in NB classifier is that the value of any feature is independent of other features. During training phase, NB classifier estimates the parameters of a probability distribution of instances of different classes. During testing phase, for an unseen test instance, it estimates the posterior probabilities of the test instance belonging to each of the classes. A class is assigned to the test instance based on the largest posterior probability. With appropriate pre-processing, NB classifier can compete with more advanced methods like support vector machines. NB classifier is very fast and requires very low storage space. If the assumed independence is correct, then NB classifier is an optimal classifier for classification problems.

3.4 Classification Analysis

Emotion classification analyses are done separately with Berlin and Enterface databases. Speaker-independent train-test partition of each database is done as explained in Section 3.1. For each database, each functional of each feature, combination of all functionals of each feature, combination of all functionals of different features are separately trained and tested with each of the three classifiers to select the best features and classifiers for the recognition of the emotion happiness as well as the overall emotions recognition rate.

4. Results and Discussion

This section focuses on evaluating the performance of features of the two databases and classifiers in recognizing emotions, in particular the emotion happiness.

4.1 Performance with Berlin Database

The classification results obtained with different features and different classifiers for Berlin database are shown in Table 1. For each feature and classifier combination, recognition rate of the emotion happiness and accuracy are tabulated. With mean features, the emotion, happiness is recognized with a maximum recognition rate of 59.26% by cochleagram features with KNN classifier, all the seven emotions are classified with a maximum accuracy of 56.7% by MFCC features with KNN classifier. Comparison of the median features show that LPC features with BDT classifier recognize the emotion happiness with maximum recognition rate of 51.85% and maximum overall accuracy of 54.93% is obtained by MFCC features with KNN classifier. In case of standard deviation features, power cepstrogram features with NB classifier provide the maximum happiness recognition rate and overall accuracy of 29.63% and 42.78% respectively. For inter-quartile range features, harmonicity feature provides the maximum happiness recognition rate of 59.26% with NB classifier and power cepstrogram features with NB classifier provide a maximum overall accuracy of 48.97%. In case of mean absolute deviation features, power cepstrogram features obtain maximum happiness recognition rate of 40.74% and overall accuracy of 48.45% with BDT and NB classifiers respectively.

Table 1 Recognition rates of happiness and all emotions obtained by different features and different classifiers for Berlin database

Functionals	Classifier	Emotion	Recognition rate in %							
			Pitch	Intensity	Formant	Harmonicity	MFCC	LPC	Cochleagram	Power cepstrogram
Mean	KNN	Happiness	25.93	11.11	25.93	11.11	48.15	11.11	59.26	7.41
		Overall	27.32	21.13	47.42	30.93	56.70	31.96	50.52	22.68
	BDT	Happiness	7.41	14.81	3.70	22.22	7.40	25.93	25.93	8
		Overall	19.07	13.40	28.87	15.98	31.96	28.35	27.32	31.96
	NB	Happiness	11.11	0	25.93	0	29.63	11.11	0	7.41
		Overall	34.54	34.02	44.85	34.54	38.14	35.57	30.93	47.94
Median	KNN	Happiness	14.81	22.22	18.52	33.33	25.93	25.93	48.15	14.81
		Overall	20.10	24.74	37.63	18.04	54.93	30.41	42.78	18.55
	BDT	Happiness	18.52	14.81	37.04	29.63	29.63	51.85	33.33	25.93
		Overall	18.04	16.49	35.57	14.43	40.72	32.47	29.91	25.26
	NB	Happiness	0	0	FT	0	25.93	7.40	14.81	11.11
		Overall	36.60	33.51	FT	31.96	41.24	32.99	33.51	48.45
Standard deviation	KNN	Happiness	18.51	7.41	18.52	3.70	25.93	18.52	25.93	14.81
		Overall	25.26	18.56	20.62	18.56	34.54	32.99	40.72	24.23
	BDT	Happiness	25.93	7.41	7.41	25.93	7.41	11.11	22.22	22.22
		Overall	17.01	9.28	22.16	15.46	23.71	25.26	25.77	25.25
	NB	Happiness	0	0	14.81	0	11.11	7.41	11.11	29.63
		Overall	28.87	22.68	41.75	30.41	33.51	29.89	34.02	42.78
Inter-quartile range	KNN	Happiness	11.11	37.03	37.04	14.81	18.52	11.11	33.33	22.22
		Overall	20.10	25.25	40.21	24.74	33.51	28.87	41.24	31.44
	BDT	Happiness	14.81	11.11	22.22	11.11	37.04	7.41	11.11	25.93
		Overall	16.49	13.40	35.05	14.43	29.38	22.16	28.87	23.71
	NB	Happiness	14.81	0	33.33	59.26	11.11	3.70	14.81	11.11
		Overall	28.35	33.51	42.78	22.16	31.96	37.63	39.69	48.97
Mean absolute deviation	KNN	Happiness	22.22	22.22	11.11	22.22	11.11	0	37.04	33.33
		Overall	24.74	24.23	29.38	18.04	30.93	26.28	38.66	31.96
	BDT	Happiness	3.70	22.22	7.41	7.40	18.52	25.93	29.63	40.74
		Overall	14.94	18.56	30.41	11.86	24.23	25.77	34.54	34.54
	NB	Happiness	0	0	25.93	0	11.11	14.81	14.81	7.41
		Overall	30.93	29.38	44.33	31.44	31.96	30.41	33.51	48.45
Combination of all functionals	KNN	Happiness	22.22	14.81	37.04	18.52	44.44	14.81	66.67	18.52
		Overall	42.27	29.91	37.63	25.77	56.19	35.56	46.91	33.51
	BDT	Happiness	25.93	7.40	22.22	11.11	37.04	25.93	18.52	22.22
		Overall	24.74	23.71	25.26	19.07	36.61	27.32	28.87	34.02
	NB	Happiness	14.81	0	FT	40.74	25.93	11.11	11.11	25.93
		Overall	32.99	34.54	FT	26.29	36.61	43.81	39.18	52.58

KNN-K nearest neighbour, BDT-Binary decision tree, NB-Naive Bayes, MFCC-Mel-frequency cepstral coefficient, LPC-Linear prediction coefficient and FT-Failed to train.

Classification with combination of all the functionals (mean, median, standard deviation, inter-quartile range and mean absolute deviation) of a feature shows the maximum happiness recognition rate of 66.67% by using cochleagram feature with

KNN classifier and MFCC feature obtains the maximum accuracy of 56.19% with KNN classifier.

From Table 1, it is clear that MFCC features are capable of effective recognition of all the seven emotions. Cochleagram and power cepstrogram features are related to the maximum recognition rate of the emotion happiness. The maximum overall accuracy is by mean feature of MFCC and maximum recognition rate of happiness is by combination of all functional of cochleagram features. It shows that for effective recognition of the emotion happiness, it is necessary to have fusion of features. The maximum classification performances are associated with the classifier KNN, the simple classification based on distance metric. It shows that the features it selves play vital role in speech emotion recognition than the parameters derived from the features.

4.1.1 Performance of Fusion of Features

The emotion classification performance of different combination of features is shown in Table 2. On evaluating the mean functional features, using KNN classifier, fusion of pitch, intensity, hormonicity and MFCC (P+I+H+MFCC) features, fusion of pitch, intensity, hormonicity, MFCC and LPC (P+I+H+MFCC+LPC) features and fusion of pitch, intensity, hormonicity, MFCC, LPC and cochleagram (P+I+H+MFCC+LPC+C) features combination performed better with happiness recognition rate of 51.85%. Fusion of pitch, intensity, harmonicity and MFCC features increases the recognition of happiness. Addition of LPC and Cochleagram features neither increases nor decreases the recognition of happiness. But the addition of power cepstrogram and formant significantly affects the recognition of the emotion, happiness. Fusion of pitch, intensity, harmonicity, MFCC and LPC (P+I+H+MFCC+LPC) recognizes all the emotions with the overall accuracy of 56.7%. Even though the inclusion of LPC has no effect on the recognition of happiness, it increases the overall emotion classification performance. In case of median features, the fusion of median of pitch, intensity, harmonicity, MFCC, LPC and cochleagram (P+I+H+MFCC+LPC+C) features with KNN classifier recognize the emotion happiness with 62.97%, which is 11.12% greater than the recognition by mean features. Even though the median features increase the recognition of happiness, it fails to increase the overall emotion recognition accuracy.

Table 2 Recognition rates of happiness and all emotions obtained by different combination of features for Berlin database

Functionals	Classifier	Emotion	Recognition rate in %					
			P + I + H	P + I + H + MFCC	P + I + H + MFCC + LPC	P + I + H + MFCC + LPC + C	P + I + H + MFCC + LPC + C + PC	P + I + H + MFCC + LPC + C + PC + F
Mean	KNN	Happiness	29.63	51.85	51.85	51.85	7.41	7.41
		Overall	35.05	56.18	56.70	48.45	22.68	22.68
	BDT	Happiness	25.93	37.04	7.40	40.74	33.33	33.33
		Overall	23.71	37.63	24.22	34.02	32.99	30.93
	NB	Happiness	0	22.22	22.22	0	7.41	7.41
		Overall	36.59	37.63	42.27	32.47	45.88	46.39
Median	KNN	Happiness	22.22	48.14	48.15	62.97	14.82	14.81
		Overall	34.54	54.12	54.12	48.45	18.56	18.56
	BDT	Happiness	33.33	29.62	18.52	37.03	18.52	18.52
		Overall	22.68	42.26	33.50	32.99	36.61	35.67
	NB	Happiness	0	25.93	22.22	14.81	33.33	FT
		Overall	38.66	41.75	42.27	35.05	50	FT
Standard deviation	KNN	Happiness	22.22	25.92	25.93	33.33	14.81	14.81
		Overall	23.19	32.47	32.99	40.72	24.23	24.23
	BDT	Happiness	7.40	18.52	22.22	33.33	37.04	33.33
		Overall	22.16	21.65	34.54	33.50	34.54	34.02
	NB	Happiness	0	14.81	18.51	14.81	18.52	18.52
		Overall	22.16	34.02	37.11	37.11	49.48	50
Inter-quartile range	KNN	Happiness	18.52	33.33	33.33	33.33	22.22	22.22
		Overall	41.24	39.69	39.69	40.72	31.44	31.44
	BDT	Happiness	14.81	25.93	11.11	14.81	18.52	18.52
		Overall	28.35	26.28	34.53	37.62	33.51	33.51
	NB	Happiness	77.78	25.93	3.70	14.81	18.52	18.52
		Overall	33.50	33.50	42..27	42.78	51.03	51.03
Mean absolute deviation	KNN	Happiness	37.04	18.52	18.52	48.15	33.33	33.33
		Overall	26.29	31.96	31.96	43.81	31.96	31.96
	BDT	Happiness	29.63	18.52	3.70	22.22	14.81	14.81
		Overall	20.10	27.32	26.28	41.23	32.47	32.47
	NB	Happiness	7.41	18.52	18.52	11.11	7.41	7.41
		Overall	28.35	32.99	36.08	39.69	47.42	48.45
Combination of all functionals	KNN	Happiness	18.51	59.26	59.26	59.26	18.52	18.52
		Overall	38.14	52.57	52.58	55.15	33.51	33.51
	BDT	Happiness	33.33	44.44	37.03	25.92	18.52	18.52
		Overall	33.50	39.69	35.05	34.02	38.66	38.66
	NB	Happiness	33.33	25.93	14.81	7.40	14.82	FT
		Overall	31.96	35.57	47.94	43.29	53.09	FT

KNN-K nearest neighbour, BDT-Binary decision tree, NB-Naive Bayes, P-Pitch, I-Intensity, H-Hormonicity, MFCC-Mel-frequency cepstral coefficient, LPC-Linear prediction coefficient, C-Cochleagram, PC-Power cepstrogram, F-Formant and FT-Failed to train.

Classification of fusion of inter-quartile range of pitch, intensity and harmonicity (P+I+H) with NB classifiers takes the happiness recognition to the maximum recognition rate of 77.78% and fusion of all other features with the above features drastically reduces the performance. Better recognition of both happiness and all emotions are achieved by the combination of all functional of pitch, intensity, harmonicity, MFCC, LPC and cochleagram (P+I+H+MFCC+LPC+C) features with KNN classifier. Table 2 explicitly shows the snag between the happiness recognition and overall emotions recognition. A feature combination that is performing better to recognize happiness is not performing better to recognize overall emotions.

4.2 Performance with Enterface Database

The classification results of different individual features derived from Enterface database are shown in Table 3. Mean of cochleagram features with NB classifier recognize happiness with the maximum recognition rate of 38.75% and the maximum overall accuracy (40.42%) in recognizing all the emotions is obtained by the mean of MFCC features with NB classifier. On evaluating the median functional, median of pitch with BDT classifier obtains the maximum happiness recognition rate of 60% and median of MFCC features with NB classifier obtain maximum overall accuracy of 38.33%. In case of standard deviation features, the maximum happiness recognition of 55% is obtained by pitch feature with NB classifier and maximum overall accuracy of 33.54% is obtained by MFCC features with NB classifier. Out of all inter-quartile features, classification of emotions by inter-quartile range of MFCC features and NB classifier achieve the maximum happiness recognition of 35% and power cepstrogram perform better with NB classifier to recognize all emotions. Mean absolute deviation of cochleagram features with NB classifier classify happiness with the maximum recognition rate of 57.5% and power cepstrogram features with NB classifier classify all emotions with the maximum accuracy of 32.5%. Fusion of all functional of an individual feature results in the maximum recognition of the emotion happiness by harmonicity features and BDT classifier. For overall emotion recognition, it is 38.96% with MFCC features classified by NB classifier.

Finally, the comparison of the performances of the three classifiers trained and tested with different combinations of the eight features show that happiness of Enterface database is recognized better by median of pitch feature with BDT classifier and the corresponding maximum recognition rate is 60%. In the recognition of all emotions, mean MFCC features with NB classifier perform better with the maximum accuracy of 40.42%.

Comparison of Tables 1 and 3 shows that a classifier which is good in recognizing emotions of one database is not good in recognizing emotions of another database. For example, KNN classifier is associated with the maximum performances of Berlin database but in case of Enterface database, the classifiers associated with maximum performances are BDT and NB classifiers. For both the databases, MFCC features recognize all the emotions with maximum accuracy and MFCC is not the feature that recognizes happiness with maximum recognition rate.

Table 3 Recognition rates of happiness and all emotions obtained by different features and different classifiers for Enterface database

Functionals	Classifier	Emotion	Recognition rate in %							
			Pitch	Intensity	Formant	Harmonicity	MFCC	LPC	Cochleagram	Power cepstrogram
Mean	KNN	Happiness	16.25	26.25	20	27.5	21.25	28.75	21.25	25
		Overall	20.21	18.33	18.96	30.83	24.17	27.08	19.58	28.75
	BDT	Happiness	21.25	18.75	25	18.75	17.5	22.5	28.75	25
		Overall	17.5	17.71	17.5	19.58	22.5	23.95	15.42	21.46
	NB	Happiness	0	1.25	6.25	27.5	20	38.75	35	21.25
		Overall	26.46	2	21.25	40.42	31.25	27.92	32.71	32.08
Median	KNN	Happiness	12.5	25	15	22.5	12.5	31.25	33.75	23.75
		Overall	19.79	21.87	15.83	29.58	21.67	26.88	18.96	27.29
	BDT	Happiness	60	11.25	36.25	17.5	23.75	8.75	28.75	36.25
		Overall	19.17	18.96	16.68	21.88	23.96	21.10	24.79	20.62
	NB	Happiness	0	0	0	42.5	10	FT	50	30
		Overall	23.54	23.13	23.13	38.33	29.38	FT	32.92	32.71
Standard deviation	KNN	Happiness	26.25	11.25	10	22.5	18.75	32.5	15	25
		Overall	19.58	16.67	19.79	27.29	20.21	25	18.75	19.38
	BDT	Happiness	21.25	25	22.5	6.25	12.5	18.75	23.75	21.25
		Overall	15.83	17.92	16.04	17.5	16.67	20.62	22.29	16.88
	NB	Happiness	55	18.75	0	30	10	55	6.25	7.5
		Overall	26.04	19.17	20	33.54	23.54	32.92	27.92	27.08
Inter-quartile range	KNN	Happiness	22.5	15	13.75	30	12.5	17.5	21.25	16.25
		Overall	19.37	17.29	17.5	24.37	18.33	25.20	18.13	21.25
	BDT	Happiness	12.5	22.5	25	25	6.25	26.25	26.25	12.5
		Overall	15	14.17	17.29	17.92	24.37	23.75	18.54	16.25
	NB	Happiness	28.75	2.5	0	35	8.75	55	33.75	27.5
		Overall	23.33	22.71	18.33	32.08	23.33	27.29	32.29	20
Mean absolute deviation	KNN	Happiness	8.75	13.75	15	33.75	17.5	25.42	13.75	21.25
		Overall	17.29	15.21	18.33	26.46	16.45	25.42	17.91	18.96
	BDT	Happiness	18.75	25	18.75	13.75	7.5	27.5	21.25	21.25
		Overall	16.46	16.25	16.88	21.46	19.10	23.96	17.70	18.54
	NB	Happiness	42.5	12.5	0	31.25	10	57.5	22.5	13.75
		Overall	23.96	17.5	20.63	34.17	22.70	30.83	32.5	23.75
Combination of all functionals	KNN	Happiness	20	18.75	20	30	13.75	31.25	13.75	27.5
		Overall	22.71	19.79	17.5	33.75	23.54	33.12	18.54	24.79
	BDT	Happiness	20	16.25	37.5	27.5	26.25	27.5	16.25	28.75
		Overall	22.29	17.08	19.58	23.54	24.16	23.75	18.75	22.92
	NB	Happiness	23.75	13.75	17.5	36.25	15	FT	36.25	28.75
		Overall	27.08	22.08	23.75	38.96	27.08	FT	18.75	31.04

KNN-K nearest neighbour, BDT-Binary decision tree, NB-Naive Bayes, MFCC-Mel-frequency cepstral coefficient, LPC-Linear prediction coefficient and FT-Failed to train.

4.2.1 Performance of Fusion of Features

The emotion classification performances of different combination of features of Enterface database are shown in Table 4.

Table 4 Recognition rates of happiness and all emotions obtained by different combination of features for Enterface database

Functionals	Classifier	Emotion	Recognition rate in %					
			P + I + H	P + I + H + MFCC	P + I + H + MFCC + LPC	P + I + H + MFCC + LPC + C	P + I + H + MFCC + LPC + C + PC	P + I + H + MFCC + LPC + C + PC + F
Mean	KNN	Happiness	18.75	31.25	31.25	40	21.25	21.25
		Overall	18.75	28.75	28.13	30.42	19.58	19.58
	BDT	Happiness	31.25	13.75	26.25	23.75	12.5	35
		Overall	20	23.75	27.92	23.75	19.17	22.08
	NB	Happiness	26.25	48.75	25	36.25	47.5	48.75
		Overall	28.54	37.50	31.46	28.12	32.08	32.21
Median	KNN	Happiness	20	27.50	31.25	26.25	33.75	33.75
		Overall	19.58	28.75	23.96	28.54	18.96	18.96
	BDT	Happiness	33.5	13.75	28.75	15	28.75	23.75
		Overall	20.62	23.75	24.38	26.88	25.83	24.79
	NB	Happiness	15	42.5	13.75	FT	FT	FT
		Overall	26.46	38.12	30.42	FT	FT	FT
Standard deviation	KNN	Happiness	22.5	27.25	25	27.5	15	15
		Overall	20.42	26.04	25.42	29.58	18.75	18.75
	BDT	Happiness	32.5	12.5	22.5	12.5	8.75	18.75
		Overall	21.04	20.83	21.67	18.12	18.75	20.42
	NB	Happiness	12.5	30	13.75	52.50	50	51.25
		Overall	23.33	36.04	26.67	36.25	38.54	38.96
Inter-quartile range	KNN	Happiness	27.5	20	22.5	25	21.25	21.25
		Overall	21.88	23.33	23.96	30.83	18.12	18.13
	BDT	Happiness	27.5	10	17.5	23.75	12.5	20
		Overall	20.21	19.17	23.12	26.88	20.08	23.33
	NB	Happiness	12.5	32.5	18.75	35	FT	FT
		Overall	23.33	32.92	26.04	36.04	FT	FT
Mean absolute deviation	KNN	Happiness	18.75	26.25	23.75	25	13.75	13.75
		Overall	22.70	27.50	27.5	28.96	17.92	17.92
	BDT	Happiness	15	17.50	10	18.75	25	16.25
		Overall	16.25	21.25	22.71	21.46	24.58	23.54
	NB	Happiness	8.75	31.25	16.25	60	53.75	52.50
		Overall	23.75	36.46	26.67	36.68	37.70	37.29
Combination of all functionals	KNN	Happiness	26.25	26.25	23.75	27.5	13.75	13.75
		Overall	26.04	31.04	29.58	30.42	18.54	18.55
	BDT	Happiness	30	32.50	25	30	27.5	26.25
		Overall	21.04	22.92	21.88	25	23.96	22.08
	NB	Happiness	36.25	38.75	15	FT	FT	FT
		Overall	32.08	40.42	29.38	FT	FT	FT

KNN-K nearest neighbour, BDT-Binary decision tree, NB-Naive Bayes, P-Pitch, I-Intensity, H-Hormonicity, MFCC-Mel-frequency cepstral coefficient, LPC-Linear prediction coefficient, C-Cochleagram, PC-Power cepstrogram, F-Formant and FT-Failed to train.

From Table 4 it is clear that the maximum recognition of happiness is achieved by fusion of mean absolute deviation of pitch, intensity, harmonicity, MFCC, LPC and cochleagram (P+I+H+MFCC+LPC+C) features classified by NB classifier. The maximum happiness recognition rate is 60%. Maximum overall emotions recognition accuracy of 42.5% is obtained by fusion of median of pitch, intensity, harmonicity and MFCC (P+I+H+MFCC) features classified by NB classifier. Fusion of features provides performance improvement in overall accuracy but no improvement is obtained in happiness recognition. The maximum happiness recognition rate is the same with the individual features as well as with fusion of features as shown in Table 3 and 4.

5 Conclusion and Future Work

In this paper, we explored happiness and overall emotions recognition performance of different individual and fusion of features and different classifiers. The proposed classification analysis has been done under speaker-independent classification system with two well-known Berlin and Enterface emotional databases. We have analyzed emotion recognition performance of pitch, intensity, harmonicity, Mel-frequency cepstral coefficient, linear prediction coefficient, cochleagram, power cepstrogram and formant features with K-nearest neighbour, binary decision tree and Naive Bayes classifiers. A hunt for finding out the optimum features and classifiers to recognize the emotion happiness is done. The combination of pitch, intensity and harmonicity features identified the emotion happiness with the maximum recognition rate of 77.78% on Berlin database and combination of pitch intensity, Mel-frequency cepstral coefficients, linear prediction coefficients, and cochleagram features predicted the emotion happiness with the maximum recognition rate of 60% on Enterface database. In future work, many modifications can be integrated within this frame work. More features, more classifiers, a proper feature selection approach and natural databases can be used in the analysis of improving the happiness emotion recognition rate.

References

[1] Wu, C. H., and Liang, W. B., 2011, "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels," IEEE Transactions on Affective Computing, 2(1), pp. 10-21.

[2] Sheikhan, M., Bejani, M., and Gharavian, D., 2013, "Modular Neural-SVM Scheme for Speech Emotion Recognition using ANOVA Feature Selection Method," Neural Computing and Applications, 23(1), pp.215-227.

[3] Gharavia, D., Sheikhan, M., Nazerieh, A., and Garoucy, S., 2011, "Speech Emotion Recognition using FCBF Feature Selection Method and GA-Optimized Fuzzy ARTMAP Neural Network," Neural Computing and Applications, 21(8), pp. 2115-2126.

[4] Iliev, A. I., Scordilis, M. S., Papa, J. P., and Falcao, A. X., 2010, "Spoken Emotion Recognition through Optimum-Path Forest Classification using Glottal Feature," Computer Speech and Language, 24(3), pp. 445-460.

[5] Yu, L., Zhou, K., and Huang, Y., 2014, "A Comparative Study on Support Vector Machines Classifiers for Emotional Speech Recognition," International Journal of Immune Computation, 2(1), pp. 35-42.

[6] Bukhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., and Weiss, B., 2005, “A Database of German Emotional Speech,” Proceedings of Interspeech, Lissabon, pp.4-8.

[7] Martin, O., Adell, J., Huerta, A., Kotsia, I., Savran, A., and Sebbe, R., 2005, “A Multimodal Caricatural Mirror,” Proceedings of the eNTERFACE Summer Workshop on Multimodal Interfaces, pp.13-20.

[8] Anagnostopoulos, C. N., Iliou, T., and Giannoukos, I., 2015, “Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000 to 2011,” Artificial Intelligence Review, 43(2), pp. 155-177.

[9] Koolagudi, S. G., and Rao, K. S., 2012, “Emotion Recognition from Speech using Source, System and Prosodic Features,” International Journal of Speech Technology, 15(2), pp. 265-289.

[10] C Zou, Yan Zhao, Li Zhao, W Zhen and Y Bao, 2007, “Emotional Recognition Using a Compensation Transformation in Speech Signal,” Computational Linguistics and Chinese Language Processing,Vol.12, No. 1, pp.79-90.

[11] Rho, S., and Yeo, S. S., 2013, “Bridging the Semantic Gap in Multimedia Emotion/Mood Recognition for Ubiquitous Computing Environment,” The Journal of Supercomputing, 65(1), pp.274-286.

[12] Lee, C. M., Narayanan, S., and Pieraccini, R., 2001, “Recognition of Negative Emotions from the Speech Signal,” Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 240-243.

[13] Ververidis, d., and Kotropoulos, C., 2006, “Emotional Speech Recognition: Resources, Features, and Methods,” Speech Communication, 48(9), pp.1162-1181.

[14] Jyoti, G. N., and Pranab, D., 2014, “Emotion Recognition from Speech Signal: Realization and Available Techniques,” International Journal of Engineering Science and Technology, 6(5), pp. 188-191.

[15] Bejani, M., Gharavian, D., and Charkari, N. M., 2014, “Audiovisual Emotion Recognition Using ANOVA Feature Selection Method and Multi-Classifier Neural Networks,” Neural Computing and Applications, 24(2), pp. 399-412.

[16] Mencattini, A., Martinelli, E., Costantini, G., Todisco, M., Basile, B., Bozzali, M., and Natale, C. D., 2014, “Speech Emotion Recognition Using Amplitude Modulation Parameters and A Combined Feature Selection Procedure,” Knowledge-Based Systems, 63, pp. 68-81.

[17] Origlia, A., Cutugno, F., and Galatà, V., 2014, “Continuous Emotion Recognition with Phonetic Syllables,” Speech Communication, 57, pp. 155-169.

[18] Zheng, W., Xin, M., Wang, X., and Wang, B., 2014, “A Novel Speech Emotion Recognition Method via Incomplete Sparse Least Square Regression,” IEEE Signal Processing Letters, 21(5), pp. 569-572.

[19] Bhargava, M., and Polzehl, T., 2012, “Improving Automatic Emotion Recognition from Speech Using Rhythm and Temporal Feature,” Proceedings of ICECIT-2012 Elsevier, pp. 139-147.

[20] Ayadi, M. M. H. E., Kamel, M. S., and Karray, F., 2007, “Speech Emotion Recognition Using Gaussian Mixture Vector Autoregressive Models,” IEEE International Conference on Acoustic, Speech and Signal Processing, IV, pp. 957-960.

[21] Yang, B., and Lugger, M., 2010, “Emotion Recognition from Speech Signals Using New Harmony Features,” Signal Processing, 90(5), pp. 1415-1423.

[22] Albornoz, E. M., Milone, D. H., and Rufiner, H. L., 2011, “Spoken Emotion Recognition Using Hierarchical Classifiers,” Computer Speech and Language, 25(3), pp. 556-570.

[23] Lee, C. C., Mower, E., Busso, C., Lee, S., and Narayanan, S., 2011, “Emotion Recognition Using A Hierarchical Binary Decision Tree Approach,” Speech Communication, 53(9-10), pp.1162-1171.

[24] Boersma, P., and Weenink, D., 2009, “Praat: Doing Phonetics by Computer,” (Computer Program), Institute of Phonetic Sciences, University of Amsterdam, Amsterdam.