

RESEARCH

Open Access



# Heart rate monitoring using human speech spectral features

Alex Pappachen James\*

\*Correspondence:  
apj@ieee.org  
School of Engineering,  
Nazarbayev University,  
Astana, Kazakhstan

## Abstract

This paper attempts to establish a correlation between the human speech, emotions and human heart rate. The study highlights a possible contactless human heart rate measurement technique useful for monitoring of patient condition from real-time speech recordings. The distance between the average peak-to-peak distances in speech Mel-frequency cepstral coefficients are used as the speech features. The features when tested on 20 classifiers from the data collected from 30 subjects indicate a non-separable classification problem, however, the classification accuracies indicate the existence of strong correlation between the human speech, emotion and heart-rates.

**Keywords:** Heart rate, Human emotions, Speech signal, Mel-frequency cepstral coefficients

## Introduction

Heart rate indicates the total number of times our heart contracts and relaxes per minute; and is expressed as the number of beats per minute (bpm) [1]. Many factors contribute to variation in heart rate such as level of physical activity, fitness, temperature, body position, emotions, body size and medication. Heart Rate also depends on the body's need to absorb oxygen and excrete carbon dioxide. Traditionally, heart rate is measured by detecting arterial pulsation. The electrical activity of the heart is measured by a non-invasive technique called electrocardiogram and is used for assessing the condition of a human heart. The use of technology is gaining significance especially in telehealth applications for the continuous monitoring for cardiac patients and for special situations like treating a burn victim or monitoring infants having the risk of sudden infant syndrome.

A device used to monitor and record the heart rate in real-time is generally referred to as heart rate monitor. The early models of heart rate monitors use electrode leads that are attached to the chest. In contrast, the modern heart rate monitors use chest strap transmitter and a wrist receiver or a mobile phone. Strapless heart rate monitors allow the user to simply touch two sensors on a wristwatch display for a few seconds to display the heart rate. The detection of heart rate from human emotions were also done based on the modelling of vowel speech signals.

In many of the biotelemetry applications, the heart rate monitoring of the ambulatory patients become extremely difficult if the mobility of the patient is restricted due to

paralysis or and/or where continuous paramedic attention and monitoring with sensors become unaffordable. Through this research, we propose to overcome such restrictions by implementing a contact less heart rate monitoring method using the real-time speech recordings based on different emotions of the patient. The system developed is based on a database of emotions and heart rate obtained from multiple subjects.

## Background

Most of the past research in heart rate detection are based on the modelling of vowel speech signals and by processing different physiological parameters. The short-time Fourier Transform (STFT) [2, 3] has been used to detect the maximum peaks of the formants in the process of detecting the heart rate.

McCarty et al. [4] related the mathematical transformation of heart rate variability with the power spectral density of speech. He founded that the positive emotions results in the alteration of heart rate variability. Kim et al. [5] developed a new emotion recognition system by processing different physiological parameters. The electrocardiogram, skin temperature and electrodermal activity were used as input signals. This particular system consisted of preprocessing, feature extraction and pattern classification stages. The characteristics of the emotion is identified from short-segment signals.

Anttonen et al. [6] developed an EMFi chair that measures heart rate. The chair is embedded with electromechanical film and traditional earlobe photoplethysmography (PPG) for measuring the heart rate. He used this setup to study the impact of emotional changes on the heart rate. Mesleh et al. [7] developed a method for the detection of heart rate from human emotions based on the modelling of vowel speech signals. The non-contact method for the detection of heart rate from human emotions is based on modelling the relationship between speech production of vowel speech signals and heart activities for humans. It uses STFT to estimate heart rates from vowel speech signals. This method is not extensively applicable in the medical field since the representation of human emotions through vowel speech cant be clearly illustrated. This method does not work for persons with artificial or transplanted hearts.

## Methodology

### Data collection

30 subjects aged between 20 and 45 had participated for the data collection. The speech and the corresponding heart rate were recorded in a sound proof room using a high quality microphone to obtain the speech signals with minimum noise. Three different emotions viz. anger, neutral and joy were selected for the study. During data collection, the subjects were asked to express specific emotions through words. The emotions were self-induced by instructing subjects to recollect the past situations. The recorded utterance was 'It Begins In Seven Hours'.

Before the start of the recording, a brief introduction was given to the subjects regarding the purpose and procedure of experiment and demonstrated how to express the emotions. The recording was conducted for 60 days and the subjects were selected randomly for the study. For one subject, three different sessions were conducted in two different days for three different emotions. The duration of a particular recording instance

is 30 seconds and 30 such instances were taken for each emotion. From a recorded instance, only ten seconds were used for feature extraction process.

1. *Hand held ECG monitor* The device used for measuring heart rate was a Handheld ECG monitor MD100 A1. This device is able to detect and display the ECG waveform non-invasively and for routine monitoring of heart rates. Strict instructions were given to the subjects about handling the device. It has two modes of display, easy mode and continuous mode. Easy mode has chosen for uncomplicated and suitable measurement. Subjects were seated in straight and stable chairs to reduce the motion of the subject. The subjects were instructed to hold the 8 metal electrodes of the device with right index finger firmly, and place the 3 electrodes against the centre of the left palm. The recorded ECGs were transferred to the computer using a software named 'Keep-It-Easy'.
2. *Recording room* Recording rooms are usually multi-room facilities that require high sound quality and less noise compared to other rooms. The more concentrated factors during the construction of an audio recording room are ceiling and walls. Porous absorbers like melamine sponges and wood are used to build the walls of the voice recording room. Concrete walls and floors are usually avoided to make the audio recording room noise free.

The speech recording room used was acoustically quiet as possible. Only the subject and the person who assist the recording will be present inside the recording room. The voice was recorded in a mobile phone (Sony Ericson Xperia Mini Pro) with hi-fi stereo headset MH 710 and a noise level of  $-89.7$  dB. The microphone was kept closer to the subjects mouth to reduce the ambient noise and to increase the sound level of the speech signals. The microphone was always being kept directly at the subjects mouth as the level of sound becomes less when the microphone moves away. Computer in the recording room is very often a disturbing factor in recording the speech signals. Hence silent processor fan and power supplies are used to reduce the sounds from the computer. A carpet helps to reduce the reflection of the sound signals.

### **Feature extraction**

The extraction of features from the emotional speech signal is one of the important tasks to understand the human speech behaviour. Mel frequency cepstrum is considered to be the robust method and widely adopted method for speech processing.

1. *Recording room Mel-frequency cepstral coefficient analysis* Mel-frequency cepstrum [8] is a mathematical representation of short term power spectrum of the speech. Mel frequency cepstral coefficients [8] are based on a standard power spectrum estimate which is first subjected to a log based transform of the Mel-frequency scale and then decorrelated by using a modified discrete cosine transform.

The process of Mel-frequency cepstral coefficient identification process includes six important steps: (1) pre-emphasis, (2) framing, (3) hamming windowing, (4) Fast Fourier transform, (5) Mel-filter bank processing, and (6) discrete cosine transforms.

*Preemphasis* is a process to boost the magnitude of higher frequencies in the signal for improving the Signal to Noise ratio. It is the initial step of noise reduction technique [9]. The recorded emotional speech is emphasized by passing through a filter to increase the magnitude of the higher frequencies of the speech signal.

$$Y[n] = X[n] - 0.95X[n - 1] \quad (1)$$

Let  $a = 0.95$ , which makes 95 % of any one sample is presumed to originate from previous sample [9].

*Framing* is the process of segmentation of speech samples originated from analog to digital converter into small frames. The speech signal is divided into frames of  $N$  samples. Neighboring frames are separated based on the value  $M$  and  $M < N$  where  $M = 100$  and  $N = 256$  [9].

*Hamming windowing* is one of the most simple window functions. It reduces the effect of leakage for the better representation of the frequency spectrum of the speech signals. The frames obtained are multiplied with the window function  $W(n)$  to reduce the discontinuities of the speech signals in the time domain. It helps in reducing the spectral artifacts of the speech signals. The window function can be denoted as  $W(n)$  [9].

$$W[n] = 0.56 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (2)$$

where,  $N$  is the number of samples per frame,  $Y[n]$  is the output signal,  $X(n)$  is the input signal, and  $W(n)$  is the Hamming window. The result of windowing signal can be represented as,

$$Y[n] = X[n] \times W[n] \quad (3)$$

*The Fast Fourier Transform* is the fastest algorithm to compute Discrete Fourier transform (DFT). The FFT algorithm is very efficient to compute the DFT calculations in less time. The DFT can be computed efficiently in matlab using the function FFT [9].

The Discrete Fourier transform of a discrete signal  $X(n)$  can be defined as,

$$X[k] = \sum_{n=0}^{N-1} X[n] e^{-j\frac{2\pi}{N}nk}, \quad k = \{0, 1, \dots, N-1\} \quad (4)$$

where,  $X[k]$  = Fourier transform of input signal  $X[n]$ ,  $f$  = frequency in Hz.

The DFT is mostly used in the area of frequency spectrum analysis since it transforms the discrete signal in the time domain into its discrete frequency domain components. Without transforming the discrete time domain signal into the discrete frequency domain signal, it is not able to compute the Fourier transform in DSP based systems [9].

*Mel filter bank processing* The range of frequencies in the FFT spectrum obtained is very broad. Since it is difficult to follow a linear scale in case of speech signals, Mel-scale is used for the filtering process. The Mel filter bank consists of a set of overlapping triangular filters applied to compute the weighted sum of filter components approximating the output to a Mel scale. The center frequencies of the Mel-filters are linearly spaced and the band width is fixed on the Mel scale [9].

The most popular equation to compute the Mel-frequency is given using

$$F(mel) = 2595 \cdot \log_{10}(1 + f/700) \quad (5)$$

*Discrete cosine transform* helps in converting the log Mel spectrum into time domain. The result of the conversion is called Mel Frequency Cepstrum Coefficient. It is a real transform and has better computational efficiency. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector [9].

The discrete cosine transform of a signal can be defined as,

$$C_x[k] = \begin{cases} \sum_{n=0}^{N-1} 2X[n] \cos(\frac{\pi}{2N}k(2n+1)) & 0 \leq n \leq N \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where,  $C_x[k]$  = output signal  $X[n]$  = input signal.

The 12 coefficients obtained after those processes corresponds Mel frequency cepstral coefficients. The sum of the Mel frequency cepstral coefficients is taken and the peaks are calculated. Each 10 s audio have minimum of three repeated utterances and for each plot of the utterance, the average of two peak to peak distance are taken and the corresponding heart rate is measured from the ECG of the corresponding 10 s audio. The peak to peak distance between the MFCC is taken as the features that correlate to the variability in the heart rate.

2. *Heart rate detection* A typical human heart rate is in the range of 60–100 bpm, but this may change according to age, sex and size of the person. The QRS complex represents the depolarization of ventricles. Ventricular rate or the heart rate is calculated as the time interval between the two QRS complexes per unit time. The duration of QRS complex is generally 0.06–0.1 s.

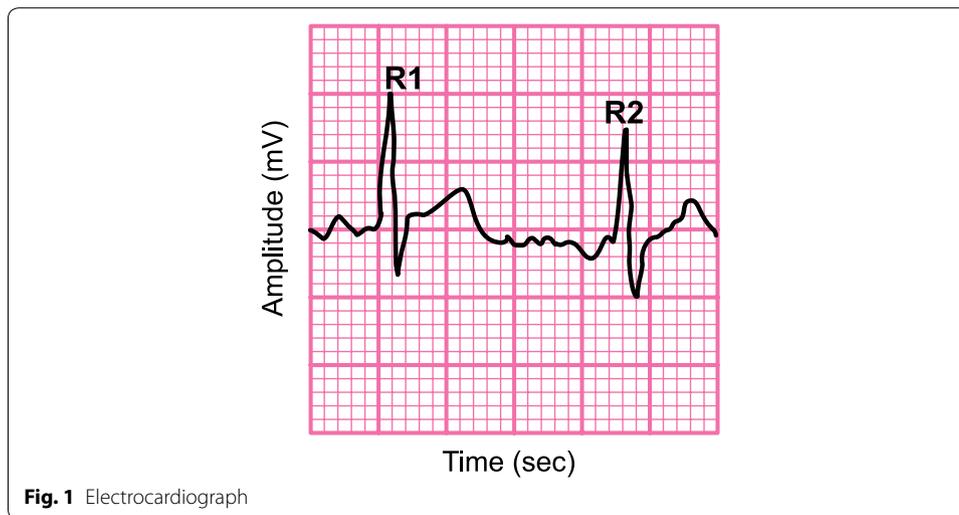
To determine the ECG, during the recording of emotional speech, MD 100 A1 hand-held ECG monitor have been used. The device has 30 s recording capacity in its easy-mode of recording. The heart rate is obtained from recorded ECGs using 1500 rule. According to basic dysrhythmias interpretation and management each horizontal box equals 0.04 s. Hence, count the number of small squares between two neighbouring R-waves and divide that number with 1500. The obtained value will be the heart rate,  $H$  of the respective emotional speech.

$$H = \frac{1500}{\text{No. of small boxes between two R-waves}} \quad (7)$$

From Fig. 1

$$H = \frac{1500}{17} = 88.23 \quad (8)$$

3. *Classification techniques* Naive Bayes Classifier: Naive Bayes classification is a supervised method of classification. The Naive Bayes classifier works on a simple concept. It



**Fig. 1** Electrocardiograph

makes use of the variables contained in the data sample, by observing them individually, independent of each other. It gives an accuracy of 40.09 and 42 % precision for 5 % train and 95 % test dataset. The Nave Bayes classifier is based on the Bayes rule of conditional probability. It makes use of all the attributes contained in the data, and analyses them individually as though they are equally important and independent of each other [10].

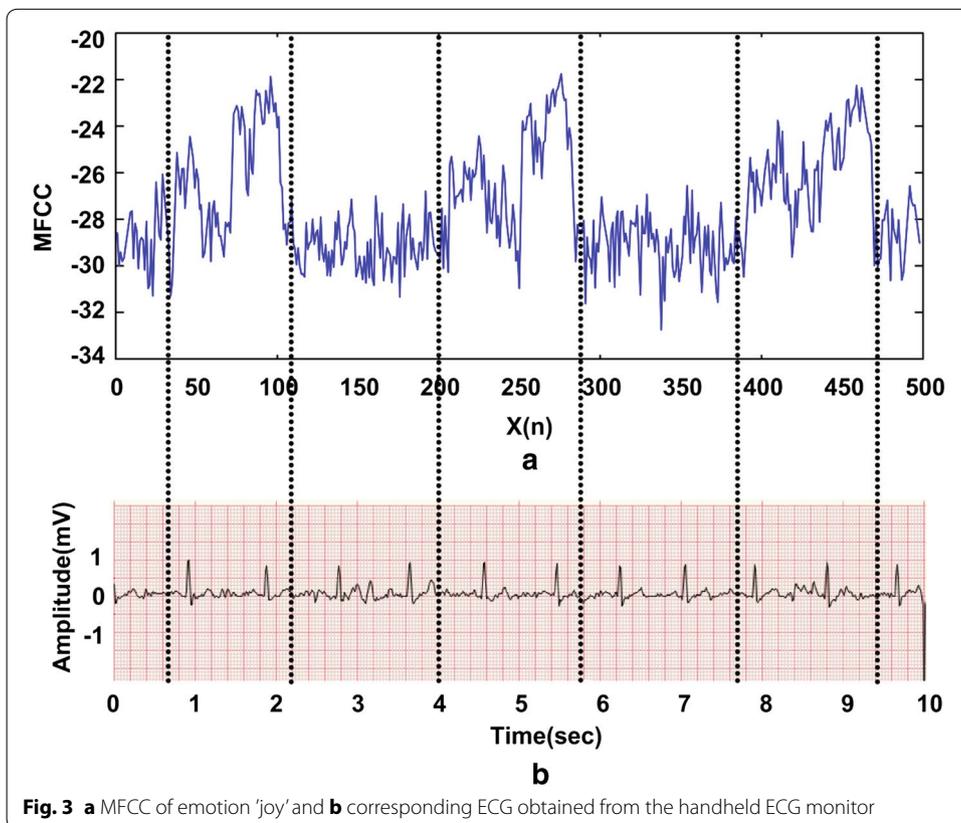
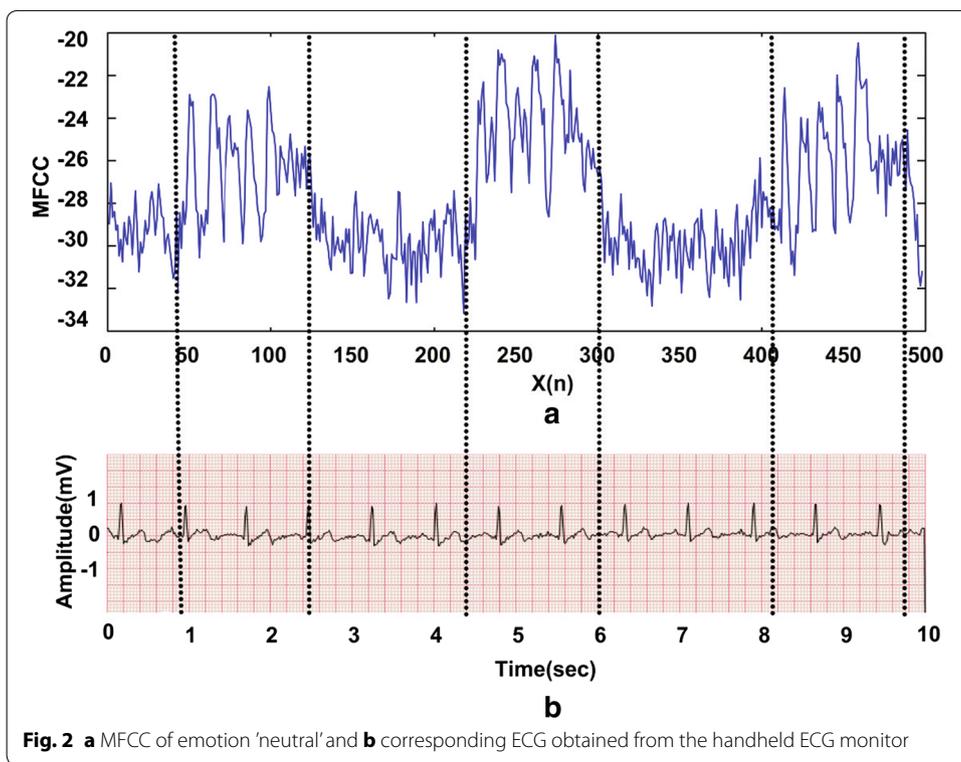
*CVParameterSelection* is a meta-classifier which can optimize over an arbitrary number of parameters. It cannot optimize on nested options. An only direct option of the base classifier is possible. This classifier sets the scheme parameters which are to be set by cross validation. The *CVParameterSelection* Classifier gives an accuracy of 33.33 % and precision 33 for 5 % training and 95 % test dataset [11].

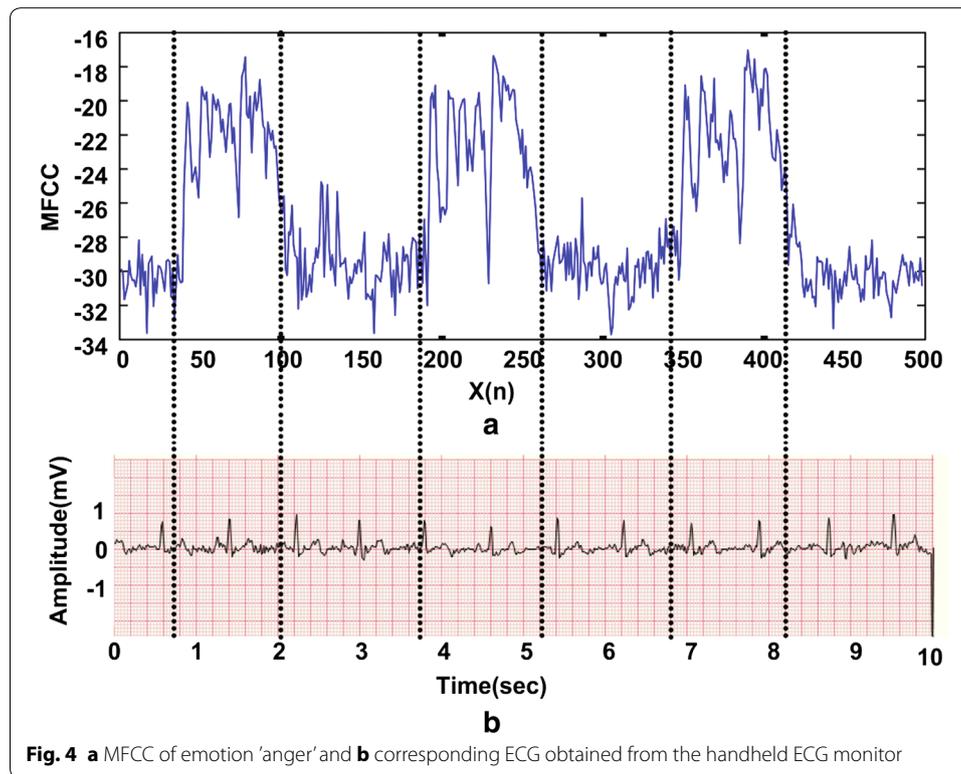
*Filtered classifier* is a class for running an arbitrary classifier on data that has been passed through an arbitrary filter. The *metaFiltered Classifier* gives an accuracy of 33.33 % and precision 33 for 5 % training and 95 % test dataset [12].

*J48*: *J48* is a classifier which is implemented by *C4.5* algorithm. *C4.5* builds decision trees from a set of training data. *J48* is an open source Java implementation of the *C4.5* algorithm in the Weka data mining tool. *C4.5* is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision trees generated by *C4.5* can be used for classification, and for this reason, *C4.5* is often referred to as a statistical classifier. The *J48 Classifier* gives an accuracy of 39.24 % and a precision of 42 % for 5 % training and 95 % test dataset [13].

## Results and discussion

The average feature difference of the Mel frequency cepstral coefficient and the heart rate measured from the recorded ECG are classified on 20 different classifiers. Figures 2, 3 and 4 shows the ECG and speech MFCC features under the common set of basic emotions recorded simultaneously. The distance features drawn from the MFCC and heart rate extracted from the ECG is correlated by the classifiers to predict the emotions. The output obtained from the classifiers for 5 % trained and 95 % tested dataset is shown in a Table 1. The results shows that there exist some correlation between the heart rate,





**Fig. 4** a MFCC of emotion 'anger' and b corresponding ECG obtained from the handheld ECG monitor

**Table 1** Comparison of different classifiers when 5 % of the class samples are used as training data and remaining 95 % of sample are used as test on emotion database

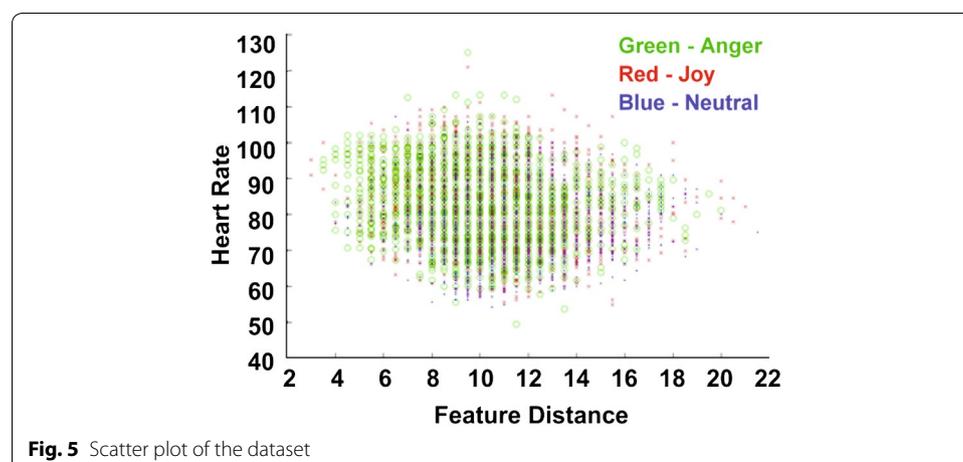
Method	% Correct	AUC	F-measure	Precision (%)	Recall (%)
BayesNet [14]	34.86 ± 2.37	0.52 ± 0.03	0.45 ± 0.12	35 ± 0.09	79 ± 0.34
ComplementaryBayes [15]	39.05 ± 0.88	0.57 ± 0.02	0.45 ± 0.11	37 ± 0.09	56 ± 0.14
NaiveBayes [10]	40.09 ± 1.12	0.60 ± 0.02	0.43 ± 0.04	42 ± 0.02	46 ± 0.08
NaiveBayesMultinomial [16]	38.62 ± 0.59	0.61 ± 0.00	0.45 ± 0.07	40 ± 0.01	53 ± 0.10
Logistic [17]	40.83 ± 1.17	0.61 ± 0.01	0.44 ± 0.05	43 ± 0.02	46 ± 0.07
MultilayerPerceptron [18]	40.27 ± 1.39	0.61 ± 0.02	0.41 ± 0.10	42 ± 0.10	43 ± 0.17
SimpleLogistic [17]	40.63 ± 1.31	0.61 ± 0.01	0.43 ± 0.07	42 ± 0.06	46 ± 0.09
IB1 [19]	37.17 ± 0.90	0.53 ± 0.01	0.37 ± 0.02	38 ± 0.01	37 ± 0.03
lbk [19]	37.21 ± 0.84	0.54 ± 0.01	0.40 ± 0.01	37 ± 0.01	42 ± 0.02
Kstar [20]	40.11 ± 1.18	0.59 ± 0.02	0.41 ± 0.03	41 ± 0.02	42 ± 0.05
LWL [21]	39.37 ± 1.22	0.59 ± 0.02	0.40 ± 0.09	43 ± 0.05	42 ± 0.17
Bagging [13]	39.54 ± 1.09	0.58 ± 0.02	0.41 ± 0.02	41 ± 0.02	41 ± 0.04
ClassificationViaClustering [22]	37.19 ± 2.31	0.53 ± 0.03	0.29 ± 0.21	25 ± 0.18	35 ± 0.25
ClassificationViaRegression [23]	40.47 ± 1.33	0.60 ± 0.01	0.42 ± 0.08	42 ± 0.07	44 ± 0.13
CVParameterSelection [11]	33.33 ± 0.00	0.50 ± 0.00	0.50 ± 0.00	33 ± 0.00	100 ± 0.00
FilteredClassifier [12]	34.86 ± 2.37	0.52 ± 0.03	0.45 ± 0.12	35 ± 0.09	79 ± 0.34
DTNB [24]	36.62 ± 2.41	0.54 ± 0.03	0.36 ± 0.16	38 ± 0.14	50 ± 0.37
JRip [25]	36.82 ± 1.95	0.53 ± 0.03	0.29 ± 0.17	37 ± 0.18	36 ± 0.35
J48 [13]	39.24 ± 1.37	0.56 ± 0.02	0.37 ± 0.11	43 ± 0.08	38 ± 0.18
NBTree [26]	38.89 ± 1.18	0.56 ± 0.02	0.35 ± 0.12	45 ± 0.09	36 ± 0.20

emotion and the human speech, that can be relevant to build a contactless speech sensor based heart-rate measurement. It can be noted that in a classification perspective, the this is a small-sample classification problem, that can prove to be a extremely challenging predication problem in real-time recording situations. The scatter diagram in Fig. 5 shows inseparable nature of the problem, and indicates that MFCC features are not robust under the variations of inter-class emotions.

Since, we have done the recordings using actors, the induced emotions may be subjective to the actors ability to mimic a emotion. In addition, the scatter plots also indicate that the emotions in its pure form are rare and often can be combination of subtle emotions. There are other emotions like boredom, disgust, fear, sad, the basic emotions to be considered. These emotions should also be taken for the classification, which would help in better analysis of ECG measurement from corresponding emotions. Table 2 shows the heart rate detection accuracy of 30 individuals who have been subjected in this study. As can be clearly seen, there is a large variation between the results on the detection accuracy, that indicates a subtle mixing of emotions leading to overlap in classes. Nonetheless it can be seen from the classification accuracies that there exist strong correlations between the emotions in the speech and heart rates.

## Conclusion

We have demonstrated the possibility for a contactless human heart rate monitoring system based on the variation in human emotions. The idea is tested using a range of well known classification techniques. The precision, accuracy, F-measure and percentage recall were also analysed. The classification is done for three different emotions and the emotional level of the subject could be identified from the corresponding heart rates. The database was created from thirty different subjects for three different emotions. The classification analysis indicated strong correlation between the heart-rate, emotion and human speech, which can be further explored to create contact-less real-time heart-rate detection devices.



**Fig. 5** Scatter plot of the dataset

**Table 2 Individual classification accuracy (%) of 30 subjects using 66 % data points for training and remaining for test**

ID	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
18	33.17	41.31	44.66	41.79	43.21	41.96	42.01	38.65	39.48	41.35	41.24	42.7	39.58	40.84	32.92	33.17	33.61	40.37	43.63	41.78
3	33.9	43.1	46.04	45.57	48.18	46.59	48.21	45.67	46.87	49.09	46.4	50.25	36.61	46.33	32.92	33.9	35.14	45.65	48.73	49.89
12	41.73	37.03	48.92	36.88	50.22	50.05	50.36	42.12	41.46	46.73	43.94	48.32	34.06	49.13	32.92	41.73	42.71	48.47	47.32	46.52
19	42.08	44.2	51.21	47.18	51.79	49.64	50.91	42.87	42.44	44.59	42.05	45.8	44.3	51.46	32.92	42.26	43.31	43.4	45.57	44.82
26	45.38	46.98	51.12	47.43	49.49	49.38	49.96	41.47	41.09	46.32	47.85	47.67	42.37	49.81	32.92	45.38	47.5	46.56	48.36	48.57
15	45.46	51.16	50.55	48.54	48.36	48.65	48.54	41.27	39.46	45.63	45.39	46.14	47.71	49.27	32.92	45.35	47.75	45.27	48.36	47.63
2	45.85	38.87	43.89	41.49	43.75	45.42	45.75	47.39	47.64	51.49	44.33	50.36	48.28	47.85	32.92	45.85	45.64	45.66	48.44	48.4
22	45.95	43.9	53.93	45.71	56	53.64	55.17	46.11	47.64	55.57	49.97	54.19	47.52	53.49	32.92	45.95	50.94	52.64	53.34	51.57
4	46.76	47.82	53.35	49.85	50.07	50	49.35	48.62	49.89	52.69	47.56	51.57	47.53	49.27	32.92	46.58	46.14	48.81	51.27	48.31
21	48.98	51.23	52.29	52.28	52.87	53.08	52.51	46.12	46.73	50.26	51.81	50.98	44.87	52.32	32.92	49.06	49.09	49.82	53.45	50.75
28	49.34	53.63	55.78	54	56.17	53.71	56.35	46.19	46.84	51.28	49.96	51.96	43.85	56.5	32.92	49.49	49.81	50.03	50.57	50.17
10	53.31	50.28	59.01	47.55	61.11	60.01	60.31	51.93	55.48	61.39	57.08	60.64	48.73	60.13	32.92	53.31	54.7	56.72	58.24	57.33
8	53.88	52.84	61.23	55.82	60.22	60.33	59.38	56.58	56.94	59.45	57.45	61.24	42.59	59.27	32.92	53.7	54.68	58.61	59.6	57.67
14	54.05	42.69	57.37	40.94	55.79	54.87	55.34	52.36	51.59	55.23	50.23	55.68	49.85	52.74	32.92	53.83	53.29	53.97	55.09	54.7
20	54.53	53.31	65.52	57.56	60.11	60.18	59.79	56.74	58.44	63.92	53.84	60.64	56.76	58.48	32.92	54.17	56.9	59.92	60.98	57.96
30	55.1	51.49	51.84	50.77	56.98	54.26	56.4	50.52	52.01	54.7	54.95	54.91	37.2	54.48	32.92	55.1	54.62	53.31	53.72	55.53
9	55.16	44.44	63.15	41.2	61.48	60.06	60.18	50.19	50.3	58.8	54.1	58.62	51.13	59.49	32.92	54.29	55.45	57.82	56.1	55.89
27	55.42	51.75	56.3	50.91	57.82	55.9	57.35	50.41	50.04	57.38	55.02	55.63	49.65	54.66	32.92	55.42	55.27	54.47	55.92	54.07
17	58.03	54.1	59.79	54.9	59.7	58.5	59.81	56.48	56.29	59.79	58.78	60.83	50.01	57	32.92	58.03	55.74	61.7	61.48	61.39
7	59.81	54.07	63.05	54.48	63.3	61.67	63.74	55.12	56.87	62.83	60.73	62.18	50.21	62.94	32.92	60.14	59.81	64.36	61.7	63.16
23	61.09	58.48	68.36	58.11	67.81	67.88	67.69	58.76	60.09	65.56	62.03	63.7	54.9	66.51	32.92	60.15	61.06	61.38	63.96	62.47
6	61.88	52.98	63.65	49.39	66.64	67.55	66.42	57.06	58	67.3	58.85	66.39	49.93	66.93	32.92	61.85	62.97	64.64	63.81	64.5
16	61.91	45.89	66.98	45.2	66.25	65.7	65.45	56.32	58.18	64.44	58.79	65.68	43.7	64.06	32.92	62.39	63.6	63.05	64.44	63.22
25	63.01	51.34	70.08	49.96	50.39	56.5	50.94	61.61	62.22	70.57	58.76	70.18	53.63	67.41	32.92	62.65	67.08	65.15	68.64	64.09
24	63.63	51.71	65.3	54.25	66.64	65.62	66.56	58.46	57.56	64.45	65.8	64.2	49.57	65.98	32.92	63.23	63.04	65.8	63.62	63.48

**Table 2 continued**

ID	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
29	68.82	42.11	68.24	42.87	64.25	65.73	64.72	63.48	63.3	69.7	65.38	69.26	63.59	68.57	32.92	68.93	68.96	69.76	70.93	68.35
13	70.16	45.13	72.84	42.87	72.88	72.12	72.51	64.87	64.53	67.94	71	68.99	59.96	72.44	32.92	69.55	69.91	68.02	68.49	70.02
1	72.17	63.62	78.05	76.3	79.76	78.16	79.43	74.37	73.5	79.86	65.99	80.06	65.3	80.08	32.92	72.1	74.3	78.01	80.27	78.01
5	80.59	57.95	83.61	63.14	80.63	82.62	79.87	78.99	79.79	82.48	75.98	81.9	64.68	82.04	32.92	80.05	80.7	81.79	81.64	81.57
11	87.35	62.94	86.66	64.51	87.21	87.6	86.84	83.64	84.74	88.37	85.78	86.92	61.11	86.55	32.92	86.69	86.73	86.7	87.39	86.69
Average	55.62	49.55	60.43	50.38	59.63	59.25	59.4	54.14	54.65	59.64	56.03	59.59	49.31	59.53	32.92	55.48	56.35	58.06	59.17	58.28

(1) BayesNet, (2) ComplementNaiveBayes, (3) NaiveBayes, (4) NaiveBayesMultinomial, (5) Logistic, (6) MultilayerPerceptron, (7) SimpleLogistic, (8) IB1, (9) IBk, (10) KStar, (11) LWL, (12) Bagging, (13) ClassificationViaClustering, (14) ClassificationViaRegression, (15) CVPParameterSelection, (16) FilteredClassifier, (17) DTNB, (18) JRip, (19) J48, (20) NBTree

**Acknowledgements**

The author will like to acknowledge Bibia Abraham from Kannur Medical College for helping with the preparation of the dataset required for this study.

**Competing interests**

The authors declare that they have no competing interests.

Received: 20 July 2015 Accepted: 27 October 2015

Published online: 11 November 2015

**References**

- Sundnes J, Lines GT, Cai X, Nielsen BF, Mardal K-A, Tveito A (2007) Computing the electrical activity in the heart, vol 1. Springer, Netherlands
- Milacic M, James AP, Dimitrijević S (2013) Biologically inspired features used for robust phoneme recognition. *Int J Mach Intell Sens Signal Process* 1(1):46–54
- Davletharova A, Sugathan S, Abraham B, James AP (2015) Detection and analysis of emotion from speech signals. *Procedia Comput Sci* 58:91–96
- McCraty R, Atkinson M, Tiller WA, Rein G, Watkins AD (1995) The effects of emotions on short-term power spectrum analysis of heart rate variability. *Am J Cardiol* 76(14):1089–1093
- Kim KH, Bang S, Kim S (2004) Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput* 42(3):419–427
- Anttonen J, Surakka V (2005) Emotions and heart rate while sitting on a chair. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp 491–499
- Mesleh A, Skopin D, Baglikov S, Quteishat A (2012) Heart rate extraction from vowel speech signals. *J Comput Sci Technol* 27(6):1243–1251
- Sigurðsson S, Petersen K, Lehn-Schiøler T (2006) Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music. In: *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*. Victoria
- Muda L, Begam M, Elamvazuthi I (2010) Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. arXiv preprint [arXiv:1003.4083](https://arxiv.org/abs/1003.4083)
- Kim S-B, Han K-S, Rim H-C, Myaeng SH (2006) Some effective techniques for naive bayes text classification. *Knowl Data Eng IEEE Trans* 18(11):1457–1466
- Staelin C (2003) Parameter selection for support vector machines. Hewlett-Packard Company, Tech. Rep. HPL-2002-354R1
- Meraoumia A, Chitroub S, Bouridane A (2012) Multimodal biometric person recognition system based on fingerprint & finger-knuckle-print using correlation filter classifier. In: *IEEE International Conference on Communications (ICC)*, 2012, pp 820–824
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Singhal A, Brown CR (1997) Dynamic bayes net approach to multimodal sensor fusion. In: *Proceedings of SPIE 3209, Sensor Fusion and Decentralized Control in Autonomous Robotic Systems*, vol. 3209. pp 2–10. doi:[10.1117/12.287628](https://doi.org/10.1117/12.287628)
- Zhang H (2004) The optimality of naive bayes. *AA* 1(2):3
- Kibriya AM, Frank E, Pfahringer B, Holmes G (2005) Multinomial naive bayes for text categorization revisited. In: *AI 2004: Advances in Artificial Intelligence*. Springer, Germany, pp 488–499
- Mitchell T (2005) Generative and discriminative classifiers: naive bayes and logistic regression. Manuscript available at <http://www.cs.cmu.edu/tom/NewChapters.html>
- Longstaff ID, Cross JF (1987) A pattern recognition approach to understanding the multi-layer perception. *Pattern Recognit Lett* 5(5):315–319
- Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. *Mach Learn* 6(1):37–66
- Buscema M, Tastle WJ, Terzi S (2013) Meta net: A new meta-classifier family. In: *Data Mining Applications Using Artificial Adaptive Systems*. Springer, New York, pp 141–182
- Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning for control. In: *Lazy Learning*. Springer, Netherlands, pp 75–113
- Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput J* 41(8):578–588
- Hou C, Nie F, Yi D, Wu Y (2013) Efficient image classification via multiple rank regression. *Image Process IEEE Trans* 22(1):340–352
- Ayu MA, Ismail SA, Matin AFA, Mantoro T (2012) A comparison study of classifier algorithms for mobile-phone's accelerometer based activity recognition. *Procedia Eng* 41:224–229
- Shahzad W, Asad S, Khan MA (2013) Feature subset selection using association rule mining and jrip classifier. *Int J* 8(18):885–896
- Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. In: *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS press, Amsterdam, pp 3–44