University of Baghdad
College of Engineering
JOURNAL OF ENGINEERING

Journal of Engineering
journal homepage: www.joe.uobaghdad.edu.iq
Number 11 Volume 25 November 2019

Electrical, Electronics and communications, and Computer Engineering

Mobile-Based Human Emotion Recognition Based on Speech and Heart Rate

Huda Majed
Computer Eng. Dep. - University of Baghdad
Email: sosoalghanem25@gmail.com

Hamid Mohammed Ali
Assist. Prof.
Computer Eng. Dep. - University of Baghdad
Email: hamidma.a@gmail.com

ABSTRACT

Mobile-based human emotion recognition is a very challenging subject, most of the approaches suggested and built in this field utilized various contexts that can be derived from the external sensors and the smartphone, but these approaches suffer from different obstacles and challenges. The proposed system integrated human speech signal and heart rate, in one system, to leverage the accuracy of the human emotion recognition. The proposed system is designed to recognize four human emotions; angry, happy, sad, and normal. In this system, the smartphone is used to record user speech and send it to a server. The smartwatch, fixed on user's wrist, is used to measure user heart rate while the user is speaking and send it, via Bluetooth, to the smartphone, which in turn sends it to the server. At the server side, the speech features are extracted from the speech signal to be classified by a neural network. To minimize the misclassification of the neural network, the user heart rate measurement is used to direct the extracted speech features to either excited (angry and happy) neural network or to the calm (sad and normal) neural network. In spite of the challenges associated with the system, the system achieved 96.49% for known speakers and 79.05% for unknown speakers

Keywords: smartphone, neural network, smartwatch, speech signal, heart rate.

تمييز المشاعر الانسانية باستخدام الهاتف الذكي عبر الانترنيت

هدى ماجد
قسم هندسة الحاسبات - جامعة بغداد

ا.م. حامد محمد علي
قسم هندسة الحاسبات- جامعة بغداد

الخلاصة

يعتبر التعرف على المشاعر المرتكز على المحمول موضوعًا شديد الصعوبة ، وقد استخدمت معظم الأساليب المقترحة والمبنية في هذا المجال سياقات مختلفة يمكن استخلاصها من المستشعرات الخارجية والهاتف الذكي , ولكن هذه الأساليب تعاني من عقبات وتحديات مختلفة. النظام المقترح دمج إشارة الكلام البشري مع معدل ضربات القلب ، في نظام واحد ، للاستفادة من دقة تمييز المشاعر البشرية. تم تصميم النظام المقترح للتعرف على أربعة مشاعر إنسانية ؛ غاضب وسعيد وحزين وطبيعي .في هذا النظام ، يتم استخدام الهاتف الذكي لتسجيل خطاب المستخدم وإرساله إلى خادم. يتم استخدام ساعة ذكية ، مثبتة على معصم المستخدم ، لقياس معدل ضربات قلب المستخدم أثناء قيام المستخدم بالتحدث وإرساله ، عبر البلوتوث ، إلى الهاتف الذكي الذي بدوره يرسله إلى الخادم .على جانب الخادم ، يتم استخراج ميزات الكلام من إشارة الكلام لتصنيفها بواسطة الشبكة العصبية. لتقليل الخطأ في تصنيف الشبكة العصبية ، يتم استخدام قياس معدل ضربات القلب للمستخدم لتوجيه

*Corresponding author
Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2019.11.05

Article received: 4/11/2018
Article accepted: 27/5/2019
Article published: 1/11/2019

ميزات الكلام المستخرجة الى الشبكة العصبية إما الشبكة العصبية المتحمس (غاضب وسعيد) أو إلى الشبكة العصبية الهدوء (حزين والعادي) .على الرغم من التحديات المرتبطة بالنظام ، فقد حقق النظام نسبة 96.49٪ للمتحدثين المعروفين و 79.05٪ للمتحدثين الغير معروفين .
الكلمات الرئيسية: الهاتف الذكي, الشبكة العصبية, الساعة الذكية, اشارة الكلام, معدل ضربات القلب.

1. INTRODUCTION

The proliferation and the advanced technology of the smartphone became an integrated part of each human in our daily life and due to the constantly increasing technical advantages of smartphones which are equipped with high computation processor power, internet connection, different types of sensors that can be utilized for remote human emotion recognition **Tarng, et al., 2010**. Many types of information that can be exploited in mobile-based human emotion recognition system, for example, image, speech, written, and even physiological. Monitoring human emotions play an important role in different fields, especially in the health environment such as psychology, neuroscience, and others. It can be used for psychiatric patients and people who live alone and need to be under surveillance **Cowie, et al., 2000**. Many approaches have been suggested and built, for efficient and accurate mobile-based emotion recognition system, each of which comes with its merits and demerits. The authors in the following approach **Rachuri, et al., 2010** proposed a mobile sensing platform, called EmotionSense system. The main features of this approach include sensing person activities, emotions, proximity, and verbal interactions between the individuals of social groups based on the mobile phone sensors and user speech to recognize human emotion. **Gimpel, et al., 2015** proposed a system called "myStress" to evaluate human stress by unobtrusively exploiting the smartphone data. In this system to deduce the users' perceived stress levels, 36 software and hardware sensors are read. **Sadat, et al., 2014** proposed an approach for detecting human emotion, based on the smartphone built-in inertial sensors and an external sensor, which is called a Force Sensing Resistor (FSR). The use of an external sensor is needed since the current phones do not have any built-in pressure sensor. The FSR sensor is used to measure the pressure's amount that put on the smartphone's screen, which its variation is infected by the user emotion. **Lee, et al., 2012,** proposed an unobtrusive approach to identify the users' emotions using the data collected from many types of smartphone sensors, after analyzing them in an inconspicuous way. The approaches mentioned above employ an enormous amount of data obtained from the hardware and software sensors, which in turn leads to an enormous time of computation and molding of the data to infer the relations between them. Also, the data obtained from the mentioned sensors could not reflect the reality of the user emotion, for example when the user is not holding the smartphone in his hand which then a wrong data concerning user movement is obtained from the smartphone hardware sensors.

Furthermore, the smartphone battery would drain quickly in such systems. The approaches of **Suk and Prabhakaran, 2014**, **Siddiqi, et al., 2017**, **Dai, et al., 2016**, **Silva, et al., 2014** depend solely on the user image obtained from the smartphone camera since the human emotions are highly correlated with facial expressions. The user emotion is detected from the facial expression presented in the user image, the main obstacle in facial emotion recognition system using smartphone devices is the camera shakiness which relatively causes head movements on the contrary to the systems that have a stationary camera, the head movements will lead to degradation of the facial emotion recognition. **Ahmed, et al., 2015** proposed a system called "SocialSense" which depends on the user mobile phone to detect user emotion during speaking, The challenges that face this approach entirely depends on the features extracted from speech signal to detect human emotion. Training speech data is a very delicate task to be performed accurately. This, in turn, affects the classifier used for this purpose. Also, the speech signal features are not standardized in the literature that which of them are perfect for emotion recognition, therefore depending solely on the speech signal to detect human emotion is not adequate without adding supporting factor to enhance the emotion recognition based on speech. The contents of the remaining sections in this paper are organized as follows: section 2 presents the system approach, section 3 presents the results and performance, section 4 presents the conclusions and lastly, and section 5 presents the suggestions for future work.

2. SYSTEM APPROACH

2.1 Overview

The emotional state affects both the speech signals and the heart rate of a human. Depending solely on each of them, for emotion recognition, is not adequate due to the difficulty in extracting clear separable features from the speech signal and the interfering of the heart rate measurements of some emotional states. The speech signal and the heart rate represent valuable indicators that reflect the user's emotion. For better emotion recognition system, this work integrated the use of the speech signal and heart rate in a unique system. Therefore, the features extracted from speech utter by the user, recorded by the smartphone, and the heart rate measurements obtained from user smartwatch are both used to detect human emotions. The main components that contribute to the whole system structure are shown in **Fig. 1**.

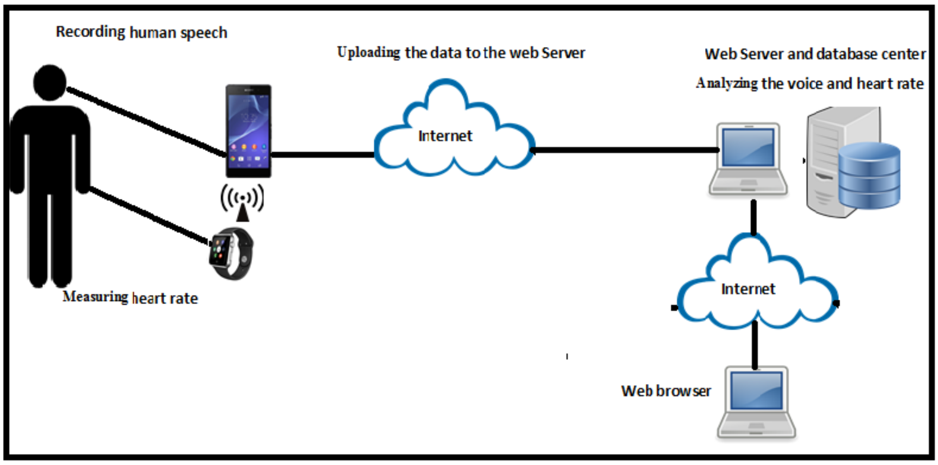


Figure 1. Overall system components.

2.2 Speech Emotion Recognition Mechanism

The shapes of the speech signals are varying due to the state of the person when a talk is carried out; hence, the speech emotion recognition follows the pattern recognition mechanism. **Fig. 2** illustrates the modules and the tasks required to recognize the emotion from the speech signal. Most of the speech emotion recognition mechanism implemented using MATLAB procedures and functions.

2.2.1 Speech input

After the speech signal has been generated at the vocal folds, it moves through the vocal tract and comes out from the human's mouth, nose and cheeks. The smartphone records the speech and transmitted to the server for analysis and emotion recognition.

2.2.2 Speech preprocessing

To acquire clear features from the speech input, the latter is manipulated to convert it into a form that is suitable for extracting effective separable features for better speech emotion recognition.

The four stages of the speech preprocessing module: normalization, silence removal, pre-emphasis, framing, and windowing.

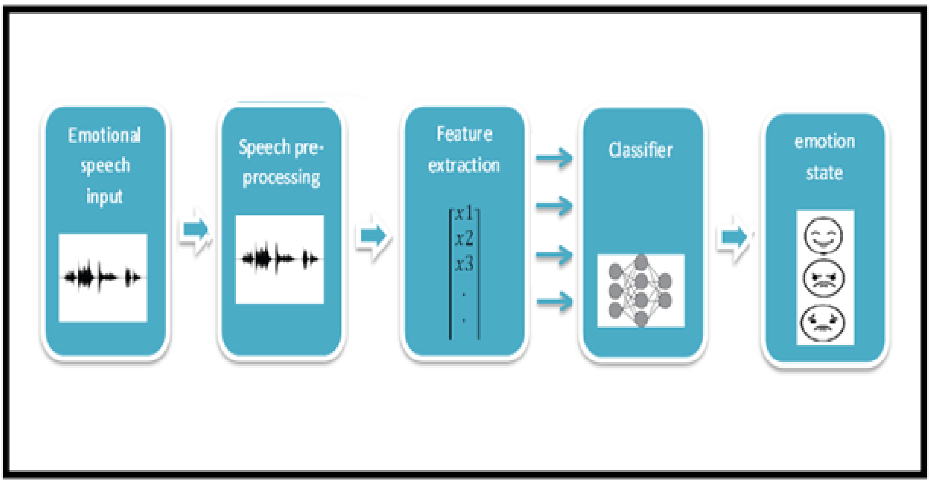


Figure 2. Basic tasks of speech emotion recognition mechanism.

2.2.3 Feature extraction

The most influential part of speech emotion recognition mechanism is extracting appropriate and accurate features that reflect the related alteration in emotion. Seven features which frequently used in emotion speech recognition, and most recommended by researchers **Anne, et al., 2016, Utane and Nalbalwar, 2013, Rong, et al., 2008, Joshi and Kaur, 2013**, are selected in this study, and are briefly explained as follow:

- Speech rate: the speed at which a person can speak, or in other words, how fast can you speak. The speech rate is strongly related to emotions like the angry person speaks faster than a sad person, **Philippou-Hubner, et al., 2012**.
- Short Term Energy (STE): the energy feature which is reflected by the variation with time of the speech signal amplitude is one of the basic features, which are linked to the emotion speech recognition.
- Zero Crossing Rate (ZCR): ZCR is the changing rate of the speech signal from positive to negative or back, it provides information about the number of zero crossings that appear in the signal **Bahargab and Talukdar, 2015**.
- Pitch: a simple definition to the pitch feature is the rate at which the vocal folds are vibrated; it is the fundamental frequency of the speech signal **Joshi and Kaur, 2013, Tarng, et al., 2010**.
- Formants: in speech science, the formants frequencies are defined as the resonance frequencies of the human vocal tract, they are the spectral peaks in the frequency spectrum of sound **Gargouri, et al., 2006, Khulage and Pathak, 2012**.
- Mel-Frequency Cepstrum Coefficient (MFCC): MFCC is based on the perceptions of the human hearing; and simulates how the human's ear reacts to voices by using mel-scale to imitate the human hearing **Muda, et al., 2010**.

- The Linear Prediction Cepstrum Coefficient (LPCC): LPCC represents specified channel characteristics of the speech signal, and a particular individual with various emotions will have various channel characteristics, **Joshi and Kaur, 2013**.

2.2.4 Classification

The classification is an important task of speech emotion recognition mechanism which detects human emotional states from the extracted features of the speech input signal. The feed forward multilayered neural network is chosen for this work due to its classification efficiency in pattern recognition, **Kumar Basu, et al., 2010**.

2.3 Heart rate-based classification

The human emotions usually accompanied by physiological changes, so monitoring these changes in an unobtrusive way can yield information that can be used to support the speech emotion recognition and increase the performance. One of the most essential physiological signals is the heart rate, it can be used to give us information about human emotions, the excited emotions (angry person and happy person) will have a higher heart rate than calm emotions (normal person and the sad person) **Science Daily**, **Ménard, et al., 2015**. The hierarchical classification of the human heart rate, shown in **Fig. 3**, is incorporated in the recognition mechanism as the first stage and the speech features as the second stage. The early stage performs classification based on the heart rate to divide the samples into excited (angry or happy) and calm (sad or normal), while the second stage performs classification based on speech features. In the second stage, the exciting category will be classified to either anger or happiness, and the calm category will be classified to either sadness or normal.

2.4 Gender-based classification

The prior knowledge of the speaker's gender can help to increase the performance of emotion speech recognition since the fundamental frequencies for males are lower than females. In average, the frequency for the male voice ranges from 62Hz to 523Hz, while female voice ranges from 110Hz to 1000Hz **Tarng, et al., 2010**. Therefore the gender property is also exploited in this work to enhance the accuracy of the system. Finally, the overall classification of the emotion recognition system, as illustrated in **Fig. 4**, clarifies how the emotional speech data are classified based on gender and heart rate in this work.

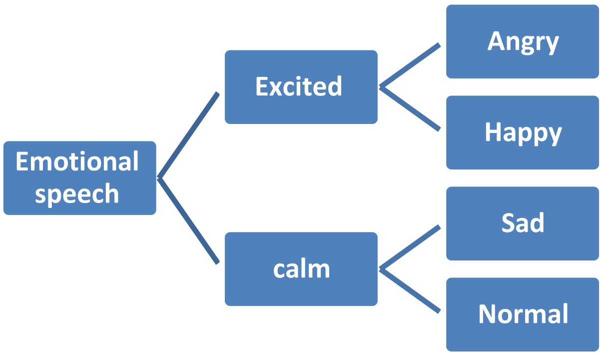
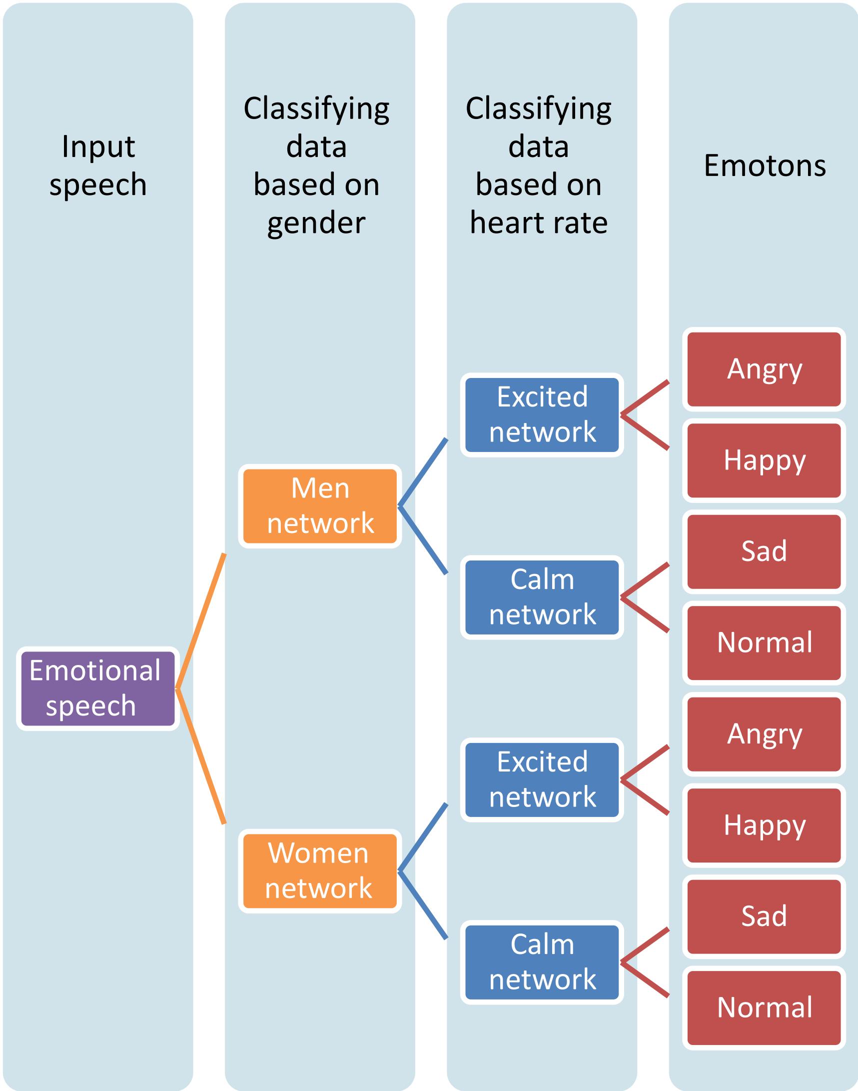


Figure 3. Two-stage hierarchical classification.



2.5 **Figure 4.** The overall classification of the emotion recognition system Emotional speech databases.

Many databases are available for emotion speech recognition systems, and in different languages **Ververidis and Kotropoulos, 2003**, however, most of them are not available for public use. In this study, two databases are used: Berlin Emotional Speech Database and Arabic emotional speech database, which is composed in this work for training and testing the whole proposed system. Berlin database is a German database created by Berlin Technical University **Burkhardt, et al., 2005**. In this database, the speech utterances were spoken by professional actors and actresses (5 males and 5 females), they are about 500 speech utterances, spoken in different 7 emotions namely joy, anger, sadness, fear, disgust, boredom and natural. Four emotions are selected from the database (anger, sad, happy, and natural), so the selected data contains 337 speech samples (151male samples and 186 female samples). The Arabic emotional speech database is obtained practically for the purpose of this work. This data which will be called Arabic speech data are generated by normal people expressing the four emotional states. Note that it wasn't an easy job preparing the Arabic speech data as it needs people to professionally acting to imitate the human emotion. The reason for creating the Arabic database is because of the lack in the availability of Arabic database.

3. Results and performance

This section presents first the practical results obtained only from the speech signal. Second, it discusses integrating the human heart rate obtained from the smartwatch to enhance the final results of human recognition system.

3.1 Speech emotion recognition performance

This subsection evaluates the performance of applying the speech emotion mechanism, adopted in this work, on two speech dataset, the German speech emotion database and the Arabic speech emotion database. The speech emotion mechanism is tested on each database separately.

3.1.1 Results of the German database

After separating the data between males and females, **Table 1** shows the data that is used for the training, while **Table 2** shows the data that is used to test the network. **Table 3** shows a summary of the results and the accuracy for both males and females.

Table 1. Utilized German dataset for the training process.

	Angry	Happy	Sad	Normal	Total
Male	46	19	19	29	113
Female	52	35	25	27	139
Total	98	54	44	56	252

Table 2. Utilized German dataset for the test.

	Angry	Happy	Sad	Normal	Total
Male	14	8	6	10	38
Female	13	9	12	13	47
Total	27	17	18	23	85

Table 3. The accuracy of German database.

	Training results	Testing results
Male	97.3%	86.8%
Female	97.1%	85.1%
Average	97.2%	85.95%

3.1.2 Results of the Arabic database

After separating the data between males and females, **Table 4** shows the data that is used to train the Arabic database while **Table 5** shows the data that is used to test the networks.

Table 6 shows a summary of the results and the accuracy for both males and females. Comparing the test and training results of the female and male networks, it is noticed that the accuracy results of the male network are lower than the female network, which is subjected to two reasons: first, generally the female utter the emotions better than males; second, the database for the female is larger than the males. Generally, the difference between the network results during the training process and the test results is due to the challenges and the difficulties in preparing the Arabic training database.

3.2 Heart rate-speech emotion recognition performance

This subsection discusses the results of using speech signal and heart rate, note that this work could not test the German dataset with heart rate because the German dataset does not contain the heart rate measurement.

After combining the results obtained from the two neural networks (excited and calm categories) for the male, the accuracy rate has improved from 91.3% to 97.67% (+6.37). While the improvement in the performance of the test data, the accuracy rate have improved from 53.3% to 80% (+26.7%). And for the female, after combining the results obtained from the two neural networks (excited and calm categories), the accuracy rate has improved from 92.6% to 95.31% (+2.71). While the improvement in the performance of the test data, the accuracy rate have improved from 64.2% to 77.3% (+13.1%). The most important progress achieved from constructing two networks, excited and calm, is preventing the misclassification between the excited emotion (angry and happy) and calm emotion (sad and normal). Furthermore, the misclassification between the excited emotions (angry and happy) themselves and calm emotions (sad and normal) themselves are reduced. **Table 7** shows a summary for the results and the accuracy for both males and females after using the heart rate.

Table 4. Utilized Arabic dataset for the training process.

	Angry	Happy	Sad	Normal	Total
Male	43	41	44	44	172
Female	64	64	64	64	256
Total	107	105	108	108	428

Table 5. Utilized Arabic dataset for the test.

	Angry	Happy	Sad	Normal	Total
Male	5	5	10	10	30
Female	12	15	13	13	53
Total	17	20	23	23	83

Table 6. The accuracy of Arabic database.

	Training results	Testing results
Male	91.3%	53.3%
Female	92.6%	64.2%
Average	91.95%	58.75%

Table 7. The accuracy of Arabic database after using heart rate.

	Training results	Test results
Male	97.67%	80.8%
Female	95.31%	77.3%
Average	96.49%	79.05%

4. CHALLENGES

Speech emotion recognition is a challenging subject, partly due to the training speech data that is produced by actors or normal people to express emotional utterance. Consequently, the training speech data likely do not reflect the exact real human emotion. Also, the features extracted from the speech signal are unclear and not standardized in the literature of what features are effective for the task. The first four points below discuss the main problems in each stage of the emotion speech recognition while the last one presents the issues in using the heart rate:

1- Emotional modeling: to describe the connections between the speech and emotions, the right methods in expressing the emotion states that speech conveys must be used. The main issue that arises in the emotion modeling is what are and how many emotions that will be recognized, there are emotions that are not clear or hard to be recognized such as boredom and disgust, other emotions might conflict with each other such as surprise vs. happy, sadness vs. disappoint.

2- Database construction: although there are many databases that are available in emotion speech recognition that can be used, however not all of them are available for the public or they might have copyright issues, but that doesn't mean it's easier to build your own data database, building a database faces problems even more than using a well-known database, and these problems are:

- The first problem that arises in building a database is the quality of the recorded samples, using samples with low quality will lead to feature distortion and classifier degradation, the results that are obtained from the classification process will have a high error rate.
- Using spontaneous or acted speech: For spontaneous speech, it's not easy to obtain a large number of samples. In certain emotions, besides the emotions that result from this speech is not clear, i.e. they are a mix of different emotions, there is also the ethical issues since spontaneous speech means getting a speech from the subjects without their knowledge. On the other hand, acted speech is more favorable to use because of the reasons mentioned above. However, the acted speech doesn't simulate real-life situation as spontaneous speech does besides it requires using very professional actors to deliver the right emotions.

- Feature selection and extraction: what is the most suitable group of features that should be used? This question itself is one of the major factors that affect the emotional speech recognition, finding the set of features that give enough information to the classifier to be able to separate between the emotions is a challenging matter in this area of research.

3- Classifier selection and application: the selection of the classifier for the speech emotion recognition is not related to fixed criterion. There are many classifiers, and each one has its issues in implantation, the other issue in classification is a diversity of emotions, and the number of speakers in training, the more distribution of emotions and the more speakers in the database means good classification results.

4- As it is mentioned earlier, the heart rate data obtained from the smartwatch can assist the emotion speech recognition in achieving better results. However, it wasn't easy to obtain the right heart rate data in certain emotion from the subjects while they are reading the given sentences.

5. CONCLUSIONS

The speech emotion recognition is a system surrounded by several obstacles and hindrances, which restrain the researchers and prevent them from accomplishing perfect results, so it is a challenging task to manage. It is noticed that different set of emotion speech features can produce different emotion recognition ratio. Also, the effectiveness of certain speech emotion features is different in languages and gender. Choosing efficient training dataset is a very important factor for feature extraction and training of the neural network. Given well-known training dataset then no matter how the speech emotion features are chosen and extracted, there will be some misclassifications between emotions. That is why this work incorporated the heart rate measurements to direct the speech emotion features to the proper neural network. In this manner, the misclassification is reduced, and the overall performance is enhanced. It is also observed that separating the neural networks based on gender, into male and female networks, improves the recognition rate. It is proven that when the number of emotions, to be identified, increase then the amount of complexity of the emotion recognition system also increase and becomes hard for the classifier to distinguish between the emotions.

6. FUTURE WORK

The complexity of the neural network depends on the extracted features. Then, it is preferable to use systematic, automatic feature selection task to select proper and most relevant features and eliminates redundant or irrelevant features. Another critical factor in speech emotion recognition is appropriately preparing the training dataset, which must reflect the reality of human emotion. Hence, building a realistic database (non–acted speech) can improve the recognition rate to fit the real-life situation. In this project, only four emotions are studied; in the future, more emotions can be included, such as surprise, fear, disgust, boredom, etc. to increase human recognition categories. Since it is possible to estimate the heart rate from the speech signal **Ryskaliyev, et al., 2016**, in the future the heart rate data measured from the speech signal can be used instead of using the smartwatch. Improving the quality of the training set and using a broader training set can contribute to enhancing the recognition rate. It is more convenient, flexible, and practical to convert the emoting recognition system from stand-alone application to the program that runs in the background continuously to monitor the patients all the time.

REFERENCES

- Ahmed, M.Y., Kenkeremath, S., Stankovic, J., 2015, *SocialSense: A Collaborative Mobile Platform for Speaker and Mood Identification,* European Conference on Wireless Sensor Networks, Vol. 8965, PP. 68-83.

- Anne, K., Kuchibhotla, S., Vankayalapati, H. D., 2016, *Acoustic Modeling for Emotion Recognition*, International Journal of Speech Technology, Vol. 19, No. 4, PP. 7-15.

- Bahargab, M. and Talukdar, P. H. 2015, *Isolated Assamese Speech, and Speaker Recognition using Neural Network*, International Symposium on Advanced Computing and Communication (ISACC), PP. 135-183.

- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., Weiss, B., 2005, *A Database of German Emotional Speech*, In Proceedings of INTERSPEECH, Lisbon, Portugal.

- Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröer, M., 2000, *FEELTRACE: An Instrument for Recording Perceived Emotion in Real Time*, ISCA Workshop on Speech and Emotion, PP. 19-24.

- Dai, D., Liu, Q., Meng 12th, H., 2016, *Can Your Smartphone Detect Your Emotion?*, International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), PP. 1-8.

- Gargouri, D., Ali Kammoun, M., Ben Hamid, A., 2006, *A Comparative Study of Formant Frequencies Estimation Techniques*, Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, PP. 15-19.

- Gimpel, H., Regal, C., Schmidt, M., 2015, MyStress: *Unobtrusive Smartphone-Based Stress Detection*, European Conference on Information Systems, at Münster, Germany, Vol. 23.

- Joshi, A. and Kaur, R., 2013, *A Study of Speech Emotion Recognition Methods*, International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 2, PP. 28-31.

- Khulage, A.A. and Pathak, Prof. B. V., 2012, *Analysis of Speech under Stress using Linear Techniques and Non-Linear Techniques for Emotion Recognition System*, International Conference of Advanced Computer Science & Information Technology, PP. 1-10.

- Kumar Basu, J., Bhattacharyya, D., Kim, T. –H., 2010, *Use of Artificial Neural Network in Pattern Recognition*, International Journal of Software Engineering and Its Applications, Vol. 4, No. 2, PP. 23-34.

- Lee, H., Choi, Y. S., Lee, S., Park, I. P., 2012, *Towards Unobtrusive Emotion Recognition for Affective Social Communication*, IEEE Consumer Communications and Networking Conference (CCNC), PP. 260-264.
- Ménard, M., Richard, P., Hamdi, H., Daucé, B., Yamaguchi, T., 2015, *Emotion Recognition based on Heart Rate and Skin Conductance*, In Proceedings of the 2nd International Conference on Physiological Computing Systems (PhyCS-2015), PP. 26-32.

- Muda, L., Begam, M., Elamvazuthi, I., 2010, *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*, Journal Of Computing, Vol. 2, PP. 138 -143

- Philippou-Hubner, D., Vitalievich Vlasenko, B., Böck, R., Wendemuth, A., 2012, The *Performance of the Speaking Rate Parameter in Emotion Recognition from Speech*, International Conference Multimedia and Expo (ICME), IEEE, PP. 296-301.

- Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., Aucinas, A., 2010, *EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research*, In Proceedings of the 12th ACM International Conference on Ubiquitous Computing, PP. 281-290.

- Rong, J., Li, G., Phoebe Chen, Y. -P., 2008, *Acoustic Feature Selection for Automatic Emotion Recognition from Speech*, In School of Engineering and Information Technology, Deakin University, 221 Burwood Highway, Melbourne, VIC 3125, Australia, Vol. 45, PP. 315-328.

- Ryskaliyev, A., Askaruly, S., Pappachen James, A., 2016, *Speech Signal Analysis for the Estimation of Heart Rates Under Different Emotional States*, In International Conference on Advances in Computing Communications and Informatics (ICACCI).

- Sadat, M., Hossain, R. B., Mahmud, H., 2014, *Recognizing Human Affection: Smartphone Perspective*, Global Journal of Computer Science and Technology, Islamic University of Technology, Bangladesh, Vol. 14.

- "*Science Daily*, "[Online]. Available: https://www.sciencedaily.com/releases/2016/09/160921093924.htm. [Accessed 3 9 2018].

- Siddiqi, M. H., Alruwaili, M., Bang, J., Lee, S., 2017, *Real Time Human Facial Expression Recognition System using Smartphone*, In International Journal of Computer Science and Network Security(IJCSNS), Vol.17, No.10, PP. 223-230.

- Silva, C., Sobral, A., Vieira, R. T., 2014, *An Automatic Facial Expression Recognition System Evaluated by Different Classifiers*, In X Workshop de Visão Computacional, at Uberlândia, Minas Gerais, Brazil, PP. 208–212.

- Suk, M., and Prabhakaran, B., 2014, *Real-time Mobile Facial Expression Recognition System - A Case Study*, In Computer Vision and Pattern Recognition Workshops (CVPRW), PP.132–137.

- Tarng, W., Chen, Y.-Y., Li, C.-L., Hsie, K.-R., Chen, M., 2010, *Applications of Support Vector Machines on Smart Phone Systems for Emotional Speech Recognition*, World Academy of Science, Engineering and Technology, Vol. 4, No. 12.

- Utane, A. S., and Nalbalwar, S. L., 2013, *Emotion Recognition through Speech*, In IJAIS Proceedings on 2nd National Conference on Innovative Paradigms in Engineering and Technology (NCIPET 2013), PP. 5-8.

- Ververidis, D. and Kotropoulos, C., 2003, *A Review of Emotional Speech Databases*, In Panhellenic Conference on informatics (PCI), PP. 560–574.