International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 6, No.3 (May-2017)
http://dx.doi.org/10.12785/ijcds/060303

On the Performance Degradation of Speaker Recognition System due to Variation in Speech Characteristics Caused by Physiological Changes

Mohammed Usman[1]

[1] *Department of Electrical Engineering, King Khalid University, Abha, Saudi Arabia*

Received 22 Feb.2017, Revised 6 Apr. 2017, Accepted 20 Apr. 2017, Published 1 May 2017

Abstract: Speaker recognition is the process of identifying a person using their speech characteristics (voice biometrics). Speech characteristics of an individual can vary due to physiological changes which may be caused by health changes, physical activity as well as emotional changes. Such changes in speech characteristics are likely to affect the accuracy of speaker recognition systems. In this paper, the performance degradation of a speaker recognition system is quantified, empirically, when the characteristics of an individual's speech change due to physiological changes caused by *'physical activity'*. The speaker recognition system used in this work is based on Mel-Frequency Cepstrum Coefficients (MFCC's) and Vector Quantization (VQ). When the speech sample of a user is obtained soon after high intensity physical activity, the changes in the individual's speech characteristics affect the accuracy of speaker recognition systems. It is necessary to understand how speaker recognition systems are affected by changes in speech characteristics in order to improve their immunity to such changes. From speech recorded after physical activity, it is found that the duration of 'voiced component' which has prominent discriminative characteristics of speech is shortened and it has an effect on the accuracy of speaker recognition system.

Keywords: Speaker recognition, Voice biometrics, MFCC, Vector quantization, Speech physiology

1. INTRODUCTION

Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style). Speaker verification has earned speaker recognition its classification as a '*behavioral biometric*' [1]. Speaker recognition finds applications in a wide range of continually growing applications. Speech is a natural biometric feature that is transportable and available as a means of verifying a claimed identity. Typically, speaker recognition technology involves identifying a speaker from a database of known speakers. A broad classification of speaker recognition systems is made in [2] as: unconstrained mode and constrained mode. In unconstrained mode, the speaker is allowed to speak any phrase (text independent), whereas, in constrained mode, the speaker is restricted to speaking a predefined, fixed phrase (text dependent).

There are several techniques available in the literature for speaker recognition such as 'dynamic time warping (DTW), Hidden Markov Models (HMM's), MFCC's, vector quantization (VQ) and Gaussian mixture models (GMM's), each having varying degrees of complexity and accuracy. Various factors, such as ambient noise, variation in hardware used during training and testing phases or even the health of the individual affect the accuracy of a speaker recognition system. In general, the more the number of constraints imposed on a speaker recognition system, the better is its accuracy [2]. Physiological as well as emotional changes in an individual result in variations in the speech produced [3] and can affect the accuracy of a speaker recognition system. It is necessary to understand these changes in order to develop speaker recognition systems which can have better immunity to such changes. While it is known that physiological changes affect the speech production process in individuals, the effect of physiological changes on the actual speech parameters remains largely unknown [4]. Hence the effect on speaker recognition systems also remains a topic of investigation and is pursued in this

E-mail:omfarooq@kku.edu.sa, musman@ieee.org

work. In [5], the authors have investigated the possibility of detecting exercise intensity level based on speech samples obtained during exercise. The work in [6] reports significant alteration of speech parameters, from a linguistic viewpoint, due to physical activity. There is published work [7-9] in the literature which has investigated the detection of physiological parameters such as heart rate, skin conductance and heartbeat pattern, based on human speech. In [10-11], the performance of speech recognition systems in the presence of noise has been presented. The rest of the paper is organized as follows: motivation for the work is explained in section 2, MFCC processing is described in section 3 and the principle of vector quantization is described in section 4. In section 5, the results of speaker recognition system based only on MFCC's are discussed in terms of the probability of false alarm and probability of false positive. The pros and cons of choosing a decision threshold either too large or too small are also presented. In section 6, the performance of speaker recognition system based on MFCC's with VQ is presented. The performance degradation of the speaker recognition system due to variations in speech characteristics induced by physical activity is discussed in section 7. Conclusions and future work are presented in section 8.

2. MOTIVATION

To the best of the author's knowledge, there is no available literature which reports the variation of speech parameters, from a signal processing viewpoint, as a result of physiological changes caused due to physical activity. Speaker recognition systems rely on signal processing techniques and therefore, there is not much information in the literature as to how physiological changes affect the performance of speaker recognition systems. The work presented in this paper shall provide a start to fill this void and is directed towards the development of speaker recognition systems which can have better immunity to physiological changes. In particular, the effect of certain physiological changes on the performance of a speaker recognition system based on MFCC's and vector quantization is presented. MFCC with VQ has been chosen due to its relatively low implementation complexity while having reasonable accuracy. Alternative features such as spectral sub-band centroids (SSC's) have been proposed in the literature [12, 13] but MFCC's outperform SSC's and therefore make a better choice [14]. The results presented in this paper will contribute to the understanding of how speaker recognition systems are affected due to certain physiological changes in individuals and in turn provide insights into developing speaker recognition systems which have better immunity to such changes. While it is difficult to develop speaker recognition system that is 100% accurate, the main advantage of using speech for verification is that it uses a signal that is natural and can be easily obtained from anywhere in the world over the telecommunication network.

3. MEL FREQUENCY CEPSTRAL COEFFICIENTS

The first step in speaker recognition system is to extract the features, i.e. to identify the components of the audio signal that are good for identifying the speaker and discarding all the unnecessary elements. MFCC's are a versatile tool in speech and audio signal processing. Speech produced by an individual depends on the shape of the individual's vocal tract and oral cavity. If the shape of an individual's vocal tract can be determined accurately, it would allow accurate representation of the sound being produced. In signal processing terms, the shape of the vocal tract is represented by the envelope of the short time power spectrum and MFCC is a tool used to accurately represent this envelope. Introduced in the 1980's by Davis and Mermelstein [15], MFCC's are widely used in automatic speaker recognition as well as speech recognition and have been state-of-the-art ever since. Figure 1 shows the steps involved in computing MFCC's.

Audio signals are semi-stationary, i.e. their statistical properties do not change much over short time intervals, typically 20 – 25 ms. The recorded speech signal is split into short frames using a Hamming window which is defined as follows:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) \quad 0 \le n \le N \qquad (1)$$

In this paper, a Hamming window of length 256 samples has been used. Since the speech signal is recorded at a sampling rate of 12500 samples per second, a window length of 256 samples corresponds to 20 ms speech frames. A 256 point Fast Fourier Transform (FFT) has been used to obtain the magnitude spectrum of the speech signal.

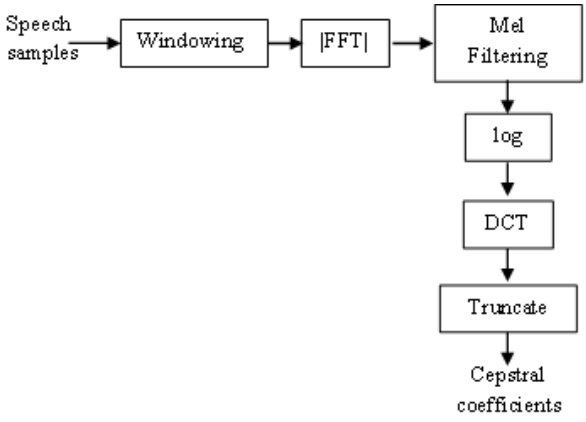

Figure 1. Computation of MFCC's

Mel filtering process is used to mimic the behaviour of human auditory system which detects different frequencies present in the signal. The human ear cannot differentiate two frequency components which are close to each other. The effect becomes more significant as the frequency increases. Therefore, rather than focusing on individual frequencies, the energy present in different frequency intervals is computed. The frequency intervals are kept smaller at lower frequencies and made increasingly wider as frequency increases. The perception of human auditory system varies linearly with frequency at lower frequencies up to about 1 kHz and logarithmically for higher frequencies [16]. A plot of the Mel filter-bank which creates frequency intervals of varying width is shown in figure 2 in which 10 filters are shown. In the actual implementation of this work, Mel filter-bank with 20 filters has been used. However, only 10 are shown in figure 2 for the sake of clarity.

The relation between Mel frequency scale and linear frequency scale is given by equations 2 and 3 and plotted in figure 3.

$$f_{mel} = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (2)$$

$$f = 700\left(exp\left(\frac{f_{mel}}{1125}\right) - 1\right) \quad (3)$$

After the Mel filtering stage, the logarithm of the mel-filtered components is computed. This gives the total energy *Y(i)* in each band of the Mel filter-bank and is represented as

$$Y(i) = \sum_{k=0}^{N/2} \log|X(n)|\, H_i\left(\frac{2\pi k}{N'}\right); i = 1,2 \ldots 20 \quad (4)$$

where, *X(n)* are the DFT coefficients obtained using FFT, $H_i(.)$ is Mel-filter band at coefficient '*i*', *N* is the frame length and *N'* is the DFT size. In this work, *N* and *N'* are both equal to 256.

The inverse DFT (IDFT) of the logarithm of the power spectrum is called as 'cepstrum'. The cepstrum may be real or complex. If only the real part is used, the real cepstrum is obtained which becomes similar to the discrete cosine transform (DCT). The cepstral coefficients are obtained as

$$c(n) = \frac{2}{N'}\sum_{i=1}^{20} Y(i) cos\left(\frac{2\pi in}{N'}\right) \quad (5)$$

After the DCT stage, the result is the Mel frequency cepstrum coefficients (MFCC's). The higher DCT coefficients represent fast changes and degrade speaker recognition performance. Therefore truncation is done to remove the higher DCT coefficients, which improves performance [17].

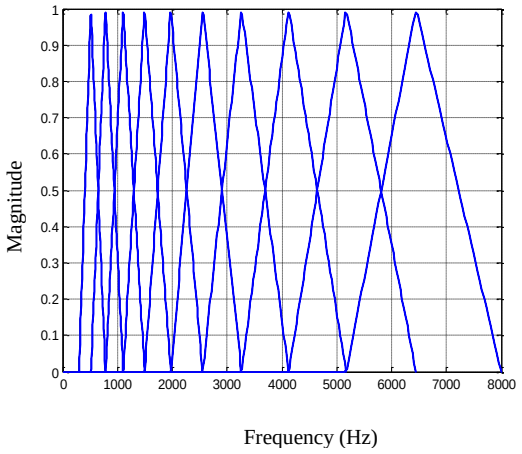

Figure 2. Mel filterbank

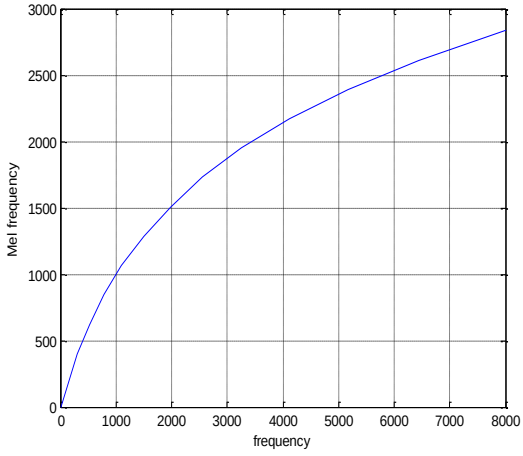

Figure 3. Frequency vs Mel frequency

4. VECTOR QUANTIZATION

Vector quantization is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its centre called a codeword or centroid. There are many algorithms to implement this clustering process. In this work, the Linde-Buzo-Gray (LBG) algorithm has been used for VQ. The LBG algorithm is a recursive algorithm which starts with a single cluster having a single centroid, splits it into two clusters, each with its own centroid. The splitting process continues until the desired number of clusters is obtained [18]. The collection of all code words is called a codebook. In [19], the authors compare different algorithms of codebook generation and conclude that the method of codebook generation is not so important. What is more important is the codebook size as it has a direct effect on the computational complexity. The distance from a vector to the closest codeword of a codebook is called a VQ-distortion and is computed as the Euclidian distance

between the vector and the codeword. The Euclidean distance between two vectors $p = [p_1, p_2, p_3 \ldots\ldots\ldots p_n]$ and $q = [q_1, q_2, q_3 \ldots\ldots\ldots q_n]$ is defined as

$$d(p,q) = \sqrt{\sum_{j=1}^{n} (q_j - p_j)^2} \qquad (6)$$

During speaker recognition, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the total VQ distortion is computed. The speaker corresponding to the VQ codebook with smallest total distortion is identified [20]. A detailed analysis of VQ and its implementation is given in [21].

5. SPEAKER RECOGNITION USING MFCC

Speaker recognition process has two phases: training phase (registration) and testing phase (recognition). During the training phase, users are registered to the database. A sample of each user's speech is recorded, its features extracted and a model is created and stored in the database. During the testing phase, any one of the registered users is taken as a test case. A sample of speech is obtained from the test user and as in the training phase, its features and model is created. The model of the test user is compared with all the models in the database and the nearest match is identified. The system is tested in two different scenarios. In the first scenario, speech from only one user is registered and its model is constructed by generating its cepstral coefficients as described in section 3. The same user is then used as 'test user' several times. Since the speech signal during the testing and training phases comes from the same user, the system should detect a perfect match. Comparison is made between the 'training phase' model and 'testing phase' model by computing the mean squared error (MSE) between training and testing phase model values. MSE is a common metric in the literature to compute the match score between the training and testing samples [22]. A better metric for computing match score in speaker recognition systems is a topic of on-going research [23]. MSE is computed according to the following expression.

$$MSE = \frac{(Y_{training} - Y_{testing})^2}{N} \qquad (7)$$

$Y_{training}$ represent the model parameter values in the training vector and $Y_{testing}$ represent the model parameter values in the testing vector. 'N' represents the number of elements in the training and testing vectors. The computed MSE value is compared with a predefined threshold 'τ' and a decision is made according to the decision hypothesis given by equation 8, in which H_0 is the hypothesis that the test sample belongs to the valid user and H_1 is the hypothesis that it belongs to an impostor.

$$\begin{aligned} MSE \leq \tau;\ & H_0 \\ MSE > \tau;\ & H_1 \end{aligned} \qquad (8)$$

It is obvious that the threshold value will affect the accuracy of the system. A small value of threshold will likely 'reject' a valid user whereas a large value of threshold may 'accept' an invalid user. The threshold value has to be chosen according to the nature of the application. For example, when using speaker recognition for banking transaction authentication, it is a serious error if an impostor is accepted as a valid person. If a valid user is rejected, he/she may try again. This type of error is called 'False Reject' or 'False Alarm'.

The performance of the system is investigated for different values of τ. It is found that the accuracy of the system varies across individuals. Table I lists the probability of false alarm of the system for 3 different MSE threshold values and for 3 different users. It is observed that for the same MSE threshold, the probability of false alarm and hence the system accuracy varies across individuals. As the training samples and test samples are collected at different times, there is some variation in the training and testing samples of the same individual which leads to a false alarm. This variation could be attributed to the individual's health, emotion as well as ambient conditions. Since the speech samples are recorded in this work under similar ambient conditions, the variations are attributed to physiological changes. It is also observed that to improve the accuracy of the system, a larger value for MSE threshold is desired. However, large value of threshold is likely to increase the probability of false positive, i.e. an impostor being accepted as a valid user. The probability of false positive is presented in table II. A large value for MSE threshold reduces the probability of false alarm but it also increases the probability of false positive.

TABLE I. PROBABILITY OF FALSE ALARM

User	Mean Squared Error threshold	Probability of false alarm
A	0.5	0.53
	1.5	0.12
	2.0	0.03
B	0.5	0.72
	1.5	0.15
	2.0	0.08
C	0.5	0.84
	1.5	0.23
	2.0	0.04

TABLE II. PROBABILITY OF FALSE POSITIVE

User	Mean Squared Error threshold	Probability of false positive
Two different users for the training and testing sessions	0.5	0.051
	1.5	0.35
	2.0	0.53

False positive and false alarm both constitute error. From tables I and II, it can be deduced that if the objective is to reduce the probability of 'false alarm', then a larger threshold value such as 2 is desired and if the objective is to reduce the probability of 'false positive', then a smaller threshold value such as 0.5 is desired. In order to improve the overall accuracy of template based speaker recognition systems such as MFCC, there are conflicting requirements for minimizing false alarm and false positive. Hence, some trade-off is involved. For many speaker recognition applications which involve authentication and authorization, it is desirable to minimize the probability of false positive.

6. SPEAKER RECOGNITION USING MFCC WITH VECTOR QUANTISATION

In the next scenario, a database of registered users is created by recording speech samples of users and creating a VQ model for each user. An arbitrary user from among the registered users is then taken as a test user and speaker identification is performed by computing the Euclidean distance, as defined in equation 6, between the test user's VQ code vector and all the code vectors of the training database. The training database consists of speech samples from 30 different users and the results have been obtained using the same 4 '*test*' users from among the 30 users registered in the training database. The number of 'test' users is limited to four in this work due to difficulty in subscribing volunteers who can perform physical activity at the desired intensity level, on a regular basis, over a period of several months. The present work shall be improved by studying more users in future. The training vector having the smallest Euclidean distance with the test vector is chosen as a match. In this work, the VQ codebook is generated using 16 centroids.

It is known from available literature [1] that the accuracy of VQ improves with more number of centroids albeit at the cost of increased computational complexity. However, increasing the number of centroids beyond a certain point leads to diminishing returns. The experimental results for speaker recognition based on vector quantization are shown in table III, which shows the VQ distortion between the test samples and training samples of the 4 test users only, for the sake of brevity. VQ-distortion is smallest along the diagonal in the table, which corresponds to a match between the test and training samples. The speaker recognition test for each user is performed 100 times over a period of around 6 months. The identification rate for the four users is shown in table IV. The identification rate of the system closely matches with that of similar systems reported in the literature [1], thus validating the system described in this paper.

7. PHYSIOLOGICAL CHANGES INDUCED BY PHYSICAL ACTIVITY

As mentioned in section 1, speech characteristics of an individual can vary due to physiological and emotional changes. The main contribution of this work is the effect physiological changes, caused by physical activity, have on the performance of a speaker recognition system. Physiological changes in the human body may be caused by many factors such as ageing, pollution, stress, health as well as physical exercise to name a few.

TABLE III. VQ DISTORTION BETWEEN TEST AND TRAINING SAMPLES

	User 1	User 2	User 3	User 4
User 1	**6.7166**	7.7412	8.1541	7.2197
User 2	4.4516	**4.1148**	5.5430	4.9358
User 3	5.7477	6.9120	**4.3346**	5.8135
User 4	5.4427	6.6022	5.9109	**3.6582**

TABLE IV. IDENTIFICATION RATE OF THE SYSTEM

User	Identification rate (%)
User 1	94
User 2	93
User 3	96
User 4	94
Average identification rate	**94.25**

In this work, the performance of a speaker recognition system due to physiological changes caused by physical exercise is investigated. The physiological response to exercise depends on the intensity, duration and frequency of exercise as well as environmental conditions [24]. The four test users performed physical exercise at an intensity level that takes their heart rate to 125-135 beats per minute. This corresponds to intensity level of 65-70% for individuals in the 25-30 years age group [25] which is the age group of the individuals who took part in this study. Speech samples were recorded immediately after the workout session which was done 4-5 times a week and applied to the speaker recognition system. One hundred such samples were collected for each of the four test users. Table V shows the identification rate of the speaker

recognition system with physiological changes induced by physical activity. It is clear that the identification rate of the speaker recognition system is lower in this scenario as compared to the identification rate using speech samples collected before physical exercise. The reason for this degradation can be established by observing how the speech production process is affected and hence the speech signal characteristics as a result of physical exercise. Speech signals consist of two components: voiced and unvoiced. Voiced components (such as vowels) are produced by the vibration of vocal cords whereas unvoiced components (such as 'ssss', 'shhh' called *fricatives*) are produced by shaping the vocal tract and directing airflow and stop sounds (such as '/k', '/t') are produced by abruptly stopping airflow. The discriminative features of an individual are known to be more prominent in voiced speech than in unvoiced speech [14]. It is a commonly observed fact that speaking becomes difficult during and soon after intense physical exercise. Speech production requires breathing and physical activity increases breathing rate for metabolic reasons. Hence, speech production is affected as a result of physical activity since breathing becomes a necessary action for both metabolic and speech production purposes simultaneously [26]. Since the process of speech production is affected, it is reasonable to conclude that some characteristics of the produced speech are also affected. These effects are subjective and depend on the fitness level of the individual. The accuracy of the speaker recognition system is found to be lower by 9%, on average, when the speech *test samples* are recorded immediately after physical exercise. While it is known that physical activity causes physiological changes which lead to variation in speech characteristics of individuals, there is not much work in the literature that quantifies such effects on speaker recognition systems. Temporal waveforms and the spectrograms of speech recorded 'before' and 'after' physical activity for one of the volunteer participating in the study are shown in figures 4 and 5, respectively. For the same sentence uttered by the same individual, there is a noticeable difference, both in the temporal waveform as well as spectrogram of the speech signals recorded 'before' and 'after' high intensity physical activity. In the speech signal recorded after physical activity, it is noticed that the duration of the voiced component of each word is shortened whereas the duration of fricatives is lengthened. There are also bursts of relatively higher amplitudes which are attributed to utterances made during heavy exhalation. Since the duration of 'voiced component' of speech, which has more prominent discriminative features is shortened, the accuracy of speaker recognition system is affected, in addition to other factors which need further investigation.

8. CONCLUSIONS AND FURTHER WORK

The characteristics of human speech change due to physiological changes. Due to this, the accuracy of speaker recognition systems is affected. In this work, the effect of changes in speech characteristics caused due to physiological changes induced by high intensity physical activity on the performance of speaker recognition systems based on MFCC and VQ have been presented. In particular, the performance degradation is quantified by testing the system extensively over a period of six months. It is found that the accuracy of speaker recognition system based on MFCC and VQ degrades by about 9%. A strong contributor for this degradation is the shortening of the prominent characteristic bearing 'voiced component' of speech. It is also found that the accuracy of the system has subtle variations across individuals, which can be due to the fitness level of each individual. While the present work involves only four test users, due to limitations described in section 6, it is desired to improve the work by conducting the study over a large number of test users. It should be noted that speech samples from popular speech corpora have not been used in this study due to specific requirements of physical activity which are not catered to in the available speech corpora. The methodology used in this work shall be used to study other speaker recognition techniques such as HMM and GMM.

TABLE V. IDENTIFICATION RATE OF THE SYSTEM WITH SPEECH SAMPLES TAKEN AFTER PHYSICAL ACTIVITY

User	**Identification rate (%)**
User 1	86
User 2	81
User 3	85
User 4	89
Average identification rate	**85.25**

By understanding the variations in speech characteristics and their effects on speaker recognition systems, it will be possible to develop techniques to make speaker recognition systems robust and more immune to such changes. Future work shall also include a more detailed analysis from a signal processing viewpoint and to establish correlation between physical activity, physiological parameters and speech signal / model / statistical parameters. All the training and testing samples used in this work have been taken from male volunteers. In future, it is planned to include speech samples from female volunteers as well.

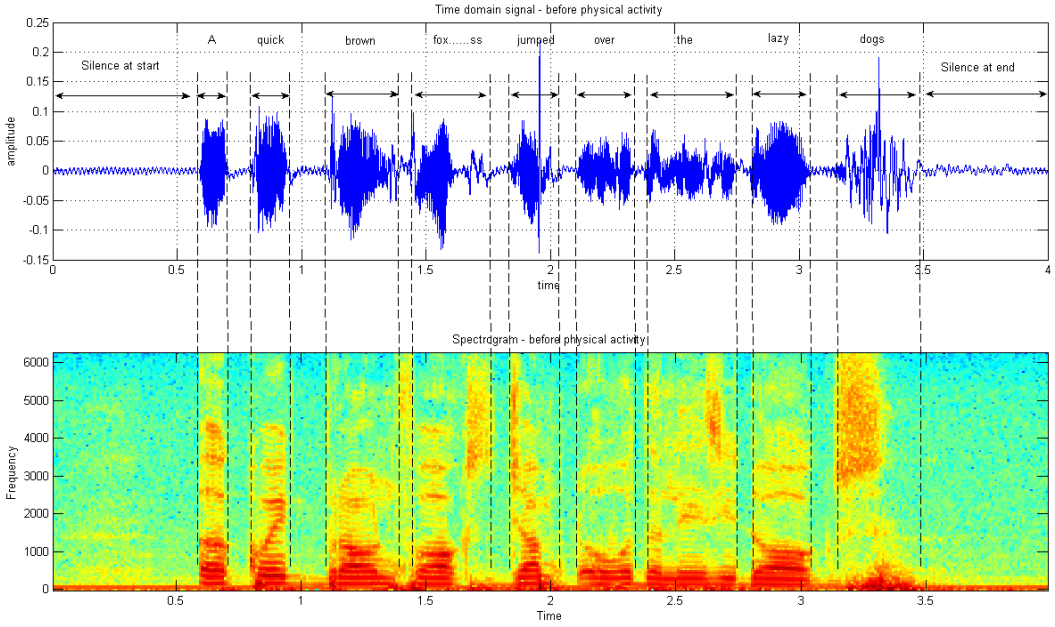

Figure 4. Time domain waveform and spectrogram – speech before physical activity

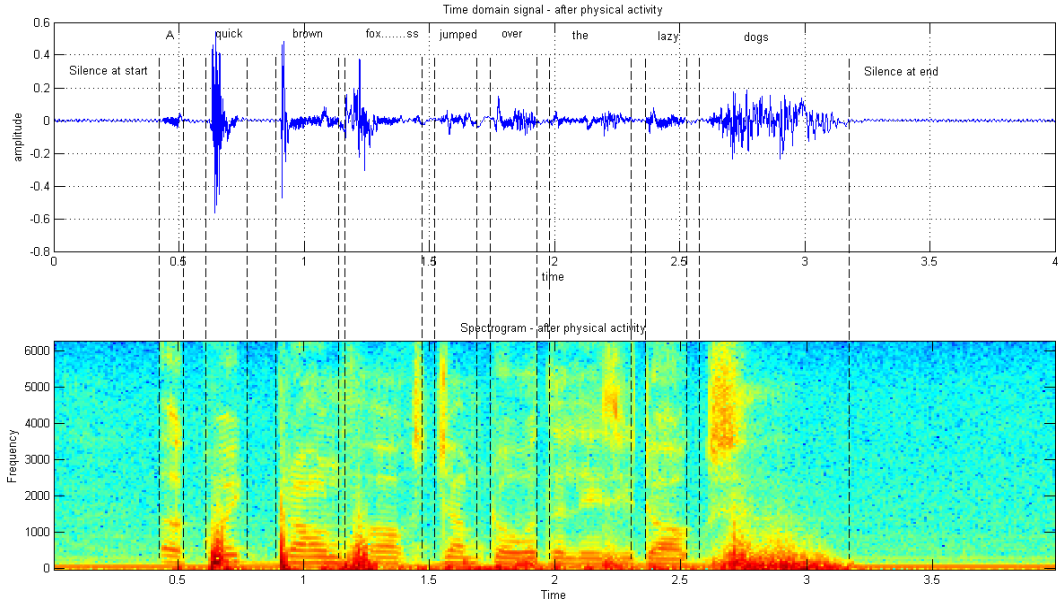

Figure 5. Time domain waveform and spectrogram – speech after physical activity

ACKNOWLEDGEMENT

The author expresses his gratitude to all volunteers who participated in this study, especially the 4 volunteers who performed physical activity at the desired intensity level and provided their speech samples.

REFERENCES

[1] G. Nijhawan, M.K. Soni, "Speaker recogntition using MFCC and vector quantisation," International Journal on Recent Trends in Engineering and Technology, vol. 11, no. 1, pp. 211–218, July 2014.

[2] D.A. Reynolds, "An overview of Automatic Speaker recognition technology," IEEE International conference on Acoustics, Speech and Signal Processing, vol. 4, pp.4072–4077. May 13-17, 2002, Orlando, Florida.

[3] Speech - The physiology of speech, http://science.jrank.org/pages/6371/Speech-physiology-speech.html, accessed January 2017.

[4] J. Trouvain, K.P. Truong, "Prosodic characteristics of read speech before and after treadmill running", 16th Annual Conference of the International Speech Communication Association , September 6-10 2015, Dresden, Germany

[5] K.P. Truong, A. Nieuwenhuys, P. Beek, V. Evers, "A database for analysis of speech under physical stress: detection of exercise intensity while running and talking, Sixteenth Annual Conference of the International Speech Communication Association., 2015.

[6] E.S. Baker, J.Hipp, A. Helaine, "Ventilation and speech characteristics during submaximal aerobic exercise", Journal of Speech, Language and Hearing research, vol. 51, issue 5, p1203, October 2008.

[7] B. Schuller, F. Friedmann, F. Eyben, "Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance", ICASSP 2013, pp. 7219 – 7223.

[8] A. Mesleh, D. Skopin, S. Baglikov, and A. Quteishat, "Heart rate extraction from vowel speech signals", Journal of Computer Science and Technology, vol. 27, no. 6, pp. 1243-1251, November 2012.

[9] D. Skopin, and S. Baglikov, "Heartbeat feature extraction from vowel speech signal using 2D spectrum representation," in Proc. 4th International Conference on Information Technology (ICIT), Amman, Jordan, June 2009.

[10] H.G. Hirsch, David Pearce, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions", Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000, Beijing, China, October 16-20, 2000

[11] A. Varga, H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems", Speech Communication, vol. 12, no. 3, pp. 247 – 252, 1993.

[12] N. Thian, Sanderson C., Bengio S., "Spectral subband centroids as complementary features for speaker authentication", International Conference on Biometric Authentication (ICBA 2004), Hong Kong, July 2004, pp. 631-639.

[13] T. Kinnunen, B. Zhang, J. Zhu, and Y. Wang, "Speaker verification with adaptive sub-band centroids", International Conference on Biometrics (ICB 2007), Aug 2007, Seoul, pp.58-66.

[14] T. Kinnunen,, H. Li, "An overview of text independent speaker recognition: from features to supervectors", Speech Communication, vol. 52, issue 1, pp.12-40, Jan 2010.

[15] S. Davis, P. Mermelstein, "Comparison of parametric representations for monosyllabic word representation in continuously spoken sentences," IEEE Transactions on Acoustics, Speech and Signal Processing. vol. 28, issue 4, pp. 357-366, Aug. 1980.

[16] J. Lyons, "MFCC Tutorial," http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/, accessed 20-April-2016..

[17] http://kom.aau.dk/group/04gr742/pdf/MFCC_worksheet.pdf, MFCC work sheet, Aalborg University.

[18] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Communications, vol. COM-28(1), pp. 84-96, Jan. 1980.

[19] T. Kinnunen, I. Sidoroff, M. Tuononen, P. Franti, "Comparison of clustering methods: a case study of text independent speaker modeling", Pattern Recognition Letters, vol. 32, issue 13, pp. 1604-1613, Elsevier, October 2011.

[20] L.R. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1st edition, 1993.

[21] Y. Linde, A. Buzo, R.M. Gray, "An algorithm for vector quantizer design," IEEE Transactions on Communications, vol.28, issue 1, pp. 84-95, Jan 1980.

[22] J. Saastamoinen, E. Karpov, V. Hautamäki and P. Fränti, "Accuracy of MFCC based speaker recognition in series 60 device", Journal on Advances in Signal Processing, EURASIP vol. 17, pp. 2816-2827, 2005

[23] C. Hanilci, F. Ertas, "Comparison of the impact of some Minkowski metrics on VQ/GMM based speaker recogntion", Journal of Computers and Electrical Engineering, Elsevier, vol. 37, issue 1, pp. 41-56, January 2011.

[24] D. A. Burton et al., Physiological effects of exercise, Oxford Journal of Continuing education in Anaesthesia, critical care and pain, vol4, No.6, 2004.

[25] American Heart Association (accessed February 2017)

http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/Target-Heart-Rates_UCM_434341_Article.jsp#.WJd2KPl97IU

[26] Y. Meckel, A. Rotstein, O. Inbar, "The effects of speech production on physiologic responses during sub-maximal exercise", Medicine and Science in Sports and Exercise, vol. 34, pp 1337-1343, 2002.

Mohammed Usman received PhD in 2008, M.Sc in 2003, both from the University of Strathclyde, Glasgow, United Kingdom and B.E in Electronics and Communication from Madras University, India in 2002. He has more than a decade of experience in academics & administration. He is a member of IEEE & IET. His research is focused on technologies for next generation wireless networks with specific interest in coding theory. As part of projects under his supervision, he has recently been involved in the research of speech and speaker recognition techniques. He has also been the Organizing Chair and TPC Chair of IEEE International conferences.