

Quality assessment on muscle locations for speech representation

Safie S.I, Rosuhana Rahim

Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Malaysia

Article Info

Article history:

Received Jun 9, 2019

Revised Aug 12, 2019

Accepted Aug 27, 2019

Keywords:

Electromyogram

Empirical mode decomposition

Kullback-Leibler Divergence

Muscle location

Total hilbert huang spectral entropy

ABSTRACT

There are more than 68 muscles, which are activated either simultaneously or sequentially during speech production. To monitor the signals from all these muscles at once, involve a lot of sensors and such system is very expensive. In the Quran therapeutic treatment applications, the use of specific muscles is very important, for the production of correct Arabic pronunciation. The proper pronunciation will improve the reader's understanding of what is being read, thus assisting the effectiveness of the therapy process. The objective of this study is to identify the most optimal muscle location, which is suitable for monitoring the quality of a recitation during the Quran's therapeutic process, based on the information content embedded in their Electromyogram (EMG) signals. Empirical Mode Decomposition (EMD) technique was used in this study to extract features of the EMG while the combination of Hilbert Huang Spectral Entropy (HHSE) and Kullback Leibler Divergence (KLD) techniques were used to quantify the information content in each feature. Combination of these techniques managed to rank ten widely used speech muscles in the literature based upon their information content. Four muscle locations have been suggested, which is believed to be sufficient in developing a low-cost self-assessment system for monitoring Quran recitation.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Safie S.I,

Malaysian Institute of Industrial Technology,

Universiti Kuala Lumpur,

1016, Jalan Sultan Ismail, Bandar Wawasan, 50250, Malaysia.

Email: sairulizwan@unikl.edu.my

1. INTRODUCTION

The Quran therapy is a psychological therapy, adopted by Muslims as a reliable alternative medicine [1]. It involves the recitation of Ruqyah (specific verses in the Quran) as a form of meditation to reduce stress and depression [2]. Ruqyah should be recited correctly, for that recitation (in Arabic) be translated into its true meaning, thus maximizing the therapy process for a patient [3-4]. Developing a Quran therapeutic system, which can automatically measure the quality of a reading, directly helps the patient rectifying its pronunciation.

Automatic Speech Recognition (ASR) specifically in Arabic has been extensively studied [5-6]. ASR works by translating a speech signal in the form of their contextual information. Ambient noise and user's speech impairment are two main challenges in developing a high-precision ASR system [7]. Nevertheless, in the Quran therapy applications, the desired system must not only be measured by its ability to recognize speech, but also the capability to assess the extent of the speech generated using the right anatomy parts of the human body. This led to the use of Electromyogram (EMG) as an alternative source for analyzing the recitation of Ruqyah, during the therapy.

Using EMG to recognize speech is not a new approach in ASR [8-9]. Researchers tend to use the term Silent Speech Recognition (SSR) when the acoustic speech signal being replaced by EMG. Several areas have been widely studied, which include:

- Specific applications; like in prosthesis control [10], oral motor treatment for speech production [11] and improve communication while subject wears a self-contained breathing apparatus (SCBA) [12].
- Feature extraction and classification techniques [13-16]
- Language or Dialect variation [17-19]
- Speech synthesis [20-22]

This study will look into another equally important factor in EMG-based speech recognition technology, namely the impact of muscles' selection towards the quality of speech. The speech production begins in the lungs as illustrated in Ture 1. Air which is exhaled by the lungs travels up through the trachea and into the larynx.

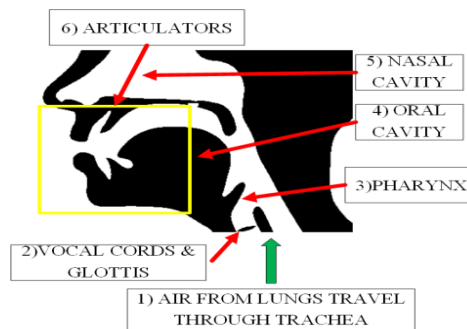


Figure 1. Speech production system

Within the larynx, there are two folds of ligament known as vocal cords. The gap between these vocal cords is called the glottis. During speech production, the vocal cords vibrate rapidly, which opens and closes the glottis. The continuous streams of air from the lungs are trimmed into periodic suction of air, and eventually generate an audible but in monotonous sound. The pharynx (throat), the oral cavity (the mouth) and the nasal cavity (the nose) then change the frequency components of the monotonous sound, which later produce different types of vowel sounds. The process of articulation will eventually reshape these sounds into a form that can be recognized by the listener, namely the speech. The articulators involves in reshaping the sounds include the lips, teeth, hard and soft palate, tongue, jaw, posterior pharyngeal wall and the inner edges of the vocal folds [23].

There are more than 68 muscles involved during speech production [24]. The frequencies and sequences of these muscles being activated are depending upon the type of pronunciation produced. To monitor all the muscles at the same time requires a large number of sensors. It is also not practical because of the cost for such systems is high.

From Figure 1, the production of speech pronunciation utilized most of the muscle located near the facial and neck. We summarized the muscle locations used for EMG based speech recognition in most literature as illustrated in Figure 2 [8-22].

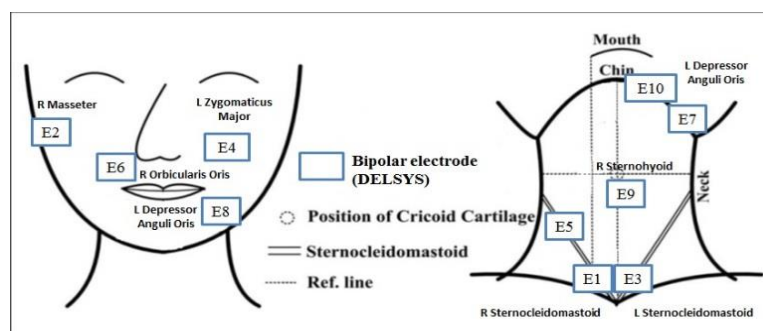


Figure 2. 10 widely used EMG location for speech recognition

The description of locations for E1 to E10 in Figure 2 is given as follows:

- (E1) 1cm to the right of the neck midline.
- (E2) 1cm to the left of the neck midline.
- (E3) Centered sternocleidomastoid at one-third from clavicle to mastoid for electrode.
- (E4) 1cm lateral to the ventral neck midline.
- (E5) 1cm lateral to submental midline.
- (E6) Below right corner of the mouth.
- (E7) Upper right corner of the mouth.
- (E8) Cantered on lateral jaw superficial to masseter muscle.
- (E9) 4cm lateral to submental midline.
- (E10) Between the left corner of the mouth and eye.

In this paper, we assess the signals from these locations in an attempt to identify the best locations that produce the highest context information, for speech recognition. Given that the level of muscle activity depends on the type of pronunciation, this study focused on muscle activity during Ruqyah recitations.

2. EMG MODELLING USING EMPIRICAL MODE DECOMPOSITION (EMD)

EMG signals from 10 muscle locations described in Section 1 will be mathematically modeled, measured and compared to assess the quality of their contextual information. In this work, Empirical Mode Decomposition (EMD) [25] will be used to model all signals while Total Hilbert-Huang Spectral Entropy H_j [26-27] and Kullback–Leibler Divergence (D_{KL}) [28] will be used for measuring the quality of information from these signals. The mathematical background of these techniques will be explained in Section 3.

Empirical Mode Decomposition (EMD) technique assumes that for any given data (or signal), at any given time, the signal can be decomposed into various coexisting simple natural oscillation called Intrinsic Mode Function (IMF), superimposing on the others. A cycle of the simple oscillating signal must have an equal number of maxima and minima peaks. However, its amplitude and frequency may vary as functions of time [29].

Consider an arbitrary discrete time signal $y(t)$, for the length of T , to be decomposed using EMD as shown in Figure 3. First, detect all local maxima, $L_{max}[n]$ of $y(t)$. Then connect these $L_{max}[n]$ by a cubic spline line to generate upper envelop, $y_{max}(t)$ for all t . Repeat the similar process for every local minima, $L_{min}[n]$ to generate lower envelop, $y_{min}(t)$ for all. The total number of local extrema (maxima and minima), n should be the same or differ at most by one.

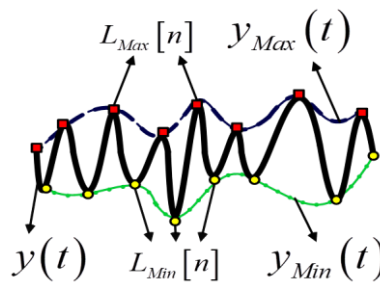


Figure 3. Empirical mode decomposition process of an arbitrary signal

The average signal, $m(t)$ between $y_{max}(t)$ and $y_{min}(t)$ is calculated as

$$m(t) = \frac{y_{Max}(t) + y_{Min}(t)}{2} \tag{1}$$

Using (1), obtain the difference, $d(t)$ between $y(t)$ and $m(t)$ using

$$d(t) = y(t) - m(t) \tag{2}$$

$m(t)$ is set to be the j -IMF, $c_j(t)$ while $d(t)$ is defined as the k -residue, $r_k(t)$ if $d(t)$ satisfy one of the following criteria [30]:

- a) $m(t)$ approaches zero
- b) The number of local extrema and the number of zero crossing of $d(t)$ differ at most one
- c) User defined maximum iteration is reached

If $d(t)$ failed to meet either of the above criteria, $y(t)$ is replaced by $d(t)$. Overall decomposition process is repeated until $c_j(t)$ is identified. This process is repeated for $j=1$ until $r_k(t)$ is either a constant, a monotonic slope, or a function with only one extremum [31]. Finally, $y(t)$ can be represented as

$$y(t) = \sum_{j=1}^k c_j(t) + r_k(t) \quad (3)$$

The quality or relevance of each $c_j(t)$ can be determined by their correlation coefficient, γ_j between $c_j(t)$ and $y(t)$ as follows:

$$\gamma_j = \frac{\int y(t)c_j(t)dt}{\sqrt{\int y^2(t)dt \int c_j^2(t)dt}} \quad (4)$$

Various types of features can be extracted from (3). The easiest features is the IMF energy, P_j which is defined as [32]:

$$P_j = \frac{\sum_{t=0}^T |c_j(t)|^2}{\sqrt{\sum_{j=1}^{k-1} \sum_{t=0}^T |c_j(t)|^2}} \quad (5)$$

Hilbert Transform (HT) can be used, to transform all $c_j(t)$ in (3) into their HT form, $h_j(t)$. Combination of $c_j(t)$ and $h_j(t)$ generates an analytical signal, $z(t)$ which can be expressed as

$$z_j(t) = c_j(t) + ih_j(t) = a_j(t)e^{i\int \omega_j(t)dt} \quad (6)$$

From (6), instantaneous amplitude, $a_j(t)$ and instantaneous frequency $f_j(t)$ can respectively calculated as follows:

$$a_j(t) = \sqrt{c_j^2(t) + h_j^2(t)} \quad (7)$$

$$f_j(t) = \frac{f_s}{2\pi} (\phi_j(t) - \phi_j(t-1)) \quad (8)$$

Where f_s =sampling frequency and $\phi_j(t) = \tan^{-1} \left(\frac{h_j(t)}{c_j(t)} \right)$.

Using (7) and (8), average amplitude, \bar{a}_j and average frequency, \bar{f}_j can be set as features. They can be respectively expressed as

$$\bar{a}_j = \frac{1}{T} \sum_{t=0}^T a_j(t) \quad (9)$$

$$\bar{f}_j = \frac{\sum_{t=0}^T a_j(t) \omega_j(t)}{\sum_{t=0}^T a_j(t)} \quad (10)$$

The frequency-time distribution of the amplitude for $z_j(t)$, designated as the Hilbert amplitude spectrum, $H_j(\omega, t)$ now can be obtained. Two types of features can be extracted from $H_j(\omega, t)$, namely marginal spectrum, $h_j(\omega)$ and instantaneous energy density level, $IE_j(t)$. They can respectively be expressed as follows:

$$h_j(\omega) = \frac{1}{T} \int_0^T H_j(\omega, t) dt \tag{11}$$

$$IE_j(t) = \int_0^{\omega_n} H_j^2(\omega, t) d\omega \tag{12}$$

The objective of this work is to inspect the quality of EMG signal, generated from various muscle locations, that is best to represent the recitation of Ruqyah. A signal is said to have more information, if the signal is closer to random and embeds a large amount of uncertainty. Measuring the uncertainty of a signal reflects the quality of that signal. Entropy is a measurement of randomness and represents the amount of energy carried by a signal. By applying the Shannon entropy concept to the $H_j(\omega, t)$, a new type of entropy was obtained and denoted as Hilbert–Huang spectral entropy, $\hat{h}_j(f)$. To simplify the representation, the $h_j(\omega)$ is written as a function of frequency (f) instead of angle frequency (ω) and normalized as follows [33]:

$$\hat{h}_j(f) = \frac{h_j(f)}{\sum_f h_j(f)} \tag{13}$$

Then, the total entropy H_j can be calculated as follows:

$$H_j = -\sum_f \hat{h}_j(f) \log(\hat{h}_j(f)) \tag{14}$$

A higher value of H_j represents a higher uncertainty of a signal.

Let P_{H_j} and Q_{H_j} represent two different probability distribution of H_j . The relative entropy of P_{H_j} with respect to Q_{H_j} can be defined as

$$D_{KL}(P_{H_j} \parallel Q_{H_j}) = \sum P_{H_j} \log \frac{P_{H_j}}{Q_{H_j}} \tag{15}$$

$D_{KL}(P_{H_j} \parallel Q_{H_j})$ is also known as Kullback–Leibler divergence from Q_{H_j} and P_{H_j} and measure the amount of information lost when Q_{H_j} is used to approximate P_{H_j} .

3. METHODOLOGY

In Islamic complementary medicine, there are many surah (or chapters) and verses from the Quran that can be used as Ruqyah. Among them is verses 1 to 7 from surah ‘Al-Fatihah’, verses 1 to 5 from surah ‘Al-Baqarah’ and verses 18 to 19 from surah ‘Al-Imran’. Some of these verses are too long, thus in our experiment, some verses are divided into smaller recordings. As a result, 17 short verses were generated, to be used as Ruqyah and recited by all participants.

In this study, a total of seven participants were selected among university students, aged between 18 and 25 years old. These participants are required to have the ability to read the verses of the Quran properly and smoothly. To ensure that these conditions are met, participants are required to read these verses to clerics. After the clerics satisfied with the recitation, the recording process began.

Two physiological signals namely, electromyogram (EMG) and speech are recorded. The speech signal is recorded using a built-in microphone of a Dell LATITUDE-E6540 laptop. Delsys Bagnoli 4-channel is used to record the EMG signals [34].

All participants will recite the selected Ruqyah while their speech and EMG are recorded. Four EMG locations are recorded simultaneously at a time. Each verse is recorded for five times. To ensure that electrode variabilities are also taken into consideration, each muscle location in Figure 2 will be repeatedly recorded using all four Delsys Bagnoli electrodes. By the end of the recording process, total of 2380 EMG and 595 speech recordings were created.

All signals then undergo pre-processing steps. First, the DC components of the EMG and Speech signals are removed. A 50Hz notch filter is applied to both signals to ensure that they are free from the power-line interference. All EMG signals are then examined in the time domain to identify any outliers. In this work, the boundary between peaks of each signal is set to ± 3 times standard deviation of overall peak-to-peak amplitudes. Any amplitudes beyond these boundaries are considered as outliers and removed. Next, bandpass filter is applied on them. The bandpass ranges for both EMG and speech signals, are from 10 to 500 Hz and from 50 to 3000 Hz respectively. Both signals are then resampled to 8000 Hz. Finally, all signals are normalized so that the amplitude between the peaks is equal to 1 [35]. All clean signals are then processed for the measurement of the H_j , calculated using (14) as described in the previous section.

4. RESULTS AND DISCUSSION

Using (14), there are two types of measurement indicators that can be used to identify the best muscle location during the recitation of Ruqyah. The first indicator is the amount of information embedded in an EMG signal. In representing the recitation of Ruqyah, a muscle is a better selection than the others, if it generates an EMG signal which embeds higher information, in this case, higher entropy value of H_j .

When an EMG signal being decomposed into k -IMF signals, it generates k number of entropy, H_j . Some IMF embeds more information compared with the others. Let the total information, TH generated from a muscle to be the sum of entropy H_j of all k -IMF calculated from the decomposed EMG signal. Mathematically, it is describe as follows:

$$TH = \sum_{j=1}^k H_j \quad (16)$$

Figure 4 illustrates the normal distribution of total information, TH calculated using (16) for 10 muscle locations. By observing Figure 4, evaluation from a total of 2380 EMG recordings indicates that each muscle location generates its own normal distribution. The mean, μ and variance σ^2 , for the TH distributions illustrated in Figure 4, for each muscle location is given in Table 1.

Table 1. Mean and Variance for the TH Distributions of Various Muscle Locations

Location	μ	σ^2	Location	μ	σ^2
E1	5.33	0.14	E6	5.06	0.08
E2	5.35	0.10	E7	5.40	0.05
E3	5.12	0.13	E8	5.12	0.04
E4	5.14	0.09	E9	5.45	0.11
E5	5.32	0.14	E10	5.38	0.07

From Table 1, the sequence from the highest μ to the lowest, for each muscle location is given as E9, E7, E10, E2, E1, E5, E4, E8, E3 and E6. In terms of consistency (the capability to embed similar information from specific muscles), the sequence from the lowest σ^2 to the highest is given as E8, E7, E10, E6, E4, E2, E9, E3, E1 and E5.

By assuming, the accuracy for the normal distribution of Figure 4, is within 6σ of each distribution, the highest value for each distribution is shown in Figure 5. The sequence from the highest maximum values to the lowest is given as 6.45(E9), 6.44 (E1), 6.43 (E5), 6.30 (E2), 6.22 (E3), 6.17 (E10), 6.10 (E7), 6.04 (E4), 5.91 (E6) and 5.72(E8).

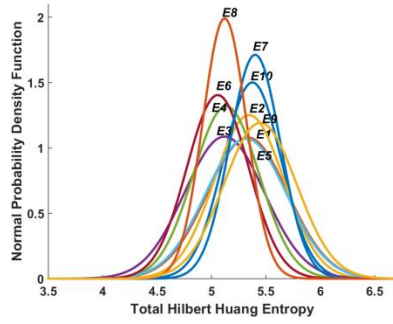


Figure 4. Total hilbert huang entropy distribution for various muscle locations

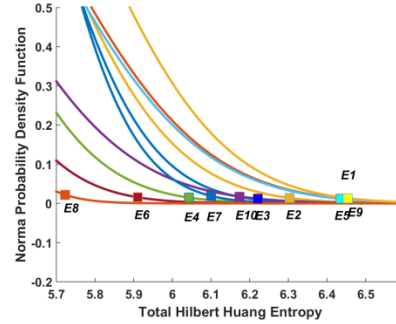


Figure 5. Total hilbert huang entropy maximum values based on the 6σ accuracy

The second indicator to measure the quality of the EMG signals generated during the recitation of Ruqyah is the difference between total information embedded in the EMG, TH_{EMG} and total information embedded in the speech, TH_{SPC} . In ideal case, TH_{SPC} must be equal to the sum of all TH_{EMG} , calculated from all muscles involved in generating the speech. Thus, the best muscle representing the recitation of Ruqyah, generates similar information between the speech and EMG. In our case, muscle which generates the smallest information difference, ΔTH is better than the other and can be described as follows:

$$\Delta TH = TH_{SPC} - TH_{EMG} \tag{17}$$

Figure 6 illustrates the normal distribution of ΔTH for all 10 muscle locations. The mean of each distribution taken from 2380 EMG recordings is given in Table 2.

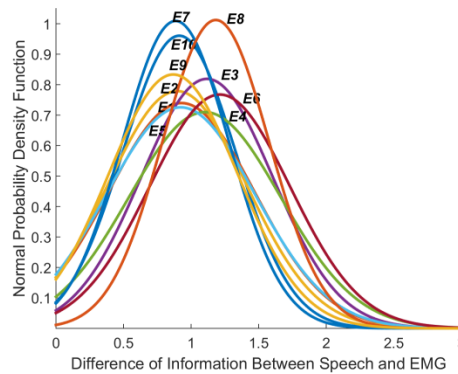


Figure 6. Different of information distribution between speech and EMG distribution on various muscles location

Based on the mean value given in Table 2, it shows that E9 generates the smallest ΔTH . This follows with E7, E2, E10, E5, E1, E4, E3, E8 and E6. From Table 2 and Figure 6, it is also observed that E5 has the highest probability to generate TH_{EMG} equal to TH_{SPC} . This follows with E2, E1, E9, E4, E10, E7, E3, E6 and E8.

Table 2. Mean (μ) of ΔTH and Normal Pdf when $\Delta TH = 0$ for Various Muscle Locations

Location	μ	$\Delta TH = 0$	Location	μ	$\Delta TH = 0$
E1	0.93	0.1678	E6	1.21	0.0505
E2	0.90	0.1682	E7	0.89	0.0820
E3	1.12	0.0589	E8	1.18	0.0112
E4	1.10	0.1039	E9	0.87	0.1612
E5	0.92	0.1772	E10	0.91	0.0860

Total information lost when specific muscles are used to represent the recitation of Ruqyah can be calculated using (15). By measuring the normal distribution between speech, $P_{HJ(spc)}$ and EMG, $Q_{HJ(EMG)}$, the $D_{KL}(P_{HJ(spc)} || Q_{HJ(EMG)})$ distance calculated for all 10 muscles location is given in Table 3.

Table 3. Total Information Loss Due when EMG Signal is used to Represent Ruqyah Recitation

Location	D_{KL}	Location	D_{KL}
E1	3.1612	E6	9.2151
E2	3.9916	E7	7.4867
E3	4.6364	E8	17.8414
E4	6.7257	E9	3.3586
E5	3.0796	E10	5.9854

Observing Table 3, it shows that the E5 generates the least loss of information, followed by E1, E9, E2, E3, E10, E4, E7, E6 and E8. Tables 1 to 3 and Figures 4 to 6 generate results for the best muscle locations, based on several different information measurements. Referring to the results, we could see that certain muscle locations repeatedly become the best muscles to represent the recitation of Ruqyah.

Table 4. The best 4 Muscle Locations in Monitoring Ruqyah Recitation

Location	Results Ranking Score						Total	Rank
	#1	#2	#3	#4	#5	#6		
E1	6	2	9	5	8	9	39	4
E2	7	5	7	8	9	7	43	2
E3	2	3	6	3	3	6	23	8
E4	4	6	3	4	6	4	27	7
E5	5	1	8	6	10	10	40	3
E6	1	7	2	1	2	2	15	10
E7	9	9	4	9	4	3	38	6
E8	3	10	1	2	1	1	18	9
E9	10	4	10	10	7	8	49	1
E10	8	8	5	7	5	5	38	5

Now, each result is ranked from 10 (best) to 1 (worst) in order to identify the overall best muscle location. Let us defined respectively results #1 to #6 as the highest mean of TH , the lowest variance of TH , the highest value of TH , the smallest difference of ΔTH , the highest probability density of ΔTH and the least of information loss D_{KL} . The ranking of results #1 and #2 are taken from Table 1, result #3 is from the observation of Figure 5, results #4 and #5 are taken from Table 2, and result #6 is taken from Table 3. The ranking score of these results for each muscle locations is given in Table 4.

From Table 4, we can conclude that muscles E9, E2, E5 and E1 are the four best muscles to represent the recitation of Ruqyah, based on the highest amount of embedded information, calculated from their generated EMG.

In the previous section, we have shown that E9, E2, E5 and E1 are among the four best muscle locations, for generating the highest quality signals with the richest embedded information, while reciting Ruqyah. All results that have been used in supporting the above finding are based upon the TH values, calculated using (16). Let's investigate visually, two different EMG signals taken from two different location, generating two different TH values as shown in Figure 7.

Both signals are taken from the same subject, while reciting the 4th verse of surah 'Al-Baqarah'. The first and second EMG signals respectively generate 6.4537 and 4.8890 as their TH values. As shown in Figure 7, the EMG signal generating higher TH is more active. This supports our argument in the previous section, that the muscles that generate higher TH are desirable, in order to monitor the Ruqyah recitation.

Although this study suggests four best muscle locations to assess the quality of Ruqyah's recitation, this does not mean, other muscle locations that are discussed throughout this paper, cannot be used for modeling a speech. Our experimental setup is purposely designed, so that there are 4-second delays, before and after verses of Ruqyah being recited as shown in Figure 8.

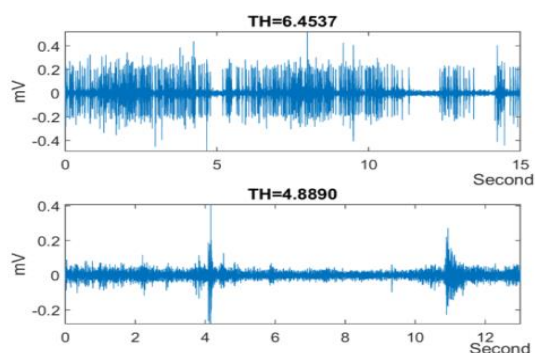
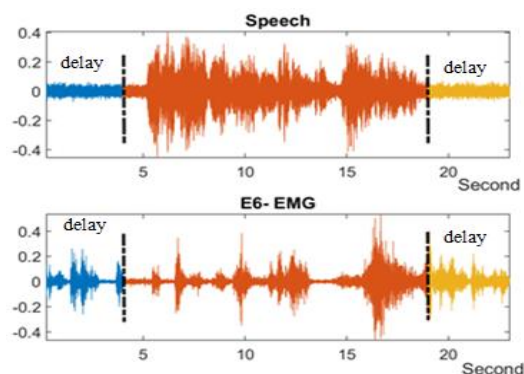
Figure 7. EMG signals of two different TH values

Figure 8. Comparison of EMG and speech activities

Normally, during speech activity, some of the muscles will contract earlier or later, before the speech is heard. An example of such muscle, is E6 as shown in Figure 8. In our studies, all calculated TH values, exclude the EMG information during the delays. In other words, the four muscle locations which were recommended in this study, generates the highest EMG activity when a speech is heard.

5. CONCLUSION

In this study, we have proposed four best muscle location, which could be used to monitor the quality of the recitation of Ruqyah. Three of these locations are on the neck while one on the face. This conclusion is the result of analyzing the information content, which is embedded in the EMG. All EMG signals were modeled using Empirical Mode Decomposition (EMD), in which Hilbert Amplitude Spectrum has been selected as the extracted features. Two techniques for measurement of information namely Total Hilbert-Huang Spectral Entropy, ΔTH and Kullback–Leibler Divergence, D_{KL} were applied on the extracted features. ΔTH measures the amount of information embedded in a signal. Five out of six results obtained in this study are based on ΔTH . Our study indicates that EMG with higher ΔTH represents a more active muscle with a strong contraction. On the other hand, D_{KL} measures the amount of information loss when an EMG from a muscle is used to represent speech. It is easier to monitor these muscles for the intent of assessing the quality of the recitation, given its strong EMG amplitude and high muscle contraction.

ACKNOWLEDGEMENTS

The research reported in this paper was supported by the Fundamental Research Grant Scheme (FRGS) No FRGS/1/2014/TK03/UNIKL/02/2 awarded by the Ministry of Higher Education of Malaysia.

REFERENCES

- [1] Furnham, Adrian, and Julie Forey, "The attitudes, behaviors and beliefs of patients of conventional vs. complementary (alternative) medicine," *Journal of clinical psychology*, vol. 50(3), pp. 458-469, 1994.
- [2] Syed, I., "Spiritual medicine in the history of Islamic medicine," *Journal of the International Society for the History of Islamic Medicine (JISHIM)*, vol. 2(4), pp. 45-49, 2003.
- [3] York, C. M., "The effects of Ruqya on a non-muslim: a multiple case study exploration," Doctoral dissertation, *Institute of Transpersonal Psychology*, 2011
- [4] Ahed M. F. Al-Sbou, "A survey of Arabic Text classification models," *International Journal of Electrical and Computer Engineering (IJECE)*, vol.8, no. 6, pp. 4352-4355, 2018.
- [5] M. Abbas and D. Megtouf, "Towards a practical and efficient WER calculation for Arabic ASR systems," 2nd International Conference on Natural Language and Speech Processing (ICNLSP), pp. 1-3, 2018.
- [6] Ali Abd Almisreb, Nooritawati Md Tahir, Ahmad Farid Abidin, Norashidah Md Din, "Acoustical comparison between /u/ and /u:/ Arabic Vowels for non-native speakers," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 11, no. 1, pp. 1-8, 2018.
- [7] Safie, Sairul, John J. Soraghan, and Lykourgos Petropoulakis. "Voice Activity Detection (VAD) using Bipolar Pulse Active (BPA) features," *Intelligent Signal Processing (WISP)*, IEEE 9th International Symposium on. IEEE, 2015.
- [8] M. Janke and L. Diener, "EMG-to-speech: direct generation of speech from facial electromyographic signals," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2375-2385, 2017.

- [9] Jingwei Too, A. R. Abdullah, Norhashimah Mohd Saad, N. Mohd Ali, H. Musa, "A detail study of wavelet families for EMG pattern recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4221-4229, 2018.
- [10] R. Ortega-Palacios et al., "Low-cost upper limb prosthesis, based on opensource projects with voice-myoelectric hybrid control," *Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE)*, pp. 1-5, 2018.
- [11] Sebastian Roldan-Vasco, Sebastian Restrepo-Agudelo, Yorhagy Valencia-Martinez, Andres Orozco-Duque, "Automatic detection of oral and pharyngeal phases in swallowing using classification algorithms and multichannel EMG," *Journal of Electromyography and Kinesiology*, vol. 43, pp. 193-200, 2018.
- [12] Nida Sae Jong, Pornchai Phukpattaranont, "A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A Thai syllable study," *Biocybernetics and Biomedical Engineering*, vol. 3(1), pp. 234-245, 2019.
- [13] N Kimura, M Kono, J Rekimoto, "*Sottovoce: An ultrasound imaging-based silent speech interaction using deep neural networks*," CHI '19 Conference on Human Factors in Computing Systems, no. 146, 2019.
- [14] GS Meltzner, JT Heaton, Y Deng, "Development of sEMG sensors and algorithms for silent speech recognition," *Journal of Neural Engineering*, vol. 15, no. 4, 2018.
- [15] Wand, Michael, Matthias Janke, and Tanja Schultz. "Tackling speaking mode varieties in EMG-based speech recognition," *Biomedical Engineering, IEEE Transactions on*, vol. 61(10), pp. 2515-2526, 2014.
- [16] Jingwei Too, A. R. Abdullah, N. Mohd Saad, N. Mohd Ali, T. N. S. Tengku Zawaw, "Application of gabor transform in the classification of myoelectric signal," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 17, no. 2, pp. 873-881, 2019.
- [17] Srisuwan, Niyawadee, Pornchai Phukpattaranont, and Chamnan Limsakul, "*Three steps of neuron network classification for EMG-based Thai tones speech recognition*," Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 10th International Conference on. IEEE, 2013.
- [18] Lyu, Ming, Caihua Xiong, and Qin Zhang, "*Electromyography (EMG)-based Chinese voice command recognition*," Information and Automation (ICIA), IEEE International Conference on. IEEE, 2014.
- [19] Mouaz Bezoui, "Speech recognition of Moroccan dialect using hidden markov models," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 1, pp. 7-13, 2019.
- [20] Srisuwan, N., Phukpattaranont, P. & Limsakul, C., "Comparison of feature evaluation criteria for speech recognition based on electromyography," *Med Biol Eng Comput*, vol. 56(6), pp. 1041-051, 2018.
- [21] Larson, Eric, Howard P. Terry, and Cara E. Stepp, "*Audio-visual feedback for electromyographic control of vowel synthesis*," Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012.
- [22] A. Rezeika, M. Benda, P. Stawicki, F. Gemblar, A. Saboor and I. Volosyak, "*30-Targets hybrid BNCI speller based on SSVEP and EMG*," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, pp. 153-158, 2018.
- [23] J. Anthony Seikel, Douglas W. King and David G. Drumright, "Anatomy & physiology for speech, language, and hearing," Delmar Cengage Learning, 5th edition, 2015.
- [24] WPP, "No. 102: Dissection of the Speech Production Mechanism," Department of Linguistics, UCLA. UCLA: Department of Linguistics, UCLA, 2002.
- [25] G. Wang, C. Teng, K. Li, Z. Zhang and X. Yan, "The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition," in *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1301-1308, 2016.
- [26] A. Humeau-Heurtier, C. Wu, S. Wu, G. Mahé and P. Abraham, "Refined multiscale hilbert–huang spectral entropy and its application to central and peripheral cardiovascular data," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2405-2415, 2016.
- [27] Li, Wei-Ling, et al., "Flow regime identification in a three-phase bubble column based on statistical, Hurst, Hilbert-Huang transform and Shannon entropy analysis," *Chemical Engineering Science*, vol. 102, pp. 474-485, 2013.
- [28] Y. Pan, T. Zheng and C. Chen, "*I-vector Kullback-Leibler divisive normalization for PLDA speaker verification*," 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, pp. 56-60, 2017.
- [29] Huang, Norden Eh, and Samuel S. Shen, "Hilbert-huang transform and its applications," vol. 5. *World Scientific*, 2005.
- [30] Y. Shrivastava, B. Singh, "A comparative study of EMD and EEMD approaches for identifying chatter frequency in CNC turning," *European Journal of Mechanics - A/Solids*, vol. 73, pp. 381-393, 2019.
- [31] Dandan Yang, Lixin Wang, Wenbin Hu, Chenkang Ding, Weibing Gan, Fang Liu, "Trajectory optimization by using EMD- and ICA-based processing method," *Measurement*, vol. 140, pp. 334-341, 2019.
- [32] Z. Liu, Q. Xie, M. Wu, W. Cao, D. Li and S. Li, "Electroencephalogram emotion recognition based on empirical mode decomposition and optimal feature selection," in *IEEE Transactions on Cognitive and Developmental Systems*, 2018.
- [33] A. Humeau-Heurtier, C. Wu, S. Wu, G. Mahé and P. Abraham, "Refined multiscale hilbert–huang spectral entropy and its application to central and peripheral cardiovascular data," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2405-2415, 2016.
- [34] <http://www.delsys.com/products/desktop-emg/bagnoli-desktop/>
- [35] S. I. Safie, M. I. Yusof, R. Rahim and A. Taib, "*EMG database for silent speech Ruqyah recitation*," 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, pp. 712-715, 2016.

BIOGRAPHIES OF AUTHORS

Safie S.I received his Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, United Kingdom. He is currently an Associate Professor and the Dean of Universiti Kuala Lumpur- Malaysian Institute of Industrial Technology. His research interest is in the big data, signal and image processing with application to biometric, psychological and physiological signal, image and video processing.



Rosuhana Rahim received her Bachelor Degree in Industrial Engineering Technology from Universiti Kuala Lumpur, Malaysian Institute of Industrial Technology (UNIKL MITEC), Pasir Gudang, Johor. She completed her Master Degree in Electrical and Electronics Engineering in Universiti Kuala Lumpur British Malaysian Institute (UNIKL BMI). His research interest is in pattern recognition systems, control systems, instrumentation and control systems, and signal and image processing techniques