**Lips Tracking Identification of A Correct Quranic letters‘ pronunciation for Tajweed teaching and learning**

Tareq Altalmas1, Muhammad Ammar Bin Jamil1, Salmiah Ahmad1, Wahju Sediono1, Momoh Jimoh e. salami1, Samsul Bahari Mohd Noor2, AND Surul Shahbudin Hassan3

1Department of Mechatronics in Enginering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.

2Department of Electrical and Electronic Engineering, Faculty of Engineering, University Putra Malaysia, Serdang, Malaysia

3Centre for Languages and Pre-University Academic Development, International Islamic University Malaysia, Kuala Lumpur, Malaysia

salmiah@iium.edu.my

***ABSTRACT:*** Mastering the recitation of the holy Quran is an obligation among Muslims. It is an important task to fulfill other ***Ibadat*** like prayer, pilgrimage and zikr. However, the traditional way of teaching Quran recitation possesses some difficulties due to the extensive training time and effort required from both teacher and learner. The first step in learning Quran recitation is to learn the correct pronunciation of the Quranic letter, which is basically also known as the Arabic letters. The Arabic letters pronunciation is based on their points of articulation and the characteristics of a particular letter. In this paper the lip’s movement identification technique was analyzed from the video signal acquired to extract the movement data of the lips while pronouncing the Quranic letters correctly. The extracted lip’s movement data were categorized into 5 groups based on the final shape of the lips. The reference data from the expert were acquired in the studio and were analyzed and identified to be in five (5) groups. The identified lip’s movement of the correct pronunciation was then tested among a public reciter and compared in terms of the percentage of similarity between the public and the professional recitation. The results show that the identified lip’s movement of the correct pronunciation is able to be used in evaluating and training the performance of the pronunciation as compared with the professional’s one. More subjects with different background will be tested in very near future with feedback instructions.

KEY WORDS: Tajweed, points of articulation, Lips reading, movement of lips, video signal

1. Introduction and Background

Al-Quran is the holy book for Muslims and it has been sent and written in the Arabic language as it is narrated in ***Surah Yusuf “Indeed, We have sent it down as an Arabic Qur'an that you might understand”.*** It is reported that our prophet ***Muhammed (s.a.w) said "Whoever recites a letter from the Book of Allah, he will be credited with a good deed, and a good deed gets a ten-fold reward. I do not say that Alif-Lam-Mim(ألم) is one letter, but Alif أ is a letter, Lam ل is a letter and Mim م is a letter''.*** The virtues of reading Al-Quran are; ten rewards for every recited letter, the people who used to recite Al-Quran are the best and last but not least Muslim position and rank in *Jannah* is determined based on the amount of Al-Quran that were memorized in his life [1]. Recitation of the Al-Quran with ***Tajweed*** is an essential task as a Muslim. In fact, it is an important means for fulfilling other worship such as praying, fasting and ***hajj***. Recitation of Al-Quran should be recited in the forms of the way that our Prophet Muhammad (s.a.w) did. To do it in such way, learning ***Tajweed*** becomes a necessary education in Islam. The word ***Tajweed*** means "to improve" or "to make better". It is also the rules and knowledge that help people to recite the Holy Quran as similar as what has been recited by the Prophet ***Muhammed*** (s.a.w) [2]. Moreover, an important part of ***Tajweed*** is to pronounce the letters from its correct articulations (***Makharij***) and by giving each letter its inherent characteristics (***Sifaat***) and dues in conditional characteristics. Characteristics (***Sifaat***) of Quranic letters differentiate letters that have the same points of articulations, and they are divided into two groups; characteristics with opposites and characteristics without opposites. The pronunciation of Quranic letters from its correct articulations points and its characteristics is considered a challenging task for people from non-Arab background. They need much effort to learn the right way of pronouncing the Quranic letters [3]. As the pronunciation of Quranic verses involves speech production process, thus, the implementation of speech recognition techniques can be beneficial for the Quranic teaching and learning.

In the recent year the utilization of computers in the process of second language teaching and learning gained a considerable attention from researchers. As an example, the systems that use computer in teaching a second language are called Computer Aided Language Learning (CALL) [4]. CALL systems have gained popularity due to its flexibility in empowering the students to develop their language more efficiency and fast as compared to the traditional ways. In fact the traditional and prevalent method of teaching ***Tajweed*** rules in Quranic teaching and learning is face-to-face based. This method, process of teaching ***Tajweed*** rules requires extensive practice sessions with a teacher where the student is trained while looking at the way the teacher pronounces it. It is quite a tough task to learn the pronunciation of any foreign language without the presence of the teacher. This process must be repeated until the student is able to recite the Quranic verses correctly. To achieve this, it requires extensive training time from both teacher and students [4]. Nowadays, computers are smart enough to mark the student’s effort based on the predetermined answers, give instant feedback, and record the student’s speech or text. However, computers should be trained in advanced to be able to do the above mentioned tasks [5], as if the computer now replaces the teacher. The process of teaching languages by using computers is improved significantly due to the combination between images, texts, video, and sound. This combination facilitates the integration of the four basic skills of listening, speaking, reading, and writing [6]. Visual-based articulatory feedback has been given to the learners of the second-language in order to correct their pronunciations. These feedbacks were the tongue position and shape [7]. On the other hand, a Computer Aided Quranic Recitation Training system has been developed to detect errors in continues reading of the Holy Quran. The system integrated both Automatic speech recognition and classifier-based approach [4]. Most of the designed systems are dealing with the full verses and words, and they are not considering the basic structure of the languages which it is the letters. In order to learn another language the person must learn how to pronounce the letters of this language correctly from the right articulation with the right characteristics.

The acoustic and visual signals (audiovisual) in the speech recognition systems lead to a promised research area Audio-visual or multimodal speech recognition system. This type of system can be used to improve the system performance in the noisy environments [8]. Naturally, the perception of human talks is a multimodal system which it consists of the ears and eyes, both the acoustic and visual information significantly improves the communication between people. Due to that, introducing the visual cues in the speech recognition systems is believed to give high impact in the mean of improving the performance and to reduce the error rates.

Lip reading and speech reading are the ability to understand the uttered words without hearing it. They are well-known practice to help people who are hearing impaired to understand the people speech from the movement of the lips. Facial expressions and gestures are used as well in this method [9][10]. Recently, the process of lip reading has been automated and opens a new wide area for the researchers. This new area comes under many names such as visual speech recognition, silent speech recognition or automated lip/speech reading[11]. Introducing visual information to the Audio-based speech recognition system results in a novel system called Audio-Visual Speech recognition system. These visual cues are used widely to enhance the performance of speech recognition system. The lip plays a vital role in speech, it moves about 80% of the time during speech, so it is considered as the best visual cues in speech perception [12]. In the recent years, the researchers have shown a significant interest in the field of Audio-Visual Speech recognition systems due to its importance in improving the accuracy of speech recognition systems[13][14][15].

This study is aimed to investigate the lip’s shape of the Quranic letters’ during pronunciation. The pronunciation was done based on the unique method of which the system used is based on the - by putting a ***SUKUN*** ( **ْ** ) on the letters and preceding it by ***Hamzah*( أ )** with ***Fathah*** (َ), [3], i.e. (**أَجْ**) for testing the letter (ج). This paper is organized as follows; Section 2 represents the Quranic letter pronunciation rules. The experiment and the implementation are discussed in Section 3. The results are discussed in Section 4 and the conclusion was discussed in Section 5.

1. The correct pronunciation of Quranic Letters

Quranic letters pronunciation relies on two major things; (1) points of articulation, exit or in Arabic, it is called ***(Makharij)*** and (2) attributes/characteristics or in Arabic it is called ***(Sifaat).*** There are five (5) main places of articulations in the vocal tract; the empty space in the mouth and throat, the throat, tongue, two lips and the nasal passage, Fig. 1 shows the chart of the articulation points and the letters that are related to these points. These articulation points have been there in the Quranic teaching and learning since many years ago and was developed based on the experiences among the previous scholars. Some letters are shared the same point of articulations or its articulations are close to each other as shown in Fig. 2, which it can be problem in the pronunciation of these letters. Therefore, the Quranic letters attribute ***(Sifaat)*** plays a key role in differentiating the letters that share the same points of articulations. Thus, the correct pronunciation of Quranic letter can be achieved by giving the letters its due by; the vowel ***“Harakat”***, the points of articulation and the manner of articulation. The Arabic letters can be categorized based on its equivalents in English language into two groups. Group 1 consists of 16 letters of that can be linked to similar letters in English language. In the other hand, Group 2 consists of 11 letters that do not have any equivalent letters in English language. In order to achieve the correct pronunciation of Quranic letters a ***SUKUN*** ( **ْ** ) should be used on the letters and it should be preceded by ***Hamzah*( أ )** with ***Fathah*** (َ)[3], i.e (**أَبْ**) for testing the letter (ب).

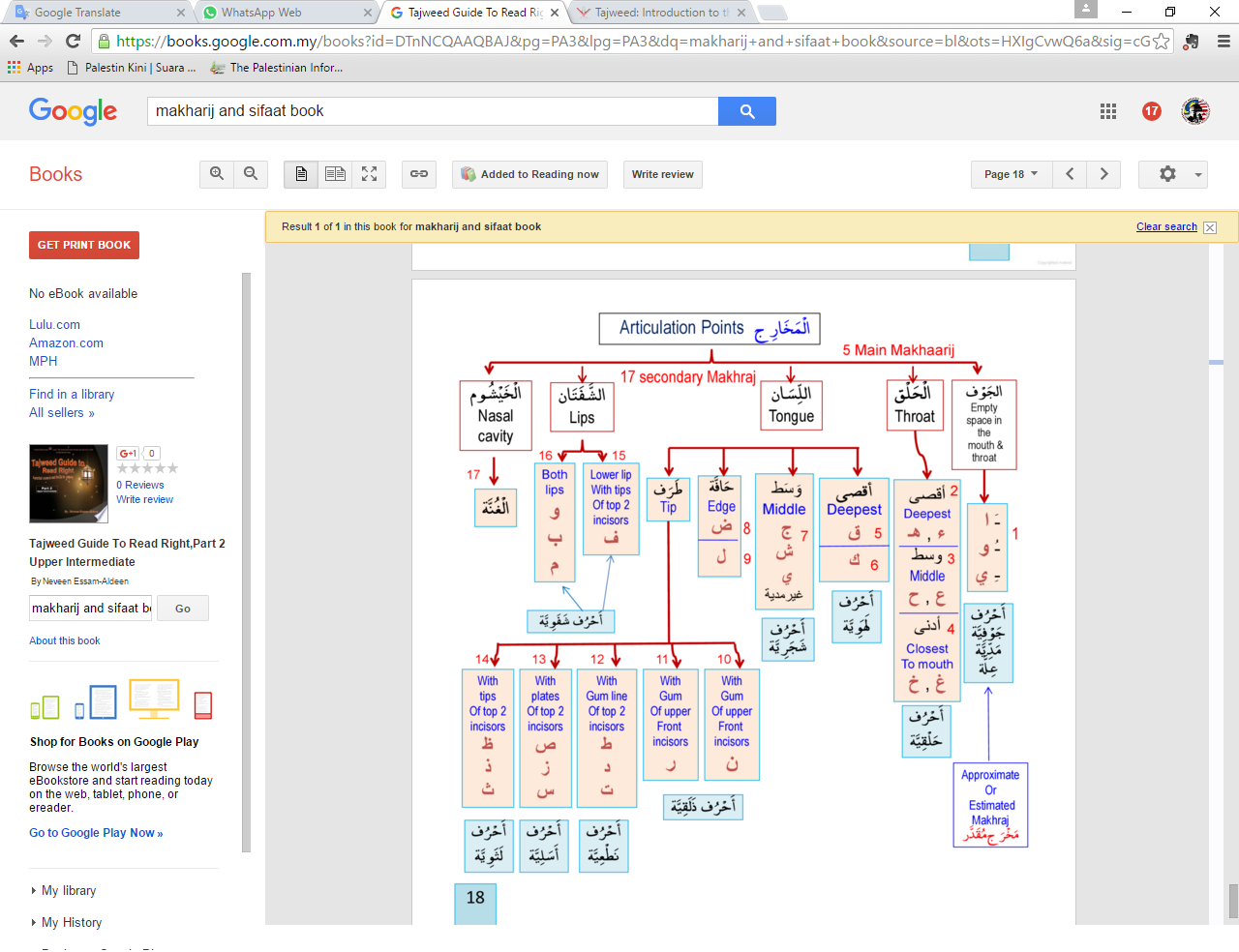


Fig. 1 chart of the articulation points and the letters that are related to these points [16]

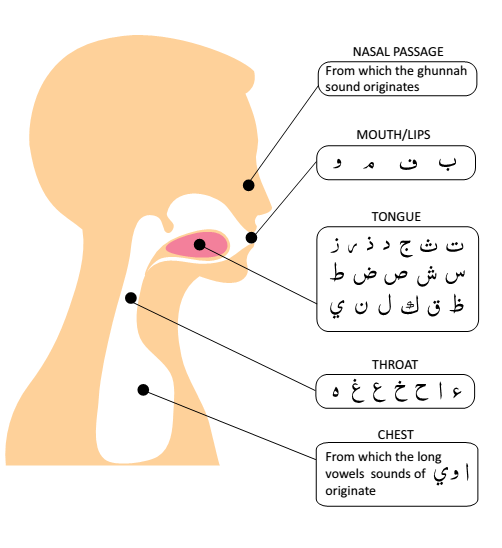
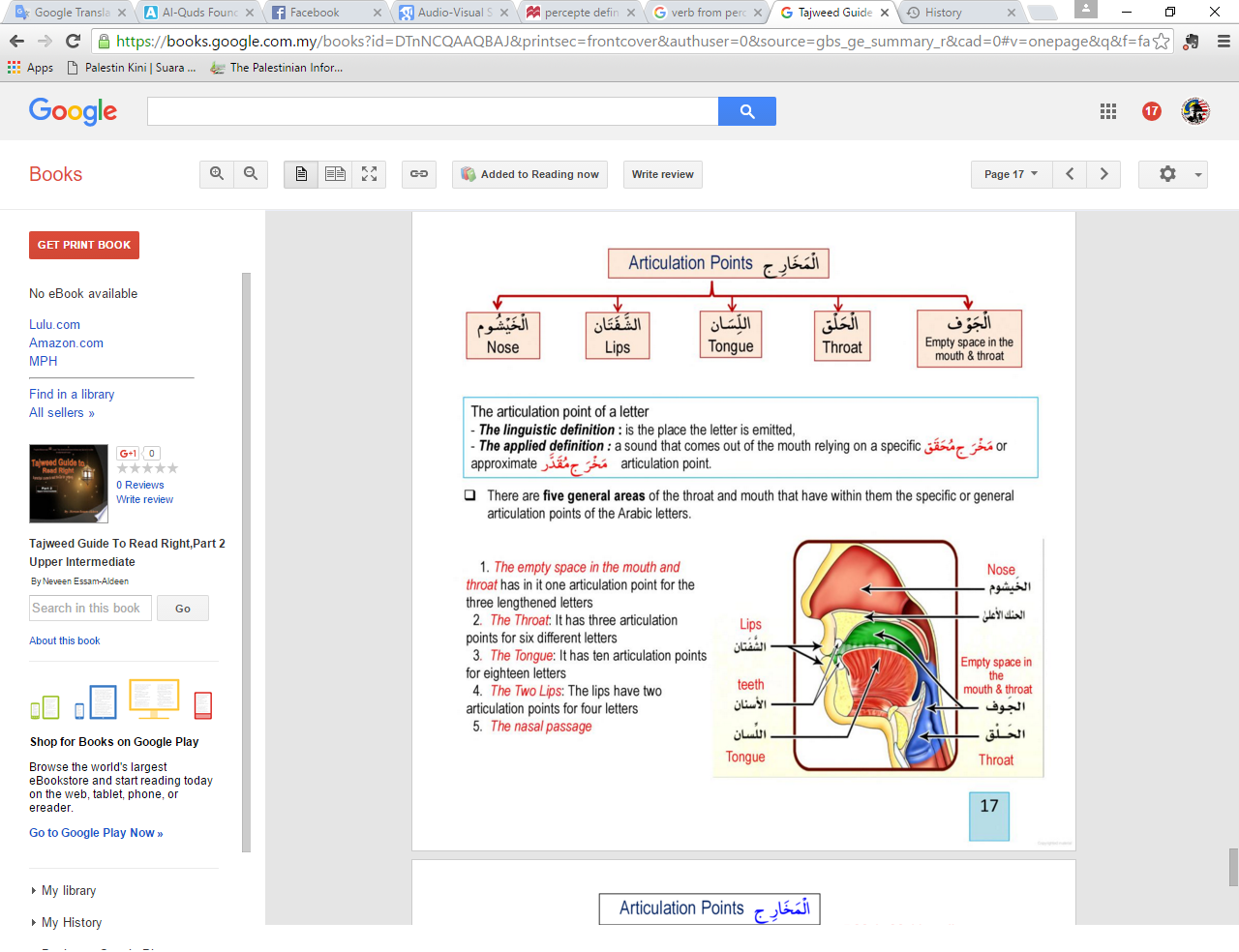


Fig. 2 chart of the articulation points and the letters that are related to these points [16]

1. Visual-Based Data Collection

As any speech analysis research, the process of data collection plays a key role in the overall system performance. In this paper the visual data are collected at an AV studio room at the Centre for Professional Development (CPD) in the international Islamic University that fulfill the standard environment for recording video and audio with almost no noise. The professional reciter’s mother tongue is Malay, and he is an expert and a Ph.D holder in Quranic teaching. The Quranic letters pronunciations have been recorded audio-visually for each letter for each Arabic vowel ***[Fathah, Kasrah, Dammah and sukun]*** ie. ***(أبْ, أبَ, أبُ, أبِ)***. But for this paper, the visual data was used to track the lip movement in pronouncing the correct Quranic letters.

The professional reciter of the Quranic letter has recorded the 28 letters (4 times) using high speed camera. High speed recording capable of recording the skin vibration of the subject face and it is able to record the voice information [17].

In this research, Casio EX-100 premium high speed camera has been used to record the face of the reciter. The recording speed of the used camera is considered as a high-frame rate between 120 to 1000 frames per second (fps), which makes it enough to record and detect the details of the movement. The camera has 5-axis image stabilizations that make it stable enough with very clear recording during the handheld recording [18]. The recording has been conducted in an equipped TV studio with sound mixer and soundproofed walls to reduce the unwanted background noise. Casio ex-100 premium high speed camera has been fixed in front of the face of the reciter at distance 1.5 m. The experimental setup of recording the data is shown in Fig. 3.



Fig. 3 Data acquisition process [Video Recording]

1. Lip Movement Data Analysis

The experiment was conducted to extract the movement data from professional reciter to be used as the reference model of correct pronunciation based on lip’s movement. The experiment was conducted first in concurrent flow for extracting the movement data, then these data sets are used to compare the pronunciation performance of other reciters. The recorded videos have been analyzed for each frame to extract the position of the points on the lips and determine the width, height, and shape of the mouth by using Matlab software. The lip movement data was first extracted from the professional reciter then from the ordinary reciter. A friendly Graphical User Interface (GUI) has been designed to be used by any user, as shown in Fig. 4.

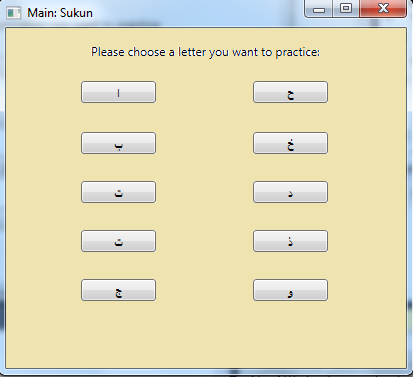
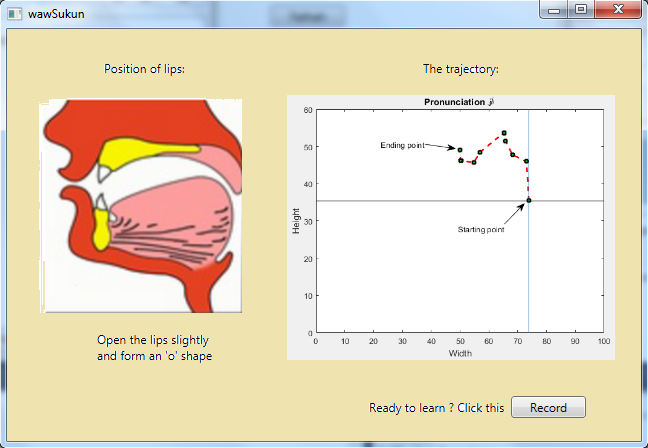
 

Fig. 4 Designed graphical user interface

Instead of the distance, displacement can also be used to measure the movement between the two aligned points. It can be used to solve various mouth shapes. Additionally, if there exist any pronunciation that is not symmetry in movement between the two parallel points, either height or width that is to pronounce. In that way, the displacement can be used to measure from a single isolated point to increase the accuracy of the performance.

There were two features studied in this paper; width and height of the lips from the starting time until the end of the pronunciation time. The width and height of the lips are selected between points 1,2 and 3,4 respectively as shown in Fig. 5. The lip movement can be extracted from the analysis of each frame and calculating the width and height for that frame. This movement provides information about the starting point, end point, and the behavior between these two points. Note that, there are limits for both the opening and closing of the mouth, which are considered as boundaries. Thus, the width and height of the lips may vary within the boundaries as can be refered to Fig. 6. The lips’ final position during the pronunciation was categorized into 4 possible shapes; (i) OpenMuzzle, (ii) OpenAgape, (iii) ClosedStretched, and (iv) Impossibleshape. The last mentioned shape is impossible to be reached when both of the width and height of the lips are decreased, as can be refered to Fig. 7.



Fig. 5 Points to be used to determine the width and height for each frame in the video.

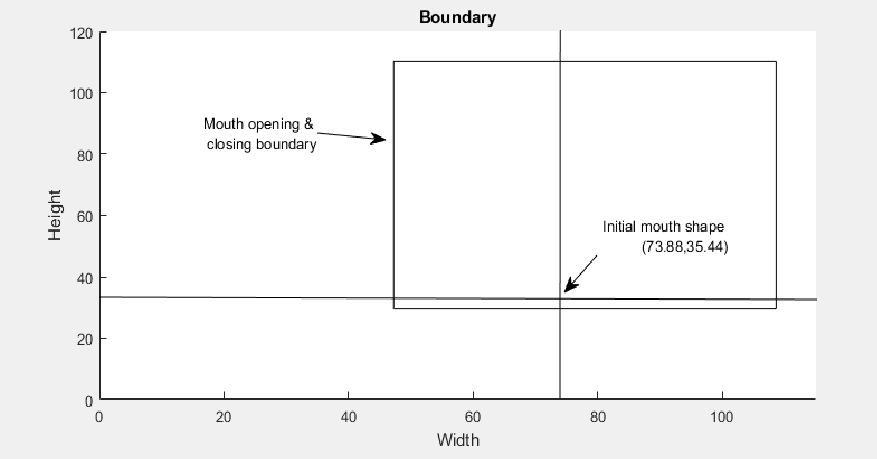


Fig. 6 Mouth opening & closing boundary

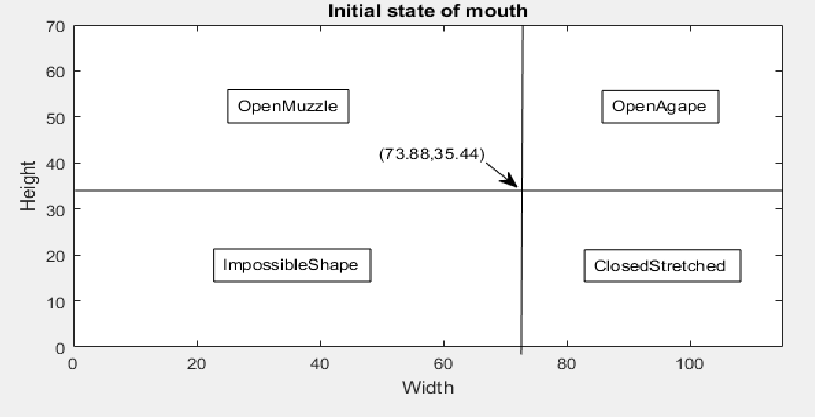


Fig. 7 Mouth shapes when pronouncing Quranic letters.

1. Results and Discussions

**5.1 The Correct Lip movement Extraction**

The first step in the analysis was to extract and plot the lips movement from the professional reciter, where this movement helps in deciding the final shape of the mouth and in categorizing the letters into 5 groups.

The selected pairs of points represent the width and height of the lips while pronouncing the Arabic letters. Therefore, by achieving the width and height of the lips, it is easy to describe the shape of the mouth of which it can be categorized into 4 basic shapes as follows; Normal, Open, Agape, and Mouth stretched based on the end position of the points. Fig. 8 shows the four possible shapes of the mouth during pronunciation of Quranic letters.

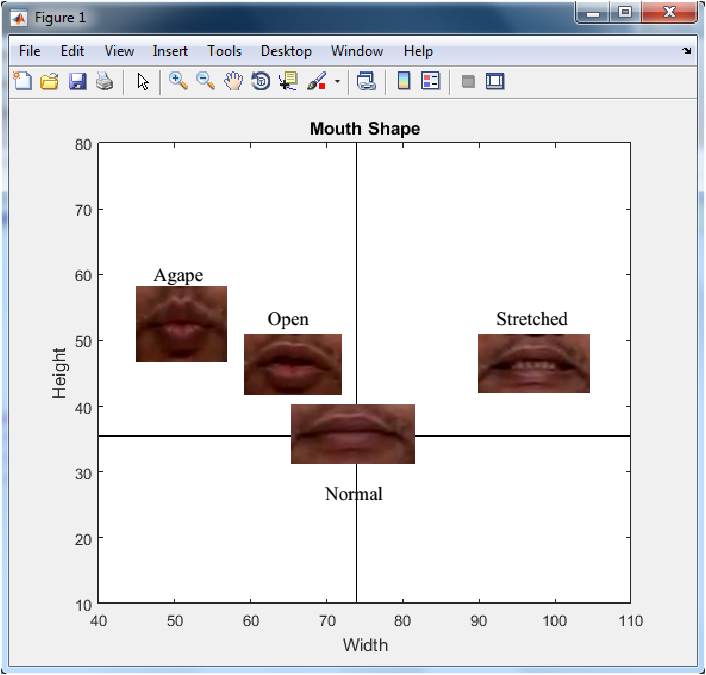


Fig. 8 Four basic Mouth shapes when pronouncing Quranic letters.

In addition to the shape of the lips, the analysis of the extracted frames from the video may provide us the position of the mouth when we start uttering the letter, the end position and the behavior of the mouth during the pronunciation. By plotting the width and the height of the lips in one graph and extract the relation between them we can create another category of the groups based on the shape of the mouth. These groups are categorized based on the movement of the lips movement and the end position. Fig. 9 shows the movement graphs of Group 1 which the height linearly increased and the width decreased and the final shape is small Open. Group 2 forms a loop and there is close during the pronunciation as shown in Fig. 10. Group 3 is where the width is decreasing while the height increased and then decreased to form O-shape as illustrated in Fig. 11. And Group 4 is where both height and width change their position. The width decreases, then increases, while the height that is opposite to the width increases and then decreases. In Group 4, the final shape of the mouth is stretched a little. Fig. 12 shows the Group 4 movement. Lastly, Group 5 follows the movement of Group 4 but with larger change which result in wider stretch mouth as shown in Fig. 13. Table. 1 illustrates the groups, final shape and the associated letters for each group.

Table 1 Groups mouth shapes and the associated letters.

|  |  |  |
| --- | --- | --- |
| **Group No.** | **Mouth behavior/shape** | **Letters** |
| 1 | Open a Bit | ن ,ل ,خ ,ع ,ح ,ق |
| 2 | Open and Close | ب ,م |
| 3 | Form an ‘o’ shape | ظ ,ط ,ص ,ر ,و , ش ,غ |
| 4 | Open the mouth and then stretch a bit | ف ,س ,د ,ج ,ت, ي ,ك |
| 5 | Open the mouth and stretch wider | ذ ,ث ,ز |

|  |  |
| --- | --- |
|  |  |
|  |  |
| Fig. 9 Lip Movement of Letters in Group 1 | |
|  | |
|  |  |
| Fig. 10 Movement of Letters in Group 2 | |
|  |  |
|  |  |
|  |  |
| Fig. 11 Movement of Letters in Group 3 | |
|  |  |
|  |  |
|  |  |
| Fig. 12 Movement of Letters in Group 4 | |
|  |  |
| Fig. 13 Movement of Letters in Group 5 | |

**5.2 Testing with a ordinary reciter**

A random, ordinary user was selected to test the developed lips tracking system. The goal is to analyze the performance of the Quranic letters pronunciation from the percentage of similarity and thus the user can evaluate rather the pronunciation is correct or not and improve. The percentage of similarity is calculated by comparing the class extracted from the plotted lip movement. Let illustrate how it works by assuming the user wants to learn how to pronounce ***أَشْ***. Fig. 14 shows the movement result for the first time testing and Table 2 shows the classes’ comparison and the percentage of similarity. Similarly, Fig. 15 and Table 3 show the result for the second attempt. Note that the experiment was analyzed using displacement measurement.

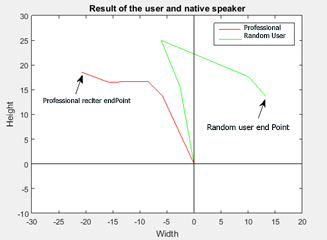


Fig. 14 First attempt of the Random user for the letter *أَشْ*

Table 2 First attempt of the Random user for the letter *أَشْ*

|  |  |  |  |
| --- | --- | --- | --- |
| **Classes detection** | **Reference speaker data** | **User data** | **Remark** |
| **1** | Loop | Loop | Correct |
| **2** | Open | Open | Correct |
| **3** | Agape | Stretched | Wrong |
|  | | | 2/3 = 66.7% |

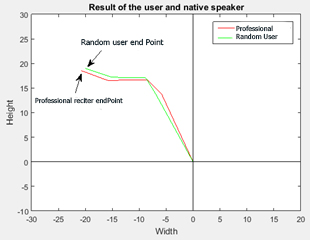


Fig. 15 Second attempt of the Random user for the letter *أَشْ*

Table 3 Second attempt of the Random user for the letter *أَشْ*

|  |  |  |  |
| --- | --- | --- | --- |
| **Classes detection** | **Reference speaker data** | **User data** | **Remark** |
| **1** | Loop | Loop | Correct |
| **2** | Open | Open | Correct |
| **3** | Agape | Agape | Correct |
|  | | | 3/3 = 100% |

1. Conclusion

The movement of the Lips while pronouncing the correct Quranic letters was extracted successfully from a recorded video of a professional reciter of Al-Quran. The displacement between frames has been calculated as from initial point until the completion of letter’s pronunciation. A comparison between the pronunciation of the professional reciter and an ordinary person has been conducted based on the displacement calculation and it shows that the system is able to detect the percentage of similarity for a correct pronunciation. The system can be improved further by conducting more frames, and by providing real time feedback to the user in the future.

1. References

[1] “Virtues of reading the Quran.” [Online]. Available: http://www.islamweb.net/en/article/109245/. [Accessed: 25-Feb-2016].

[2] “Tajweed introduction.” [Online]. Available: http://www.readwithtajweed.com/tajweed\_Intro.htm. [Accessed: 25-Feb-2016].

[3] “letters\_Makhaarij.” [Online]. Available: http://www.readwithtajweed.com/tajweed\_Makhaarij.htm. [Accessed: 23-Feb-2016].

[4] H. M. A. Tabbaa and B. Soudan, “Computer-Aided Training for Quranic Recitation,” *Procedia - Soc. Behav. Sci.*, vol. 192, pp. 778–787, Jun. 2015.

[5] C. Meskill, *Teaching and Learning in Real Time: Media, Technologies, and Language Acquisition*. Athelstan, 2002.

[6] L. Wang, Y. Qian, M. Scott, G. Chen, and F. Soong, “Computer-assisted audiovisual language learning,” *Computer (Long. Beach. Calif).*, vol. 45, no. 6, pp. 38–47, 2012.

[7] P. Badin, A. Ben Youssef, G. Bailly, F. Elisei, T. Hueber, and GIPSA, “Visual articulatory feedback for phonetic correction in second language learning,” *Actes de SLATE*, pp. 1–10, 2010.

[8] M. Anusuya and S. Katti, “Speech recognition by machine: A review,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 6, no. 3, pp. 181–205, 2009.

[9] “Hearing loss - lipreading,” 2012. [Online]. Available: https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/hearing-loss-lipreading. [Accessed: 22-Feb-2016].

[10] T. Chen, “Audiovisual speech processing,” *Signal Process. Mag. IEEE*, vol. 18, no. 1, pp. 9–21, 2001.

[11] A. B. Hassanat, “Visual Words for Automatic Lip-Reading,” Sep. 2014.

[12] W. Ur Rehman Butt and L. Lombardi, “A survey of automatic lip reading approaches,” in *8th International Conference on Digital Information Management, ICDIM 2013*, 2013, pp. 299–302.

[13] A. B. A. Hassanat, *Speech and Language Technologies*. InTech, 2011.

[14] K. Iwano, T. Yoshinaga, S. Tamura, and S. Furui, “Audio-visual speech recognition using lip information extracted from side-face images,” *Eurasip J. Audio, Speech, Music Process.*, vol. 2007, 2007.

[15] E. S. Salama, R. A. El-khoribi, and M. E. Shoman, “Audio-Visual Speech Recognition for People with Speech Disorders,” *Int. J. Comput. Appl.*, vol. 96, no. 2, pp. 51–56.

[16] N. Essam-Aldeen, *Tajweed Guide To Read Right,Part 2 Upper Intermediate*. Lulu.com.

[17] M. Akutsu, Y. Oikawa, and Y. Yamasaki, “Extract voice information using high-speed camera,” *J. Acoust. Soc. Am.*, vol. 133, no. 5, p. 3297, May 2013.

[18] “Casio EX-100 Review and Specs.” [Online]. Available: http://cameradecision.com/review/Casio-Exilim-EX-100. [Accessed: 06-Mar-2016].