English speaking proficiency assessment using speech and electroencephalography signals

Abualsoud Hanani, Yanal Abusara, Bisan Maher, Inas Musleh

Department of Electrical and Computer Engineering, Birzeit University, Birzeit, Palestine

Article Info

ABSTRACT

In this paper, the English speaking proficiency level of non-native English speakers was automatically estimated as high, medium, or low performance. For this purpose, the speech of 142 non-native English speakers was recorded and electroencephalography (EEG) signals of 58 of them were recorded while speaking in English. Two systems were proposed for estimating the English proficiency level of the speaker; one used 72 audio features, extracted from speech signals, and the other used 112 features extracted from EEG signals. Multi-class support vector machines (SVM) was used for training and testing both systems using a cross-validation strategy. The speech-based system outperformed the EEG system with 68% accuracy on 60 testing audio recordings, compared with 56% accuracy on 30 testing EEG recordings.

Corresponding Author:

Abualsoud Hanani
Department of Electrical and Computer Engineering, Birzeit University
1 Marj Road, Birzeit, West Bank, Palestine
Email: ahanani@birzeit.edu

1. INTRODUCTION

In addition to be the language of science, English becomes the medium of instruction and communication across the world. It is very important for English learners to get proficiency assessment, as the output of such an assessment influences their academic careers. Therefore, providing reliable and instant English proficiency assessments are most important in determining their academic progress. One of the most common problems facing non-native speakers when learning English is the inability to pronounce the sounds of English words properly. Quality of pronunciation is the major difference between native and non-native English speakers that will be noticed. Learning proper English pronunciation helps learners to communicate effectively with the native English speakers. From this point, having an automatic system which provides an immediate assessment and feedback for English learner will help and improve the English quality of learners. Currently, there are some systems which can automatically measure the English pronunciation quality of speaker automatically and give an instant feedback at word and phoneme levels. This kind of systems depends on the speech signal for extracting acoustic features of each phone and compare them with a standard pronunciation models (mono-phones or tri-phones), which are usually trained on English native speech.

Computer assisted language learning (CALL) systems [1], [2] are proven to be helpful and effective for learning non-natives pronunciation details of a new language, especially in the starting phase of learning and for pronunciation training. Computer assisted pronunciation training system (CAPT) [3] is another learning systems which are always available and can be used everywhere, and allow the learner to make mistakes without loss of self-confidence because it is one to one teaching process. This gives the learner a positive experience

in learning process [3]. Most computer-assisted pronunciation training system are based on automatic speech recognition (ASR) techniques. CALL system use ASR for offering new perspective for language learning [4]. Language learners "are known to perform best in one-to-one interactive situations in which they receive optimal corrective feedback" [3]. Due to the lack of time, in most cases, it is not always possible to provide individual corrective feedback. Therefore, ASR-based CALL systems can recognize what a person actually uttered, to detect pronunciation and language errors, and to provide feedback spontaneously. According to [4], the system's accuracy is higher for native speech than for non-native, and that the speech recognition technology is still at an early stage of development in terms of accuracy. Although there are some commercial systems for non-native English speakers across the world. As it can be seen that the general ASR application for English learning may not work satisfactory for Arab pronunciation learners, because "the former requires the ASR in general to be forgiving to allophonic variation due to accent [5]. Most of available English proficiency assessment systems depend on the speech signal for extracting acoustic features of each sound and compare them with a standard pronunciation models for each phone, or tri-phones, which are usually trained on English native speech. The accuracy of such systems is acceptable but since they depend on speech signal only, the performance of these systems is quietly affected by the quality of recorded speech signal. In other words, background noise, which is inevitable, degrades the efficiency of these systems dramatically. In order to overcome this limitation and to improve the system accuracy, in this framework, we are using EEG signals which reflect neurons firing inside the learner brain and use it, beside the speech signal to measure the proficiency and confidence level of English learner. The EEG signal has been successfully used in many applications in various fields such as emotion detection, robot controlling and other appliances by thinking, typing characters by thinking and many other interesting applications. We believe, by combining speech signal and EEG signals, that the performance of automatic systems for estimating English proficiency and confidence level can be improved significantly, and the problems of depending only on speech signal will be partially or completely solved. To our knowledge, this is the first attempt to use multimodality (voice and EEG) for assessing speaking English quality and confidence level of non-native speakers. Such system is very important for giving an immediate and instant feedback for English learner, especially, when assessing speaking quality. The traditional way of assessing speaking quality and level of confidence is by setting up an interview with an expert in English. Therefore, developing an automatic system for such a task will come back with many benefits for both language learners and assessors.

The rest of paper is organized as follow: literature review is discussed and presented in section 2, dataset collection and description is presented in section 3.1. Sections 3.2 and 3.3 describe the audio and EEG based systems, respectively. Experiments and results are presented and discussed in section 4. Conclusion and future work are presented in section 5.

2. LITERATURE REVIEW

Automatic speech processing technology has been applied to many different fields over the past two decades. For example, preparation for English proficiency tests required for the higher education institutions [6], foreign-based English skills call center agents evaluation [7], and aviation English evaluation [8] are heavily dependent on the speech technology. For more examples and more details [9].

In most of these systems, the participants need to speak to the system for language proficiency evaluation. Read aloud is the most common type, where the participant reads out loud one sentence or a set of sentences. In order to make these systems more interactive and communicate with the participants by speech, automatic speech recognition (ASR) systems, which converts speech into text, are used, even with heavily accented non-native speech. Different features types representing non-native speakers when producing English sounds and speech patterns are extracted from the participants responses and used in the English proficiency evaluation. Some of the most successfully used features include phone's spectral match to native speaker acoustic models [10] and a phone's duration compared to native speaker models [11], fluency features, such as the rate of speech, average pause length, and number of disfluencies [12] and prosody features, such as pitch and intensity slope [13].

Although most of the applications elicit restricted speech, some applications have used automated scoring for non-native spontaneous speech, in order to make speaker's communicative competence is fully evaluated (e.g., [6] and [14]). In such systems, the same types of features based on the prosody, fluency, and pronunciation are extracted. Furthermore, features related to additional aspects of a speaker's proficiency in the non-native language can be extracted, such as vocabulary usage [15], syntactic complexity [16], [17], and

topical content [18].

In other related studies, the fundamental frequency (F0) and pitch contours were used to assess the oral reading proficiency of non-native speakers, automatically. For example, Tepperman *et al.* [19] developed canonical contour models of F0 and compared non-native speakers model with the native speakers. The contours were modeled at the word level and then the prosody scores were computed based on a combination of the contour features, energy features, and duration features. A correlation of 0.80 between these scores and human ratings were reported. Moreover, based on an autocorrelation maximum posteriori approach, authors in [19] took advantage of pitch floor and ceiling values to capture aspects of pronunciation quality not seen at the segment level and achieved a maximum accuracy of 89.8% in a classification task discriminating between native and non-native speakers.

Electroencephalography (EEG) is an electrophysiological monitoring method to measures and records the electrical activity of the brain. It is a readily available test that provides evidence of how the brain functions over time. It is typically noninvasive, with the electrodes placed along the scalp, using the standardized electrode placement scheme, known as 10-20 international system [20]. Wires attach these electrodes to a machine, which records the electrical impulses. It is proven that EEG signals can be used for new methods of communication besides the well-known clinical applications. It has been proven that it is feasible to link between speech recognition and EEG, that is to use EEG for the recognition of normal speech [21].

3. SYSTEM DESIGN
3.1. Database description

To our knowledge, there is no available dataset which includes both speech and EEG for the purpose of estimating English language proficiency level for non-native English speakers. Therefore, we have collected our own dataset. For this purpose, a short interview has been designed which consists of two parts. First, each participant is asked to talk about himself/herself in two minutes, then he/she asked to describe a picture presented on a paper in front of him/her in around 2 minutes. With the help of an English expert from the English department at Birzeit University, 142 participants (university students learning in English and instructors teaching in English at Birzeit University) with different level of English proficiency, have been made recorded interviews in English. A high quality close microphone was used for recording audio signals of all participants (142) in a quiet environment. Sampling frequency of 44.1 KHz was used and recordings were saved in wav format. The average length of speech files is 2.5 minutes. During speech recording, an Emotiv Epoc headset device [22] was used for recording EEG signals for all male participants (58) and saved in csv file format. Fourteen soft electrodes were attached to the participant's scalp with special adhesive electrode gel, located according to the international 10/20 system [20].

Because of long hair and covered head of most of female participants, EEG signals could not be recorded for all females (84). Therefore, EEG signals were recorded for males only. With help of English expert from English department at Birzeit University, an evaluation criteria was used for evaluating English proficiency level of each participant. English expert listened to all recorded interviews and do the assessment by assigning 1 to 10 for each participant using a predefined assessment criteria.

Based on the results of expert evaluation, all participants had been divided into three skill levels; participants with average scores 8-10 are classified as high proficiency (HP), participants with average scores of 5-7 are classified as medium proficiency (MP), and participants with average scores 1-4 are classified as low proficiency (LP). According to this criteria, 20 participants were classified as HP, 47 participants as MP and the remaining 75 were classified as LP. More details about the participants and the result of human expert evaluation are found in the Table 1.

Table 1. Information details of the participants

Total no of participants	142
Gender	58 Males + 84 Females
Age	20 – 56 years old
Avg. years of studying English	2-5 years
Living in English speaking countries	10 participants lived for more than 6 months in USA and UK
English expert classification	20 HP (11M+9F), 47 MP (13M+34F), and 75 LP (25M+50F)

3.2. Speech-based system

3.2.1. Speech features

To build an automatic assessment system based on speech, speech recordings of each skill level (HP, MP and LP) were used to train a classification system which can classify skill level of English speaker into one of the three classes; HP, MP or LP.

− Short frame energy: The speech signal is divided into short frames of 20 ms length at rate of 10 ms (i.e. frames overlap is 50%). A hamming window is multiplied by each frame and then the short frame energy is computed in decibel, as shown in (1), for each frame and used as an audio feature for our proposed system.

$$E = \sum_{n=0}^{N-1} 10 log 10 s(n)^2 w(n) \tag{1}$$

Where, s(n) is the speech signal and w(n) is Hamming window, with a formula shown in (2).

$$w(n) = 0.54 - 0.46 cos(\frac{2\pi n}{n-1}), 0 \geq n \leq N - 1 \tag{2}$$

− Short frame zero-crossing rate (ZCR): In order to make the speech signal unbiased, the signal average (dc component) is computed and subtracted from the signal. Then, the number of time-axis crossings is computed for each short frame. These counts are then divided by the total number of zero-crossings of the whole utterance, as shown in (3).

$$ZCR = \frac{1}{2} \sum_{n=0}^{N-1} |sgn(s(n)) - sgn(s(n-1))| \tag{3}$$

Where, $sgn()$ is sign function which gives 1 for positive values and -1 for negative values.

− Mel frequency cepstral coefficients (MFCCs): MFCC features are the most commonly used in the speech processing applications. They represent the general shape of power spectrum for each frame with low dimensional feature vectors (12). More details about MFCC technique can be found in [23]. The first 12 MFCCs of each frame are appended to the audio feature vectors of our audio-based system.

− Short-time frame pitch: Pitch refers to the fundamental frequency of the voiced speech. Pitch is an important feature contains speaker specific information. It is a property of vocal folds in the larynx and is independent of vocal tract. A single pitch value is determined from every windowed frame of speech. There is a number of algorithms for estimating pitch form speech signal. Among these, one of the most popular algorithms is the robust algorithm for pitch tracking (RAPT) proposed by Talkin [24]. This algorithm was used to extract pitch for use in all experiments reported in this paper.

− Formant frequencies: The general shape of the vocal tract is characterized by the first few formant frequencies. Praat toolkit [25] has been used to estimate the first three formant frequencies and their gains and appended to the acoustic feature vectors.

− Phoneme rate: Speaking rate has been used as a feature in numerous speech processing applications. In this work, speaking rate was estimated from the number of phonemes in each 0.5 s window. The publicly available English phone recognizer developed in [26] has been used to generate phonetic labels.

− Pauses: Pauses in speech have a meaning. There are two types of pauses; empty pauses which are silent intervals in the speech signal, in which speaker is usually thinking of the next utterance. The second type of pauses is the filled pauses with vocalizations, which do not have a lexical meaning. Usually, non-native speakers need more time (pauses) for thinking of the proper and suitable words and for producing meaningful sentences while speaking. These pauses are relatively longer than the natural pauses made by the native speakers when they are speaking. Therefore, the length of the pauses and the frequency of pauses may carry an important information about the speaker proficiency in the foreign language. Based on the short frame energy and zero-crossing rates, a simple algorithm has been developed for estimating the length and the number of pauses occurred in each utterance. Low energy and high zero-crossing rate frames are usually silence frames and, hence, classified as pauses frames. Whereas, the frames with high energy and relatively low zero crossing rates are for normal

speech, hence, classified as speech frames. If a number of successive pause frames exceeds a practically specified threshold, they are considered as a pause. Therefore, if the pause is exceeds a certain duration time or if it is occurred many times while speaking, this may indicate low language proficiency. Each of the above 24 audio features (energy, ZCR, 12 MFCCs, pitch, 6 formant frequencies with their gains, phoneme rate, average pauses length, number of pauses) is represented by average, minimum and maximum values for each utterance. This results in 72-dimensional vector for each utterance. Using cross-validation method (as described earlier), these feature vectors are used to train and test a three-class support vector machines (SVM) system. The results are presented in section 4.

3.3. EEG-based system
3.3.1. EEG pre-processing

In addition to language thinking while speaking, recorded EEG signals include multiple sources of actions such as eye blinking, eye movement, head movement, and muscle movement. which are known as artifacts. In our case, we are interested in second language thinking information. Therefore, the other artifacts are considered as noises for our system and need to be removed. Numerous techniques proposed for avoiding, rejecting and removing artifacts from EEG signals [27]. More recently, independent component analysis (ICA) technique has been used successfully to remove EEG artifacts [28], and it has been demonstrated to be more reliable than other artifact-removal methods.

There are many implementations for ICA artifacts removal which differ on the independence of the components and estimation of the mixing matrix. However, recent study has showed that there is no significant difference in the performance of these algorithms. Moreover, it is shown that ICA reliability depends more on pre-processing, such as raw data filtering, than algorithm type [29].

3.3.2. EEG feature extraction

Feature extraction is the process of finding appropriate representative features from the raw EEG signals which can be used for classification of different brain activity patterns. The most common set of features for raw EEG signal processing are temporal, frequential and time-frequency [30].

3.3.3. EEG classification

There are many classification techniques applied in EEG processing. The most successfully used classification techniques include SVMs, k-nearest neighbour (KNN) and naïve bayesian (NB) [31]. In order to remove the low and high frequency noise from the recorded data, signals were band-pass filtered (bandwidth 1 to 40 Hz). In order to separate and remove sources associated with artifacts from EEG, the ICA algorithm (refer to as Fast ICA) [28], was applied to the filtered data. Features are extracted from the pre-processed EEG channels. The EEG signal of each channel is divided into frames of 1.5 s length with 0.5 s overlap. A set of frequential features were obtained by a 128-point fast fourier transform (FFT). The sum of the spectral power lying in delta (1 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz) and beta (13-20 Hz) bands and relative intensity ratio of each band are used as the features. The eight features extracted from each single channel, of the fourteen channels, are concatenated together to form the final feature vector of length 112.

4. RESULTS AND DISCUSSION
4.1. Speech-based system

As mentioned earlier, the leave one out cross-validation technique was used for training and testing our two sub-systems. For audio system, speech of 60 participants (20 for each group) are used for training and testing SVM classifier. With the 62 tests, the system accuracy is 68%. The confusion matrix of the system is shown in Table 2. It is clear from the confusion matrix that there is large confusion between high and medium performance groups and very small confusion with low performance group. A possible explanation for this, is that the average English proficiency level of LP classified participants is near the lowest border of low scale. On the other hand, the average level of HP classified participants is near low border. This makes it difficult to distinguish between high performance and medium performance skill levels.

4.2. EEG-based system

Recall that the amount of EEG data is much less compared with speech data. According to the English expert evaluation, the number of participants, who have EEG recordings, in each of the three skill levels are as

2506 ❑ ISSN: 2088-8708

follow: 10 for HP group, 11 for MP group and 21 for the LP group. To keep data size of each class balanced, EEG recordings of 10 subjects from each group were used for training and testing, using leave one out strategy and using multi-class SVM. The EEG system accuracy is 56%. The confusion matrix is shown in Table 3.

Table 2. Confusion matrix of speech-based system

		Recognized		
		HP	MP	LP
True	HP	13	7	0
	MP	8	12	2
	LP	1	3	16

Table 3. Confusion matrix of EEG-based system

		Recognized		
		HP	MP	LP
True	HP	6	2	2
	MP	3	4	3
	LP	1	2	7

Similar to audio-based system, EEG-based system presents a high confusion between the high and medium performance groups. This motivates us to combine these two groups into one group (high) and re-train the two systems with two-class SVMs (high vs low). The audio-based system performance is increased to 83% and EEG-based system is increased to 78%.

5. CONCLUSION

In this paper, two English proficiency level estimation systems were proposed. One system uses audio features extracted from speech recordings and the other uses features extracted from EEG signals. In the audio system, each utterance is represented by 72 audio features, whereas, in EEG-based system, each EEG recording is represented by 112 features. For this purpose, 142 volunteers made recorded English speech. EEG signals of 58 of them were recorded during English speech, using Emotiv EPOC headset. With a help of English expert, each participant had been evaluated and categorized into three different English skill levels. With 1-fold cross-validation, 20 audio recordings from each level were used for training and testing audio-based system. Similarly, 10 EEG recordings from each skill level were used for training and testing EEG-based system. The audio-based system outperformed EEG-based system with 68% accuracy compared with 56% accuracy for the EEG system.

In the future work, more data to be recorded for more participants, specifically EEG data. The two-subsystems will be combined together at the feature level, i.e. concatenating audio features and the EEG features into one feature vector. It will be also interesting to combine the sub-systems at the model level, i.e. train a back-end classifier which combines scores of the two sub-systems and compare its result with the feature level concatenation and also with each individual sub-system.

REFERENCES

[1] T. Kawahara and M. Minematsu, "Computer-assisted language learning call system," in *Proceedings of 13th annual conference of the international speech communication association ISCA,* 2012.

[2] B. Granström, "Speech technology for language training and e-inclusion," in *Ninth European Conference on Speech Communication and Technology,* 2005.

[3] B. Mak *et al.*, "Plaser: Pronunciation learning via automatic speech recognition," in *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing,* 2003, pp. 23–29.

[4] H. Strik, A. Neri, and C. Cucchiarini, "Speech technology for language tutoring," *Proceedings of LangTech-2008,* 2008, pp. 73-76.

[5] A. Neri, C. Cucchiarini, and H. Strik, "Automatic speech recognition for second language learning: how and why it actually works," in *Proc. ICPhS,* 2003, pp. 1157–1160.

[6] L. Chen, K. Zechner, and X. Xi, "Improved pronunciation features for construct-driven assessment of non-native spontaneous speech," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics,* 2009, pp. 442–449.

[7]   A. Chandel *et al.*, "Sensei: Spoken language assessment for call center agents," in *2007 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU),* 2007, pp. 711–716.

[8]   "Pearsoneducation, inc.versanttmaviationenglishtest," Accessed: Sep 30, 2018. [Online]. Available: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/versant-test/Presentations/VAET-ATP2008.pdf.

[9]   M. Eskenazi, "An overview of spoken language technology for education," *Speech Communication,* vol. 51, no. 10, pp. 832–844, 2009.

[10]   S. M. Witt *et al.*, "Use of speech recognition in computer-assisted language learning," Ph.D. dissertation, University of Cambridge Cambridge, United Kingdom, 1999.

[11]   H. Franco, L. Neumeyer, V. Digalakis, and O. Ronen, "Combination of machine scores for automatic grading of pronunciation quality," *Speech Communication,* vol. 30, no. 2-3, pp. 121–130, 2000, doi: 10.1016/S0167-6393(99)00045-X.

[12]   C. Cucchiarini, H. Strik, and L. Boves, "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology," *The Journal of the Acoustical Society of America,* vol. 107, no. 2, pp. 989–999, 2000, doi: 10.1121/1.428279.

[13]   F. Hönig, A. Batliner, and E. Nöth, "Automatic assessment of non-native prosody annotation, modelling and evaluation," in *International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT),* 2012, pp. 21–30.

[14]   C. Cucchiarini, H. Strik, and L. Boves, "Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech," *The Journal of the Acoustical Society of America,* vol. 111, no. 6, pp. 2862–2873, 2002, doi: 10.1121/1.1471894.

[15]   L. Chen and S.-Y. Yoon, "Application of structural events detected on asr outputs for automated speaking assessment," in *Thirteenth Annual Conference of the International Speech Communication Association,* 2012.

[16]   J. Bernstein, J. Cheng, and M. Suzuki, "Fluency and structural complexity as predictors of l2 oral proficiency," in *Eleventh annual conference of the international speech communication association,* 2010.

[17]   S. Xie, K. Evanini, and K. Zechner, "Exploring content features for automated speech scoring," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics,* 2012, pp. 103–111.

[18]   J. Cheng, "Automatic assessment of prosody in high-stakes english tests," in *Twelfth Annual Conference of the International Speech Communication Association,* 2011.

[19]   J. Tepperman, T. Stanley, K. Hacioglu, and B. Pellom, "Testing suprasegmental english through parroting," in *Speech Prosody 2010-Fifth International Conference,* 2010.

[20]   H. H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.,* vol. 10, pp. 370–375, 1958.

[21]   A. Porbadnigk, M. Wester, J. P Calliess and T. Schultz, "EEG-based speech recognition: impact of temporal effects," *Pcoceedings of the 2nd International Conference on Bio-Inspired System and Signal Processing,* 2009, pp. 376-381.

[22]   "Emotiv EPOC headset," [Online]. Available: https://www.emotiv.com. (accessed May 10, 2020).

[23]   F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of mfcc," *Journal of Com- puter science and Technology,* vol. 16, no. 6, pp. 582–589, 2001, doi: 10.1007/BF02943243.

[24]   D. Talkin, "A robust algorithm for pitch tracking (rapt)," *Speech coding and synthesis,* vol. 495, 1995.

[25]   P. Boersma, "Praat: doing phonetics by computer," [Online]. Available: http://www.praat.org/, (accessed Mar, 2020).

[26]   P. Schwarz, P. Matejka, L. Burget, and O. Glembek, "Phoneme recognizer based on long temporal context," Brno University of Technology. [Online]. Available: http://speech.fit.vutbr.cz/software, 2006, (accessed Feb, 2020).

[27]   M. Van den Berg-Lenssen, C. Brunia, and J. Blom, "Correction of ocular artifacts in eegs using an autoregressive model to describe the EEG; a pilot study," *Electroencephalography and clinical Neurophysiology,* vol. 73, no. 1, pp. 72–83, 1989, doi: 10.1016/0013-4694(89)90021-7.

[28]   A. Mayeli, V. Zotev, H. Refai, and J. Bodurka, "An automatic ica-based method for removing artifacts from eeg data acquired during fmri in real time," in *2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC),* 2015, pp. 1–2, doi: 10.1109/NEBEC.2015.7117056.

[29]   Z. Zakeri, S. Assecondi, A. Bagshaw, and T. Arvanitis, "Influence of signal preprocessing on ica-based eeg decomposition," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013,* 2014, pp. 734–737, doi: 10.1007/978-3-319-00846-2_182.

[30]   S. Sanei, "Adaptive processing of brain signals," John Wiley & Sons, 2013.

[31]   A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN neuroscience,* vol. 2014, 2014, doi: 10.1155/2014/730218.

BIOGRAPHIES OF AUTHORS

Abualsoud Hanani [ID] [icons] is an assistant professor at Birzeit University, Palestine. He obtained Ph.D. Degree in Computer Engineering from the university Birmingham (United Kingdom) in 2012. His researches are in fields of speech and image processing, signal processing, AI for health and education and AI for software requirement Engineering. He is affiliated with IEEE member. In IEEE Access journal, IJMI journal, Computer Speech and Language journal, IEEE ICASSP conference, InterSpeech conference , and other scientific publications, he has served as invited reviewer. Besides, he is also involved in different management committees at Birzeit University. He can be contacted at email: abualsoudh@gmail.com.

Yanal Abusara [icons] is a graduate student at Birzeit University. She obtained Bachelor Degree in Computer Engineering from Birzeit University in 2017. Her researches are in fields of digital systems, signal processing, and speech processing. Recently, muli-modality. She is affiliated with IEEE as student member. She is also involved in NGOs, student associations, and managing non-profit foundation. She can be contacted at email: yanal1993-as@hotmail.com.

Bisan Maher [icons] is a graduate student at Birzeit University. She obtained Bachelor Degree in Computer Engineering from Birzeit University in 2017. Her researches are in fields of digital systems, signal processing, and speech processing. Recently, muli-modality. She is affiliated with IEEE as student member. She is also involved in NGOs, student associations, and managing non-profit foundation. She can be contacted at email: bissan007@hotmail.com.

Inas Musleh [icons] is a graduate student at Birzeit University. She obtained Bachelor Degree in Computer Engineering from Birzeit University in 2017. Her researches are in fields of digital systems, signal processing, and speech processing. Recently, muli-modality. She is affiliated with IEEE as student member. She is also involved in NGOs, student associations, and managing non-profit foundation. she can be contacted at email: inassamir16@gmail.com.