Speaker Identification: A Novel Fusion Samples Approach

Article *in* International Journal of Computer Science and Applications · July 2016

Speaker Identification: A Novel Fusion Samples Approach

Qeethara Kadhim Al-Shayea
MIS Department
Al-Zaytoonah University of Jordan
Amman, Jordan

Muzhir Shaban Al-Ani
Computer Science Department
Al-Anbar University
Anbar, Iraq

***Abstract*— Speaker identification systems are an important part of the biometric techniques. Many speaker identification systems were designed and implemented during the last few years and these systems depend on different techniques. This paper presented a simple speaker identification approach based on fusion via samples and statistical approach to generate the adequate features. This approach describes a simple method that employs statistical approach to generate feature vectors that were defined each speaker.**

Keywords-component; speaker identification; speaker recognition; feature extraction; windows and fusion approach.

I. INTRODUCTION

Password and identification card are well known traditional security systems that can be broken easily. Using of biometrics features have added many benefits on the traditional verification methods and avoid many of their threats. Biometric Recognition systems make use of biometric features in both forms Biological and Behavioral to establish biometric personal identification. Biometrics is the oldest known documents in the history of China, Egypt and Ancient Babylon. Recently biometric recognition systems become very important field of processing for their huge amount of applications especially in security [1,2]. Biometrics can be categorized into two types [1,2,3]: Biological biometric including fingerprint, face, hand geometry and iris, retina recognition etc. Behavioral biometric including voice, signature, gait, handwriting and speaker recognition etc.

A typical biometric system operates by recovering and acquiring of biometric data from a human, extracting features from the input data, and comparing these features with the template features that stored in the biometric database. Speaker identification is one of the important behavioral biometric that has been widely used in security issues. Some of speaker's identification are speech dependent and others are speech dependent. Most speaker recognition system have three important phases; the feature extraction phase, the speaker model phase, and the matching algorithm phase [4,5].

In this work, we proposed speaker identification system based on fusion via samples and statistical approach. This proposed approach describes a simple method to generate feature vectors that were defined each speaker.

II. RELATED WORK

As we mentioned above that speaker identification play an important role in our life for their wide range of fields. One of the most important field is security systems. After their study in [3], Jain, Nandakumar and Ross mentioned that despite the challenges that remain, the biometrics community can celebrate its accomplishments over the past 50 years. In addition, many works are published in this field, some of these works are explained briefly as below:

Hasan et al. [6] presented a security system based on speaker identification. Mel frequency Cepstral Coefficients have been used for feature extraction and vector quantization technique is used to minimize the amount of data to be handled. The feature matching techniques used in speaker recognition include, Dynamic Time Warping (DTW), Hidden Markov Modeling (HMM), and Vector Quantization (VQ). VQ is used to minimize the data of the extracted feature.

Busso et al. [7] presented a smart room approach for speaker localization and identification. The proposed scenario applied using two meeting with four participants acting at real time for five-minute long. The proposed algorithm for participant's localization needed about 3 seconds per participant to converge during the start of the meeting. The implemented system demonstrated that complementary modalities can increase the general participant identification and localization including the active speaker identification and localization.

Al-Ani et al. [8] proposed a hybrid speaker identification system using neural networks and discrete wavelet transform. The system is divided into four main components including data collection, preprocessing, discrete wavelet transform and model generating for speakers. The obtained results indicated that the evaluation criteria parameters obtained were; FAR=14.5% and FRR=24.5%.

Gudnason and Brookes [9] explained a novel method of speaker recognition that is based on the voice source signal. They applied closed-phase LPC analysis for feature extraction process in order to estimate the vocal tract transfer function. The proposed approach was applied on two databases; TIMIT and YOHO. The misclassification rate for TIMIT and YOHO were reduced from 1.51% to 0.16% and from 13.79% to 10.07% respectively.

Avci [10] presented an expert speaker identification system for speaker identification using Turkish speech signals. A discrete wavelet adaptive network based fuzzy inference system (DWANFIS) model is used in this study. The rate of correct classification is about 90.55% for the sample speakers.

Tzagkarakis and Mouchtaris [11] explained two methods for noise-robust text-independent speaker identification are described and compared against a baseline method for speaker identification based on the Gaussian Mixture Model. These methods are statistical approach based on the Generalized Gaussian Density (GGD) and Sparse Representation Classification (SRC) method. The results indicated that the SRC approach significantly outperforms the other two methods under the short test and training sessions restriction.

Wu and Tsai [12] developed speaker identification system based EMD for feature extraction and classification using artificial neural network. As a classifiers BPNN and GRNN were used. They mentioned that GRNN is much faster, also, GRNN performed well in the recognition rate. The experimental results in their paper revealed that the proposed EMD method with GRNN can accomplish speaker identification in a short time and achieve a satisfactory recognition rate.

Rossi, Amft and Troster [13] proposed a collaborative personal speaker identification approach. This approach can be generally applied in different use scenarios. They concluded that allowing unknown speakers in a conversation does not hamper system performance and gains achieved through collaboration. They found that their collaborative fusion provides benefits even in situations, where only one system knows the actual speaker. In this situation, collaborating systems indirectly elevate the correct speaker by returning low matching scores for their models.

Herbig, Gerl and Minker [14] presented a novel approach for joint speaker identification and speech recognition. identification.

Mhadse and Wani [15] explained speaker identification based automation system to control lights and electrical appliances in a home or office using voice commands. A combination of speech processor and MATLAB® software are used to perform and recognize speech processing.

Schmidt et al. [16] implemented a system based on I vectors, a current approach for speaker identification, and locality sensitive hashing, and an algorithm for fast nearest neighbor search in high dimensions. The implemented algorithm is faster than a linear search while maintaining the identification accuracy of an I vector-based system.

III. Speaker Identification

If the internal evidence suggests that it is 'fully possible' that two samples were spoken by the same individual, we can only go further, that is, make the leap towards probability, if their shared phonetic characteristics are unlikely to be found in combination in any other individual in the relevant population [17,18]. Various factors affecting the reliability of identification and the experimental results showed that the error rate ranging from 6% to 63% [19,20,21]. Voiceprint identification techniques and their effects on changing phonetic context was implemented [22,23]. A linear predictive vector code books and hidden Markov models for speaker recognition and speech signal was applied [24,25]. An index audio based on content words using either large vocabulary speech recognition or keyword spotting [26,27,28,29]. In [30,31] a text independent speaker identification based on Gaussian mixture modeling approach was presented.

IV. Speaker Identification Proposed Approach

A. *Database Construction*

The database is constructed of conversational speech from 20 male speakers of 5 samples each of 7 words. The duration of each passage ranges from 3.5 to 4.5 seconds. The speech from speakers are recorded via high quality microphone. All of the passages were recorded at 44.100 kHz, 16 bits, in mono PCM wave.

B. *System Phases*

The speaker identification proposed approach based on two phases: Enrollment Phase (Learning Phase) in which the samples of speech are collected from different speakers and then trained to construct the speaker database as shown in Fig. 1.

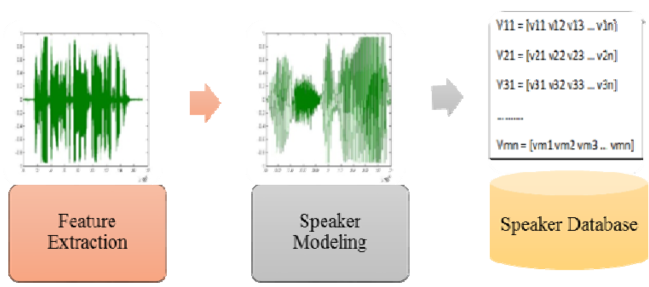


FIGURE 1.ENROLLMENT PHASE

Identification Phase (Testing Phase) in which a test sample from an unknown speaker is compared with the speaker database as shown in Fig. 2. Both phases include the same initial step, feature extraction, which is used to extract speaker dependent characteristics from speech.

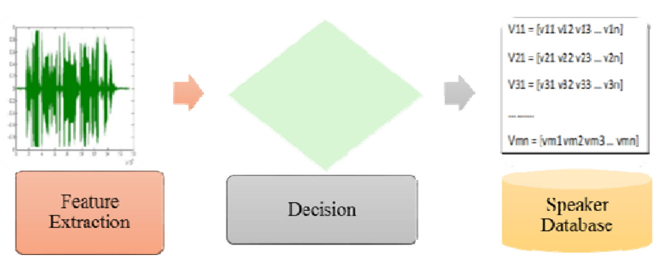


FIGURE 2. IDENTIFICTION PHASE

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

C. System Design and Implementation

The main components of the implemented system are shown in Fig. 3:

a) Digitizing, including converting the speech data into digital form.

b) Preprocessing, including noise removal, windowing and resizing.

c) Feature extraction, including principle component analysis (PCA).

d) Data fusion, including fusion of the extracted data to generate the adequate feature vector.

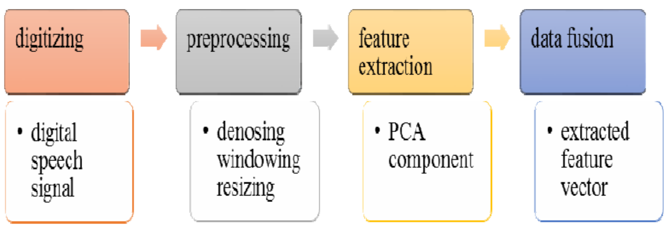


FIGURE 3.COMPONENTS OF THE PROPOSED SYSTEM

V. RESULTS AND ANALYSIS

A. Data Acquisition and Preprocessing

The speaker samples are recorded and saved, then these samples are resizing to be adequate for processing. The processing speech for each speaker is segmented into the indicated words as shown in Fig. 4.

FIGURE 4.PREPROCESSING OF SPEECH SIGNAL

B. Performance Evaluation

The tested speech is preprocessed and resized, then analyzed to extract a sequence of feature vectors. The sequence of feature vectors is divided into overlapping segments of feature vectors:

$$1S_{ID} = \{x1, x2, x3 \ldots xn\} \quad (1)$$

$$2S_{ID} = \{x1, x2, x3 \ldots xn\} \quad (2)$$

and so on.

The performance evaluation is computed as the percentage of the correctly identified segments over all tested segments

$$PC_{ID} = \frac{\text{number of correctly identified segments}}{\text{total umber of segments}} \quad (3)$$

According to the tested speech data we indicated that the identified segments are 19 over 20 means that the recognized PC_{ID} is 95% that indicated a good recognition.

C. Statistical Measures

The statistical measures dedicated to compare two wave signals as shown in table I:

PSNR: Peak Signal to Noise Ratio.

MSE: Mean Square Error.

MAXAE: Maximum Absolute Error.

ERATIO: Energy Ratio.

S: Number of Speakers.

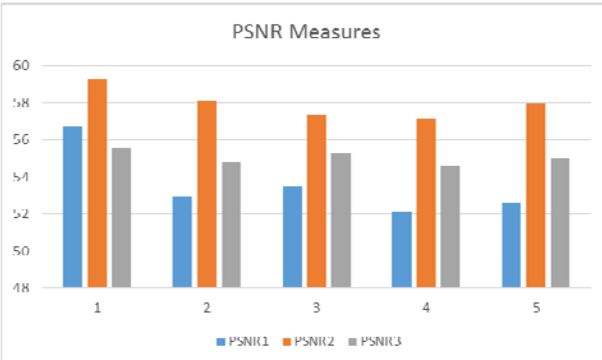
W: Number of Words.

TABLE I. STATISTICAL MEASURES

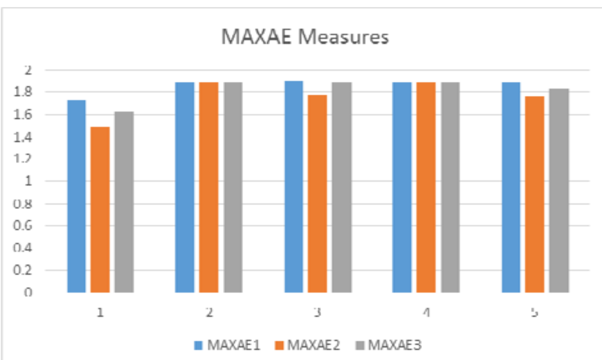
Original Word	Tested Word	PSNR	MSE	MAXAE	ERATIO
S1_W1	S1_W1	∞	0	0	1
S1_W1	S1_W2	56.6941	0.1392	1.7287	0.9789
S1_W1	S1_W3	52.9226	0.3318	1.8932	0.9276
S1_W1	S1_W4	53.4852	0.2914	1.8977	1.1879
S1_W1	S1_W5	52.0971	0.4012	1.8953	1.0691
Average		**52.6505**	**0.3564**	**1.8954**	**1.0634**
S2_W1	S2_W1	∞	0	0	1
S2_W1	S2_W2	59.2624	0.0771	1.4871	0.2536
S2_W1	S2_W3	58.0847	0.1011	1.8849	0.8203
S2_W1	S2_W4	57.3250	0.1204	1.7799	0.5749
S2_W1	S2_W5	57.1570	0.1251	1.8837	0.7915
Average		**57.9573**	**0.1059**	**1.7589**	**0.6101**
S1_W1	S2_W1	54.5623	0.2274	1.8915	0.4695
S1_W1	S2_W2	55.5461	0.1813	1.6230	0.1190
S1_W1	S2_W3	54.7831	0.2162	1.8885	0.3851
S1_W1	S2_W4	55.2568	0.1938	1.8898	0.2699
S1_W1	S2_W5	54.6477	0.2230	1.8882	0.3716
Average		**54.9592**	**0.2083**	**1.8362**	**0.3230**

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

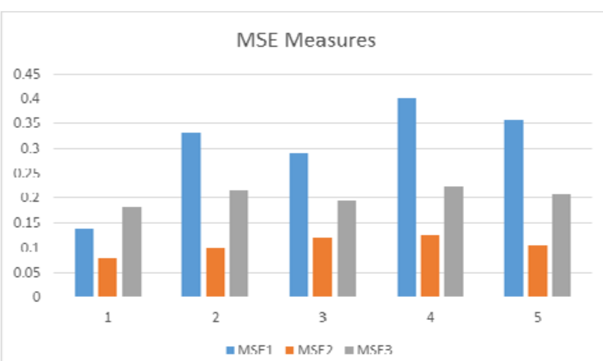
The measures are implemented between three set of testing; first speaker-first speaker, second speaker-second speaker and first speaker-second speaker respectively as shown in Fig. 5. Part (a) indicates the peak signal to noise ratio, part (b) indicates mean square error, part (c) indicates maximum absolute error and part (d) indicates energy ratio. From the extensive check of Fig. 5 we can denote that the histogram of the similar speaker gives nearest distance of that values.



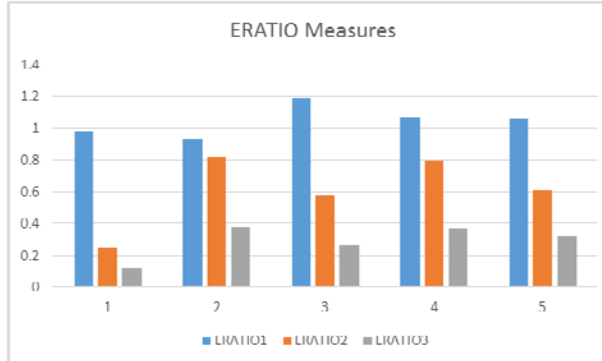
(A) PEAK SIGNAL TO NOISE RATIO



(B) MEAN SQUARE ERROR



(C) MAXIMUM ABSOLUTE ERROR



(D) ENERGY RATIO

FIGURE 5. STATISTICAL MEASURES

VI. CONCLUSIONS

The speaker identification deals with the process that intent to extract speaker information from a sequence of spoken words. This process required an adequate acoustic environment and word repetition of speakers. In our approach we try to repeat each sentence five times for the same speaker in order to generate the required features. The implemented approach for speaker identification depends on statistical approach. These measures are peak signal to noise ratio, mean square error, maximum absolute error and energy ratio. The obtained results indicated a good recognition for similarity reached to 95%.

REFERENCES

[1] G. I. Raho, M. S. Al-Ani, A. H. Al-Hamami, R. Kanaan, "Universal Developing of Persons Identification Based on RFID", Journal of Emerging Trends in Computing and Information Sciences, vol. 6, no. 10, pp. 592-597 ,October 2015.

[2] R. S. Prasad, M. S. Al-Ani, S. M. Nejres , "Hybrid Fusion of Two Human Biometric Features", International Journal of Business and ICT, vol. 2, no. 1-2, pp. 19-27, March-June 2016.

[3] A. K. Jain, K. Nandakumar and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities", Pattern Recognition Letters, vol. 79, pp. 80-105, 2016.

[4] G. I. Raho, M. S. Al-Ani, A. Al-Alosi and L. A. Mohammed, "Signature Recognition Using Discrete Fourier Transform", International Journal of Business and ICT, vol. 1, no. 1-2,pp. 17-26, March-June 2015.

[5] Q. Al-Shayea and M. Al-Ani (2016), "Efficient Window Approach of FIR Filter Design (MSK2)", IJCSNS International Journal of Computer Science and Network Security, vol. 16, no. 2, pp. 63-68, February 2016.

[6] M. R. Hasan, M. Jamil, M. G. Rabbani and M S. Rahman, "Speaker Identification Using Mel Frequency Cepstral Coefficients", 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.

[7] C. Busso, S. Hernanz, C. Chu, S. Kwon, S. Lee, P. G. Georgion, I. Cohen and S. Narayanan, "Smart Room: Participant and Speaker Localization and Identification", IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, pp. 18-23, March 2005.

[8] M. S. Al-Ani, T. S. Mohammed and K. M. Eljebory , "Speaker Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform", Journal of Computer Science, vol. 3, no. 5, pp. 304-309, 2007.

[9] J. Gudnason and M. Brookes, "Voice Source Cepstrum Coefficients for Speaker Identification", IEEEI CASSP, 2008.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[10] D. Avci, "An expert system for speaker identification using adaptive wavelet sure entropy", Expert Systems with Applications, vol. 36, pp. 6295-6300 ,2009.

[11] C. Tzagkarakis and A. Mouchtaris, "Robust Text-Independent Speaker Identification Using Short Test and Training Sessions", 18th European Signal Processing Conference (EUSIPCO-2010), 2010.

[12] J. Wu and Y. Tsai, "Speaker identification system using empirical mode decomposition and an artificial neural network", Expert Systems with Applications, vol. 38, pp. 6112-6117, 2011.

[13] M. Rossi, O. Amft and G. Troster, "Collaborative personal speaker identification: A generalized approach", Pervasive and Mobile Computing, vol. 8, pp 415–428, 2012.

[14] T. Herbig, F. Gerl and W. Minker, "Self-learning speaker identification for enhanced speech recognition", Computer Speech and Language, vol. 26, pp. 210–227, 2012.

[15] P. V. Mhadse and A. C. Wani, "Speaker Identification Based Automation System Through Speech Recognition", International Journal of Emerging Technology and Advanced Engineering, vol. 3, Issue 1, January 2013.

[16] L. Schmidt, M. Sharifi and I. L. Moreno, "Large-Scale Speaker Identification", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1650-1654, 2014.

[17] B. M. Arons, Speech Skimmer: Interactively Skimming Recorded Speech. Ph.D. thesis, MIT Media Laboratory, 1994.

[18] D. Roy and C. Schmandt, NewsComm: A Hand-Held Device for Interactive Access to Structured Audio. Master's thesis, MIT Media Laboratory,1995.

[19] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett and K. M. Stevens , "Identification of a Speaker by Speech Spectrogram", American Association for the Advancement of Science, vol. 166, pp. 338-343, October 1969.

[20] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett and K. M. Stevens, "Identification of a Speaker by Speech Spectrogram: A Scientist Views of its Reliability for Legal Purposes", Journal of Acoustic Society America, vol. 47, no. 2, pp. 597-612, February 1970.

[21] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett and K. M. Stevens, "Speaker Identification by Speech Spectrogram", The Journal of the Acoustical Society of America, vol. 54, no. 2, pp. 531-534, 1973.

[22] B. N. Hazen, "Speaker Identification Using Spectrograms mad on different spectrographs", M. S. Thesis. State University of New York at Buffalo, 1972.

[23] B. N. Hazen, "The effects of Changing Phonetic Context on the Voiceprint Identification Technique", Ph. D. Thesis. State University of New York at Buffalo, 1972.

[24] R. E. Helms, "Speaker Recognition Using Linear Predictive Vector Code Books", Ph. D. Thesis. Southern Methodist University, 1981.

[25] A. B. Boritz, "Linear Predictive Hidden Markov Models and the Speech Signal", Proceeding IEEE ICASSP, pp. 1291-1294, May 1982.

[26] D. A. James, "A System for Unrestricted Topic retrieval from Radio News Broadcasts", In Proc. of ICASSP, IEEE, Atlanta, May 1996.

[27] D. Roy, NewsComm: A Hand-Held Device for Interactive Access to Structured Audio. Masters thesis, MIT Media Laboratory, 1995.

[28] D. Roy and C. Malamud, "Speaker Identification based Text to Audio Alignment for an Audio Retrieval System", In Proc. ICASSP, IEEE, Munich, 1997.

[29] G. J. F. Jones, J. T. Foote, K. S. Jones, and S. J. Young, "Robust Talker-Independent Audio Document Retrieval", In Proc. ICASSP, pp. 311-314, IEEE, Atlanta, 1996.

[30] D. A. Reynolds, "A Gaussian Mixture Modeling Approach to Text Independent Speaker Identification", Ph. D. Thesis, Georgia Inst. Of Technology, Sep. 1992.

[31] D. A. Reynolds, "PC based TMS320C30 Implementation of a Gaussian Mixture Model Text Independent Speaker Recognition System", Proceedings Int. Conf. Signal Processing Applications Technology, Nov. 1992, pp. 967-973.

AUTHORS PROFILE

Qeethara Kadhim Abdulrahman Al-Shayea is Associate Professor in Department of Management Information Systems Faculty of Economics & Administrative Sciences Al-Zaytoonah University of Jordan. She has received Ph. D. in Computer Science, Computer Science Department, University of Technology, Iraq, 2005. She received her M.Sc. degree in Computer Science, Computer Science Department from University of Technology, Iraq, 2000. She has received her High Diploma degree in information Security from Computer Science Department, University of Technology, Iraq, 1997. She has received B. Sc. Degree in Computer Science Department from University of Technology, Iraq, 1992. She is interested in Artificial Intelligent, Business Intelligence, Image Processing, Computer Vision, Signal Processing and Information Security. She has already published many papers in international journals and conferences.

Muzhir Shaban Al-Ani has received Ph. D. in Computer & Communication Engineering Technology, ETSII, Valladolid University, Spain, 1994. Assistant of Dean at Al-Anbar Technical Institute (1985). Head of Electrical Department at Al-Anbar Technical Institute, Iraq (1985-1988), Head of Computer and Software Engineering Department at Al-Mustansyria University, Iraq (1997-2001), Dean of Computer Science (CS) & Information System (IS) faculty at University of Technology, Iraq (2001-2003). He joined in 15 September 2003 Electrical and Computer Engineering Department, College of Engineering, Applied Science University, Amman, Jordan, as Associated Professor. He joined in 15 September 2005 Management Information System Department, Amman Arab University, Amman, Jordan, as Associated Professor, then he joined computer science department in 15 September 2008 at the same university. He joined in August 2009 College of Computer Science, Al-Anbar University, Al-Anbar, Iraq, as Professor.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500