Automatic Speaker Recognition System

Article Info

Received: 10/1/2016
Accepted:15/2/ 2016
Published online: 1/3/2016

Feras E. AbuAladas, Akram M. Zeki

International Islamic University Malaysia
Ajloun National University
Firas4_ads@yahoo.com, akramzeki@iium.edu.my

ISSN 2231-8844

Abstract

Speech contains many features that can be used to determine both gender and speaker identity and it is a natural form of communication between humans. Speech processing has been one of the most sensational areas of signal processing. Human voice is considered one of the main biometrics that could be used to identify person where the unique characteristics of one or more of biometrics for each person can be used. One of the biometrics that a person can be distinguished by his\here voice. Thus, we point to speaker recognition systems as those technologies, which used human speech to recognize. This paper presents an overview of automatic speaker recognition system (ASR) technique used for feature extraction such as MFCC, LPCC, and wavelet transformation.

Keywords: **Speaker recognition; biometrics; verification; identification; Feature extraction**

1. Introduction

Pattern recognition discriminate signals, image or object, depending on a given set of parameter called features. The term ‘pattern’ denotes the n-dimensional data vector X=(x1,x2,…xn)T of measurements, whose components xi (i=1..n) are the object features. The features are variables (specified by the researcher) that are considered significant for classification (object discrimination). In discrimination, assume that there exist C groups or classes (G1, G2,…, Gc), each pattern x is associated with a categorical variable z that indicates the class or group membership; that is, if z =j, then the pattern belongs to Gj (j=1, 2,…,C). Pattern recognition is regarded as a basic attribute of human beings, as well as other living organisms (Webb, 2003).

Aauthentication of personal identity through biometrics is one of application, fall under pattern recognition. Bbiological characteristic used by of Biometric technology should be, unique for each person where there is little potential that other individual can replace these features or stolen or forged). Human voice is one of the biometrics where person can be distinguished by is his/here voice, thus we refer to voice recognition systems as those technologies which utilize human speech to recognize each individual from other (Togneri & Pullella, 2011).

Speaker recognition is the process of recognize the speaker based on features that extracting from his speech, were all us have different voices and cannot be exactly duplicated. (Malode & Sahare, 2012)

Automatic speaker recognition is divvied in two categories speaker identification and speaker verification. In speaker verification system the compression is made only (one to one data set) and person is authenticated if he /she is the one who she/he claims to be figure (1) shows the basic structure of speaker verification.

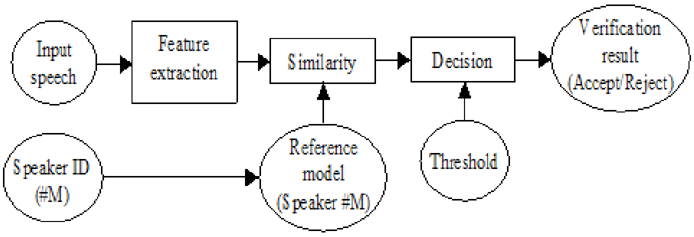

Fig. 1. Basic structure of speaker Verification

On the other hand speaker identification system comparison is made (one to many data set) to find the one that match, where speaker identification is the process of finding the identity of an unknown voice, figure (2) shows the basic structure of speaker identification (Powar & Patil) Speaker recognition has two component, feature extraction techniques and feature matching. Feature extraction is the task of extract a tiny set of data from the person voice that can be represent whole speech signal and used later to represent every speaker, in addition to that Feature matching include the procedure to classify the unknown speaker where the extracted features from his/her voice input comparing with everyone from a set of known speakers.(Singh & Khan, 2015).

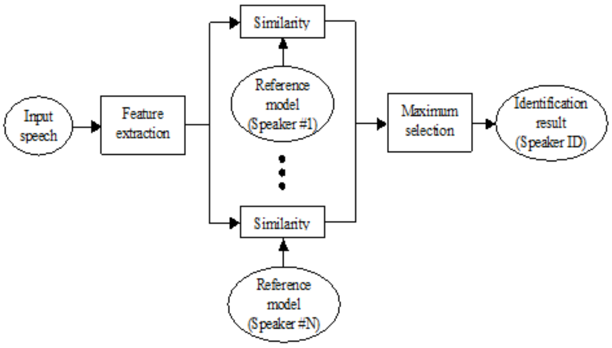
Similarity
Reference model (Speaker #1)
Input speech
Feature extraction
Maximum selection
Identification result (Speaker ID)
Similarity
Reference model (Speaker #N)

Fig. 2. Basic structure of speaker identification

In general speaker recognition system consist of three phases:the preprocessing phase were the computer records the voice, feature extraction phase to analysis and extract the main feature of voice, recognition phase (discrimination algorithm or classifier). (Rashed & Bahgat, 2013).

II. Related work

Some strategies for speaker recognition have been proposed in the past few years. These methods can achieve high performances on good quality data which are captured in controlled conditions. Generally, some of these researches are summarized below with different techniques.

AboElenein, Amin, Ibrahim, & Hadhoud, (2016) Proposed mel frequency cepstral coefficient MFCC and Vector Quantization (VQ) techniques to obtained vector of features without lost in information. Gaussian Mixture Model (GMM) with gender detection used as classification to train and test features, where each speaker spoke six sentences , 4 for training and 2 for testing .Experimental results showed that the accuracy of proposed algorithm is 91% in comparative with VQ and GMM where the accuracy is 88% and reduce the time processing about 50%.

Ramakrishnan et al., (2015) proposed integrated linear prediction residual (ILPR) with pitch synchronous (PS) mix with discrete cosine transform (DCT) as an altrnative way of characterizing the voice source (VC) and regarded as an features for speaker identification (SID), as speakers models, Gaussian mixture model (GMM) used to capture the change in feature from speaker tp speaker , three differents data base used TIMIT, YOHO and NIST 2003 databases used, the expermental resulets showed that MFCC features improves the identification accuracy by 12% in absolute terms, and proving that the proposed DCTILPR has good promise as a feature for SID studies.

Djemili et al., (2015) they proposed an algorithms to improving the performance of a speaker identification system based on a frame level scoring, as a features-set Mel-frequency cepstral coefficients (MFCC) considered and Gaussian mixture (GMM) used for modeling speaker identification system, The speaker proposed algorithm for speaker identification applied on IVIE corpus by selecting 120 speakers randomly from TIMIT database and use an identification error rate to measured the final performance. Experimental results based on IVIE showed that a relative reduction in error rates of 24.4 while in TIMIT is 37.3%.the final performance based on IVIE and TIMIT are 3.4% and 5.2% respectively.

Al-Hmouz et al., (2015) Investigated different multimodal speaker identification approach to show the effectiveness of the multimodal system in minimize the limitations correlated with any single biometric method such as reduced in accuracy, finite security, noisy measurements. TIMIT dataset used for evaluate multimodal speaker identification, Mel-frequency cepstral coefficients (MFCC), Linear predictive coding(LPC) and discrete wavelet based linear predictive coding(DWLPC) are used as a features extraction techniques, these features are examined, combined as hybrid approach and modeled by Gaussian mixture model (GMM) for speaker identification, Experimental results show that The best performance achieved when combining MFCC and DWLPC when the speaker identification is evaluated in the noisy environments, and all features extraction methods there is an improvement in classifications rate.

Nandyal (2015) presented a robust approach for text dependent voice recognition. where the speaker can speak only fixed text, for speaker identification/verification system, for represent feature Mel Frequency Cepstrum Coefficients (MFCC) are used, these features are used for train ANN classifier in enrolment phase, Experimental results show that recognize accuracy achieved by using multilayer neural network (feed-forward back propagation) as a classifier is around 92% and obtained 12% false rejection rate.

Mathur & Sharma, (2015) Proposed Mel frequency Cepstrum Coefficient (MFCC) as an algoriths to extract features for speaker identification system, these features are quantized where each vector is represented by a number of centroids by applying vector quantization algorithm, speaker identified according to the minimum distance between centroids in testing phase and MFCC's in training phase, Experimental results showed that the maximum performance of using MFCC's at 32 filter for speaker recognition is 80%.

Yadav & Bhalke (2015) in this paper they proposed system for speaker identification by sing acoustic characteristics in speech signal, the signal is pass through preprocessing phase to detect and remove the silent from speech, signal is decomposed at two level by using discrete wavelet transform, Traditional MFCC and discrete wavelet transform based (MFCC) are used as a features-set, different of 15 speakers is used from TIMIT Database to determine a speaker identity, Vector quantization is used. Experimental results showed that the use of traditional MFCC give an 80% accuracy while used MFCC based DWT give 85% accuracy and the maximum of accuracy achieved when use MFCC based DWT, Vector Quantization technique using LBG algorithm as afeatures extraction technique is used.

Nijhawan & Soni (2014) In thi paper feature extraction technique Mel frequency Cepstral coefficients (MFCC) is used and quantized using Vector Quantisation-Linde, Buzo, and Gray (VQLBG) algorithm for speaker recognition system (SRS), under noisy condition Voice Activity Detector (VAD) has been used to distinguish between silence and voice activity to improves the performance of SRS. for speaker identification recognition Euclidean distance approach is used to comparing the features of new recorded voice against database. The Experimental results showed that the accuracy of recognition around 95% for 256 numbers of centroids and no false recognition.

Shah & Ahsan (2014) Proposed an automatic speaker identification system to discriminate (Quran reciter) of Arabic Language.Improving identification accuracy done by using Discrete wavelet transformation (DWT) and linear predicative code (LPC) as a fetures extraction techniques were these features used individually (one at a time) and combind to train Random Forest (RF) classifier, three approches are used with differents number of sample and to investigate the performance of classifier. The experimental results showed that using combination features for train a calssifer gives best performance and can improve the recognition accuracy in comparative of use features one at a time were 90.90% identification accuracy was achieved.

Nagaraja & Jayanna 2013) They study the effect of combined the features extracted from Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC) in comparative of use features individualy for speaker identification system in context mono, cross and multilingual, created data of 30 speakers were each one recored his/here voice in three different langauges (english, hindi and kannada) langauges. The experimental results showed that the number of speakers identified by MFCC is 18 and 20 speakers by LPC while the number of speakers when combination of features (MFCC and LPC) is 22. This concludes that use MFCC and LPC features combined together instead of using (one at time) improving the speaker identification performance about 30% for created dataset.

Dash et al., (2012) Proposed a speaker recognition system by using Mel Frequency Cepstrum Coefficient (MFCC) features, these features are trained and tested by BPNN for identification of speaker after quantized using vector quantization (VQ) algorithm, were minimum Euclidean distance used to identify the speaker, to find the preferable implementation number of filters of MFCC were it changed to (12, 22, 32, and 42) and type of window are considered. The experimental results showed that the maximum performance was reached at 32 number of filter were the efficiency is 85% while the using hamming window with same filter decrease the efficiency to 75% .

Kumar et al. (2011) they made a comparative study of Mel Frequency Cepstral Coefficient (MFCC) and Linear Prediction Coefficient (LPC) features, these features are extracted from clean and noisy database for speaker identification , noisy database wase preperd by adding speech and F16 noises, MFCC and LPC features trained and tested by Gaussian mixture model (GMM), determine the recognized speaker depends on maximum log likelihood of each testing feature vector with the (GMM). experimental results showed that the performance of identification of clean database is 96.65% with MFCC and 93.65% with LPC for both noises (speech and F16) MFCC and LPC reached 88.02, 82.07 respectively. This concludes that the use of MFCC increased the performance of identification.

Harrag et al. (2011) presented a feature selection algorithm based on genetic algorithm optimization for Arabic speaker recognition system. The proposed algorithm adopts classifier performance and the number of the selected features as heuristic information and selects the optimal feature subset in terms of smallest feature set size and the best performance of classifier, feature vectors containing Mel-Frequency Cepstral Coefficients (MFCCs) are used and K-Nearest-Neighbor (KNN) classifier performance and the length of selected feature vector are considered for performance evaluation. The experimental results, showed that our GA is able to select the more informative features without losing the performance and can obtain better classification accuracy with a smaller feature set for various speakers ,these features is crucial for real time application and low resources devices, for automatic speaker recognition (ASR) the features vectors containing Mel-Frequency Cepstral Coefficients (MFCCs) where the size of features reduced over 60% that led to a less complexity of our ASR system and reduce the error rate (ER) to 25%.

Wu & Lin (2009) Proposed a system for speaker identification depends on the speaker utterances, discrete wavelete transformation and transform and wavelet packet transform (WPT) used for features extraction , these features fed to a general regressive neural network (GRNN) to evaluated the effective of features . Experimental results of DWT, conventional WPT and WPT in Mel Scale methods showed that the recognition rate was increased while extraction time is constant were the recognition rate of proposed irregular decomposition of WPT for speaker identification system is is 96.6 % with 57 features this demonstrated the effective of using fewer features in compartive with DWT, 5-level WPT, 6-level WPT, 7-level WPTand WPT in Mel scale 70.8, 71.6, 94.6, 97.8, 94.4, respectively, were 7-level WPT used 128 features.

Chen et al. (2004) Used multi-resolution property of the wavelet transform to improved the speaker identification performance, the linear predictive cepstral coefficients (LPCCs) is used for feature extraction after decomposed speech signal into various frequency band, these features calculate from each band and fed to train a classifiers were Gaussian mixture model (GMM) is used. Two identifier approaches for speaker identifications system, feature combination Gaussian mixture model (FCGMM) and likelihood combination Gaussian mixture model (LCGMM), FCGMM combined LPCC features that extracted from each subband to build a single feature vector to train GMM. LGMM recombined LPCC features that extracted from each band and fed it to different GMM classifier, KING database used to evaluate these two approaches. The experimental results showed that LGMM is better and more effective than FCGMM and both approaches are more effective than MFCC and conventional GMM using full-band LPCC, were the best performance achieved by LGMM is 94.96% in clean environments.

III.CONCLOUSION

This paper has reviewed the researches done in the area of automatic speaker recognition. Several techniques for feature extraction and classification have been discussed. Some techniques are preferred over other such as MFCC in feature extraction, in other hand these technique can be integrated with each other to increase the accuracy of speaker recognition system such as DCTILPR & MFCC, DWT&LPC, LPC&MFCC were the development can be occurs in this stage that concentrated on reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy.

Table1. Comparison of different feature extraction technique.

Ref	**Technique**	**Finding**
1	MFCC	MFCC
14	(DCTILPR), MFCC DCTILPR & MFCC	DCTILPR&MFCC
5	MFCC	MFCC
2	LPC, MFCC DWLPC LPC&MFCC DWLP&MFCC	LPC,DWLPC
11	MFCC	MFCC
9	MFCC	MFCC
21	MFCC,DWMFCC	DWMFCC
12	MFCC	MFCC
16	DWT, LPC DWT&LPC	DWT&LPC
10	LPCC MFCC&LPCC	MFCC&LPCC
4	MFCC	MFCC
7	MFCC, LPC	MFCC
6	MFCC	MFCC
20	WPT- Mel Scale DWT	WPT- Mel Scale
3	LPCC, MFCC	MFCC

REFERENCES

[1] AboElenein, N. M., Amin, K. M., Ibrahim, M., & Hadhoud, M. M. (2016). *Improved text-independent speaker identification system for real time applications.* Paper presented at the 2016 Fourth International Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC).

[2] Al-Hmouz, R., Daqrouq, K., Morfeq, A., & Pedrycz, W. (2015). *Multimodal biometrics using multiple feature representations to speaker identification system.* Paper presented at the 2015 International Conference on Information and Communication Technology Research (ICTRC),.314-317

[3] Chen, W.-C., Hsieh, C.-T., & Lai, E. (2004). Multiband approach to robust text-independent speaker identification. *Journal of Computational Linguistics and Chinese Language Processing, 9*(2), 63-76.

[4] Dash, K., Padhi, D., Panda, B., & Mohanty, S. (2012). Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(4).

[5] Djemili, R., Korba, A., Bourouba, H., & O'Shaughnessy, D. (2015). *Boosting speaker identification performance using a frame level based algorithm.* Paper presented at the 2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 1-6

[6] Harrag, A., Saigaa, D., Boukharouba, K., Drif, M., & Bouchelaghem, A. (2011). *GA-based feature subset selection: Application to Arabic speaker recognition system.* Paper presented at 2011 11th International Conference on the Hybrid Intelligent Systems (HIS),.

[7] Kumar, P., Jakhanwal, N., & Chandra, M. (2011). *Text dependent speaker identification in noisy environment.* Paper presented at the 2011 International Conference on Devices and Communications (ICDeCom),.

[8] Malode, A. A., & Sahare, S. (2012). Advanced speaker recognition. *International Journal of Advances in Engineering & Technology, 4*(1), 443-455.

[9] Mathur, R., & Sharma, M. S. N. (2015). Performance Comparison of Speaker Identification using Vector Quantization by MFCC Algorithm.

[10] Nagaraja, B., & Jayanna, H. (2013). Combination of Features for Multilingual Speaker Identification with the Constraint of Limited Data. *International Journal of Computer Applications, 70*(6), 1-6.

[11] Nandyal, S. (2015). MFCC Based Text-Dependent Speaker Identification Using BPNN.

[12] Nijhawan, G., & Soni, M. (2014). Speaker Recognition Using MFCC and Vector Quantisation. *Int. J. on Recent Trends in Engineering and Technology, 11*(1).

[13] Powar, S. M., & Patil, V. Study of Speaker Verification Methods.

[14] Ramakrishnan, A., Abhiram, B., & Prasanna, S. M. (2015). Voice source characterization using pitch synchronous discrete cosine transform for speaker identification. *The Journal of the Acoustical Society of America, 137*(6), EL469-EL475.

[15] Rashed, A., & Bahgat, W. M. (2013). Modified Technique for Speaker Recognition using ANN. *International Journal of Computer Science and Network Security (IJCSNS), 13*(8), 8.

[16] Shah, S. M., & Ahsan, S. N. (2014). *Arabic speaker identification system using combination of DWT and LPC features.* Paper presented at the 2014 International Conference on Open Source Systems and Technologies (ICOSST),.

[17] Singh, N., & Khan, R. (2015). Speaker Recognition and Fast Fourier Transform. *International Journal, 5*(7).

[18] Togneri, R., & Pullella, D. (2011). An overview of speaker identification: Accuracy and robustness issues. *Circuits and Systems Magazine, IEEE, 11*(2), 23-61.

[19] Webb, A. R. (2003). *Statistical pattern recognition*: John Wiley & Sons.

[20] Wu, J.-D., & Lin, B.-F. (2009). Speaker identification using discrete wavelet packet transform technique with irregular decomposition. *Expert Systems with Applications, 36*(2), 3136-3143.

[21] Yadav, S. S., & Bhalke, D. (2015). Speaker Identification System using Wavelet Transform and VQ modeling Technique. *International Journal of Computer Applications, 112*(9).