



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Modular Arithmetic and Wavelets for Speaker Verification

¹E.F. Khalaf, ²K. Daqrouq and ²M. Sherif

¹Department of Computer Engineering,

²Department of Communication and Electronics Engineering, Philadelphia University, Jordan

Abstract: The aim of this study is to concentrate on optimizing dimensionality of feature space by selecting the number of repeating the remainder (modular arithmetic) applied for a speech signal with Wavelet Packet (WP) upon level three features extraction method. The functions of features extraction and classification were performed using the modular arithmetic, wavelet packet and three verification functions (MWVS) expert system. This was accomplished by decreasing the number of feature vector elements of individual speaker obtained by using modular arithmetic and wavelet packet method (MWM) (285 elements). To investigate the performance of the proposed MWVS method, two other verification methods were proposed: Gaussian mixture model based method (GMMW) and K-Means clustering based method (KMM). The results indicated that a better verification rate for the text-independent system was accomplished by MWVS and GMMW. Better result was achieved (91.36%) in case of the speaker-speaker verification system. In case of white Gaussian noise (AWGN), it was observed that the MWVS system is generally more noise-robust in case of using approximate discrete wavelet transform sub-signal instead of the original signal. The system works in real time. This was performed by the recording apparatus and a data acquisition system (NI-6024E) and interfacing online with Matlab code that simulates the expert system. A major contribution of this study is the development of a less computational complexity speaker verification system with modular arithmetic capable of dealing with abnormal conditions for relatively good degree.

Key words: Wavelet, modular arithmetic, gaussian mixture, K-Means, gaussian noise, speech signal

INTRODUCTION

Automatic Speaker Verification (ASV) refers to the mission of verifying speaker's identity by means of the speaker-specific information contained in speech signal. Speaker verification methods are absolutely divided into text-dependent and text-independent applications. When the same text is used for both training and testing, the system is called to be text-dependent but for text-independent process, the text used to train and test of the ASV system is totally unconstrained. The text-independent speaker verification necessitates no restriction on the type of input speech. In contrast, the text-independent speaker verification generally gives less performance than text-dependent speaker verification, which requires test input to be the same utterances as training data (Nemati and Basiri, 2010; Xiang and Berger, 2003).

Speaker verification has been the topic of active research for many years, and has many important applications where propriety of information is a concern (Lamel and Gauvain, 2000). Applications of speaker

verification can be found in biometric person authentication such as an extra identity check in credit card payments over the Internet while, the potential applications of speaker identification may be utilized in multi-user systems. For example, in speaker tracking the task is to locate the segments of given speaker (s) in an audio stream (Kwon and Narayanan, 2002; Lapidot *et al.*, 2002; Martin and Przybocki, 2001). It has also potential applications in an automatic segmentation of teleconferences and helping in the transcription of courtroom discussion.

There has been a wide spectrum of proposed approaches to speaker verification starting with very simplistic models such as those based on long term statistics. The most sophisticated methods rely on large vocabulary speech recognition with phone-based HMMs.

Feature extraction is a key stage in speaker verification systems. Speech extracted features used in a speaker verification system drop within two classes based on their related space. One class includes features defined in an unconditional or absolute and irrelative space, while the other includes features defined in a relative space. For

the first class, depiction of a speaker in the feature space is not related to any reference speaker (Naini *et al.*, 2010). While there is a momentous body of literature on features in the absolute space, very little research has been conducted for studying the properties of features extracted in the relative space. Mel-frequency cepstral coefficient (MFCC), Linear Prediction Cepstrum Coefficient (LPCC), wavelet coefficients (Afshari, 2011; Avci and Akpolat, 2006), etc., are among the most common speech features in absolute space. In recent times, Campbell *et al.* (2006) used Maximum A Posteriori (MAP) adapted GMM mean super vectors as an absolute feature with Support Vector Machine (SVM) as a discriminative model for speaker verification (Campbell *et al.*, 2006). For features defined in a relative space, each speaker in the feature space is described relative to some reference speakers. As well as, extracted features in the relative space can be applied in conjunction with any other set of techniques from the verification phase menu that are deemed more suitable. Kuhn *et al.* (2000) introduced the eigenvoices concept and represented each new speaker relative to eigenvoices (Kuhn *et al.*, 2000; Thyse *et al.*, 2000). Afterwards, other researchers used a different approach where they introduced the idea of space of anchor models to represent enrolled speakers in verification systems and to verify a test speaker in a relative feature space (Naini *et al.*, 2010; Mami and Charlet, 2002, 2003, 2006).

Speech features are often extracted by Fourier Transform (FT) and short time Fourier transform (STFT). Unfortunately, they accept signals stationary within a given time frame and may therefore lack the ability to represent localized events properly. Recently, Wavelet Transform (WT) has been proposed for feature extraction. The particular benefit of wavelet analysis possesses is the characterizing signals at different localization levels in both time and frequency domains (Derbel *et al.*, 2008; Wu and Ye, 2009; Zheng *et al.*, 2002). Furthermore, the WT is well suited to the analysis of non-stationary signals (Daqrouq *et al.*, 2010; Daqrouq, 2011). It provides an alternative to classical linear time-frequency representations with better time and frequency localization characteristics. In earlier studies, these properties were applied in speaker recognition, particularly Wavelet Packet Transform (WPT) (Wu and Lin, 2009; Lung, 2004; Avci, 2009).

Artificial neural network performance is depending mainly on the size and quality of training samples (Visser *et al.*, 2003). When the number of training data is small, not representative of the possibility space, standard

neural network results are poor. Incorporation of neural fuzzy or wavelet techniques can improve performance in this case, particularly, by input matrix dimensionality decreasing. Artificial Neural Networks (ANN) are known to be excellent classifiers but their performance can be prevented by the size and quality of the training set (Qureshi and Jalil, 2002).

The specific aim of the present study was to address this question by developing and evaluating the number of repeating the remainder (by modular arithmetic) for speech based text-independent verification of the speaker from imposters. For better investigation two other verification methods are proposed; Gaussian Mixture Model method and K-Means clustering method.

FEATURES EXTRACTION BY WAVELET PACKET

Wavelet packet: The wavelet packet method is a generalization of wavelet decomposition that offers a richer signal analysis. Wavelet packet atoms are waveforms indexed by three naturally interpreted parameters: position and scale as in wavelet transform decomposition and frequency. In the following, the wavelet transform is defined as the inner product of a signal $x(t)$ with the mother wavelet $\psi(t)$:

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) \tag{1}$$

$$W_{\psi}x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right) dt \tag{2}$$

where, a and b are the scale and shift parameters, respectively. The mother wavelet may be dilated or translated by modulating a and b .

The wavelet packets transform performs the recursive decomposition of the speech signal obtained by the recursive binary tree (Fig. 1). Basically, the WPT is very similar to Discrete Wavelet Transform (DWT) but WPT decomposes both details and approximations instead of only performing the decomposition process on approximations. The principle of WP is that, given a signal, a pair of low pass and high pass filters is used to yield two sequences to capture different frequency sub-band features of the original signal. The two wavelet orthogonal bases generated from a previous node are defined as:

$$\psi_{j+1}^{2p}(k) = \sum_{n=-\infty}^{\infty} h[n]\psi_j^p(k-2/n) \tag{3}$$

$$\psi_{j+1}^{2q}(k) = \sum_{n=-\infty}^{\infty} g[n]\psi_j^p(k-2/n) \tag{4}$$

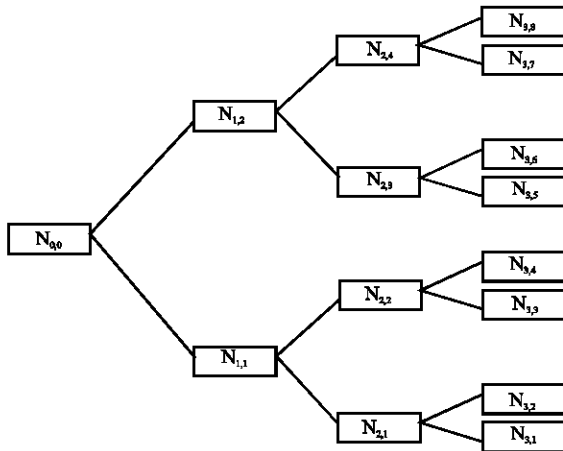


Fig. 1: Wavelet packet at depth 3

where, $h[n]$ and $g[n]$ denote the low-pass and high-pass filters, respectively. In Eq. 3 and 4, $\psi[n]$ is the wavelet function. Parameters j and p are the number of decomposition levels and nodes of the previous node, respectively (Wu and Lin, 2009).

Modular arithmetic: Modular arithmetic is referenced in number theory, group theory, ring theory, knot theory, abstract algebra, cryptography, computer science, chemistry and the visual and musical arts. Modular arithmetic provides less computational complexity (Cormen *et al.*, 2001). It performs efficiently on large numbers. Wong and Blow (2006) presented a logical design of an all-optical processor that performs modular arithmetic. Modular arithmetic could be used to process the header of all-optical packets on the fly (Wessing *et al.*, 2002).

In mathematics, modular arithmetic is a system of arithmetic for integers, where numbers wrap around after they reach a certain value. Time-keeping on a clock introduces an example of modular arithmetic (If the time is 7:00 now, then 8 h later it will be 3:00). The Swiss mathematician Leonhard Euler pioneered the modern method to congruence about 1750, when he explicitly introduced the idea of congruence modulo a number N . Modular arithmetic was developed by Carl Friedrich Gauss in his book *Disquisitiones Arithmeticae*, published in 1801 (Encyclopedia Britannica, 2010).

In fact, the notion of modular arithmetic is related to that of the remainder in division. The operation of extracting the remainder is sometimes referred to as modulo operation. Modular arithmetic can be handled mathematically by introducing a congruence relation on the integers that is well-matched with the operations of the ring of integers: addition, subtraction and

multiplication. For a positive integer n two integers a and b are assumed to be congruent modulo, written:

$$a \equiv b \pmod{n} \tag{5}$$

if their difference $a-b$ is an integer multiple of n . The number n is called the modulus of the congruence.

In computer science discipline, it is the remainder operator that is usually indicated by either "%" (e.g. in C, Java, JavaScript, Perl and Python) or "Mod" (e.g. in BASIC, SQL, Haskell and Matlab). These operators are commonly pronounced as "mod", however, it is specifically a remainder that is computed. The remainder operation can be represented using the floor function.

If $a \equiv b \pmod{n}$, where $n > 0$, then if the remainder b is calculated:

$$b = a - \left\lfloor \frac{a}{n} \right\rfloor \times n \tag{6}$$

where, $\left\lfloor \frac{a}{n} \right\rfloor$ is the largest integer less than or equal to $\frac{a}{n}$ then $a \equiv b \pmod{n}$ and $0 \leq b < n$

If instead a remainder b in the range $-n \leq b < 0$ is required, then:

$$b = a - \left\lfloor \frac{a}{n} \right\rfloor \times n - n \tag{7}$$

In this work, the WP tree consists of three stages and therefore has $G_{High} = 2^3$ high pass nodes (with original signal node) and $G_{Low} = 2^3$ low pass nodes with original signal node. More generally, for a q -stage tree, there are $G = (G_{High} + G_{Low}) = 2^{q+1} - 1$ nodes.

For a given orthogonal wavelet function, a library of wavelet packet bases is generated. Each of these bases offers a particular way of coding signals, preserving global energy and reconstructing exact features. The wavelet packet is used to extract additional features to guarantee higher recognition rate. In this study, WPT is applied at the stage of feature extraction, but these data are not proper for classifier due to a great amount of data length. Thus, it is essential to seek for a better representation for the speech features. Previous studies showed that the use of entropy of WP as features in recognition tasks is efficient. Avci and Akpolat (2006) proposed a method to calculate the entropy value of the wavelet norm in digital modulation recognition. In the biomedical field, Behroozmand and Almasganj (2007) presented a combination of genetic algorithm and wavelet packet transform used in the pathological evaluation and the energy features are determined from a group of wavelet packet coefficients. Kotnik *et al.* (2003) proposed

a robust speech recognition scheme in a noisy environment by means of wavelet-based energy as a threshold for de-noise estimation. In the study of Wu and Lin (2009) the energy indexes of WP were proposed for speaker identification. Sure entropy is calculated for the waveforms at the terminal node signals obtained from DWT (Avci, 2009) for speaker identification. Avci (2007) proposed features extraction method for speaker recognition based on a combination of three entropies types (sure, logarithmic energy and norm). In this study, the modular arithmetic to decrease the number of features in each WP node is proposed. For a speech signal in a WP node.

$\{u(t)\} = \{u(t_1), u(t_2), \dots, u(t_M)\}$, where M is the length of $u(t)$, the remainder b is written as $\text{mod}(n)_{u(t)} = 0, 1, 2, \dots, n-1$, for $n > 2$. In this study, the number of repeating the remainder $\text{mod}(n)_{u(t)}$ is utilized as a modular arithmetic wavelet speech signal feature vector:

$WR(n) = r(0), r(1), r(2), \dots, r(n-1)$. Where r is the number of repeating the remainder $\text{mod}(n)$ applied for a speech signal without WP and R denotes the same but without WP (Fig. 2).

Verification: Depending on the application, the universal area of speaker recognition is divided into two particular tasks: identification and verification. In speaker identification, the goal is to decide which one of a group of known voices best matches the input voice sample. This is also referred to as closed-set speaker identification. Applications of pure closed-set identification are limited to cases where only enrolled speakers will be encountered but it is a useful means of studying the separability of speakers' voices or finding similar sounding speakers, which has applications in speaker-adaptive speech recognition. In verification, the task is to decide from a voice sample if a person is who he or she claims to be. This is sometimes referred to as the

open-set problem, because this task requires distinguishing a claimed speaker's voice known to the system from a potentially large group of voices unknown to the system (i.e., imposter speakers) (Reynolds *et al.*, 2000).

The basis for presented verification systems is the modular arithmetic wavelet method MWM used to represent speakers. More specifically, the feature vectors extracted from a person's speech is modeled by speaker model. For a D -dimensional feature vector denoted as x , the model for speakers is defined as the average vector calculated for fifteen different utterances of a same speaker. This average vector is used to present the hypothesized speakers to be compared to background speakers, which present the imposter speakers. Background speakers are defined as the average vector calculated for fifteen different utterances of a many speakers.

Hypothesized Speaker Model: The basis for verification systems is to create feature vector pattern used to represent speakers. Therefore, in this study, WR for hypothesized speaker model presenting is utilized. More specifically, the average of WR feature vectors extracted from fifteen different person's utterances is used to represent each speaker pattern. This feature vector is compared to claimed speaker WR using verification functions.

Background speaker model: The background speakers should be selected to signify the population of expected imposters, which is in general application specific. Two issues that come up with the use of background speakers are the selection of the speakers and the number of speakers to use. Ideally the number of background speakers should be as large as possible to better model the imposter population but practical considerations of

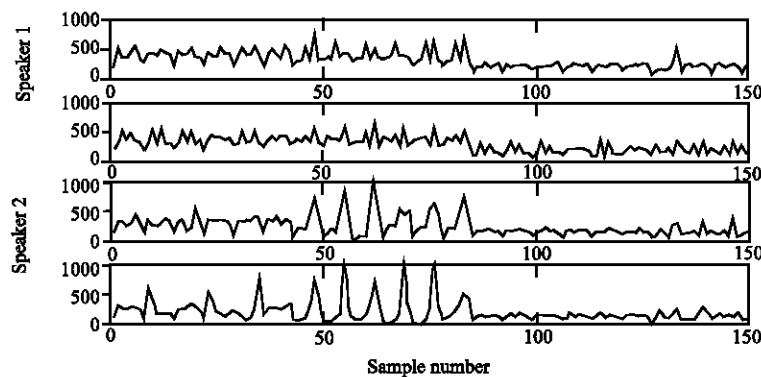


Fig. 2: First 150 coefficients of the feature vector extracted by MWM for two speakers

computation and storage state a small set of background speakers. In the verification experiments, the number of background speakers is set to three. In proposed scenario, it is assumed that imposters will attempt to gain access only from similar sounding or at least same-sex speakers.

The single-speaker detection task can be defined as a basic hypothesis test between:

- H_0 : Y is from the hypothesized speaker S and
- H_1 : Y is not from the hypothesized speaker S

The optimum test to choose between these two hypotheses is a following ratio test given by:

$$\frac{V(Y, H_0)}{V(Y, H_1)} = \begin{cases} \geq \theta & \text{Accept} \\ < \theta & \text{Reject} \end{cases} \quad (8)$$

where, $V(Y, H_i), i = 0, 1$, is the verification function for the hypothesis H_i evaluated for the observed speech segment Y.

Verification functions: In this study three verification functions are proposed:

- Percentage Root Mean Square Difference Score PRDS: This verification function is based on the distance concept given by:

$$PRDS(Y, H_i) = 100 - [100 \times \sqrt{(\sum (MW_Y - (MWM_{H_i})') / \sum (MW_Y'))^2}]^{1/2} \quad (9)$$

where, MW_Y is WR taken for observed speech segment Y, at level three of WP and MWM_{H_i} is the average of WR vectors calculated for fifteen different utterances of a same speaker (hypothesized speaker H_0) or background model (H_1). The decision threshold for accepting or rejecting is determined by:

$$\frac{PRDS(Y, H_0)}{PRDS(Y, H_1)} = \begin{cases} \geq 1 & \text{Accept} \\ < 1 & \text{Reject} \end{cases} \quad (10)$$

- Logarithmic Claimed to Signal Ratio Score CSRS:

$$CSRS(Y, H_i) = 1 - \left(\left| \log_{10} \frac{MW_Y}{MWM_{H_i}} \right| / 100 \right) \quad (11)$$

The decision threshold for accepting or rejecting is determined by:

$$\frac{CSRS(Y, H_0)}{CSRS(Y, H_1)} = \begin{cases} \geq 1 & \text{Accept} \\ < 1 & \text{Reject} \end{cases} \quad (12)$$

- Correlation Coefficient Function: This function is denoted by CC calculated between MW_Y and MWM_{H_i} . The decision threshold is obtained by:

$$\frac{CC(Y, H_0)}{CC(Y, H_1)} = \begin{cases} \geq 1 & \text{Accept} \\ < 1 & \text{Reject} \end{cases} \quad (13)$$

Verification by gaussian mixture model: Gaussian Mixture Model GMM recently has become the dominant approach in text-independent speaker identification and verification. One of the influential attributes of GMMs is their capability to form smooth approximations to arbitrarily formed densities. As a typical model based approach, GMM has been utilized to characterize speaker's voice in the form of probabilistic model. It has been reported that the GMM approach outperforms other classical methods for text-independent speaker recognition. In the following paragraph briefly mathematical development of the GMM based speaker verification scheme is introduced:

Given $Y = \{Y_1, Y_2, \dots, Y_k\}$ where, $Y = \{y_{t1} = T_1, y_{t2} = T_2, \dots, y_{tj} = T_j\}$ is a sequence of T_j feature vectors in j th cluster R^d , the complete GMM for speaker model λ is characterized by the mean vectors, covariance matrices and mixture weights from all component densities. The parameters of the speaker model are denoted by: $\lambda = \{P_{j|t}, \mu_{j|t}, \Sigma_{j|t}\}, i = 1, 2, \dots, M_j$ and $j = 1, 2, \dots$. Then, the GMM likelihood can be written as:

$$p(Y|\lambda) = \prod_{t=1}^T p(y_{t1}|\lambda) \dots p(y_{tk}|\lambda) \quad (14)$$

Equation 14, $P(y_t|\lambda)$ is the Gaussian mixture density for j th cluster and defined by a weighted sum of M_j component densities (Lung, 2007). The Expectation Maximization (EM) algorithm to create an object of the Gaussian mixture distribution class restraining maximum likelihood estimates of the parameters in a Gaussian mixture model with k components for data in the n -by- d matrix X , where n is the number of observations and d is the dimension of the data is used. In this study, verification is performed by building Gaussian mixture model by EM with 2 components of MWM vectors of two speakers feature vectors GMMW. Then the GMM likelihood is used as the verification decision whether accept or reject. This is accomplished by determining empirical threshold for decision performing.

Verification by K-mean clustering method: In this section, a brief outline of K-Means clustering algorithm

and verification by this method is presented. Clustering in N dimensional Euclidean space R^N is the process of partitioning a given set of n points into a number, say K, of clusters based on some similarity metric which establishes a rule for assigning patterns to the domain of a particular cluster centroid as seen at Fig. 3. Let the set of n points $\{x_1, x_2, \dots, x_n\}$ be represented by the set S and the K clusters is represented by C_1, C_2, \dots, C_k (Bandyopadhyay and Maulik, 2002). Then $C_i \neq \phi$ for $i = 1, \dots, k$, $C_i \cap C_j = \phi$ for $i = 1, \dots, K, j = 1, \dots, K$, where:

$$\bigcup_{i=1}^K C_i = S$$

K-Means (Hemalatha and Vivekanandan, 2008; Wagstaff *et al.*, 2001) is one of the commonly used clustering techniques, which is an iterative hill climbing algorithm. It consists of the following steps:

- Choosing K initial cluster centroids Z_1, Z_2, \dots, Z_k randomly from the n points $\{x_1, x_2, \dots, x_n\}$
- Assigning point $x_i, i = 1, 2, \dots, K$ to cluster $C_j, j \in \{1, 2, \dots, k\}$ where, $\|x_i - z_j\| \leq \|x_i - z_p\|, p = 1, 2, \dots, k$ and $j \neq p$
- Calculating new cluster centroids: $z_1^*, z_2^*, \dots, z_k^*$, where

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j$$

$i = 1, 2, \dots, K$, where n_i is the number of elements belonging to cluster C_i

- If $z_i^* = z_i, \forall i = 1, 2, \dots, K$ then end. Otherwise continue from 2

K-means is a common clustering algorithm that has been used in a variety of application disciplines, such as

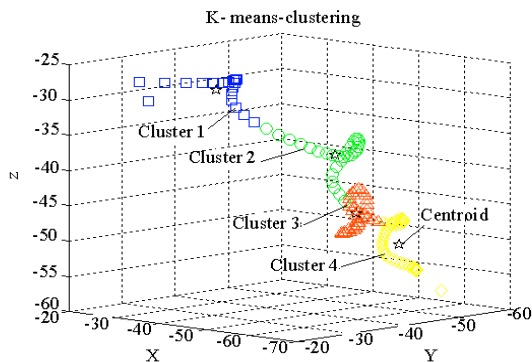


Fig. 3: K-Means data clustering with K=4 (This paragraph with equations and Fig. 3 are essential for understanding of contribution of the paper)

image clustering and information retrieval, as well as speech and speaker recognition. Different types of clustering algorithms that are based on K-Means, are mentioned by Wagstaff *et al.* (2001), such as the modified version for background knowledge, a genetic algorithm and the syllable contour that is classified into several linear loci that serve as candidates for the tone-nucleus using segmental K-Means segmentation algorithm.

Here is an investigation of a new speaker verification system that based on K-Mean feature extraction method KMM taken from speech signals. More specifically, the presented verification method by K-Means clustering consists of two main stages:

- Partitions the points in the N-by-P data matrix X (two original signals for two speakers) into two clusters. Then the two cluster centroid locations in the 2-by-P matrix is extracted. For each speaker four columns (8 coefficients) are preserved
- Partitions the points in the N-by-P data matrix X into four clusters. Then coefficients is extracted as follows:
- Distances from each point to every centroid in the N-by-4 matrix D, afterwards four coefficients: mean value, standard deviation, maximum and variance are determined
- Four cluster centroid locations in the 4-by-P matrix C (16 coefficients)
- Sums of point-to-centroid distances in the 1-by-4 vector M (4 coefficients)
- The first 32 elements of N-by-1 vector I containing the cluster indices of each point

In total, 64 coefficients vector are extracted by this method for each speaker for five signal frames of 1 second duration (totally 320 coefficient is extracted for each speaker). The verification decision is taken based on the method presented in by Eq. 10, 12 and 13.

RESULTS AND DISCUSSION

A testing database was created from Arabic language. The recording environment is a normal office environment via PC-sound card, with spectral frequency 4000 Hz and sampling frequency 16000 Hz. These Arabic utterances are Arabic spoken digits from 0 to 15. In addition, each speaker read ten separated 30 sec different Arabic texts. Total 84 individual speakers (19 to 40 years old) who are 54 individual male and 30 individual female spoken these Arabic words and texts for training by the method (creating hypothesized speaker model for

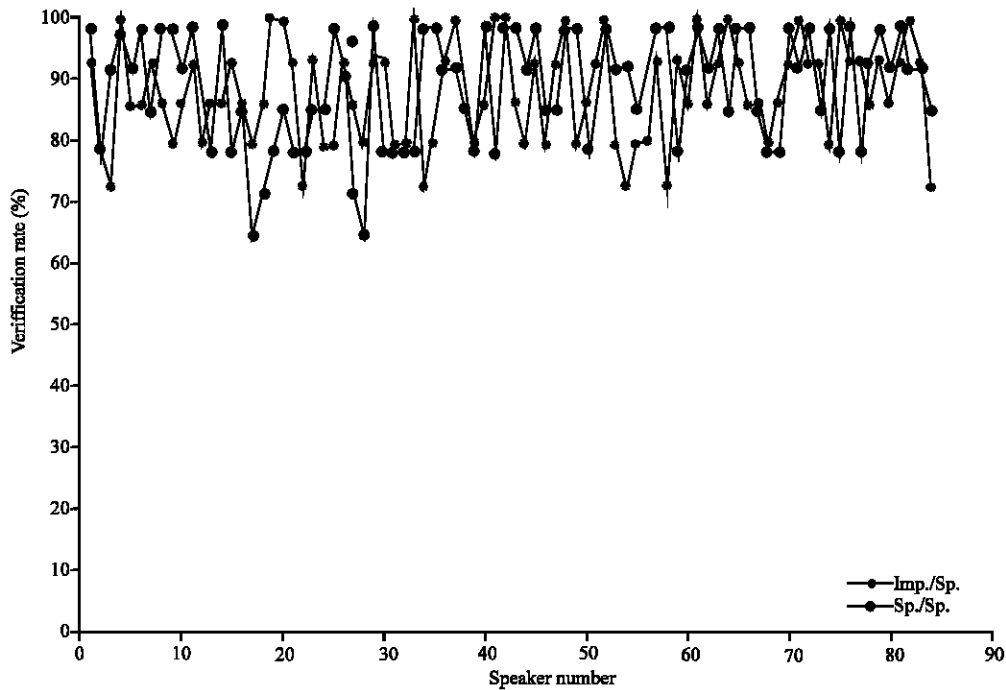


Fig. 4: Speaker verification rate results for 84 speakers

each individual speaker and background speaker model). The total number of tokens considered for training was 2100.

Experiments were conducted on a subset of recorded database consist of 54 male and 30 female speakers of different spoken words and texts. At first, feature vector was created by extracting MWM from silence-removed data for each frame. Finally, verification process was performed using the three verification functions. For accepting, at least two of three of the three verification functions should be more than the threshold.

Speaker verification is a binary decision task to state whether a test utterance belongs to a speaker or not (hence, an outside imposter). Evaluations were carried out on the pool of 84 speakers, with the individual speaker features constructed using 100% of the data, and the imposter speaker model obtained from 100% of the utterances belonging to all speakers. In case of individual speaker to same speaker of different utterances verification (speaker-speaker system), 25 trials are applied for each speaker. In case of individual imposter to speaker verification (imposter-speaker system), 25 trials are also applied for each speaker. The verification rates for the 84 speakers are illustrated at Fig. 4. All experiments were applied according to the text-independent system.

A single run of speaker verification task consists of scoring test files against the speaker model and background model. If the score ratio is greater than a

threshold for at least two of the three verification functions using specified by Eq. 10, 12 and 13, the test file is categorized as the target speaker, otherwise when the score is less than this certain threshold for at least two of the three verification functions, the test file is classified as an outside imposter. The performance of presented verification system according to speaker-speaker verification system and imposter-speaker verification system for independent-text platform taken for different arithmetic modulus denoted by Mod: seven, eleven and nineteen were reported at Table 1. Better result was achieved (91.61%) in case of the speaker-speaker verification system for Mod19.

In the next experiment, the performances of the MWVS systems in the speaker-speaker and speaker-imposter platforms were compared with the same of GMM and MWM method GMMW presented in section 3.1 and K-Means method KMM presented in section 3.2, under the recorded database. The results of these experiments via recorded database are summarized in Table 2. These results indicate that under similar condition, MWVS and GMMW provide a better platform for speaker verification than KMM. Moreover, the speaker-speaker system provides more accurate results than the imposter-speaker results for the all three methods.

Subsequent to assessment in the normal condition, experiments to assess the speaker verification system in the speaker-imposter platform under abnormal noisy were

Table 1: Verification rate results for different modular arithmetic

Mod 19	Mod 11	Mod 7	Verification system
91.61	82.92	78.63	Sp.-Sp.
91.11	82.03	77.51	Imp.-Sp.

Table 2: Speaker verification rate results for GMMW, KMM and MWVS

Verification rate (%)	Imp.-Sp. system Ver. rate (%)	Sig.-Sig. system Ver. rate (%)	Number of signals	Verification method
91.365	91.03	91.70	2100	GMMW
77.640	77.23	78.05	2100	KMM
91.360	91.11	91.61	2100	MWVS

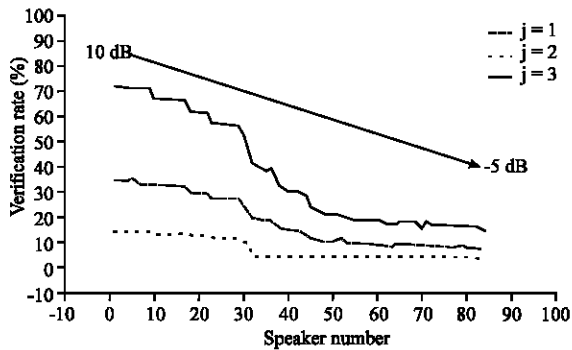


Fig. 5: Verification rates in the speaker-imposter platform under AWGN using DWT for 84 speakers

conducted. To implement this experiment, additive white Gaussian noise (AWGN) was added to the verified signals only, with SNR from -5 to 10 dB. Afterwards were applied to presented method, where no important recognition rate was noticed. In this experiment, the performance of the speaker verification system in speaker-imposter was improved by using the approximate sub-signals of DWT at level (j) 1, 2 and 3. The results of this experiment are illustrated in Fig. 5. These results indicate that the proposed verification system can tackle additive white Gaussian noise condition more robustly if approximate sub-signal of DWT at level 3 is used instead of the original signal.

CONCLUSION

Modular arithmetic and wavelet packet based speaker verification system is proposed in this study. This system was developed using a performing three verification methods. In this study, effective feature extraction method for text-independent system is developed, taking in consideration that the computational complexity is very crucial issue. The experimental results on a subset of recorded database showed that feature extraction method proposed in this paper is appropriate for text-independent verification system. Two other verification methods are proposed GMMW and KMM. The results of the

experiments conducted in this study demonstrated a better performance of MWVS in text-independent verification task. Finally, the developed speaker verification system was employed with data obtained under abnormal conditions where AWGN noisy was added. In this case it was observed that the MWVS system can tackle additive white Gaussian noise condition more robustly if approximate sub-signal of DWT at level 3 is used instead of the original signal. Another major contribution of this research is the development of a less computational complexity speaker verification system with modular arithmetic capable of dealing with abnormal conditions for relatively good degree.

REFERENCES

Afshari, M., 2011. An algorithm to analyze of two-dimensional function by using wavelet coefficients and relationship between coefficients. *Asian J. Applied Sci.*, 4: 414-422.

Avci, D., 2009. An expert system for speaker identification using adaptive wavelet sure entropy. *Exp. Syst. Appl.*, 36: 6295-6300.

Avci, E. and Z.H. Akpolat, 2006. Speech recognition using a wavelet packet adaptive network based fuzzy inference system. *Expert Syst. Appl.*, 31: 495-503.

Avci, E., 2007. A new optimum feature extraction and classification method for speaker recognition: GWPNN. *Expert Syst. Appl.*, 32: 485-498.

Bandyopadhyay, S. and U. Maulik, 2002. An evolutionary technique based on K-means algorithm for optimal clustering in RN. *Inform. Sci.*, 146: 221-237.

Behroozmand, R. and F. Almasganj, 2007. Optimal selection of wavelet packet-based features using genetic algorithm in pathological assessment of patients speech signal with unilateral vocal fold paralysis. *Comput. Biol. Med.*, 37: 474-485.

Campbell, W.M., D.E. Sturim and D.A. Reynolds, 2006. Support vector machines using GMM super vectors for speaker verification. *Signal Process. Lett.*, 13: 308-311.

Cormen, T.H., C.E. Leiserson, R.L. Rivest and C. Stein, 2001. *Introduction to Algorithms*. 2nd Edn., McGraw-Hill Book Company, Cambridge and New York.

Daqrouq, K., 2011. Wavelet entropy and neural network for text-independent speaker identification. *Eng. Appl. Artificial Intell.*, 10.1016/j.engappai.2011.01.001

Daqrouq, K., I.N. Abu-Isbeih, O. Daoud and E. Khalaf, 2010. An investigation of speech enhancement using wavelet filtering method. *Int. J. Speech Technol.*, 13: 101-115.

- Derbel, A., F. Kallel, M. Samet and A.B. Hamida, 2008. Bionic wavelet transform based on speech processing dedicated to a fully programmable stimulation strategy for cochlear prostheses. *Asian J. Scientific Res.*, 1: 293-309.
- Hemalatha, M. and K. Vivekanandan, 2008. A semaphore based multiprocessing k-mean algorithm for massive biological data. *Asian J. Scientific Res.*, 1: 444-450.
- Kotnik, B., Z. Kacic and B. Horvat, 2003. Noise robust speech parameterization based on joint wavelet packet decomposition and autoregressive modeling. *Proceedings of the 8th European Conference on Speech Communication and Technology*, Sep. 1-4, Geneva, pp: 1793-1796.
- Kuhn, R., J.C. Junqua, P. Nguyen and N. Niedzielski, 2000. Rapid speaker adaptation in eigenvoice space. *IEEE Trans. Speech Audio Process.*, 8: 695-707.
- Kwon, S. and S. Narayanan, 2002. Speaker change detection using a new weighted distance measure. *Proceedings of International Conference on Spoken Language Processing, (ICSLP'2002)*, Denver, CO, pp: 2537-2540.
- Lamel, L.F. and J.L. Gauvain, 2000. Speaker verification over the telephone. *Speech Commun.*, 31: 141-154.
- Lapidot, I., H. Guterman and A. Cohen, 2002. Unsupervised speaker recognition based on competition between self-organizing maps. *IEEE Trans. Neural Networks*, 13: 877-887.
- Lung, S.Y., 2004. Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition. *Pattern Recognition*, 37: 1543-1544.
- Lung, S.Y., 2007. Efficient text independent speaker recognition with wavelet feature selection based multilayered neural network using supervised learning algorithm. *Pattern Recognition*, 40: 3616-3620.
- Mami, Y. and D. Charlet, 2002. Speaker identification by location in an optimal space of anchor models. *Int. Conf. Spoken Language Process.*, 2: 1333-1336.
- Mami, Y. and D. Charlet, 2003. Speaker identification by anchor models with PCA/LDA, post-processing. *IEEE Int. Conf. Acoustics, Speech Signal Process.*, 1: 180-183.
- Mami, Y. and D. Charlet, 2006. Speaker recognition by location in the space of reference speakers. *Speech Commun.*, 48: 127-141.
- Martin, A. and M. Przybocki, 2001. Speaker recognition in a multi-speaker environment. *Proceeding of the 7th European Conference on Speech Communication and Technology, (CSCTL'01)*, Aalborg, Denmark, pp: 787-790.
- Naini, A.S., M.M. Homayounpour and A. Samani, 2010. A real-time trained system for robust speaker verification using relative space of anchor models. *Comput. Speech Language*, 24: 545-561.
- Nemati, S. and M.E. Basiri, 2010. Text-independent speaker verification using ant colony optimization-based selected features. *Exp. Syst. Appl.*, 38: 620-630.
- Qureshi, I.M. and A. Jalil, 2002. Object recognition using ANN with backpropagation algorithm. *J. Applied Sci.*, 2: 281-287.
- Reynolds, D.A., T.F. Quatieri and R.B. Dunn, 2000. Speaker verification using adapted gaussian mixture models. *Digital Signal Process.*, 10: 19-41.
- Thyes, O., R. Kuhn, P. Nguyen and J.C. Junqua, 2000. Speaker identification and verification using eigenvoices. *Int. Conf. Spoken Language Process.*, 2: 242-245.
- Visser, E., M. Otsuka and T.W. Lee, 2003. A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments. *Speech Commun.*, 41: 393-407.
- Wagstaff, K., C. Cardie, S. Rogers and S. Schrodl, 2001. Constrained K-means clustering with background knowledge. *Proceedings of the 18th International Conference on Machine Learning, (ICML'01)*, San Francisco, CA, USA, pp: 577-584.
- Wessing, H., H. Christiansen, T. Fjelde and L. Dittmann, 2002. Novel scheme for packet forwarding without header modifications in optical networks. *IEEE J. Lightwave Technol.*, 20: 1277-1283.
- Wong, W.M. and K.J. Blow, 2006. Design and analysis of an all-optical processor for modular arithmetic. *Optics Commun.*, 265: 425-433.
- Wu, J.D. and B.F. Lin, 2009. Speaker identification using discrete wavelet packet transform technique with irregular decomposition. *Expert Syst. Appl.*, 36: 3136-3143.
- Wu, J.D. and S.H. Ye, 2009. Driver identification based on voice signal using continuous wavelet transform and artificial neural network techniques. *Exp. Syst. Appl.*, 36: 1061-1069.
- Xiang, B. and T. Berger, 2003. Efficient text-independent speaker verification with structural *Gaussian mixture* models and neural network. *IEEE Trans. Speech Audio Proc.*, 11: 447-456.
- Zheng, H., Z. Li and X. Chen, 2002. Gear fault diagnosis based on continuous wavelet transform. *Mech. Syst. Signal Process.*, 16: 447-457.