An Efficient Process of Human Recognition Fusing Palmprint and Speech features

Mahesh P.K.
JSS Research Foundation
Research Scholar, E&C Dept.
SJCE, Mysore

M.N. ShanmukhaSwamy
SJCE, Mysore
Electronics and Communication Dept.,
SJCE, Mysore

ABSTRACT
This paper presents fusion of two biometric traits, i.e., palmprint and speech signal, at matching score level architecture uses weighted sum of score technique. The features are extracted from the pre-processed palm image and pre-processed speech signal. The features of a query image and speech signal are compared with those of a database images and speech signal to obtain matching scores. The individual scores generated after matching are passed to the fusion module. This module consists of three major steps i.e., normalization, generation of similarity score and fusion of weighted scores. The final score is then used to declare the person as genuine or an impostor. The system is tested on database collected by the authors for 120 subjects and gives an overall accuracy of 98.47% with FAR of 1.36% and FRR of 0.87%.

Keywords
Multimodal biometrics, Palmprint, Speech signal, score normalization and fusion.

1. INTRODUCTION
Multimodal biometric systems consolidate the evidence presented by multiple biometric modalities and typically provide better recognition performance compare to single biometric modality. Due to its promising applications as well as theoretical challenges, multimodal biometric has drawn more and more attention in recent years [1]. Although information fusion in a multimodal system can be performed at various levels, integration at the matching score level is the most common approach due to the case in accessing and combining the score generates by different matchers. Since the matching scores output by the various modalities are heterogeneous, score normalization is needed to transform these scores into a common domain, prior to combining them.

We proposes, multimodal biometric system for identify verification using two modalities, i.e. Palmprint and speech. The proposed system is designed for applications where the training data contains palmprint and speech. Integrating the palmprint and speech features increases recognition performance of person authentication.

The final decision is made by fusion at matching score level. Multimodal system is developed through fusion of palmprint verification and speaker verification. We extract the features using Haar Wavelet transform method for palmprint and Subband based Cepstral Parameters (SBC) technique for speech. Integrating these two features at fusion level, which gives better performance and better accuracy.

The rest of this paper is organized as fallows. Section 2 presents the System overview, which is used to increase recognition quality. Section 3 and 4 presents algorithms for calculation of palmprint and speech features using Haar Wavelet transform method and SBC technique respectively. Section 5, the individual traits are fused at matching score level using weighted sum of score technique. The experimental results are given in section 6. Finally Conclusions are given in the last section.

2. SYSTEM STRUCTURE
The block diagram of a multimodal biometric system using two (palm and speech) modalities for human recognition system is shown in Figure 1. It consists of three main blocks, that of Preprocessing, Feature extraction and Fusion. Preprocessing and feature extraction are performed in parallel for the two modalities. The preprocessing of the audio signal under noisy conditions includes signal enhancement, tracking environment and channel noise, feature estimation and smoothing [2]. The preprocessing of the palmprint typically consists of the challenging problems of detecting and tracking of the palm and the important palm features.

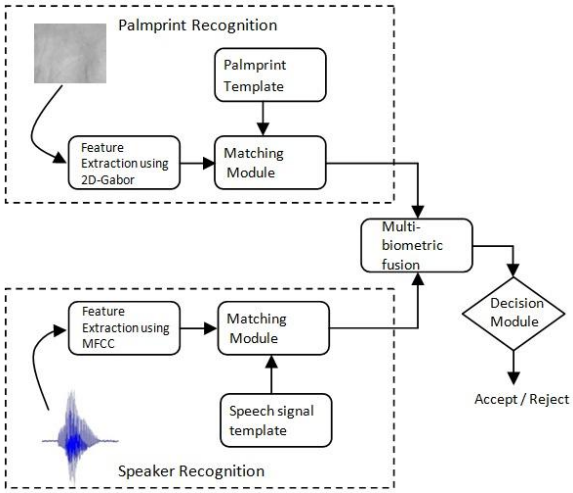

Figure 1 Block diagram of speech signal and palmprint multimodal biometric system

Further, features are extracted from the training and testing images and speech signal respectively, and then matched to find the similarity between two feature sets. The matching scores

generated from the individual recognizers are passed to the decision module where a person is declared as genuine or an imposter.

3. FEATURE EXTRACTION USING HAAR WAVELET

Features are the attributes or values extracted to get the unique characteristics from the image and speech signal.

3.1 Palmprint feature extraction methodology

3.1.1 Identify hand image from background

Our designed system is such that palmprint images are captured using contact-less without pegs, keeping the image background relatively uniform and relatively low intensity when compared to the hand image. Using the statistical information of the background, the algorithm estimates an adaptive threshold to segment the image of the hand from the background. Pixels with intensity above the threshold are considered to be part of the hand image.

3.1.2 Locate region-of-interest

The palm area is extracted from the binary image of the hand. After translating the original image into binary image, we find two key positioning points in the palmprint image using automatic detecting method. The first valley in the graph is the gaps between little finger and ring finger, Key Point 1. The third valley in the graph is the gaps between middle finger and index finger, Key Point 2. The key point is circled in Figure 2. The hand image is rotated by θ degrees. The hand images are rotated to align the hand images into a predefined direction. θ is calculated using the key points as shown in the Figure 2. Since the size of the original image is large, a smaller hand image is cropped out from the original hand image after image alignment using key points. Figure 3 shows the proposed image alignment and ROI selection method.

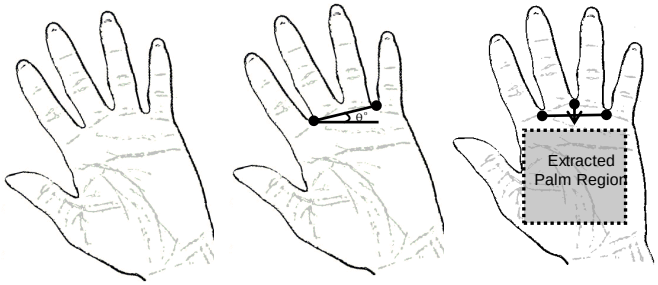

Figure 2 Schematic diagram of image alignment

Figure 3 Segmentation of ROI

3.2 Feature extraction

This paper introduces a more complex form of use of palmprint biometrics by manipulating the palmprint image. Each of the detail images is divided into several square cells. We divide the palmprint into M^2 cells as shown in Figure 4. On taking this approach we have a number of M^2 cell much smaller in size, thus each cell contains the necessary unique information in order to authenticate the user. One of the advantages of this approach is that if one cell is corrupted, for any reason, we still can get better authentication result.

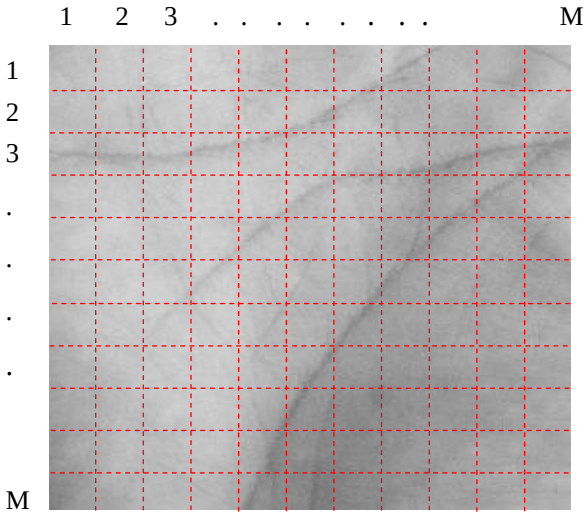

Figure 4 Segmentation of ROI with M^2 cells

Firstly, a 2-D lowpass filter is applied to the image. The result is subtracted from the image to minimize the non-uniform illumination effect. Secondly, a Gaussian window is used to smooth out the image since Haar wavelet, due to its rectangular wave nature, is sensitive to noise.

A 1-level decomposition of the image by the Haar wavelet is carried out. For each of the three detail images obtained, i.e. image consisting of the horizontal, vertical and diagonal details, a smoothing mask is applied to remove noise. It was found that most of the low frequency components are attributable to the redness underneath the skin and should preferably be excluded from features for identification. Thus, pixels with frequency values within one standard deviation are set to zero. Values of the rest of the pixels are projected onto a logarithm scale so as to minimize the absolute differences in the magnitude of the frequency components between two images. That is,

$$I(x_i, y_i) = \begin{cases} 0, & if \quad |I(x_i, y_i)| \le std(I(x, y)) \\ \ln(\ |I(x_i, y_i)| - std(I(x, y)) + 1\ ), & o.w. \end{cases} \quad (1)$$

where $I(x_i,y_i)$ is the frequency value in a detail image. The processed image is shown in Figure 5.

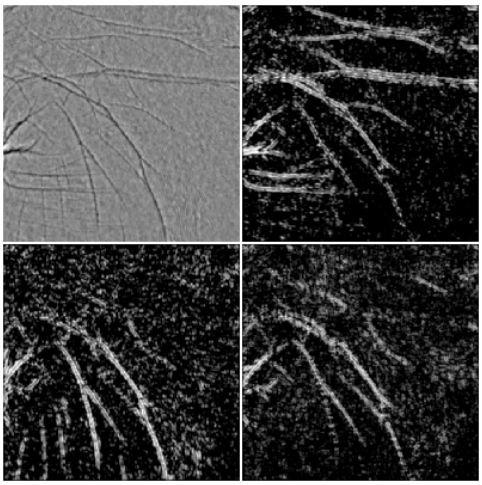

Figure 5 Haar wavelet transform of Palmprint

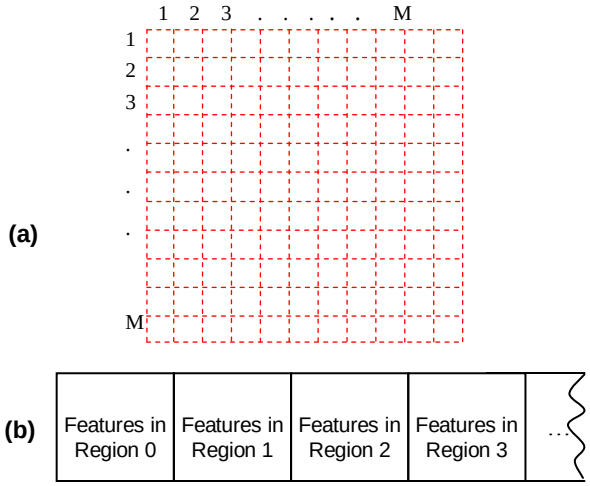

Figure 6 Feature vector

3.3 Matching score calculation

Since the palm images under process are divided into square cells of same widths regardless of the size of the original image, different palm sizes will result in feature vectors of different lengths. Due to the possibility of having variations in the extent the hand is stretched, the resultant maximum palm area may vary within the same subject. Therefore, the distance measure used must be able to fairly compare two feature vectors with unequal dimension.

The score is calculated as the mean of the absolute difference between two feature vectors. If $featureV_i$ represents a feature vector of N_i elements, the score between two images is given as:

$$Score(i,j) = \frac{\sum_{n=1}^{\min(N_i,N_j)} | \; featureV_i(n) - featureV_j(n) \; |}{\min(N_i, N_j)} \quad (2)$$

4. FEATURE EXTRACTION USING SUBBAND BASED CEPSTRAL PARAMETERS

4.1 Subband Decomposition via Wavelet Packets

A detailed discussion of wavelet analysis is beyond the scope of this paper and we therefore refer interested readers to a more complete discussion presented in [3]. In continuous time, the Wavelet Transform is defined as the inner product of a signal x(t) with a collection of wavelet functions $\psi_{ab}(t)$ in which the wavelet functions are scaled(by a) and translated (by b) versions of the prototype wavelet $\psi(t)$.

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right)dt \quad (3)$$

$$W_\psi x(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \quad (4)$$

Discrete time implementation of wavelets and wavelet packets are based on the iteration of two channel filter banks which are subject to certain constraints, such as low pass and/or high pass branches on each level followed by a sub sampling-by-two unit. Unlike the wavelet transform which is obtained by iterating on the low pass branch, the filterbank tree can be iterated on either branch at any level, resulting in a tree structured filterbank which we call a wavelet packet filterbank tree. The resultant transform creates a division of the frequency domain that represents the signal optimally with respect to the applied metric while allowing perfect reconstruction of the original signal. Because of the nature of the analysis in the frequency domain it is also called subband decomposition where subbands are determined by a wavelet packet filterbank tree.

4.2 Wavelet Packet Transform Based Feature Extraction Procedure

Here, speech is assumed to be sampled at 8 kHz. A frame size of 24msec with a 10msec skip rate is used to derive the Subband based Cepstral Parameters features, whereas a 20msec frame with the same skip rate is used to derive the MFCCs. We have used the same configuration proposed in [4] for MFCC. Next, the speech frame is Hamming windowed and pre-emphasized.

The proposed tree assigns more subbands between low to mid frequencies while keeping roughly a log-like distribution of the subbands across frequency. The wavelet packet transform is computed for the given wavelet tree, which results in a sequence of subband signals or equivalently the wavelet packet transform coefficients, at the leaves of the Tree. In effect, each of these subband signals contains only restricted frequency information due to inherent bandpass filtering. The complete block diagram for computation of Subband based Cepstral Parameters is given in Figure 7. The energy of the sub-signals for each subband is computed and then scaled by the number of transform

coefficients in that subband. The subband signal energies are computed for each frame as,

$$S_i = \frac{\sum_{m \in i} \left[(W_\psi)(i), m \right]}{N_i} \quad (5)$$

W_ψ *: Wavelet packet transform of signal x,*

i :subband frequency index (i=1,2...L),

N_i *: number of coefficients in the* i^{th} *subband.*

4.3 Subband based Cepstral Parameters

As in MFCCs the derivation of parameters is performed in two stages. The first stage is the computation filterbank energies and the second stage would be the decorrelation of the log filterbank energies with a DCT to obtain the MFCC. The derivation of the Subband Based Cepstral parameters follows the same process except that the filterbank energies are derived using the wavelet packet transform rather than the short-time Fourier transform. It will be shown that these features outperform MFCCs. We attribute this to the computation of subband signals with smooth filters. The effect of filtering as a result of tracing through the low-pass/high-pass branches of the wavelet packet tree, is much smoother due to the balance in time-frequency representation. We believe that this will contribute to improved speech/speaker characterization over MFCC. These parameters have been shown to be effective for speech recognition in car noise [5] and for classification of stressed speech. Subband Based Cepstral parameters are derived from subband energies by applying the Discrete Cosine Transformation:

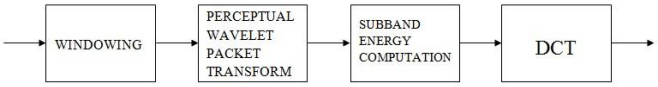

Figure 7 Block diagram for Wavelet Packet Transform based feature extraction procedure

$$SBC(n) = \sum_{i=1}^{L} \log S_i \cos\left(\frac{n(i-0.5)}{L} \pi \right), n = 1, ...n' \quad (6)$$

where n' is the number of SBC parameters and L is the total number of frequency bands. Because of the similarity to root-cepstral [6] analysis, they are termed as subband based cepstral parameters.

4.4 The Gaussian Mixture Model

In this study, a Gaussian Mixture Model approach proposed in [7] is used where speakers are modeled as a mixture of Gaussian densities. The use of this model is motivated by the interpretation that the Gaussian components represent some general speaker-dependent spectral shapes and the capability of Gaussian mixtures to model arbitrary densities.

The Gausssian Mixture Model is a linear combination of M Gaussian mixture densities, and given by the equation,

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \quad (7)$$

Where $\vec{x}$ is a D-dimensional random vector, $b_i(\vec{x})$, *i=1,...M* are the component densities and p_i, i=1,…M are the mixture weights. Each component density is a D-dimensional Gaussian function of the form

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\sum_i|^{1/z}} \exp\left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^T \sum_i^{-1} (\vec{x} - \vec{\mu}) \right\} \quad (8)$$

Where $\vec{\mu}$ denotes the mean vector and $\sum_i$ denotes the covariance matrix. The mixture weights satisfy the law of total probability, $\sum_{i=1}^{M} p_i = 1$. The major advantage of this representation of speaker models is the mathematical tractibility where the complete Gaussian mixture density is represented by only the mean vectors, covariance matrices and mixture weights from all component densities.

5. FUSION

No individual trait can provide 100% accuracy. Further, the results generated from the individual traits are good but the problem arises when the user is not able to give his speech signal correctly due problem in throat and due to background noise. Similarly, in palmprint due to noise the image may deteriorate. Thus in such a situation an individual cannot be recognized using the speech signals and the biometric system comes to a standstill. Similarly, the problem faced by palmprint recognition system is the presence of scars and cuts. The scars add noises to the palmprint image which cannot be enhanced fully. Thus, the system takes noisy palmprint as input which is not able to extract the features correctly and in turn, leads to false recognition of an individual. Thus to overcome the problems faced by individual traits of speech signal and palmprint, a novel combination is proposed for the recognition system. The integrated system also provide anti spoofing measures by making it difficult for an intruder to spoof multiple biometric traits simultaneously. Scores generated from individual traits are combined at matching score level using weighted sum of score technique. Let MS_{Speech} and MS_{Palm} be the matching scores obtained from Speech signal and palmprint modalities respectively. The steps involved are:

5.1.1 Score Normalization

This step brings both matching scores between 0 and 1 [8]. The normalization of both the scores are done by

$$N_{Speech} = \frac{MS_{Speech} - \min_{Speech}}{\max_{Speech} - \min_{Speech}} \quad (9)$$

$$N_{Palm} = \frac{MS_{Palm} - \min_{Palm}}{\max_{Palm} - \min_{Palm}} \quad (10)$$

where $\min_{Speech}$ and $\max_{Speech}$ are the minimum and maximum scores for speech signal recognition and $\min_{Palmprint}$ and $\max_{Palmprint}$ are the corresponding values obtained from palmprint trait.

5.1.2 Generation of Similarity Scores

Note that the normalized score of palmprint which is obtained through Haar Wavelet gives the information of dissimilarity between the feature vectors of two given images while the normalized score from speech signal gives a similarity measure. So to fuse both the score, there is a need to make both the scores as either similarity or dissimilarity measure. In this paper, the normalized score of palmprint is converted to similarity measure by

$$N'_{Palm} = 1 - N_{Palm} \quad (11)$$

5.1.3 Fusion

The different biometrics systems can be integrated at multi-modality level to improve the performance of the verification system. The following steps are performed for fusion:

1. Given a query image as input, features are extracted by the individual recognition
2. The weights *a* and *b* are calculated using FAR and FRR.
3. Finally, the sum of score technique is applied for combining the matching score of two traits i.e. speech signal and palmprint. Thus the final score MS_{Final} is given by,

$$MS_{FINAL} = \frac{1}{2}\ a * MS_{SPEECH} + b * MS_{PALM} \quad (12)$$

Where *a* and *b* are the weights assigned to both the traits. The final matching score (MS_{Final}) is compared against a certain threshold value to recognize the person as genuine or an imposter.

6. EXPERIMENTAL RESULTS

The results are tested on speech signals and palmprint images collected by the authors. The database consists of six palm images (120×6) and six speech signals (120×6) per person with total of 120 persons. The palm images are acquired using CCD camera with uniform light source. However, speech signals are acquired using a microphone with uniform background noise. For the purpose allowing comparisons two levels of experiments are performed. At first level palmprint and speech signal algorithms are tested individually. At this level the individual results are computed and an accuracy curve is plotted as shown in Figure 8. At this level the individual accuracy for palmprint and speech signal is found to be 93.79% and 95.21% respectively as shown in Table 1.

However in order to increase the accuracy of the biometric system as a whole the individual results are combined at matching score level. At second level of experiment the matching scores from the individual traits are combined and final accuracy graph is plotted as shown in Figure 9. Table 1 shows the accuracy and error rates obtained from the individual and combined system. The overall performance of the system has increased showing an accuracy of 98.47% with FAR of 1.36% and FRR of 0.87% respectively.

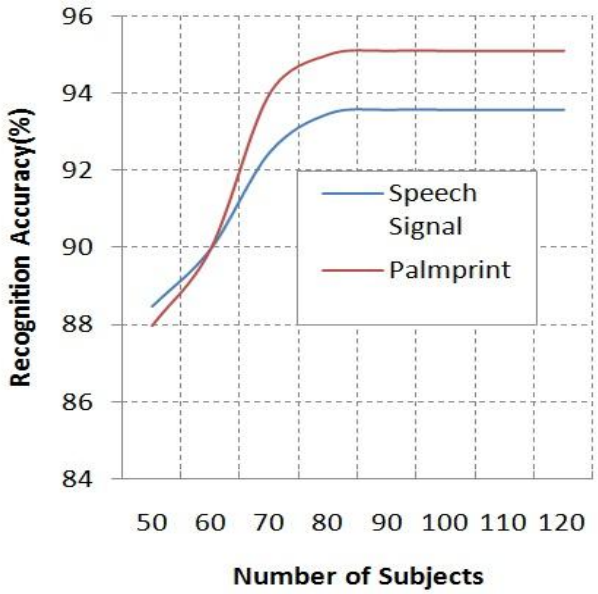

Figure 8 Accuracy plots of individual recognizers

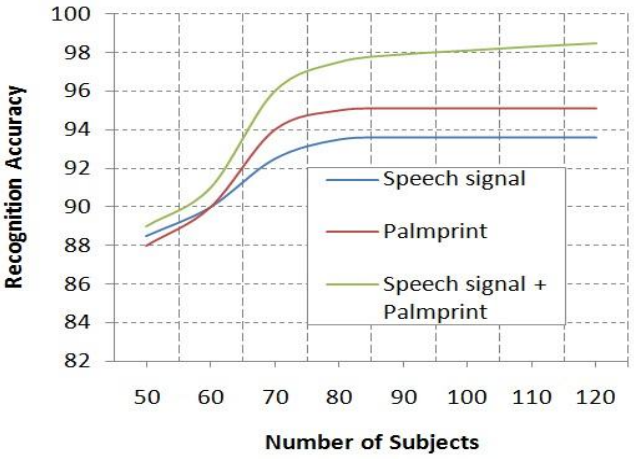

Figure 9 Accuracy graph for combined classifier

Table 1
Figures showing individual and combined accuracy.

Trait	Algorithm	Accuracy (%)	FAR (%)	FRR (%)
Palmprint	Haar Wavelet	93.79	2.72	4.73
Speech signal	SBC+GMM	95.21	5.87	1.35
Fusion	Haar + SBC	98.47	1.36	0.87

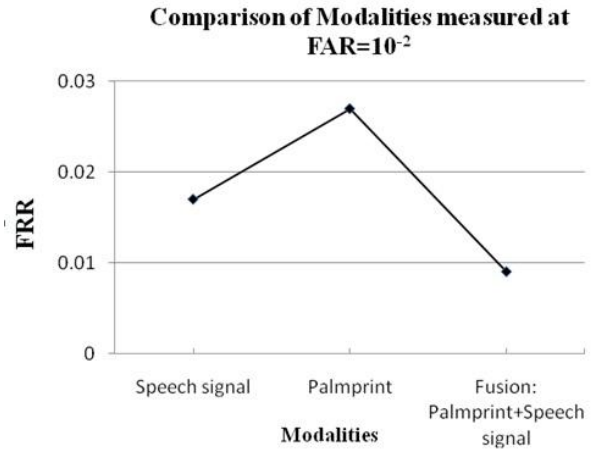

Figure 10 Comparison of Modalities measured at FRR=10^{-2}

7. CONCLUSION

Biometric systems are widely used to overcome the traditional methods of authentication. But the unimodal biometric system fails in case of biometric data for particular trait. This paper proposes a new method in selecting and dividing the ROI for analysis of palmprint. The new method utilizes the maximum palm region of a person to attain feature extraction. More importantly, it can cope with slight variations, in terms of rotation, translation, and size difference, in images captured from the same person. Feature vectors are arranged such that point-wise comparison is matching features from the same spatial region of two different palms. Thus the individual score of two traits (speech & palmprint) are combined at classifier level and trait level to develop a multimodal biometric system. The performance table shows that multimodal system performs better as compared to unimodal biometrics with accuracy of more than 98%.

8. REFERENCES

[1] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiomtrics*. Springer-Verlag, 2006..

[2] S. Prabhakar,,. K. Jain, Decision-level fusion in fingerprint verification, *Pattern Recognition* 5 (4) (2002) 861-874.

[3] O. Rioul and M. Vetterli, "Wavelets and Signal Processing, "IEEE Signal Proc. Magazine, vol. 8(4), pp. 11-38, 1991.

[4] D. A. Reynolds and R. C. Rose, "Robust Text_Independent Speaker Identification Using Gaussian Mixture Speaker Models" IEEE Transactions on SAP, vol.3, pp, 72-83, 1995.

[5] E. Erzin, A. E. Cetin and Y. Yardimci, "Subband analysis for speech recognition in the presence of car noise," ICASSP-95, vol. 1, pp.417-420,1995.

[6] P. Alexandre and P. Lockwood, "Root cepstral analysis: A unified view: Application to speech processing in car noise environments," Speech Communication, v.12, pp. 277-288,1993.

[7] D. A. Reynolds, "Experimental Evaluation of Features for Robust Speaker Identification," IEEE Transactions on SAP, vol. 2. Pp. 639-643,1994.

[8] A. K. Jain, K. Nandakumar, & A. Ross, Score Normalization in multimodal biometric systems. *The Journal of Pattern Recognition Society, 38*(12), 2005, 2270-2285