Feature Selection by Independent Component Analysis for Robust Speaker Verification

Ahmet Şentürk and Fikret S. Gürgen
Computer Engineering
Boğaziçi University
Istanbul, 34342, Turkey

Abstract

A robust approach that unifies independent component analysis (ICA) subspace feature selection in connection with the speaker verification (SV) is proposed. ICA subspace provides statistically independent basis that spans the input space of corrupted speech, then the selected independent components are applied to a vector quantizer (VQ) for SV purpose. The Euclidean distance in the feature space is kept invariant by using ICA and is also used in the VQ based SV system as a matching choice. In the feature selection stage, a batch-mode FastICA algorithm and two adaptive algorithms EGLD-ICA and Pearson-ICA are employed for two-microphone case. As a result, the selected features provide a lower classification error and a better generalization in real environments. The performance of the approach is demonstrated with YOHO database [8] in cocktail party effect and ambient noise cases.

1 Introduction

Independent and principal components are well-known techniques of pattern recognition and intelligent systems. For independent component analysis (ICA), statistically independent features are selected from the data. With principal component analysis (PCA), maximally variant or diagonalized covariance features are selected. Both ICA and PCA were employed in various applications of these areas [1-3]. Data sets with Gaussian distributions can be represented with principal components, while data sets with non-gaussian, sparse-nature statistics can be decomposed to independent components. For the problem of speaker verification (SV) [5-7, 9-11] in real environments, undesired components with unknown, mostly non-gaussian distributions such as ambient noise, a competing speaker or music as in a cocktail party interfere with the speaker's voice. Decomposing the interfered, independent components of speech becomes an issue (is described also as blind source separation (BSS)) in building robust SV systems as one choice thus, ICA approach gains importance here instead of PCA approach of selecting the most representative attributes.

With PCA, features are selected from maximally variant eigen directions and principal components are used to reduce the dimension of the feature space. For the ICA case, the near-orthogonality property of the decomposed ICA basis causes a performance that is almost identical to an eigenvector basis with Euclidean distance classifiers [3]. Other properties of attribute selection are maximum non-gaussianity and employing a floating search under the objective of maximum performance of a classifier [4].

Speaker verification (SV) systems typically involve patterns which have some measure of variance across representative elements in a speaker (or class). For example, speaker's voice is often under variable environmental conditions. Additionally, the voice of speaker is often a variant in the database due to the time, sickness and various recording conditions. Verifying speakers under these variants represents a challenging problem. Vector quantizers (VQs) provide an optimal decision hyperplane by employing Euclidean distance based codevectors or projecting the data into high-dimensional space clusters. VQs have been shown to be very effective classifiers and provide ability to deal with SV (also with speaker identification (SI)) for text dependent (TD) and text independent (TI) cases for large databases [11-12]. Their properties such as good accuracy and simplicity in computation make them a good choice in SV near the models such as Gaussian Mixture Model (GMM) [13].

A recent application is to prevent impostors from gaining access to small labtop, handheld devices. This generates a need for security and security measures should be incorporated to these devices [9-10]. As a result, the integration of two biometric techniques, voice and face identification, into handheld devices gains importance. In the case of voice identification e.g. SV, handheld devices offer a serious challenge: the mobility of a device ensures that the environmental conditions that the device will be in highly variable background noises with potentially very low signal-to-noise ratios (SNR).

The study proposes independent components data representation before a VQ based speaker verifier. Environmental variants are considered to increase the intra-speaker (intra-class) variance. The raw input data are projected to a subspace to reduce the effect of the variants through the selection of the independent components by preserving Euclidean distance measurement. This enhances the feature space before the verification. Then, the VQ approach uses the independent components for verification thus employing a measure of invariance for the features for various environments.

YOHO database which is prepared under tightly constrained conditions is used to illustrate the performance of the ICA-VQ verifier. The cocktail party effect and ambient noise conditions are simulated with various scenarios: interfering speaker, music, and white noise and the performance degradation of the database is illustrated. Then, the ICA implementations [14-16] for two inputs (microphones), batch-mode FastICA and two adaptive algorithms EGLD-ICA and Pearson-ICA, are demonstrated in BSS problem with a number of utterances from the database. Finally, independent components of sound mixtures from separate sources are utilized to achieve better performance or lower error rates with the verifier.

2 Speaker Verification with Independent Components

2.1 Subspaces for Principal and Independent Components

A set of mean-adjusted data elements **X** are described by [$\mathbf{x_1}$ $\mathbf{x_2}$ $\mathbf{x_3}$ .... $\mathbf{x_n}$] where $\mathbf{x_i}$ is the i^th pattern in the set, as a column vector. It is generally possible to construct a decomposition of the data by a set of basis vectors that are maximally decorrelated by using a matrix **W**

$$\mathbf{S=WX} \quad (1)$$

where the columns of **S** are decorrelated. The decorrelated space S can be used as a basis for a low-dimensional representation:

$$\mathbf{x_{LDi}} = \mathbf{S}^{\mathbf{T}}\, \mathbf{x}_{\mathrm{i}} \quad (2)$$

each data vector $\mathbf{x_i}$ is decomposed to a subset of the columns of **S** to represent significant features in the data. Both PCA and ICA approaches may be used to compute **W**.

For the principal components from ***n*** observations on ***p*** variables *(n>p)*, a *pXp* covariance matrix $\mathbf{XX^T}$ can be computed (sometimes called as *R analysis*. When *p>n,* $\mathbf{X^TX}$ becomes covariance matrix and it is called as *Q analysis*). When singular value decomposition (SVD) on **X** is used to decompose **X** as $\mathbf{X = U\Sigma V^T}$, the covariance matrix

$$\mathbf{XX^T = U\Sigma V^T\, V\Sigma^T U = U\Sigma^2 U^T} \quad (3)$$

This is an eigendecomposition on $\mathbf{XX^T}$ where **U** are eigenvectors and Σ^2 are eigenvalues. The eigenvectors are scaled to unity so that the resulting subspace

$$\mathbf{X_{LD} = U^T X} \quad (4)$$

will have the same variance as **X**. For independent components, **X** observations are decomposed by finding a **W** such that **S** components become decorrelated and statistically independent. In one way, the mutual information between the components of the random variable **s** (column of **S**) becomes a measure for the degree of independence:

$$\mathbf{I(s)} = \int \mathbf{f(s)} \log [\mathbf{f(s)} / \prod_k \mathbf{f_k(s_k)}]\, \mathbf{ds} \qquad (5)$$

where f(**s**) is the joint probability of **s** and $f_k(\mathbf{s_k})$ are the marginal densities. It has been shown that [3] if a nonlinear mapping **y** = g(**s**) is applied so that the **y** marginal densities become uniformly distributed, the mutual information is obtained from the entropy by
$\mathbf{I(y)}= \int \mathbf{f(y)}\mathit{log}\, \mathbf{f(y)}\, \mathbf{dy}$ and **I(y)** can be minimized by taking derivation according to $\mathbf{W_{ij}}$,
$\partial\mathbf{I(y)}/\partial\mathbf{W_{ij}} = (\mathbf{W}^T)^{-1} + \mathbf{E}[g(\mathbf{Wx})\mathbf{x}^T]$ where **E[.]** is expected value operator. If we multiply each side of $\partial\mathbf{I(y)}/\partial\mathbf{W_{ij}}$ equation by $\mathbf{W}^T\mathbf{W}$, we obtain a natural gradient algorithm:

$$\Delta\mathbf{W} \approx (\mathbf{I} + \mathbf{E}\, [g(\mathbf{Wx})\mathbf{x}^T])\, \mathbf{W} \qquad (6)$$

where **ΔW** is incremental **W** matrix and **I** is unit matrix. The iterations of **ΔW** allows for the discovery of directions in the data which provide good generalization.

In ICA, which is defined by **S=WX** model (**W** may also be called as *unmixing matrix)*, a white, orthonormal **S** matrix is found by SVD if **S** is statistically independent. Thus, ICA approach searches an orthonormal rotation in the whitened space by a $\mathbf{W_0}$ orthonormal matrix, $\mathbf{S} = \mathbf{W_0 U} = \mathbf{W_0}\, \mathbf{X\Sigma V^{-1}}$ and independent basis is an orthonormal transformation of the PCA basis in the whitened space $\mathbf{E[SS^T]} = \mathbf{W_0}\, \mathbf{E[UU^T]W_0^T} = \mathbf{W_0 W_0^T} = \mathbf{I}.$ Thus, independent component basis for a low dimensional subspace is

$$\mathbf{X_{LD}} = \mathbf{S^T X} = (\mathbf{W_0 U^T})\mathbf{X} \qquad (7)$$

As a result, the main difference between the PCA and ICA basis is an orthonormal transformation that makes ICA basis decorrelated (eqn. (4) and eqn. (7)). This implies the invariance of Euclidean distance measurement in both PCA and ICA subspaces since Euclidean distance is preserved by an orthonormal transformation.

A description of the proposed approach using independent components with a vector quantizer classifier is provided. Modification of the bases is used to improve the generalization of the classifier. Environmental variants are considered to increase the intra-class variance. However, various implementations of the ICA algorithm that model the source distributions adjust the positions of the basis vectors thus providing a measure of invariance to the features.

2.2 Vector Quantization Algorithm for Speaker Verification

To construct centroids of partitions over speaker's feature space, the average error or distortion of the feature vectors $\{\mathbf{x}_t,\ 1< t <T\}$ of length T with the speaker **k** codebook is computed by

$$\mathbf{e_k} = \mathbf{1/\,T\, \Sigma^T_{t=1}\, min_{1<=j<=N}\, [d(x_t, C_{k,j})]} \quad \mathbf{1<= k <= L} \qquad (8)$$

d(. , .) is a general distance function between two vectors. $\mathbf{C_{k,j}} = (\mathbf{c_{k,,j,1}}, \mathbf{c_{k,,j,2}}, ..., \mathbf{c_{k,,j,M}})$ is the jth code of dimension **M**. **N** is the codebook size. **L** is the total number of speakers in the database. The VQ algorithm of SV can be implemented by Euclidean distance based LBG algorithm [12] to generate codebooks. In this case, the VQ attempts to find centroids of partition of each speaker by the uttered word using minimization of the Euclidean distance $\mathbf{d(x_t, C_{k,j})} = (\mathbf{\Sigma^M_{i=1}\, (x_{t,i} - C_{k,j,i})^2})^{1/2}$. As a result, the proposed ICA-VQ algorithm becomes an optimally matching pair for robust SV conditions due to the spanning of ICA subspace in adverse conditions and preserving the Euclidean distance, applying a modification to this subspace and employing the same Euclidean measure for verification decision.

3 ICA Approaches

To summarize, ICA searches an orthonormal rotation in the whitened space that also makes **S** statistically independent and the choice of ICA approaches can be between batch-mode (block) and adaptive algorithms. In the batch-mode, various efficient approaches are available: among these, tensor-based methods are used for small dimensions but a very popular one is FastICA approach. It is a fixed-point iteration that maximizes both one-independent component and multiple-independent components in the objective functions including likelihood. Here, we implement FastICA approach [14-15].

In the adaptive case, the approaches are derived by stochastic gradient methods. In the case of the simultaneous estimation of all independent components or one-component estimation, various approaches with different objective functions were studied. The popular choices become the gradient ascend of likelihood algorithms with constant functions, skewness of distributions with underlying source distributions such as the extended generalized lambda distribution model (EGLD) or Pearson system, related infomax objective functions and stochastic gradient methods that maximize negentropy and its approximations. In our study, we use two adaptive approaches: the EGLD-ICA [14, 17-18] and Pearson-ICA [14, 19].

1) FastICA approach. One can use batch (block) algorithms such as FastICA in an environment where no adaptation is needed. This is the case in many practical situations such as for the stationary interferences (such as periodic voices) to a SV system. The usage of FastICA offers a fast convergence. For one-independent component and whitened data case, fixed-point iteration FastICA with generalized objective function can be described as follows:

$$\mathbf{w(k)} = E\{\mathbf{x\ g(w(k\text{-}1)^T\ x)}\}\text{-}\ E\{\mathbf{g'(w(k\text{-}1)^T\ x)}\}\ \mathbf{w(k\text{-}1)} \qquad (9)$$

where ***E{.}*** is expectation operator, **w** is weight vector which is also normalized to unit form after every iteration, and **g** function is the derivative of the objective function **G**. The expectations are estimated by using sample averages over a sufficiently large sample of the input data. By combining the expectations of the algorithm, we estimate several independent components one-by one or in parallel.

The FastICA is a batch-mode, neural algorithm. It is implemented in parallel and distributed, but it is not adaptive. Instead of using every data point immediately for learning, it uses sample averages calculated over larger samples of data and finds all non-Gaussian independent components one at a time using a fixed-point iteration. For more details, we refer to Hyvärinen *et al.* [14-15]. The performance of FastICA approach in changing acoustic conditions is expected to be limited due to the nonadaptive nature of it.

2) EGLD-ICA approach. It is an adaptive maximum likelihood estimation ICA method which models source distributions by taking into account the skewness of the distributions. The estimated sources are modeled using the extended generalized lambda distribution model (EGLD). The modeling scheme provides a useful connection between practical estimator and theoretical measure of independence. The score function of the EGLD is used as an objective function of ICA that we maximize. Natural gradient and fixed-point algorithms that employs EGLD model are proposed for maximization.

The Generalized Lambda Distribution (GLD) is defined by the inverse distribution function

$$\mathbf{F^{-1}(p) = \lambda_1 + (p\ \lambda_3 - (1\text{-}p)\ \lambda_4) / \lambda_2} \qquad (10)$$

where $\mathbf{0 <= p <= 1}$ and λ_1, λ_2, λ_3 and λ_4 are the parameters of the distribution. GLD is valid when λ_2 / (λ_3 $\mathbf{p}^{\lambda 3\text{-}1}$ – λ_4 $\mathbf{(1\text{-}p)}^{\lambda 4\text{-}1}$) => **0**. The relationships between parameters λ_1, λ_2, λ_3 and λ_4 and moments α_1, α_2, α_3 and α_4 is established by four nonlinear equations that can be solved numerically. Since the density function of the GLD is not available in a closed form, a score function is derived from the inverse distribution function by **p = F(y)** where **F(y)** is the distribution function of a GLD. The value of score function for observation y is computed by numerically solving for solving for the value of p from eqn. (10) and then applying the score function **φ(p)** formula [12]. The coverage area of data of generalized beta distribution (GBD) overlaps to the GLD except a small region. We refer to Eriksson *et al.* [12] for details and summarize the EGLD-ICA procedure that repeats until convergence:

Step 1. Compute the third and fourth sample moments α_3 and α_4 for current data $\mathbf{y_k = W_k\, x}$ and choose GLD if $\alpha_4 > \mathbf{2.2 + 2{*}\alpha_3^2}$ and else GBD.

Step 2. Estimate parameters for GLD and GBD by method of moments and compute scores $\boldsymbol{\varphi(y_k)}$.

Step 3. Compute $\mathbf{W_{k+1}}$ using natural gradient or relative gradient algorithm

$$\mathbf{W_{k+1} = W_k + \eta\, (I - \varphi(y)\, y^T)\, W_k} \qquad (11)$$

where η is learning rate.

As a summary, maximum likelihood based EGLD-ICA uses source distribution adaptive objective function that is derived using EGLD as the model. The EGLD family is an extension of GLD and GBD source models that are generated by the computer and fitted to the empirical source data. The EGLD family includes the values for all the most important distribution including normal, uniform, gamma and beta distributions. This makes the approach separate a wide class of source signals sub and super-Gaussians, and skewed distributions with zero kurtosis. As a result, EGLD-ICA approach is expected to give a better performance in changing acoustic mismatch conditions.

3) Pearson-ICA approach. In this adaptive maximum likelihood based ICA method, the Pearson system is employed for modeling source distributions. The Pearson system also covers an extensive range of different values of kurtosis and skewness and includes many distributions with practical importance. A fixed objective function is used to improve the speed and stability of the Pearson-ICA in the cases where sources are easily separable. The Pearson system is defined by the differential equation

$$\mathbf{f'(x) = (x\text{-}a)\, f(x) \,/\, (b_0 + b_1\, x + b_2\, x^2)} \qquad (12)$$

where a, b_0, b_1 and b_2 are parameters of the distribution. It is shown that if the source distributions are known, the score functions are the optimal choice for the objective function. The score function of the Pearson system is easily solved from eqn (12), $\boldsymbol{\varphi(x) = -\, f'(x) \,/\, f(x) = (x\text{-}a) \,/\, (b_0 + b_1\, x + b_2\, x^2)}$. The parameters a, b_0, b_1 and b_2 may be estimated by the method of moments. The simplicity of score makes Pearson system particularly appealing for ICA.

In our experiments we employ fixed point algorithm and hyperbolic tangent contrast $\boldsymbol{\varphi(y) = \tanh(2y)}$ for both clearly sub-gaussian (e.g. a sine wave) and clearly super-gaussian sources (e.g. a synthetic ECG). We refer to Karvanen *et al.* [9, 14] for details and summarize the Pearson-ICA procedure that repeats until convergence:

Step 1. Compute the third and fourth sample moments α_3 and α_4 for current data $\mathbf{y_k = W_k\, x}$ and choose the Pearson system or fixed (**tanh**) objective (contrast).

Step 2. Once the Pearson system was chosen, estimate parameters of the distribution by method of moments.

Step 3. Compute scores $\boldsymbol{\varphi(y_k)}$ for the Pearson system or fixed objective.

Step 4. Compute $\mathbf{W_{k+1}}$ using natural gradient or relative gradient algorithm described in equation (12).

In the computation, both moment estimators for parameters and score function are simple rational functions and they can even be combined (Step 2 and Step 3) during the estimation stage, thus Pearson-ICA becomes faster.

As a fast and adaptive approach with a good model distribution, Pearson-ICA is also expected to perform reasonably well in changing acoustic mismatch conditions.

4 Simulation of Mismatch Conditions and Implementation of ICA approach

Independent components approach converges to BSS with example of the interfering signals [16-17], in which the observed values of **x** correspond to an m-dimensional discrete time signals x(t), t = 1,2,... recorded from the ith microphone and the independent components $s_i(t)$ correspond to original, uncorrupted original source signals. In SV with interfering signals case, assume that speaker is speaking simultaneously to each microphone with interfering voices such as other speakers as in a cocktail party, ambient noise and musical signals. Then the problem converges to separating the voices of different sources, using recordings of several (in our case two) microphones in the same environment.

Mismatch conditions in the SV experiments are modeled as ambient noise, music, interfering speakers and signals are represented in Figure 1. Since SV data is taken from standard YOHO database, we need to describe **x** samples of random observations through digitized samples in the computer. For this purpose, we produce the samples of corrupted speech by using ambient noise samples from white noise source, a playing music and the voices of other speakers of the database, then we combine these signals to speaker data. As a result, each of the sound mixtures in Figure 2 represents the ambient noise and cocktail party phenomena, and corresponds to a recording from two microphones in the environment.

Figure 3 shows the proposed system that consists of two modules: ICA and SV. Each module is implemented independently, and the input and output waveforms can also be observed individually.

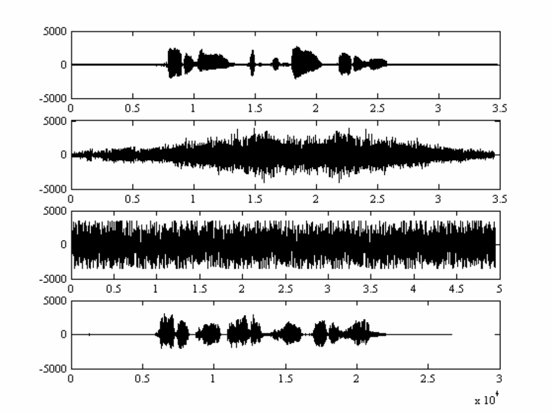

Figure 1. Original, uncorrupted source signals of original speaker, music, noise and interferer speaker.

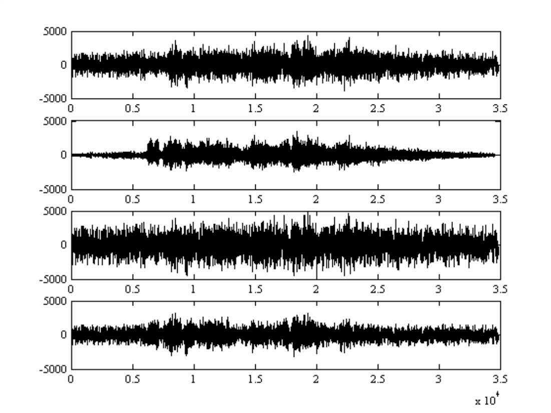

Figure 2. Linear mixtures of source signals as recordings from two microphones.

In our simulations with the ICA module, we assume two microphone inputs located with d distance as shown in Figure 4. Each microphone picks up signals from speech source and interference source which can be ambient noise, music or another speaker's voice. Ambient noise was simulated with additive white Gaussian noise with various levels of signal to noise (SNR) ratio. The other speakers' voices and various music signals are added to the original speaker's voice.

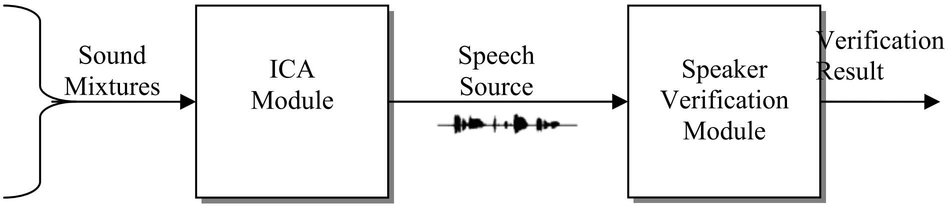

Figure 3. The proposed ICA-SV system

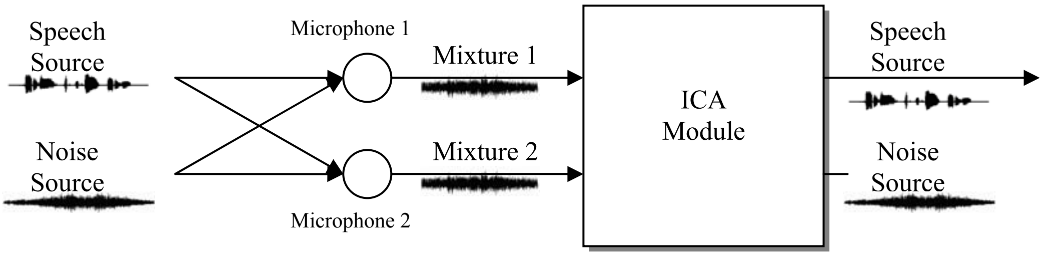

Figure 4. ICA module

5 Speaker Verification in Adverse Conditions

As it is known, speaker recognition (SR) involves which voice model from a known set of voice models best characterizes a speaker, task of speaker identification (SI), and the goal to decide whether a speaker corresponds to a particular known voice or to some other unknown voice, task of speaker verification (SV). SR methods can also be divided into text-dependent (TD) and text-independent (TI) methods [5-7]. The former requires the speaker to issue a predetermined utterance, whereas the latter do not rely on a specific text being spoken. In general, because of the higher acoustic-phonetic variability of the TI input, more training material is necessary to reliably characterize a speaker than with TD methods. For SV, input utterances with distance values (scores) to the reference template smaller than the threshold are accepted as being utterances of the registered speaker, while input utterances with distances larger than the threshold are rejected as being of those of a different speaker or impostors. With SI the registered speaker, whose reference template is nearest to the input utterance between all the registered speakers is selected as being speaker of the input utterance.

The performance of a verification system is often shown as a Receiver Operating Characteristics (ROC) or Detection Error Trade-off graph [5]. The ROC is obtained by varying the decision threshold and obtaining an operating point in terms of False acceptance [(FA(%)] and false rejection [(FR(%)] rate. While the ROC is useful for finding the discrimination ability of the system, it doesn’t convey information on how the system will perform in real life applications. In mismatch conditions it can easily be observed that the distribution of impostor and true claimant scores change, hence the threshold found for a particular operating point on the training data corresponds to a different operating point on test data in adverse conditions. This is an essential source of performance degradation that is interpreted as operating point shift on the ROC.

In this paper, we investigate the ICA based feature selection to a VQ-based SV system with three types of implementations: the FastICA, the EGLD-ICA and Pearson-ICA. The selected independent components provide a feature space modification scheme to improve the generalization of the classifier. They decompose the speech with environmental variants into new, modified basis vectors. Then, these selected vectors are used in the VQ SV system. The VQ includes the following stages: feature extraction, VQ preparation or training and the measurement of the distance between speakers and unknown speaker or testing. In the feature extraction stage, mel-cepstrum (MFCC) coefficients are computed from independent components. They are selected as commonly used representation with decorrelating property and compensation property for convolution channel distortion [5-7]. In VQ preparation or training stage, feature vectors are averaged over distinct sound classes to form series of vectors, known as codewords, from a training set, thus a non-parametric data reduction is applied to the speakers' data. In the testing stage, a matching operation is performed to match the features measured from the waveform of a test utterance, the test data of a speaker, against speaker models obtained during training [5, 6].

6 Experimental Setup and Results

The VQ-based verification system and speech pre-processing used for experiments are similar to standard systems [9-11]. Original and corrupted speech was analyzed every 10ms with a frame width of 20ms. Some of the important specifications are as follows:

i. 19 MFCC coefficients are used as feature vectors. We have found little improvement in using more features.
ii. In the VQ implementation, the codebook size is selected as 32. This was an optimum performance giving codebook size. A total of 55 randomly selected speakers from YOHO database are used to construct 55 codebooks. We have used 5281 recordings for training of VQ and 165 recordings for testing of the system.
iii. We have selected a value of 0.7 as the equal error rate (EER) performance or the decision threshold which minimizes FA and FR to a reasonable rate for the verification system as shown in Figure 5 and we have achieved 95% verification success rate.

The experiments were performed using the YOHO database prepared by Higgins *et al.* [8]. YOHO database is a large scale, scientifically controlled and collected, high-quality speech database for speaker authentication testing at high confidence levels. It includes combination lock phrases, like 34-56-78, recordings with 8 kHz sampling collected over 3 month period. We used 55 randomly selected speakers from YOHO database and used 5281 recordings for training and 165 recordings for testing.

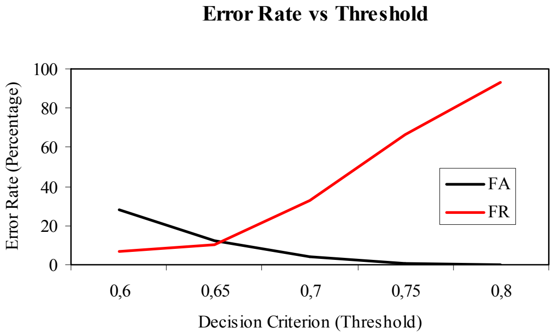

Figure 5. The relationship between verification error rate and ROC decision threshold.

To reliably test the performance of each ICA approach with the SV system we artificially generated four types of test sets each simulates the ambient noise and cocktail party conditions. In each test set we linearly mix the sounds shown below:

(i) **SS.** Mixture of an original speaker and one interferer speaker.
(ii) **SN.** Mixture of an original speaker and ambient noise.
(iii) **SM.** Mixture of an original speaker and music.
(iv) **SNMS.** Mixture of an original speaker and all of the signals which are interferer speaker, music and ambient noise.

Original source signals, that are used to generate sound mixtures, are all uncorrupted and known to be independent. Also for each speaker to test, we used the same ambient noise, music and competing speaker signals and generated the linear mixtures using the same mixing matrix. Without using any ICA module, the verification performance of the system decreases to 0% for each test sample due to the mismatch conditions as we expected.

The choice of a comparative measure for the performance of each specific case and method was also a primary issue. The experiments were conducted with various interferences such as competing speaker, ambient noise, music and a combination of all of them, thus we prefer *overall performance improvement (OPI)* for each case and method in the same conditions. Another objective measure was SNR improvement but this would not be suitable for interfering speaker and music cases because SNR was not specific to the nature of signal. SNR was only suitable for ambient noise interferences. Also, subjective listening tests were subjective measure for only output of ICA part they would not use for the overall system performance.

To observe the effects of operating point shift it would be ideal to report the performance in terms of both FA and FR. However, due to our primary interest we have quantified the performance into a single measure OPI = FA + FR. It is found for all the experiments, the FR increased as interferences existed, while FA decreased slightly. Hence for highly interfered cases the dominance of FR is observed.

Table 1 shows the performance improvements of the SV system with three types of preprocessing: FastICA, EGLD-ICA and Pearson-ICA and for four mismatch conditions of testing environment: **SS**, **SN**, **SM** and **SNMS**. The speakers' codebooks were generated from clean speech in the training part, then the threshold was chosen for impostors and true claimants. Finally the test part was used for final evaluation. For each speaker, his/her 3 test utterances were used separately as true claimants resulting in 165 true claimant tests. Impostor claims were simulated by using utterances from speakers other than the claimed speaker and his/her background speakers, resulting in 5281 impostor access tests. From the OPI results, we observe superiority of the EGLD-ICA approach. The EGLD-ICA approach has OPI scores that are better than the other methods. One reason is the adaptive nature of the algorithm that fits all adverse conditions. The other reason is the inclusion of generalized lambda distribution that models almost all types of disturbance sources. No clear conclusive result of superiority of OPI for the other methods is observed.

Table 1. The overall performance improvements (OPI) (%) of SV system after ICA preprocessing

	SS	SN	SM	SNMS
FastICA OPI	66	34	32	18
EGLD-ICA OPI	63	34	58	50
Pearson-ICA OPI	37	18	32	47

Our experiments were described in a login scenario that may be a part of multi biometric user verification process with the speaker and face identification components. When "logging on" to the handheld device, users spoke their name, and then spoke a prompted lock combination of randomly selected a few digit numbers near the frontal view of their face. The system recognizes the spoken name to obtain the "claimed identity". It may then perform face verification on the face image and speaker verification on the prompted lock combination phrase. Users were "accepted" or "rejected" based on the combined scores of the two biometric techniques. In this scenario, the effective usage of a preprocessing approach depends on its computational efficiency. It is observed that the batch mode FastICA becomes the fastest among all. But adaptive approaches EGLD-ICA and Pearson-ICA are also fast enough to be used for verification tasks in real applications.

We have also listened to the quality improvement of each separated signal over the mixed signals thus we subjectively observed the effectiveness of each algorithm. Subjectively, the outputs of EGLD-ICA also give a better listening quality.

7 Discussion of Results

A number of significant results are illustrated by these experiments:

- Enhanced generalization performance and lower error rates can be obtained by the selected features through ICA basis. This is also interpreted as the modification of the input feature space. These features are then used for the VQ based SV system.
- Independent components decompose the environmental variants from speaker's voice thus provides a better input to VQ system. Principal components are not very useful in this case since they are used to lower dimensionality of the feature space.
- Independent and principal components have an implicit relationship and they both preserve the Euclidean distance. In our study, it is claimed that ICA-VQ system becomes a matching pair since Euclidean distance between speaker's utterances is kept invariant during the ICA stage, then it is employed in the VQ algorithm.
- Adaptive EGLD-ICA approach gives better OPI. As a result, the EGLD-ICA becomes more effective with the adaptive nature and with the ability to model the interference sources. Extended generalized lambda distribution covers the interferences used in the simulations better. In fact, it is known that its coverage includes sub and super Gaussian distributions. This is also proven with our experiments.
- Generally it is interpreted that ICA preprocessing compensates the operating point shift on the ROC in the adverse conditions.

8 Conclusion

The study proposes a two-module system that includes an ICA-SV system used for mismatch conditions of training and test environments. The ICA processing becomes important in various conditions near the other robust SV methods such as the usage of robust features, classifiers, etc. It provides a feature space modification and selection scheme to improve the generalization of the classifier. It decomposes the corrupted speech into independent components through ICA basis vectors. Speaker verifier uses the new speech data that are derived from these projections into the modified and selected basis vectors. Principal components or PCA become ineffective for environmental variants. One main advantage of the usage of independent real-time ICA module is that it can be used with other methods that improve the effectiveness of the system. On the other hand, a drawback of it may be the usage of an input mechanism with two (or more) microphones for processing but this can be easily implemented in mobile devices with today's technology. The duty of second microphone is to present a second input to the unmixing matrix of ICA approach that is vital for separation of sources. Finally, our study also points that adaptive EGLD-ICA algorithm models the experimented interference conditions better than the other approaches. In various changing acoustic conditions such as the cocktail party conditions, ambient noise, music or in the combination of them, the EGLD-ICA approach introduces a substantial amount of improvement.

Acknowledgement

We express special thanks to BAP-05A101 project of Boğaziçi University.

References

1. A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, John Wiley & Sons. (2001)
2. E. Alpaydın, Introduction to Machine Learning, The MIT Press. (2004)
3. J. Fortuna, D. Capson, Improved Support Vector Classification using PCA and ICA Feature Space Modification, Pattern Recognition, Vol. 37, pp. 1117-1129. (2004)
4. P. Pudil, J. Novovicova, J. Kittler, Floating Search Methods in Feature Selection, Pattern Recognition Letters. Vol. 15(11), pp. 1119-1125. (1994)
5. S. Furui, Digital Speech Processing, Synthesis, and Recognition, Marcel Dekker Inc. (2001)
6. Thomas F. Quatieri, Discrete-Time Speech Signal Processing, Prentice Hall PTR.(2002)
7. X. Huang, A. Acero, and H. W. Hon, Spoken Language Processing, Prentice Hall PTR. (2001)

8. A. Higgins, J. Porter and L. Bahler, YOHO Speaker Authentication Final Report. ITT Defense Communications Division. (1989)
9. T. J. Hazen, E. Weinstein, and A. Park, Towards Robust Person Recognition On Handheld Devices Using Face and Speaker Identification Technologies, ICMI'03, pp. 289-292. (2003)
10. A. Park and T. J. Hazen, ASR Dependent Techniques for Speaker Identification, ICSLP 2002 Proc., pp. 1337-1340. (2002)
11. J. He, L. Liu, G. Palm, A Discriminative Training Algorithm for VQ-based Speaker Identification, IEEE Trans. On Speech and Audio Proc., Vol. 7, No 3, pp. 353-356. (1999)
12. N. Fan and J. Rosca, Enhanced VQ-based Algorithms for Speech Independent Speaker Identification, Audio and Video based Biometric Person Authetication (AVBPA) 2003 Proc. LNCS Springer, pp. 462-469 (ISBN 3-540-40302-7). (2003)
13. A. D. Reynolds, C. R. Rose, Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models, IEEE Trans. On Speech and Audio Proc., Vol. 3, No 1, pp. 72-83. (1995)
14. A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, John Wiley & Sons. (2001)
15. A. Hyvärinen and E. Oja, A Fast Fixed-Point Algorithm for Independent Component Analysis, Neural Computation 9:1483-1492. (1997)
16. J. Karvanen, J. Eriksson, and V. Koivunen, Pearson System Based Method for Blind Separation, Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, pp. 585-590. (2000)
17. J. Eriksson, J. Karvanen, and V. Koivunen, Source Distribution Adaptive Maximum Likelihood Estimation of ICA Model, Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, pp. 227-232. (2000)
18. Z.A. Karian, E.J. Dudewicz, and P. McDonald, The Extended Generalized Lambda Distribution System for Fitting Distributions to Data: History, Completion of Theory, Tables, Applications, the `Final Word' on Moment Fits, Communications in Statistics: Simulation and Computation, Vol. 25, No. 3, pp. 611-642. (1996)
19. J. Karvanen, J. Eriksson, and V. Koivunen, Pearson System Based Method for Blind Separation, Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, pp. 585-590. (2000)