World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:2, No:7, 2008

Spectral Analysis of Speech: A New Technique

Neeta Awasthy, Member IEEE, J.P.Saini and D.S.Chauhan

Abstract— ICA which is generally used for blind source separation problem has been tested for feature extraction in Speech recognition system to replace the phoneme based approach of MFCC. Applying the Cepstral coefficients generated to ICA as preprocessing has developed a new signal processing approach. This gives much better results against MFCC and ICA separately, both for word and speaker recognition. The mixing matrix A is different before and after MFCC as expected. As Mel is a nonlinear scale. However, cepstrals generated from Linear Predictive Coefficient being independent prove to be the right candidate for ICA. Matlab is the tool used for all comparisons. The database used is samples of ISOLET.

Keywords - Cepstral Coefficient, Distance measures, Independent Component Analysis, Linear Predictive Coefficients.

I. Introduction

ONE of the most common approaches in speech recognition involves frame based approach where development of front end feature measurements coefficients or spectral analysis is done. This involves feature detection in frequency space as in FFT or in time space as in LPC or in cepstral space in MFCC.

Pattern recognition technique is purely statistical for which the analog pattern must be digitized by sampling. Speech being highly dynamic in time, some methods must be utilized to find a short time stationary signal. This is done by forming small comparable block/frames and then window it to avoid the discontinuity at the end. This signal is processed.

Comparison of two samples in time domain in terms of amplitude is not possible and there is a huge variation in results as the same wave shape can be represented from infinite numbers of amplitudes loosing uniqueness in comparison results. There are a good number of techniques which can be used for comparison of two physical signals in frequency domain. Some popular techniques are : Autocorrelation, Discrete Fourier Transform(DFT), Fast Fourier Transform(FFT), Linear Predictive Coefficients(LPC) and Mel Frequency Cepstrum Coefficient(MFCC). For all simple purposes FFT gives the fastest results as the number of computations reduce drastically. FFT has its upper hand due to immunity to noise, but as the speech signal is non-stationary, its FFT is not possible. Further its STFT(Short Time Fourier Transform) may be calculated for a small frame, duly windowed. Further, it may be taken to wavelet transforms. In 1978 FFT lost its research track in speech recognition due to continuously failing in the constantly rising expectations of complexity.

The most thorough test of various front ends was done two decades back by Davis in his famous paper[1]. In this paper Linear Predictive Coefficients(LPC), Linear Predictive Cepstral Coefficient (LPCC), Linear Freq. Cepstral Coefficient(LFCC), Reflection Coefficient (RF) and Mel Frequency Cepstral Coefficient (MFCC) were tested for effects on accuracy of word recognition in a template based, dynamic time warping speech recognizer.

The basic inference drawn out of this work was that MFCC parameters give the best performance with six coefficients (or ten coefficients with increased performance of Automatic Speech Recogniser) addressing to most information, which is relevant for speech recognition. Even now his basic findings stand unchallenged [1].

Later, Joseph W. Picone[2] studied the works in Speech Recognition and discussed signal-modeling techniques in four sub categories:

[1]Spectral Shaping
[2]Spectral Analysis
[3]Parametric transformation
[4]Stastical modeling

It was highlighted that the difference in time of processing between various signal-modeling approaches was a small percentage of the total processing time. It was underlined that the focus was in maintaining high performance and minimizing the degrees of freedom. All techniques were compared in these four spheres and conclusion was:

1. Neural Network based systems tend to use filter bank amplitudes directly
2. Cepstral Coefficients are the dominant acoustic measurement
3. FFT-derived-mel-scaled cepstral coefficients are the most common form of cepstral analysis used
4. FFT is immune to noise

Here again, authority of MFCC stood unchallenged. However, it was an early age in Speech Recognition and there were many more things to come[2].

During that time ICA (Independent Component Analysis) was in its young stage. Herault and Jutter gave a general algorithm derived from PCA (Principal Component Analysis) which deals with IInd order statistics. In 1983 the problem of ICA was addressed and a real time iterative algorithm was proposed based on neuro-minetic architecture. In 1986, the name ICA came into horizon. However, at those times, the higher order statistics and cumulants were not introduced explicitly. PCA forms an orthogonal set of axes pointing in the direction of maximum variance, so it forms a representational basis that projects in the direction of maximum variability. ICA is a generalization of PCA that decorrelates the high order moments of the features up to infinite order in theory. Early approaches showed an infinite number of cross cumulants and a maximum of fourth order independence. The most popular fixed-point algorithm was given by Hyvarinen [4], known as FASTICA. Giannakis et.al.[10], in 1987 addressed the issue of identifiability of ICA in 1986 using third order cumulants. Cardoso focused on algebraic properties of fourth order cumulants and interpreted them as linear operators acting on matrices[11]. Later he investigated other algebraic approaches using only fourth order cumulants[12,13]. In this series the most recent research has been carried out with third and fourth order cumulants named as CuBICA by Blaschke and Wiskott [5].

II. LPC ANALYSIS

i) Preemphasis – The digitized speech is passed through a first order low pass filter. This process flattens the signal and makes it less susceptible to finite precision effects later in the signal processing. To average the transmission conditions and backgrounds, or, even to average the signal spectrum, the preemphasizer is made to adapt slowly.
The first order preemphasis network is defined as :

$$H(z) = 1-az-1 \qquad 0.9 \le a \le 1.0 \qquad (1)$$

Thus, the output of preemphasis network s~(n) is related with the input s(n), by the difference equation:

$$s'(n) = s(n)- as'(n-1) \qquad (2)$$

the values of a commonly used is 15/16=0.9375 or 0.95
One possibility to choose a in (2) is to choose a'(n) = $r_n(1)/r_n(0)$

In this paper, a one step forward linear prediction has been used i.e. the value of x(n) by a weighted linear combination of past values x(n-1), x(n-2),…,x(n-p) has been predicted. Hence the linearly predicted value of x(n) is

$$X'(n) = -\sum_{k=1}^{p} a_P(k) X(n-k) \qquad (3)$$

where $a_p(k)$ represents the weights in the linear combination. The difference between the value x(n) and the predicted value x'(n) is called the forward prediction error fp(n):

$$f_P(n) = X(n) - X'(n) = X(n) + \sum_{k=1}^{p} a_P(k) x(n-k) \qquad (4)$$

ii) Frame Blocking- The preemphasized speech signal, is blocked into the frames of N samples, with adjacent frames being separated by M samples. Let the first frame consist of first N samples. The second frame begins M samples after the first frame, and overlaps N-M samples.. This process continues until all the speech samples are accounted for in one or more frames. A common technique is to have small frame size with wide overlapping. It is easy to see that M≤N, then the adjacent frames overlap, and the resulting LPC spectral estimates will be correlated frame to frame. If M<<N, the LPC spectral estimates from frame to frame will be smooth with a trade off with complexity of computation. On the other hand, if M>N, then there will be loss of information and the speech will not be recognized or reconstructed. A typical result is an analysis frame of 300 samples and no. of samples shift between frames to be 100

iii) Windowing- For speech processing we want to assume that the signal is short-time stationary and perform a Fourier transform on these small blocks. A simple solution is to multiply the signal by a window (gate) function that is zero outside some definite range. Then the resulting windowed signal is defined as:

$$x_l'(n) = x_l(n)w(n) \qquad 0 \le n \le N-1 \qquad (5)$$

This can generate the discontinuities.
One way to avoid these discontinuities and taper the signal at the beginning and the end of each frame is use of a Hamming Window for auto correlation method of LPC due to its raised cosine structure. It is defined as follows:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad 0 \le n \le N-1 \quad (6)$$

iv) Autocorrelation Analysis- In this step, each frame of windowed signal is auto correlated to give:

$$r_l = \sum_{n=0}^{N-1-M} \tilde{x}_l(n)\, \tilde{x}_l(n+m) \qquad m=0,1,2,\ldots.,p, \quad (7)$$

where highest correlation value, p, is the order of LPC analysis. Typically, values of p from 8 to 16 are used. It is interesting to note that the zeroth autocorrelation, $R_l^{(0)}$, is the energy of lth frame. The frame energy is an important parameter for speech-detection.
In our case,

$$\gamma_{xx}(l) = \sum_{n=-\infty}^{\infty} x(n)x(n-l) = \sum_{n=-\infty}^{\infty} x(n+l)x(n)$$

$$\text{for } l=0,\pm1, \pm2,\ldots \qquad (8)$$

v) LPC Analysis- The next processing is LPC analysis, which converts each frame of p+1 autocorrelations into an 'LPC parameter set'. In our case, it is LPC coefficients. It can also be the reflection (PARCOR) coefficients, the log area ratio coefficients, the cepstral coefficients, or any desired transformation of the above. We have used Levinson/Durbin Algorithm:

1. $E^{(0)} = r(0)$ (9)

2. $$k_i = \frac{r(i) - \sum_{j=1}^{L-1} \alpha_j(i-1) r(|i-j|)}{E^{(i-1)}} \qquad 1 \leq i \leq p \qquad (10)$$

3. $\alpha_i^{(i)} = k_i$ (11)

4. $\alpha_j^{(i)} = \alpha_i^{(i-1)} - k_i \, \alpha_{i-j}^{(i-1)}$ (12)

5. $E^{(i)} = (1-k_i^2)E^{(i-1)}$ i=j=i-1 (13)

Where the summation of is omitted for i=1. On solving the equations recursively for i=1,2,3,…p, the final solution is given as:

a_m = LPC coefficients = $\alpha_m^{(p)}$ $1 \leq m \leq p$ (14)

vi) LPC Parameter Conversion to Cepstral Coefficients- A very important LPC parameter set, directly derived from LPC coefficients is the LPC cepstral coefficients, c(m). The recursion used is:

$$c_o = \ln \sigma^2 \qquad (15)$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \qquad 1 \leq m \leq p \qquad (16)$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \qquad m > p \qquad (17)$$

where σ^2 is the gain term in the LPC model. The cepstral coefficients, which are the coefficients of the Fourier transform representation of the log magnitude spectrum, have been shown to be more robust, reliable feature set for speech recognition than other set of coefficients.

vii) Parameter Weighting- The low-order cepstral coefficients are sensitive to overall spectral slope and the high-order cepstral coefficients are sensitive to noise and other forms of noise like variability. Thus, a standard technique is to weigh the cepstral coefficients by a tapered window so as minimize these sensitivity. One option is to take Fourier representation of the log magnitude spectrum and the differentiated log magnitude spectrum, such that:

$$\log |S(e^{jw})| = \sum_{m=-\infty}^{\infty} c_m e^{-j\omega m} \qquad (18)$$

$$\frac{\partial}{\partial \omega}[\log |S(e^{jw})|] = \sum_{m=-\infty}^{\infty} (-jm) c_m e^{-j\omega m} \qquad (19)$$

The differential log magnitude spectrum has the property that any fixed spectral slope in the log magnitude spectrum becomes a constant. Any prominent spectral peak in the log magnitude spectrum or formants is well preserved as a peak in the differentiated log magnitude spectrum. Hence, by considering the multiplication by (-jm) in the representation of the differentiated log magnitude spectrum as a form of weighting, we get

$$\frac{\partial}{\partial \omega} \log |S(e^{jw})| = \sum_{m=-\infty}^{\infty} \hat{c}_m \, e^{-j\omega m} \qquad (20)$$

$$\hat{c}_m = c_m(-jm) \qquad (21)$$

To achieve the robustness for large values of m (low weight near m=Q) and to truncate the infinite computation of these equations, we consider a more general weighing form

$$\hat{c}_m = w_m c_m \qquad 1 \leq m \leq Q \qquad (22)$$

where an appropriate weighing is the band pass lifter is given as :

$$w_m = [1 + \frac{Q}{2} \sin(\frac{\pi m}{Q})] \qquad 1 \leq m \leq Q \qquad (23)$$

This weighing function truncates the computation and de-emphasizes cm around m=1 and around m=Q.

III. INDEPENDENT COMPONENT SOLUTION

A. Definition:

The independent components are latent variables, they can not be directly observed. ICA is used to find unknown waveforms out of the mixtures/ sources given. The observered m values of x correspond with n constituent source signal :

$X_1(t)=a_{11}S_1+a_{12}S_2+\ldots+a_{1n}S_n$
$X_2(t)=a_{21}S_1+a_{22}S_2+\ldots+a_{2n}S_n$
.
.
.
$X_m(t)=a_{m1}S_1+a_{m2}S_2+\ldots+a_{mn}S_n$ (24)

This problem assumes that $S_1(t)$ and $S_2(t)$ are statistically independent at any time instant. ICA is used to estimate a_{ij} based on the information of the independence of the source, which allows us to separate the signals out of their observed mixtures.

The N element vector is the mixture input to ICA which is actually the Fourier coefficients of input signals. There are n such mixtures of m independent sources so that m<n. The basic mixing model of ICA without noise is presented as:

$$x = As \qquad (25)$$

or

$$x = \sum_{i=1}^{n} a_i s_i \qquad (26)$$

Noise is generally considered gaussian to simplify a problem, but in ICA it will further complicate the problem, as it has already been proved that only one of the sources can be gaussian.

B. Assumptions

The assumptions are :

1. The components s_i are statistically independent.
2. Independent components must have non-gaussian distribution because gaussian mixtures can not be separated.
3. As supported by theorem 11, in combination to theorem 10,of the famous paper of Pierre Comon; only one of the mixtures can be Gaussian[7].

Let us check each assumption at a time.

1. The components s_i are statistically independent:
Consider two scalar-valued random variables s1 and s2. they have no information about each other, but not x1 and x2, which are mixture variables. When defined in probability density; If $p_1(y_1)$ is the probability density function of y_1 alone and If $p_2(y_2)$ is the probability density function of y_2 alone, then independence means that their joint probability is factorisable:

$$p_1(y_1,y_2)= p_1(y_1).\ p_2(y_2) \qquad (27)$$

or given two functions, h_1 and h_2, we always have their expectations:

$$E\{h_1(y_1)\ h_2(y_2)\}=E\{h_1(y_1)\}\ E\{h_2(y_2)\} \qquad (28)$$

Independence implies uncorrelated ness. However, uncorrelated ness does not imply Independence- two random variables y_1 & $y_2$are said to be uncorrelated , if the covariance is zero:

$$\text{i.e.}E\{y_1y_2\}=E\{y_1\}E\{y_2\} \qquad (29)$$

$$\text{whereas, } E\{y_1^2y_2^2\}\neq E\{y_1^2\}E\{y_2^2\} \qquad (30)$$

Thus, ICA methods constrain the estimation procedure so as to get uncorrelated independent components estimates. This simplifies the problem of ICA by reducing the number of free parameters.

2. Independent components must have non-gaussian distribution because gaussian mixtures can not be separated:

Say A is orthogonal and sources S_i are gaussian. Then mixture x_1 & x_2 are gaussian , uncorrelated and of unit variance.
Then probability density function becomes:

$$f(x_i)=\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2),-\infty < x < \infty \qquad (31)$$

Where $-\infty<\mu<\infty$ is mean and $\sigma>0$ is standard deviation. The distribution function becomes

$$F(x)=\frac{1}{2\pi}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt \qquad \text{with } \sigma=1 \text{ and } \mu=0 \qquad (32)$$

Then, their joint densities are given by

$$p(x1,x2)=\frac{1}{2\pi}\exp(\frac{-x_1^2+x_2^2}{2}) \qquad (33)$$

This is a completely symmetric circle.
It does not contain information on the direction of the columns of mixing matrix A, Thus, A can not be estimated.

3. Only one of the sources can be Gaussian-
For X=AS .Let :

a. X and S be two random vectors.
b. A is a rectangular matrix
c. S has independent components.
d. X has pairwise independent components.

If A has two non-zero entries in the same column j, then S_j is either gaussion or deterministic.This is a direct consequence of Darmois Theorem stated in 1953 as:
Define the two random variables X1 and X2 as –

$$X1=\sum_{i=1}^{N}a_ix_i \qquad X2=\sum_{i=1}^{N}b_ix_i \qquad (34)$$

Where x_i are independent random variables. Then if x_1 and x_2 are independent, all variables x_j for which $a_j\ b_j\neq 0$ are gaussian. Also , Let S be a vector with independent components, of which at most one is Gaussian, and whose densities are not reduced to a point like mass. Let A be an orthogonal NXN matrix and X the vector X=AS. Then the following properties are equal.

a. The components X_i are pair wise independent.
b. The components X_i are mutually independent
c. $C=\wedge P$, Where, $\wedge$ is scaling diagonal matrix, P is permutation.

Note that implications iii) ⇒ ii) and ii) ⇒i) are quite obvious. Last one i) ⇒iii) has to be proved.
Assume X has pair wise independent components, and let A is not of form $\wedge P$. since A is orthogonal, it is necessarily has two non –zero entries in at least two different columns. Then using the entries proof twice, S has at least two Gaussian components, which is contrary to the original hypothesis, that A is not of form $\wedge P$. So if A is $\wedge P$ then only one of the source can be Gaussian. Converse of this is equally true i.e. If the distribution is non gaussian it is independent. The central limit theorem , states that sums of independent random variables tends to be normally distributed even though its summands are not. Thus the sum of two independent random variables usually has a distribution that is closer to gaussian than any of the two original random variables.Now a very natural question arises here is measure of gaussianity or non gaussianity, which gives rise to an estimation principal of ICA.

C. Principals of ICA Estimation:

Two major principals are used for ICA estimation:
a. The information theory approach to ICA is minimization of mutual information ie finding most non gaussian direction. Non gaussian is Independent.

Mutual Information is defined as:
Let x be a random variable with values in R^N and denoted by $p_x(u)$ its probability density function. Vector x has mutually independent components if and only if

$$p_x(u) = \prod_{i=1}^{N} p_{xi}(u_i) \quad (35)$$

So , a natural way of checking whether x has independent components is to measure a distance between both sides

$$\delta(p_x, \prod_{i=1}^{N} p_{x_i}) \quad (36)$$

So the average mutual information of x is :

$$I(p_x) = \int p_x(u) \log(\frac{p_x(u)}{\prod p_{x_i}(u_i)}).du, u \in C^N \quad (37)$$

If we recall kullback- leibler divergence defined for probability density form

$$I(p_x) = \int p_x(u) \log(\frac{p_x(u)}{p_z(u)}).du \quad (38)$$

This equation satisfies

$$\delta(p_x, p_z) \geq 0$$

And it satisfies the equality if and only if $p_x(u) = p_z(u)$ almost everywhere. Thus, (37)holds if and only if the variables x_i are mutually independent.

In information theory, mutual information is defined between m scalar random variables, y_i, I = 1…..m, as follows

$$I_{(y_1, y_2, \dots y_m)} = \sum_{i=1}^{m} H(y_i) - H(y) \quad (39)$$

where $H(y_i)$ is length of codes with y_i when these elements are coded separately and H(y) is the length of code when y is coded as random vector. (39) is always positive and zero, if and only if the components are statistically independent.

Non gaussianity can be used to estimate the Independence of the components. The measure of non gaussianity is kurtosis or fourth order cumulant defined by

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (40)$$

Properties of kurtosis are:
1. Kurtosis may be positive or negative: so that square of kurtosis are mod(kurtosis) can be used as a measure of non gaussianity.
2. Kurtosis is linear :

$$kurt(x_1 + x_2) = kurt(x_1) + kurt(x_2)$$

and $kurt(\alpha x) = \alpha^4 kurt(x)$ (41)

So, independent components can be formed by kurtosis minimization.
The additive property of kurtosis states:

For $y = W^T x = z_1 s_1 + z_2 s_2$

$$kurt(y) = kurt(z_1 s_1) + kurt(z_2 s_2)$$

$$= z_1{}^4 kurt(s_1) + z_2{}^4 kurt(s_2) \quad (42)$$

and since variance is unity thus

$$E(y^2) = z_1{}^2 + z_2{}^2 = 1$$

It is easy to show by deflation approach of adaptive blind separation of independent sources that maxima of z_1^4kurt(S_1) + z_2^4kurt(S_2) – are at the points when one of the elements of vector Z is zero and other nonzero. Because of unit circle constrain one element is 1or –1 .Thus, kurtosis can very well be theoretically used as optimization criterion for ICA.
Disadvantages of kurtosis:
Kurtosis can be very sensitive to outliers. Its value may depend on only a few observation in the tails of distribution. While may be erroneous or irrelevant observations.

b. Another measure of non gaussianity is negative entropy. Entropy is coding length of random variable. The more random/ unpredictable or unstructured the variable is , the more positive entropy is then,
For discrete random variable, Entropy is defined by

$$H(y) = -\sum_i P(Y = a_i) \log P(Y = a_i) = -\sum_i P \log P \quad (43)$$

Similarly the differential entropy for continuous random vector y with density f(y) is defined by

$$H(y) = -\int f(y) \log f(y) dy \quad (44)$$

The gaussian variable has the largest entropy among all random variables of equal variance. So negentropy is difference between maximum entropy , gaussian and entropy of the given random variable defined as

$$J(y) = H(y_{gaussian}) - H(y) \quad (45)$$

Hence negentropy can only be positive, it is actually the relative entropy. Defined as negative entropy or negentropy. Both gaussian feature and natural independence can be characterized by with the help of negentropy. The only problem of using negentropy lies is its computation complexity, so its simpler approximations are used.

D. *Preprocessing for ICA*

1.Centering the data-
The most basic and necessary preprocessing is to center the data. This is done by subtracting its mean vector m=E{x} so as to make x a zero mean variable. This implies that now s is also zero-mean. It can be shown by taking expectations on left and right side basic equation of ICA. However, after ICA this mean vector is added back to centered estimates of s.

2. Whitening-
Next is to whiten the observed variables by transforming the observed vector x linearly so that its components are uncorrelated and their variances are equal to unity, so that E{x’x’^{T}}=I. It is done by eigen value decomposition. This reduces the number of parameters to be estimated. Instead of n2 parameters of Ã only n(n-1)/2 parameters of orthogonal matrix A is found. Discarding the eigen value dj of E{xx^T} will further reduce the dimension of data.

3. Filtering-
The performance of ICA for a given set of data may depend upon some band pass filtering. Let $x^*_i(t)$ be the linearly filtered output of observed signals $x_i(t)$.The ICA model still holds for $x^*_i(t)$, with the same mixing matrix. In this paper it is extended whether this holds for linear cepstrals which are independent also.

E. 'FastICA'

A very efficient and famous algorithm of 'FastICA' was given by Hyvarian. This allows a preliminary 'whitening' step for the zero mean mixture signals, which improves the convergence speed of the ICA procedure. The process is as follows:

A. Initialize nonzero weights W.
B. Iterate till it converges:

1.For outputs p=1,…,n: perform steps(2-4):
2.Vector update:

$$w_p^+ = E\{xg(w_p^T x)\} - E\{g'(w_p^T x)\}w_p \quad (46)$$

where g is a nonlinear function, g' its derivative with time.

3..Normalize to a unitary-length vector:

$$w_p = \frac{w_p^+}{\| w_p^+ \|} \quad (47)$$

4.De-correlation of current vector (by a Gram-Schmidt orthogonal) against the previous vector set :

$$w_p = w_p - \sum_{j=1}^{p-1} w_p^T w_j w_j$$

$$w_p = \frac{w_p}{\| w_p \|} \quad (48)$$

F. Measure of Dissimilarity

For two feature vectors defined on a vector space χ, We define a metric or distance function 'd' on the vector space χ as a real-valued function on the Cartisian product $\chi x \chi$ such that it qualifies the following properties:
1. Positive Definiteness –
$0 \le d(x,y) < \infty$ for $x,y \in \chi$and $d(x,y)=0$ if and only if x=y
2. Symmetry – $d(x,y) = d(y,x)$ for $x,y \in \chi$
3. Triangular Inequality – $d(x,y) \le d(x,z)+d(y,z)$ for $x,y,z \in \chi$
4. Distance Function – $d(x+z,y+z) = d(x,y)$
Distance function is included because speech is a subjective data and something like loudness cannot be measured with only first three properties.

For speech recognition, only MSE does not fulfill the purpose of distance with subjective meaningfulness [8]. Thus, two components for Identification /Faulty rejection are taken. They are:

1. Rejection of correct data (utterance/ speaker wise)
2. Identification of wrong data (utterance/ speaker wise)
So, the performance index must have two components as shown in eqn.(7).

Distance has been assigned as sum of square of distance and the error is considered as inverse of the same. MSE[$-y_i,s_i$] and MSE[y_i,s_i] are computed and lower value is found. These MSE(s) form a matrix $E=[a_{i,j}]_{nxn}$; where each element

$$a_i = \frac{1}{\sqrt{MSE[y_i, s_i]}} \quad (49)$$

One more error element is added to this error which is a penalty to wrong identification i.e. if a single reference component is matched with more than one tested component; so that the total error becomes positive and its magnitude is large enough to be compared as Performance Index (PI).

$$PI = \frac{1}{n}[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{a_{ij}}{\max_i(a_{ij})} - n] + \frac{1}{n}[\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{a_{ij}}{\max_j(a_{ik})} - n] \quad (50)$$

G. Applications
Applications of ICA can be prominently divided into three parts:-
a) Depending upon the type of mixing matrix X:
i) If mixing matrix A is Toeplitz and triangular then ICA is only a deconvolution problem.
ii) if A is not triangular, the filter is allowed to be non-causal.
iii) If A is non-toeplitz, the filter is allowed to be non-stationary.Thus, blind deconvolution is a constrained ICA problem.

b) Blind Source Separation:
During 1989-91 the problem of separation of two and more than two sources was addressed by various researches. In his doctoral thesis, Fety, [13-5,24] addressed the problem of identifying the dynamic model of y(t) = F.z(t) using second order cumulants. This was addressed as signal operation problem. This problem was later addressed by Tond in 1991 [14-5,61]. This has applications in famous 'Cocktail Party problem' or used in channel identification. Other applications are in antenna array processing, in estimation of radiating sources form the unknown arrays [5-20], or jammer rejection or in noise reduction or in two stage localization procedure, if array are perturbed/ noisy or ill- calibrated[4].

c) Feature Extraction:
The second application of ICA is feature extraction, on which this work has been carried out. This is a fundamental problem in digital signal processing where it is a challenge to find out suitable signature or characteristic features for image, audio, channel properties and for data compression and denoising.
Independent Component Analysis is a statistical method for transforming an observed multidimensional random vector into components that are statistically independent from each

other[6]. Thus feature extraction turns out to be an application of ICA[4]. For successful extraction of features a lot of techniques were developed as Fast Fourier Transform, Bank of filters front end processor, Short Term Power Spectrum, Linear Predictive Coefficients, Discrete Cosine Transform (DCT), MFCC to name a few most respected ones. It is interesting to note that linear cepstrals being independent can form the same output matrix.

IV. EXPERIMENTS

The three approaches tested are:
1. Standard MFCC features
2. ICA for speech feature detection
3.Weighted LPC cepstrals passed through ICA for speech feature detection
The two set of data is :
1. Single speaker multiple utterances: Word Identification
2. Single utterance multiple speakers : Speaker Identification

Here the responses on database of utterances from one speaker based on samples of ISOLET database has been investigated. The ISOLET speech database of spoken letters of the English alphabet. The speech is high quality (16 kHz with a noise canceling microphone). 150 speakers x 26 letters of the English alphabet twice in random order. The original frame based approach has been accepted. First all the components are rescaled in<-1,1> range. The endpoints are detected, then a frame of 10 ms with a hamming window with 50% overlap is created. The features are extracted only beyond a certain threshold level. The features extracted were three formant frequencies from LPC analysis and 15 MFCCs with one endpoint detection component. These features were stored in a reference matrix per spoken word r(i). Now, the test input is generated. Its features are extracted say y(i) and compared with reference input components say s(i). Later all features are extracted from LPC analysis. For better results the LPC cepstrals are weighted.

Then the same database is used for ICA technique. The templates of ICA are created and compared. Then on the same database its LP cepstrals are taken and this vector is applied to ICA. All the comparisons are done after Dynamic Time Warping (DTW).

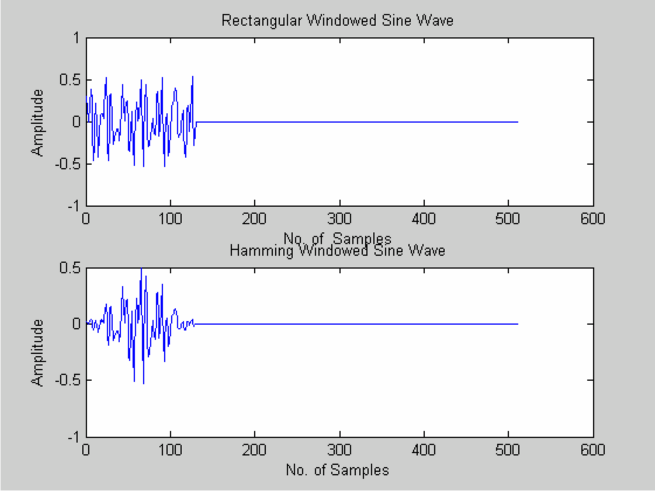

Fig.1: Windowing of signals (Hamming window reducing discontinuity)

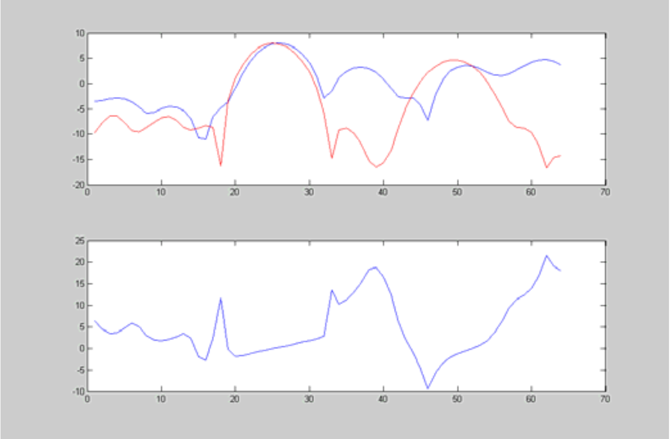

Fig. 2: LPC of two utterances ‘A’ and ‘B’

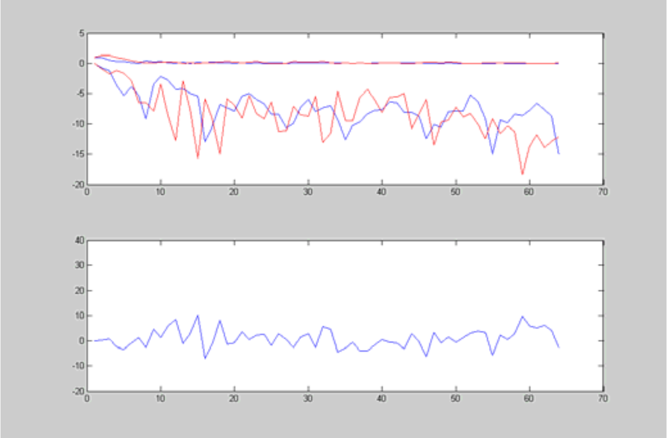

Fig.3.:Weighted LPC Cepstrals giving a smoother curve

Ist Set: Single speaker multiple utterances:

Table I: Comparison of error using formant frequencies and MFCC as features.

Tested / Reference Letter	A1	A2	B1	B2
A1	5.39	10.31	205.81	203.02

A2	12.67	5.26	189.66	203.09
B1	201.33	203.45	5.38	10.73
B2	199.64	204.84	11.65	5.14

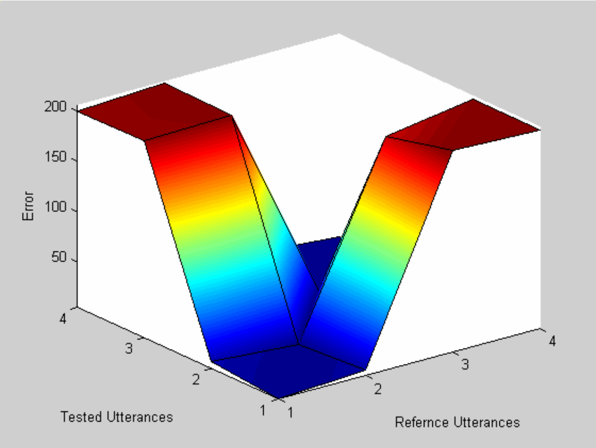

Fig.4: Plot of Error in utterances using MFCC

Table II: Comparison of error using ICA as features.

Tested/ Reference Letter	A1	A2	B1	B2
A1	6.43	12.87	230.91	224.98
A2	12.92	6.86	189.66	214.53
B1	214.87	231.09	7.03	13.05
B2	220.61	219.01	12.79	6.76

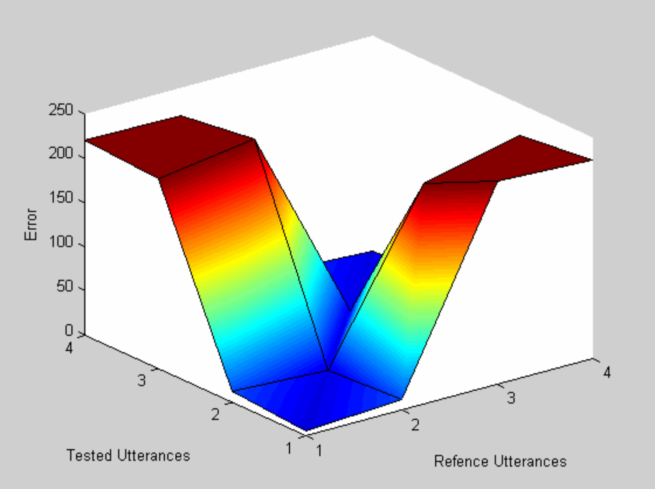

Fig.5: Plot of Error in utterances using ICA

Table III: Comparison of error using weighted LPC applied to ICA as features.

Tested/ Referenc e Letter	A1	A2	B1	B2
A1	2.90	11.78	220.71	223.02
A2	11.63	3.01	230.83	220.83
B1	235.95	229.06	2.96	12.00
B2	220.93	220.94	11.09	3.04

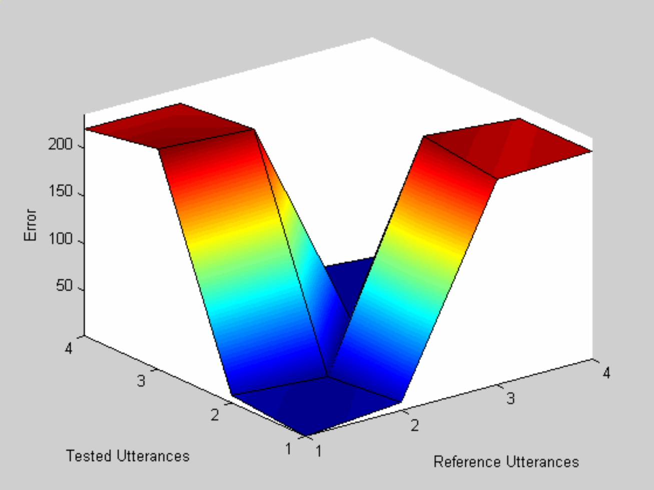

Fig.6: Plot of Error in utterances using Linear Predictive Cepstral applied to ICA

IInd Set: Multiple speakers same utterance('B' for this case):

Table IV: Comparison of error using formant frequencies and MFCC as features.

Tested / Reference Speaker	Male1	Male2	Female1	Female2
Male1	3.26	15.62	145.421	137.82
Male2	18.24	3.45	196.30	159.28
Female1	139.25	150.28	3.49	20.21
Female2	169.35	123.85	13.58	3.82

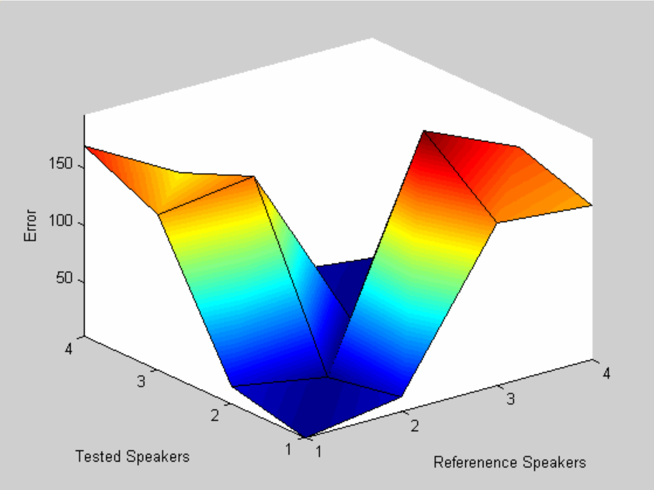

Fig.7: Plot of Error in speakers using MFCC

Table V: Comparison of error using ICA as features.

Tested/ Reference Speaker	Male1	Male2	Female1	Female2
Male1	3.12	12.87	230.91	224.98
Male2	12.56	2.99	189.66	214.53
Female1	110.25	146.21	2.86	13.00
Female2	156.38	112.69	12.82	2.81

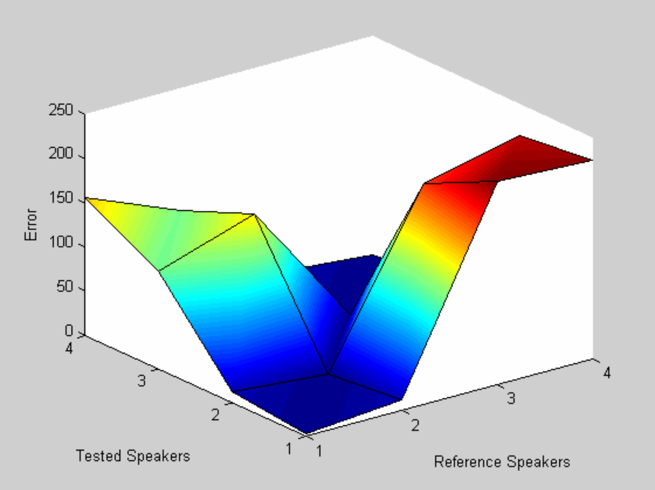

Fig.8: Plot of Error in speakers using ICA

Table VI Comparison of error using weighted LPC applied to ICA as features.

Tested/ Reference Speaker	Male1	Male2	Female1	Female2
Male1	1.48	13.92	222.70	224.58
Male2	14.85	2.04	241.59	223.59
Female1	270.28	210.79	2.20	16.02
Female2	198.26	165.93	14.69	2.31

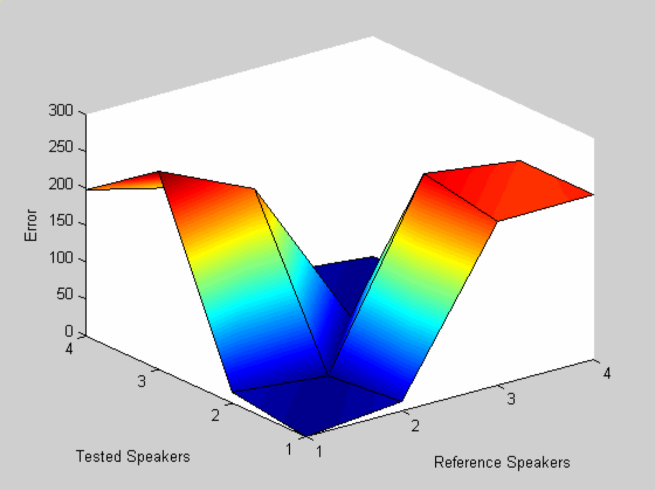

Fig.9: Plot of Error in speakers using Linear Predictive Cepstral applied to ICA

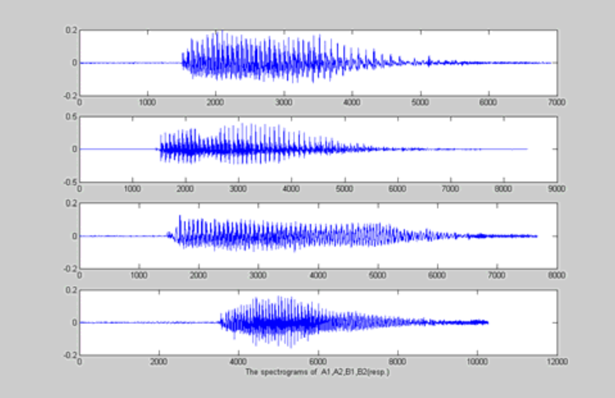

Fig.10: Two wave forms of 'A' and two waveforms of 'B'

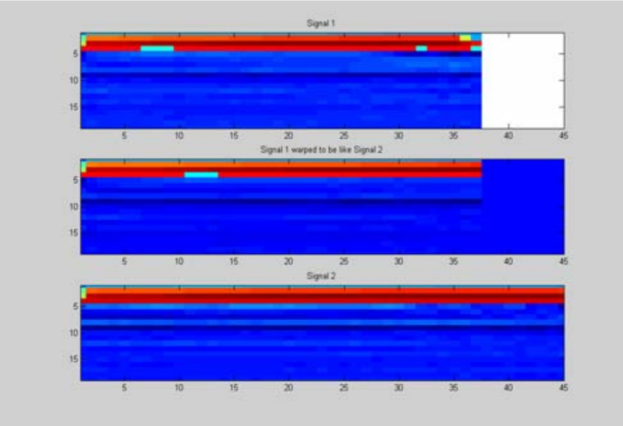

Fig.11: DTW of MFCC of 'A1' and 'A2'

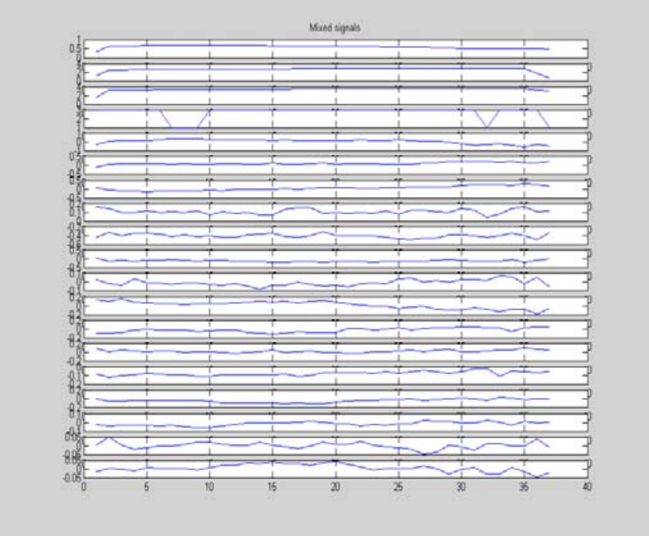

Fig.12: Plot of all Weighted LP Cepstral applied to ICA for 'A1'.

V. RESULTS AND CONCLUSION

In Fig.1, the first plot is signal multiplied with rectangular window and the second plot is of signal multiplied with hamming window. It is evident that in case of rectangular window, there are sharp discontinuities at beginning and end of the window; whereas, for hamming window, the discontinuity is not so sharp. Fig. 2 shows the LPC coefficient fo two utterances and their comparisons. Fig. 3 shows weighted cepstrals and their smoothened comparative results.

During this work the mixing matrix A was checked up with MFCC and without MFCC. They are different as expected becaure MFCC is on a Mel scale which is a non linear scale. Thus LPC was taken and The performance index was calculated as described in Section II. In Table I, comparison of same alphabet gives minimum error. In Table II, the comparisons were made out of ICA coeff. Result is worse than the previous (MFCC case). In Table III, LPC coefficients were applied to ICA. It was found that in this particular case of ISOLET database, the performance is better than the earlier two algorithms (MFCC and ICA separately); i.e. for identification of same or similar alphabet, the error has reduced , whereas for different alphabet the error is larger than the earlier two approaches. Thus the classification is better. Fig. 4,5,6 are the illustrations of table I, II and III respectively.

Similarly, for speaker identification, same / similar speakers gave lesser error with MFCC as compared to ICA(Table IV and V) but weighted LP Cepstrals applied to ICA technique gave better performance than the earlier two approaches(Table VI). In case of same speaker the PI (actually error) got reduced, whereas, for different speakers the PI became exceptionally large(Table VI), so classification is even better. In our future experiments we will find out optimum neural architecture in this case to improve comparison. Fif. 7,8,9 are the illustrations for table IV, V and VI. Fig. 10 shows two sample wave forms of two utterances. Figure 11 illustrates that the difference in time duration of two utterances (say 'A') is taken care of by Dynamic Time Warping of two signals. Fig12 shows the plot of all 19 weighted cepstrals applied to ICA.

REFERENCES

[1] Davis S. and P.Mermelstein," Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. ASSP* 28,pp.357-366,1980.

[2] Joseph W. Picone, "Signal Modeling Techniques in speech recognition," *Proceedings of the IEEE,* vol.81, no.9, pp.1215-1247,1993.

[3] Jutten C. and Herault, "Blind Separation of Sources, Part I: An adaptive algorithm based on a neuromimetic architecture," Signal Process., vol.24, no.1, pp.1-10,1991.

[4] Hyvarinen A., "A family of fixed-point algorithms for Independent Component Analysis," *ICASSP*, pp 3917-3920, 1997.

[5] Blaschke and Laurenz Wiskott, "CuBICA: Independent component analysis by simultaneous third and fourth order cumulant diagonalization," *IEEE Trans. on Signal Processing,* vol.52, no.3, pp.1250-1256,2004.

[6] Hyvarinen A. and Erkki Oja, "Independent Component Analysis: Algorithms and Applications", *http://www.cis.hut.fi/projects/ica/*

[7] Pierre Comon, "Independent Component Analysis, A new concept?," *Signal Processing*, 36, pp.287-314,1994.

[8] Lawrence Rabiner & Biing-Hwang Juang, *Fundamentals of Speech Recognition*. Pearson Education, 2003.

[9] The software generated for this purpose may be referred by sending a mail at drneetaa@gmail.com.

[10] Kishore S. Trivedi, 'Probability and Statistics with Reliability, Queing & Computer Science Applications', *PHI*, 1999.

Prof. D.S. Chauhan was born on 18th Sept. 1949 . He obtained B.Sc. Engineering in year 1972, from BHU, M.E. in year 1978 from Madras University, (studied at REC Tiruchirappalli) and Ph.D. in year 1986 from IIT Delhi under QIP program. He did his post–doctoral work at GSFC/NASA Maryland during 1988-91 under Dr. John M. Vranish an NRC fellowship programmer. He developed "Skin sensor for collision avoidance while robots are working together on a mission Craft". He became Professor in year 1998 on open position and Director, KNIT Sultanpur, from June, 1999–July 2000. He published 47 papers to his credit and guided 6 Ph.D. awarded in year 2000-2004 at BHU and UP Technical University. 15 M. Tech. dissertation and 15 M. Tech. projects. Presently two additional Ph.D. are completing in next six months.

Dr. Chauhan became the founder Vice Chancellor, of U. P. Technical University in year 2000 and continuing on extension.

Saini J. P. – was born in Jhansi, India on June 26, 1966. He received the B. Tech. degree in Electronics Engineering from Dr. RML Avadh University Faizabad, India in 1987, M. Tech. degree from I.I.T. Kanpur, India in 1995 and Ph. D. degree from Dr. RML Avadh University Faizabad, India in 2001. From 1987 to 1999 he was Lecturer at the Department of Electronics Engineering, KNIT Sultanpur, India. He is currently Reader at Bundelkhand Institute of Engineering & Technology, Jhansi, India. He has also served with U. P. Technical University, Lucknow, India as additional Examination Controller. He is member of IETE India and ISTE India. His main research interests include digital communication, signal and speech processing, channel equalization, and neural networks,

Neeta Awasthy was born in Kanpur in 1965. She did her B.Tech. from HBTI, Kanpur in 1988. She worked for major electronics and computer projects . She came back to academics in 1999. She did her masters papers in digital communication and registered in Ph.D. in 2001. Presently working as faculty of Electronics for UPTU. She is member of IEEE, ISTE, CSI, and IETE. Her areas of interest include Digital Signal Processing, Artificial Neural Networks, and Biomedical Electronics & Instrumentation.