Statistical Extraction of Speech Features Using Independent Component Analysis and Its Application to Speaker Identification

Gil-Jin Jang*, Yung-Hwan Oh*

*Department of Computer Science, Korea Advanced Institute of Science and Technology

(Received 18 January 2002; revised 30 June 2002; accepted 26 August 2002)

Abstract

We apply independent component analysis (ICA) for extracting an optimal basis to the problem of finding efficient features for representing speech signals of a given speaker. The speech segments are assumed to be generated by a linear combination of the basis functions, thus the distribution of speech segments of a speaker is modeled by adapting the basis functions so that each source component is statistically independent. The learned basis functions are oriented and localized in both space and frequency, bearing a resemblance to Gabor wavelets. These features are speaker dependent characteristics and to assess their efficiency we performed speaker identification experiments and compared our results with the conventional Fourier-basis. Our results show that the proposed method is more efficient than the conventional Fourier-based features, in that they can obtain a higher speaker identification rate.

Keywords: *Feature extraction, Independent component analysis, Generalized gaussian mixture model, Speech coding, speaker identification*

I. Introduction

Speaker identification is a process of selecting the best-matched speaker among the enrolled speakers, with the personal identity information extracted from speech signals. Currently, one of the main focuses in speaker identification research is based on finding efficient features for speech signals. The standard Fourier basis approach has taken the leading role by decomposing the speech signals into a superposition of a finite number of sinusoids. Thereafter the spectral analysis techniques such as linear predictive coding (LPC), filter bank analysis, cepstral analysis and their variants have been shown fairly good performances[1]. Different efforts are also made to find the new features adopting statistical approaches. Using independent component analysis (ICA)[2,3], a technique of linearly transforming cepstral vectors for the performance improvement of a speaker identification system was proposed[4]. In[5], linear discriminant analysis (LDA) was used to enhance the discriminability of the cepstral features with respect to the given data. However, because these approaches are still involved in the Fourier-based representation, they are not necessarily able to express the statistical structure of the speech signals but assume that all the signals are infinitely stationary and that the significances of all the Fourier basis functions are equal.

In the other aspect, ICA has suggested statistical ways of constructing a basis for a given set of speech signals

Corresponding author: Gil-Jin Jang (jangbal@speech.kaist.ac.kr)
Department of Computer Science, Korea Advanced Institute of Science and Technology, Gusong-dong, Usong-gu, Daejon 305-701, South Korea

[6]. This paper investigates a statistical speech analysis method based on ICA for modeling the differences in the statistical features among the speakers. The ICA filters maximize the amount of information in the transformed domain, so that the adapted individual basis functions obtained by ICA can model the distribution of the individual speaker. In estimating the probability density functions for the sources, previous works[6-8] adopted fixed priors such as Laplacian or strong super-Gaussian which have a more sharpened point at the peak than the Gaussian distribution. However, since we do not want to impose a certain density on the sources we employ the generalized form of Gaussian functions[9], also called the generalized exponential power distributions, which can model the wide range of distributions. We compare the ICA-based features with the Fourier-based and PCA features by performing speaker identification experiments on 20 speakers chosen from the TIMIT database. In our approach, the source coefficients for each basis function are modeled by the generalized Gaussian density, then the speaker is classified by the one with the highest likelihood given all the basis functions for each class. Given the results, we demonstrate that the proposed features are effective in describing the statistical structures of speakers.

II. Learning Basis Functions

This section presents the generalized Gaussian ICA learning algorithm for adapting statistically optimal basis features of the speech sounds. The learned basis outsteps the sinusoidal decomposition that the Fourier basis inherently subsumes. The analysis procedure follows the basics of the Fourier transform: it is assumed that the speech signals are constructed by the linear combinations of the basis functions in the time domain. Instead of the sinusoidal basis functions ICA infers the time-domain basis functions generating the most probable coefficients given the learning data.

2.1. A Model for Speech Signal Representation

The ICA algorithm was initially proposed to solve the blind source separation problem, i.e. given only mixtures of a set of underlying sources, the task is to separate the mixed signals and retrieve their original sources[3,10] without knowing how the signals were mixed nor the distribution of the sources. In contrast to correlation-based transformations such as principal component analysis (PCA), ICA not only decorrelates the signals (using 2 nd -order statistics) but also reduces higher-order statistical dependencies, thus making the output signals as statistically independent as possible. Furthermore, the ICA bases are not restricted to be orthogonal.

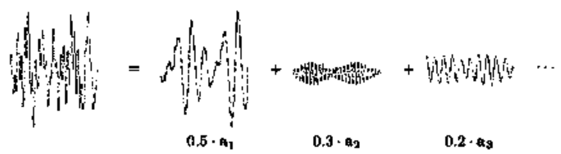

Figure 1. Decomposition of a speech signal into scaled basis functions. The speech signal (leftmost) is reconstructed by the complete basis functions with the activations (0.5, 0.3, 0.2, ...).

ICA assumes an unknown source vector $\mathbf{s}$ with components s_i that are statistically independent to each other. The sources are not observed directly but linear combinations of the sources such that

$$\mathbf{x} = \mathbf{A}\,\mathbf{s} = \sum_{i=1}^{M} \mathbf{a}_i s_i \qquad (1)$$

where $\mathbf{A}$ is a $N \times M$ scalar matrix. The columns of $\mathbf{A}$, $\mathbf{a}_i$, are called the basis functions generating the observed segments of the speech signal in the real world whereas $\mathbf{W} = \mathbf{A}^{-1}$ refers to the ICA filters that transform the segments into activations or source coefficients $\mathbf{s} = \mathbf{W}\,\mathbf{x}$. $\mathbf{A}$ has to be square and full rank in order to be a complete basis. The goal of ICA is to find the basis functions by adaptation or learning given only the observed data $\mathbf{x}$. During the adaptation process a cost function such as the mutual information function is minimized. Once the minimum is achieved, the sources s_i will be as statistically independent as possible.

Figure 1 illustrates the linear decomposition of speech signals as in equation 1. The speech segment can be represented as a linear superposition of the basis functions

{ $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, ...}, scaled by the corresponding scalar activations {0.5,0.3,0.2,...}. The speech segment is represented mostly by 3 basis functions, while other coefficients are almost zero. For Fourier basis, each $\mathbf{a}_i$ is a complex sinusoid with its own frequency and unit magnitude, resulting in mutual exclusion (orthonormality) with the other sinusoids. ICA basis is different in that the basis functions are real and not necessarily orthogonal, and the sources are statistically independent.

Our primary interest is in learning efficient codes, i.e. the activations $\mathbf{s}$ should be as sparse as possible. Sparseness in this case assumes that the input data are encoded in the sources in such a way that the distributions of $\mathbf{s}$ are peakier (having a more sharpened point at the peak) than those of $\mathbf{x}$: assuming the variances of $\mathbf{s}$ and $\mathbf{x}$ to be equal, for some short interval centered at the mean (e.g. $[\mu-\sigma_s, \mu+\sigma_s]$), the percentage of being inside increases more for $\mathbf{s}$ than $\mathbf{x}$ as the interval shrinks, although $p(\mathbf{s})$ should have a longer but thinner tail to retain the equal variances. For the very small-sized interval, the encoding errors occur mostly outside, so we need to encode and decode only small percentage of those informative tails for peaky (sparse) coefficients $\mathbf{s}$, and therefore the adapted basis functions generate efficient codes.

What matters is how well matched the model distribution is to the true underlying distribution. We are interested in inferring the distribution that results in maximally independent coefficients for the sources. We therefore use the generalized Gaussian distribution to model the underlying distribution of the source coefficients. The generalized Gaussian prior[9], also known as exponential power distribution, whose simplest form is $p(s) \propto \exp(-|s|^q)$, can describe Gaussian, platykurtic, and leptokurtic distributions by varying the exponent q. The optimal value of q for given data can be determined from the maximum a posteriori value and provides a good fit for the distributions. In the following sections we present an ICA learning algorithm using generalized Gaussian as a flexible prior that estimates the distributions of the sources.

2.2. Statistical Learning of Speech Basis

The goal of ICA is to adapt the basis functions by estimating $\mathbf{s}$ so that the individual components s_i are statistically independent, and this adaptation process minimizes the mutual information between the components s_i. In our experiments, we used the infomax learning rule [3] with natural gradient extension and updated the basis functions by the following learning rule:

$$\Delta \mathbf{A} \propto \mathbf{A}[\mathbf{I} - \varphi(\mathbf{s})\, \mathbf{s}^T]. \qquad (2)$$

where $\mathbf{I}$ is the identity matrix, $\varphi(s) = -\frac{\partial p(s)/\partial s}{p(s)}$ and s^T denotes the matrix transpose of $\mathbf{s}$. We assume that $\mathbf{A}$ is square (the number of sources are equal to the number of sensors). $\Delta\mathbf{A}$ is the change of the basis functions that is added to $\mathbf{A}$ and will converge to zero once the adaptation process is complete. Calculating $\varphi(\mathbf{s})$ requires a multivariate density model for $p(\mathbf{s})$. Because the components are supposedly independent, the likelihood of the source vector factorizes into component densities:

$$p(\mathbf{s}) = \prod_i^N p_i(s_i). \qquad (3)$$

The parametric density estimate $p_i(s_i)$ plays an essential role in the success of the learning rule in 2. Local convergence is assured if $p_i(s_i)$ is an estimate of the true source density[11]. Note that the global shape of $p_i(s_i)$ was fixed in previous works[6-8].

2.3. The Generalized Gaussian Distributions

The success of the ICA learning algorithm for our purpose depends highly on how closely the ICA density model describes the true source density. The better the source density estimation, the better the basis functions or speech features in turn are responsible for encoding the speech signals. The generalized Gaussian distribution models density functions that are peaky and symmetric at the mean, with a varying degree of normality in the following general form:

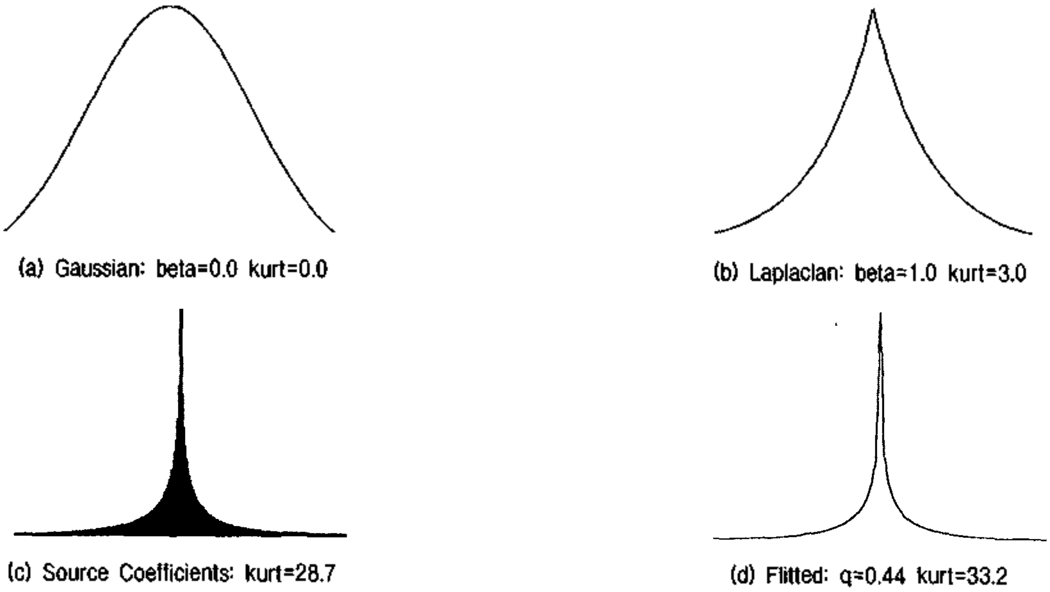

Figure 2. The distribution of source coefficients for actual speech compared to the Gaussian and Laplacian distributions. (a) Gaussian distribution of kurtosis 0. (b) Laplacian distribution of kurtosis 3, peakier than Gaussian. (c) The distribution of the source coefficients, extracted from 200,000 samples of speech data from the TIMIT database. (d) The maximum-likelihood estimated exponential power distribution from the distribution of (c). The x-axis is the value of the source coefficients.

$$p(x|\mu,\sigma,q)=\frac{\omega(q)}{\sigma}\exp\left[-c(q)\left|\frac{x-\mu}{\sigma}\right|^{q}\right], \quad (4)$$

where $\mu=E[x]$, $\sigma=\sqrt{E[(x-\mu)^2]}$, $c(q)=\left[\frac{\Gamma[3/q]}{\Gamma[1/q]}\right]^{q/2}$, and $\omega(q)=\frac{\Gamma[3/q]^{1/2}}{(2/q)\,\Gamma[1/q]^{3/2}}$. The exponent q controls the distribution's deviation from normality. The Gaussian, Laplacian, and strong Laplacian ---speech signals--- distributions are modeled by putting $q=2$, $q=1$, and $q<1$ respectively. By substituting $p_i(s_i)$ in equation 3 with equation 4, we obtain the following score function:

$$\varphi_i(s_i)=-\eta_i|s_i|^{q_i-1}q_i c(q_i)\sigma_i^{-q_i}, \quad (5)$$

where $\eta=\mathrm{sign}(s_i)$, c and σ_i are as defined above. Gradient ascent learning is applied in order to estimate the parameters that maximize the log likelihood. The detailed derivation of learning algorithm is found in[9].

Figure 2 compares the distribution of the source coefficients for the actual speech segments (c), with several kinds of the exponential power distributions; (a), (b), and (d). In (c), it is sharply peaked around zero and has heavy and long tails; there is only a small percentage of informative values (non-zero coefficients) in the tails and most of the data values are around zero, i.e. the data is sparsely distributed. From a coding perspective this implies that we can encode and decode the data with only a small percentage of the coefficients. The peakiness of a distribution can be calculated by the standard kurtosis measure, $K(s)=E[(s-\bar{s})^4/\sigma_s^4]-3$, which is proportional to the peakiness. The calculated kurtosis of the coefficients for speech signals is 28.7, and ideally those of Gaussian (a) and Laplacian (b) are 0 and 3. Hence, neither Gaussian nor Laplacian prior distribution can describe the source distribution correctly. From the actual speech data, we obtained the maximum a posteriori value of the exponent using the algorithm formerly presented[9]. The distribution is plotted in (d), with the exponent 0.44, and kurtosis increased to 33.2.

III. Comparing Learned Bases

The speaker-specific bases were learned with separate data for each speaker, and used in obtaining the features of the speech signals, while the same bases were used regardless of the speakers at the Fourier bases. We used

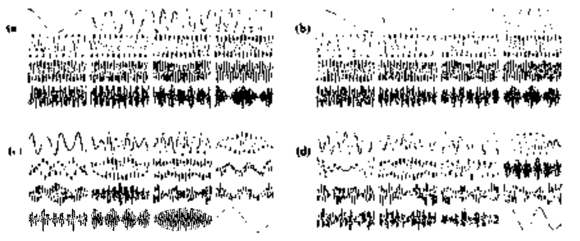

Figure 3. Comparison of DFT and PCA basis functions. (a) 16 examples of DFT basis functions out of 64, (b) DCT, (c) PCA basis functions of male speakers 'mjeb1', and (d) PCA basis functions of female speakers 'fkfb0' from the TIMIT speech database. The waveforms of both DFT and DCT basis functions are inherently pure cosines. Each DFT basis function has 2π cyclicity (first and last samples in the segment have the same values) but each DCT basis function has π. The PCA basis functions have no such restricted waveforms.

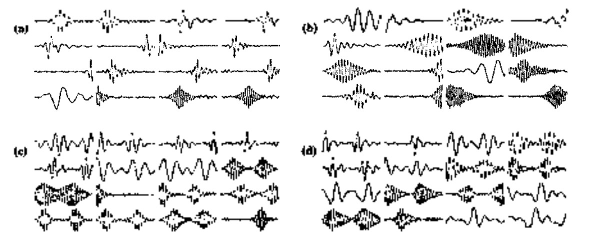

Figure 4. Example plots of learned ICA basis functions. (a), (b): male speakers 'mjeb1' and 'mdpk0'; (c), (d): female speakers 'fkfb0' and 'fecd0' from the TIMIT database. Each basis function is up-sampled by 3 to remove artifacts from sample aliasing. Only 16 basis functions out of 64 are shown. They are obtained by the generalized Gaussian ICA learning algorithm from the 64-sample speech segments.

the TIMIT database to learn the basis of each speaker, and the features ---coefficients--- were compared in terms of the sparseness and the speaker identification rate.

3.1. Data Description

From the TIMIT database, 20 speakers are chosen in order not to be biased to any special case. The speakers in the TIMIT database are classified by 8 dialects of the US English, and therefore we selected speakers to have the same (but not exactly) number of dialects. Regarding the text of the sentences, we selected the sentences from the SX (phonetically-compact) and the SA (dialect) set. For each speaker the SX set consists of 5 sentences randomly chosen out of 450 different sentences that were designed to provide a good coverage of pairs of phones frequently used in English. The SA set consists of 2 utterances of 2 designated sentences that were same for all the speakers. 4 of the SX and SA utterances are assigned for training, the rest 3 for testing with no intersection. The unit of speaker identification testing was 1 utterance. The average length of an utterance is about 3 seconds, so the amount of training data is 12 seconds (3×4 utterances) on the average. We down-sampled the original 16 kHz-sampled data to 8 kHz and applied pre-emphasis with the filter $1-0.95z^{-1}$ to complement the energy decrease in the high frequencies of human speech. Those processes reduce the redundancy and prevent a low-frequency component from dominating the gradient. Because every utterance is labeled with a speaker ID, only the target speaker's data are used for learning of the speaker-specific bases in a supervised manner. For the given speech signals we employed every window of length 64 samples (8 ms) starting at every other sample as learning data $\mathbf{x}$. The number of datapoints for each speaker's basis was roughly 96,000 segments because the length of the training data was 12 seconds (96,000 samples).

3.2. Characteristics of Basis Functions

In equation 1 we assumed the basis functions has the real values only, so we adopted the real parts of the short-time (windows of 64 samples) discrete Fourier transform (DFT) basis functions as a conventional Fourier basis. The discrete cosine transform (DCT) which gives only real coefficients is closely related to DFT and adopted for comparison:

$$\text{DFT:}\quad s(k)=\sum_{n=1}^{N}x(n)\cos\frac{2\pi(k-1)}{N}(n-1), \tag{6}$$

$$\text{DCT:}\quad s(k)=\sum_{n=1}^{N}x(n)\cos\frac{\pi(k-1)}{2N}(2n-1),\quad k=1,\ldots,N \tag{7}$$

The DFT traditionally assumes that the data $x(n)$ is periodically continued with a period of N. Due to this cyclicity assumption underlying Fourier transform, samples along two opposite borders of the segment are expected to have the similar values. Figure 3 compares the waveforms of the DFT and DCT bases with the PCA basis. They look similar in that the basis functions are spread all over the time axis. However PCA basis is data driven and exhibits less regularity. The DFT and DCT basis functions are completely stationary (figure 3 (a)-(b), composed of 1 sinusoid of unique frequency for each) as defined in euqation 7.

In contrast to these conventional bases, subsets of learned basis functions are presented in figure 4. The adaptation of the generalized Gaussian ICA learning started from 64×64 square PCA basis functions, and the gradients of basis functions were computed on a block of 500 waveform segments. The parameter q_i for each $p(s_i)$ was updated every 10 gradient steps. Figures 4 (a)-(b) are examples of the male basis functions, and (c)-(d) represent the female ones. Unlike the PCA and Fourier bases, they are represented by the superposition of sinusoids of different magnitude and some of them reside only in confined ranges in the time domain. Male basis functions have one Gaussian-modulated sinusoid generally, and female ones have a few, generally two, or cover all time axis like the Fourier basis. In frequency domain, each basis function covers the different portions from high to low frequencies. Their structures are similar to Gabor-like filters, especially for male basis functions. More analysis on the difference of the male and female basis functions can be found in our early work[12].

3.3. Comparing Basis Coefficients

We measured statistical parameters and the schematic diagrams to quantify the effectiveness of the basis functions in comparison with conventional methods. The parameters of the generalized Gaussian mixture model are estimated for the real DFT, DCT, PCA, and ICA coefficients of the same training data. Figure 5 compares the log-scaled histograms of the coefficients. The distribution of the ICA coefficients is slightly peakier and has a longer tail, and this spreading of the tail yields greater sparseness. This is reflected in the kurtosis measure of 37.3 for the ICA, compared to 25.2, 26.7 and 24.9 for the PCA, DCT and DFT bases. Figure 6 illustrates the statistical dependency of the coefficients schematically in two-dimensional plots. Fig. 6-(a) shows that the nearby DFT coefficients are highly correlated, and fig. 6-(b) shows that this correlation decreases as the adjacency reduces. In fig. 6-(c) and (d), both PCA and ICA bases succeed in decorrelation, moreover ICA basis is sparser

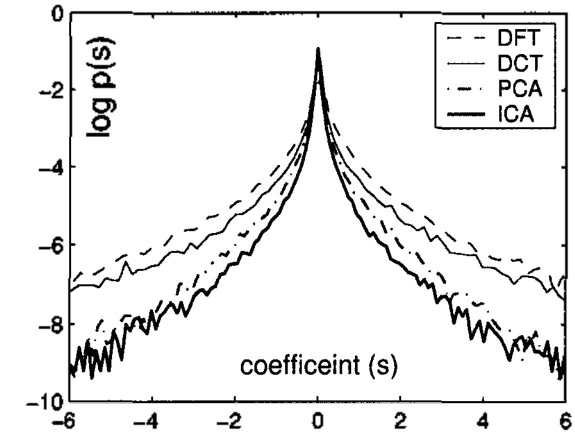

Figure 5. The univariate statistics of the coefficients of the DFT, DCT, PCA, ICA bases, which are depicted by the histograms with the dashed, thin solid, dash-dotted, and thick solid line. The data are from the male speaker 'mgrl0' in TIMIT. Note that the y-axis is a log scale.

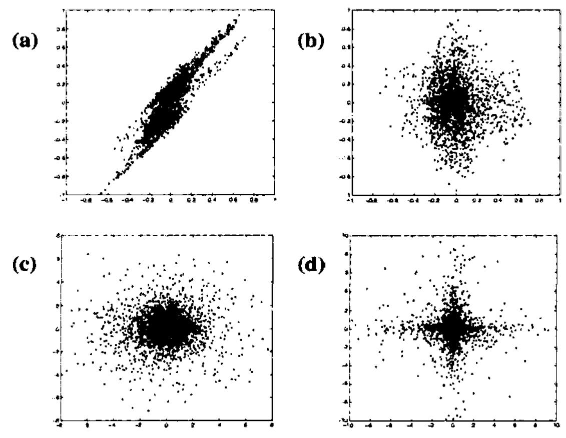

Figure 6. Schematic view of correlation: 2-dimensional plots of the selected two coefficients for each basis. The data are from the male speaker 'mgrl0' in the TIMIT. (a), (b) DFT [2 versus 3] and [2 versus 20]; (c) PCA[1, 2]; (d) ICA [1,2]. The DFT basis has high correlation between adjacent coefficients as shown in (a), and ICA basis best satisfies the sparseness criterion that the distribution is peaky and has long tail.

than the other three bases.

3.4. Supervised Classification of Speakers

We performed speaker identification experiments using the generalized Gaussian mixture model[9], which is an extension of the Gaussian mixture model with ICA and the generalized Gaussian density model, on estimating the density functions of the coefficients of each basis. The mixture model assumes that the observed data can be categorized into mutually exclusive classes whose components are generated by a linear combination of independent sources. The source densities are modeled by generalized Gaussians. The generalized Gaussian mixture model using ICA infers for each class the source parameters and the basis functions.

To compute the probability density of a speech signal, $x(n)$, we extract the coefficients for all the speaker basis. We first construct the speech segments $\mathbf{x}_t$'s sampled every shift length P:

$$\mathbf{x}_t = [x(n)\ x(n+1)\ \ldots\ x(n+N-1)]^T \quad (8)$$

where $n = P(t-1)+1$ and N is the length of each basis function. The shift P is half of N in our experiment, $P = \lfloor N/2 \rfloor$. The generalized Gaussian and the basis function matrix obtained for each speaker, obtained from the speaker-specific learning data, comprise each component of the mixture model. The number of components is therefore K, the number of speakers. The priori probabilities of speakers are assumed to be equal, i.e. for all k, $p(C_k) = 1/K$, because the models are trained separately in a supervised manner. Identification is done by calculating a posteriori probability by the maximum likelihood Baysian rule:

$$\begin{aligned} k^* &= \arg\max_k \prod_t p(C_k \mid \mathbf{x}_t, \theta_k) \\ &= \arg\max_k \prod_t \frac{p(\mathbf{x}_t \mid C_k, \theta_k) p(C_k)}{p(\mathbf{x}_t)} \\ &= \arg\max_k \prod_t p(\mathbf{x}_t \mid C_k, \theta_k), \end{aligned} \quad (9)$$

where $\theta_k = (\mathbf{A}_k, \mu_k, \mathbf{q}_k)$ is the set of parameters describing the generalized Gaussian p.d.f., which are transformation matrix, mean vector, and exponents of the source distributions obtained by the ICA learning algorithm. In the case of ICA and PCA, the multivariate p.d.f. $p(\mathbf{x}_t|\theta_k)$ is calculated by the product of the marginalized distribution of the source coefficient $s_{t,i}$,

$$p(\mathbf{x}_t|\theta_k) = \frac{\prod_{i=1}^{M} p(s_{t,i}|q_i)}{|\det \mathbf{A}|} \quad (10)$$

The univariate density $p(s_{t,i}|q_i)$ is calculated by a generalized Gaussian density function. For Fourier basis, since the transformation matrix ---Fourier transform--- is the same for all the speakers, the denominator in equation 10 is omitted. The results shown in table 1 suggest that the ICA features are effective in discriminating speakers. The identification rate of the DCT was less than the DFT, because it decomposes the speech signals in a stricter manner ($2N$ periodicity), which results overfitting speaker distributions.

Table 6. Identification Rates and the mean value of Kurtosis for each basis.

	DFT	DCT	PCA	ICA
ID rate	82.2%	79.8%	84.1%	87.5%
Kurtosis	24.9	26.7	25.2	37.3

IV. Discussion

This paper presented the use of ICA as a speech analysis technique of extracting the statistically optimal basis features of the speech sounds. This approach has already been introduced in many early works. Bell and Sejnowski (1996) applied this method to extracting basis features from natural sounds and speech sounds, and Lewicki and Sejnowski (1999) discussed the efficiency of coding speech sounds using the learned ICA basis features. Lee et al. (2000) proposed a technique of applying the basis features to speech recognition. In this work the likelihood of a given speech segment is calculated by a hidden Markov model whose observation probabilities are modeled by Gaussian mixture density functions of the log-scaled magnitude of the inferred coefficients.

Our work extended the previous ones to the problem of the speaker identification. In the learning steps of the basis functions, we applied the generalized Gaussian density modeling for the distributions of source coefficients, and applied the corresponding ICA learning algorithm presented by Lewicki (2000), based on the observation that the distribution of the source coefficients of speech signals are highly super-Gaussian as illustrated in figure 2. Using the generalized Gaussian mixture models[9] in recognizing the unknown input speech signals, the proposed method takes advantage of the accurate modeling of the distributions of the underlying source coefficients. The basis itself describes the statistical structures of the speech signals of a given speaker without transforming them into another domain such as cepstral space.

By the experimental results we have shown that learned bases have better speaker identification performance than the commonly used spectral representations such as the DFT (discrete Fourier transform) and the DCT (discrete cosine transform). Our data set consists of spontaneously uttered sentences that give each speaker's speech characteristics. In these examples, the learned basis resulted in increased identification rates of 5.3% and 7.7% compared to the DFT and the DCT.

V. Conclusion

We applied ICA to speech signals from individual speakers to extract a set of optimal basis functions. The basis functions were adapted using the generalized Gaussian ICA model resulting in basis functions and source coefficient statistics that were characteristical features for the individual speaker. Most basis functions were localized in time and frequency resembling Gabor-like wavelet filters. The corresponding source coefficients were extremely sparse resulting in efficient codes. The generalized Gaussian ICA model is embedded into a mixture model allowing classification of the individual speakers based on the basis functions models for each speaker class. Our initial identification rates suggest superior performance compared to the Fourier or PCA based method. This can now serve as a baseline to further investigate and optimize the identification procedure. Currently, we are optimizing the system parameters to compare our results to state of the art speaker identification systems.

References

1. R. J. Mammone, X. Zhang, and R. P. Ramachandran, "Robust speaker recognition: a feature-based approach," *IEEE signal processing magazaine*, 58-71, 9, 1996.
2. P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, **36**, 287-314, 1994.
3. A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, **7** (6), 1004-1034, 1995.
4. G.-J. Jang, S.-J. Yun, and Yung-Hwan, "Feature vector transformation using independent component analysis and its application to speaker identification," *In Proceedings of Eurospeech, (Budapest Hungary)*, 767-760, Sept 1999.
5. H. Hermansky, S. Sharma, and P. Jain, "Data-drived non-linear mapping for feature extraction in HMM," *In Proceeding of the Workshop on Automatic Speech Recognition and Understanding, (Keystone, CO., USA)*, December 1999.
6. J.-H. Lee, H.-Y. Jung, T.-W. Lee, and S.-Y. Lee, "Speech feature extraction using independent component analysis," *In Proc. ICASSP*, **3**, (Istanbul, Turkey), 1631-1634, Jun 2000.
7. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive-field properties by learning a sparse code for natural images," *Nature*, **381**, 607-609, 1996.
8. A. Hyvaerinen, "Sparse code shrinkage: denoising of non-gaussian data by maximum likelihood estimation," *Neural Computation*, 11 (7), 1739-1768, 1999.
9. T.-W. Lee and M. S. Lewicki, "The generalized Gaussian mixture model using ICA," *In International Workshop on Independent Component Analysis (ICA'00), (Helsinki)*, 239-244, Jun 2000.
10. C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, **24**, 1-10, 1991.
11. D. T. Pham and P. Garrat, "Blind source separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. on Signal Proc.*, **45** (7), 1712-1725, 1997.
12. T.-W. Lee and G.-J. Jang, "The statistical structures of male and female speech signals," *In Proc. ICASSP, (Salt Lake City, Utah)*, May 2001.

[Profile]

• Gil-Jin Jang

The Journal of the Acoustical Society of Korea, Vol. 21, No. 4E.

• Yung-Hwan Oh

The Journal of the Acoustical Society of Korea, Vol. 21, No. 4E.