International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

Advancement Is Independent Component Analysis Speech Enhancement Process

Kasagani Satyanarayana[1]& Mr.M.K.Linga Murthy[2]

[1.]M.Tech Student of Lakireddy Bali Reddy College of Engineering(Autonomous),Maylavaram,AP-India,

[2]Assistant Professor, ECE Dept., Lakireddy Bali Reddy College of Engineering(Autonomous),Maylavaram,AP-India

Email: [1]satya786sriram@gmail.com, [2]lingamurthy413@gmail.com.

ABSTRACT

Speech enhancement is a process of increasing the quality, intelligibility of the speech signal which is one dimensional. So, by removing the external world noise or artefacts this can be improve and reject the background interference in the form of additive background noise which helps in the improvement of performance of communication system. The enhancement of speech signal consists of 3 major objectives: A) Sound is to be clear to the listener. B) Improve robustness. C) Lossless perception is required which improves the accuracy. For loss enhancement subspace methods were used such as Spectral Subtraction method, Time domain method, Transform method. The signal enhancement process is going to be perform by using sub band decomposition method under low and high pass filter collaboration i.e., WAVELET (DWT) transformation. The existing technique KLT sub space is a loss transformation and performed enhancing signal but theoretically it can be improved by Wavelet transformation and will perform a lossless transformation technique.

Key words: Speech Enhancement; Intelligibility; Noise; ICA ; KLT – Karhuenen Loeve Transform; DWT–Discrete Wavelet Transformation

I. INTRODUCTION

The performance of speech communication systems in applications such as hands-free telephony, degrade considerably in adverse acoustic environments. The presence of noise can cause loss of intelligibility as well as the listener's discomfort and fatigue. Speech enhancement methods seek to improve the performance of these systems and to make the corrupted speech more pleasant to the listener. These methods are also useful in other applications such as automatic speech recognition. In this paper we focus on the signal subspace approach (SSA) for speech enhancement [1]. This technique is based on the decomposition of the noisy signal vector space into two orthogonal subspaces called the noise subspace and the signal subspace. In this context, the signal subspace decomposition can be achieved either using the Karhunen-Loeve transform (KLT) via Eigen value decomposition (EVD) of the data covariance matrix [1]–[4], or using the singular value decomposition (SVD) of a data matrix [5]–[7]. The discrete cosine transform (DCT) has also been proposed as an approximation to the KLT [8], [9]. In the SSA, enhancement is obtained by removing the noise subspace as a first step. Then the clean speech is recovered in the remaining signal subspace by optimally weighting the signal coefficients in this subspace. The different SSA methods vary according to the weighting scheme used [6]. The SSA can also be interpreted as a filter bank with the weighting coefficients serving as the sub band filters [10].

As in most single channel speech enhancement methods such as spectral subtraction [11], the signal subspace methods suffer from the annoying residual noise known as *musical noise*. Tones at random frequencies, resulting from poor estimation of the signal and noise statistics, are at the origin of this artefact. In spectral subtraction and its variants, modifications using a human hearing model were proposed to reduce the prominence of the musical noise [12]–[16]. This technique, which was first introduced in audio coding [17], is based on the fact that the human auditory system is able to tolerate additive noise as long as it is below some *masking threshold*. Methods to calculate the masking threshold are developed in the frequency domain according to critical band analysis and the excitation pattern of the basilar membrane in the inner ear [18].

Recently, a DCT based SSA imitating the human hearing resolution was proposed [9]. However, no algorithm which employs a sophisticated hearing model with a KLT based SSA is available. The reason is that the SSA do not operate in the frequency domain where the available hearing models are developed. In this paper, we present a frequency to eigendomain transformation (FET) which provides a way to calculate a perceptually based eigenfilter. This is done by estimating an eigenvalue decomposition based power spectral density (PSD) from
which a masking threshold is calculated. This threshold is transformed to the speech signal eigendomain using the FET allowing to design the perceptual eigenfilter. This filter yields better residual noise shaping from a psychoacoustic perspective. We provide an analysis of the FET and show how it can be incorporated in the SSA to improve its performance.We also show how the method can be modified to cover the more general case of colored noise. Informal as well as formal subjective listening test results show that the proposed new method outperforms the conventional SSA. The results also show that our method provides better noise shaping in the sense that for a given speech signal, the residual noise has relatively similar characteristics in different noisy environments.

Estimation of Covariance

The calculation of the KLT is typically performed by finding the eigenvectors of the covariance matrix, which, of course, requires an estimate of the covariance matrix. If the entire signal is available, as is the case for coding a single image, the covariance matrix can be estimated from n data samples as

$$[\widehat{\mathbf{C}}]_x = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T ,$$

where xi is a sample data vector. If only portions of the signal are available, care must be taken to ensure that the estimate is representative of the entire signal. In the extreme, if only one data vector is used then only one nonzero eigen value exists, and its eigenvector is simply the scaled version of the data vector. For typical images, it is rarely the case that their covariance matrix has any zero eigen values. For a data vector of dimension N, a good rule of thumb is that at least 10 × N representative samples from the various regions within an image be used to ensure a good estimate if it is not feasible to use the entire image.

In contrast to a Fourier series where the coefficients are fixed numbers and the expansion basis consists of sinusoidal functions (that is, sine and cosine functions), the coefficients in the Karhunen–Loève theorem are random variables and the expansion basis depends on the process. In fact, the orthogonal basis functions used in this representation are determined by the covariance function of the process. One can think that the Karhunen–Loève transform adapts to the process in order to produce the best possible basis for its expansion.

In the case of a *centered* stochastic process $\{X_t\}_{t\in[a,b]}$ (*centered* means $\mathbf{E}[X_t] = 0$ for all $t\in[a,b]$) satisfying a technical continuity condition, X_t admits a decomposition

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

where Z_k are pairwise uncorrelated random variables and the functions e_k are continuous real-valued functions on $[a, b]$ that are pairwise orthogonal in $L^2([a, b])$. It is therefore sometimes said that the expansion is *bi-orthogonal* since the random coefficients Z_k are orthogonal in the probability space while the deterministic functions e_k are orthogonal in the time domain. The general case of a process X_t that is not centered can be brought back to the case of a centered process by considering $X_t - \mathbf{E}[X_t]$ which is a centered process.

Moreover, if the process is Gaussian, then the random variables Z_k are Gaussian and stochastically independent. This result generalizes the *Karhunen–Loève transform*. An important example of a centered real stochastic process on [0, 1] is the Wiener process; the Karhunen–Loève theorem can be used to provide a canonical orthogonal representation for it. In this case the expansion consists of sinusoidal functions.

The above expansion into uncorrelated random variables is also known as the *Karhunen–Loève expansion* or *Karhunen–Loève decomposition*. The empirical version (i.e., with the coefficients computed from a sample) is known as the *Karhunen–Loève transform* (KLT), *principal component analysis*, *proper orthogonal decomposition (POD)*, *Empirical orthogonal functions* (a term used in meteorology and geophysics), or the *Hotelling transform*.

- Throughout this article, we will consider a square-integrable zero-mean random process X_t defined over a probability space $(\Omega, F, \mathbf{P})$ and indexed over a closed interval $[a, b]$, with covariance function $K_X(s, t)$. We thus have:

$$\begin{aligned}&\forall t \in [a,b] \qquad X_t \in L^2(\Omega, F, \mathbf{P}),\\ &\forall t \in [a,b] \qquad \mathbf{E}[X_t] = 0,\\ &\forall t, s \in [a,b] \qquad K_X(s,t) = \mathbf{E}[X_s X_t].\end{aligned}$$

- We associate to K_X a linear operator T_{KX} defined in the following way:

$$\begin{cases} T_{K_X} : L^2([a,b]) \to L^2([a,b]) \\ f \mapsto \int_a^b K_X(s,\cdot) f(s) ds \end{cases}$$

Since T_{KX} is a linear operator, it makes sense to talk about its eigenvalues λ_k and eigenfunctions e_k, which are found solving the homogeneous Fredholm integral equation of the second kind

$$\int_a^b K_X(s,t) e_k(s)\, ds = \lambda_k e_k(t)$$

The Karhunen–Loève expansion minimizes the total mean square error

In the introduction, we mentioned that the truncated Karhunen–Loeve expansion was the best approximation of the original process in the sense that it reduces the total mean-square error resulting of its truncation. Because of this property, it is often said that the KL transform optimally compacts the energy.

More specifically, given any orthonormal basis $\{f_k\}$ of $L^2([a, b])$, we may decompose the process X_t as:

$$X_t(\omega) = \sum_{k=1}^{\infty} A_k(\omega) f_k(t)$$

where

$$A_k(\omega) = \int_a^b X_t(\omega) f_k(t)\, dt$$

and we may approximate X_t by the finite sum

International Journal of Research

International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015

Available at http://internationaljournalofresearch.org

$$\hat{X}_t(\omega) = \sum_{k=1}^{N} A_k(\omega) f_k(t)$$

for some integer *N*.

II. Independent Component Analysis

ICA finds the independent components (also called factors, latent variables or sources) by maximizing the statistical independence of the estimated components. We may choose one of many ways to define independence, and this choice governs the form of the ICA algorithm. The two broadest definitions of independence for ICA are

1. Minimization of mutual information
2. Maximization of non-Gaussianity

The Minimization-of-Mutual information (MMI) family of ICA algorithms uses measures like Kullback-Leibler Divergence and maximum entropy. The non-Gaussianity family of ICA algorithms, motivated by the central limit theorem, uses kurtosis and negentropy.

Typical algorithms for ICA use centering (subtract the mean to create a zero mean signal), whitening (usually with the eigenvalue decomposition), and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm. Whitening and dimension reduction can be achieved withprincipal component analysis or singular value decomposition. Whitening ensures that all dimensions are treated equally *a priori* before the algorithm is run. Well-known algorithms for ICA include infomax, FastICA, and JADE, but there are many others.

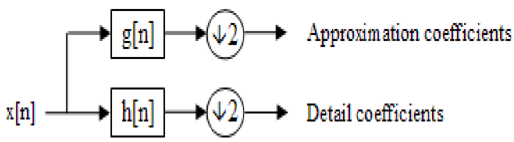
In general, ICA cannot identify the actual number of source signals, a uniquely correct ordering of the source signals, nor the proper scaling (including sign) of the source signals.

ICA is important to blind signal separation and has many practical applications. It is closely related to (or even a special case of) the search for a factorial code of the data, i.e., a new vector-valued representation of each data vector such that it gets uniquely encoded by the resulting code vector (loss-free coding), but the code components are statistically independent.

Application

ICA can be extended to analyze non-physical signals. For instance, ICA has been applied to discover discussion topics on a bag of news list archives.

Some ICA applications are listed below optical Imaging of neurons neuronal spike sorting face



recognition modeling receptive fields of primary visual neurons predicting stock market prices mobile phone communications colour based detection of the ripeness of tomatoes removing artifacts, such as eye blinks, from EEG data.

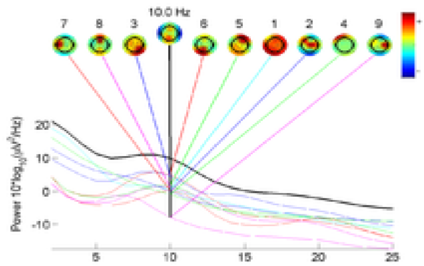


Fig: Independent Component Analysis in EEG LAB

III. DWT

One level of the transform

The DWT of a signal x is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response g resulting in aconvolution of the two:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n-k]$$

The signal is also decomposed simultaneously using a high-pass filter h. The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter.

Fig : Block diagram of filter analysis

However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then subsampledby 2. In the next two formulas, the notation is the opposite: g- denotes high pass and h- low pass as is Mallat's and the common notation:

$$y_{\mathrm{low}}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k]$$

$$y_{\mathrm{high}}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k]$$

This decomposition has halved the time resolution since only half of each filter output characterises the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled.

With the subsampling operator $\downarrow$

$$(y \downarrow k)[n] = y[kn]$$

the above summation can be written more concisely.

$$y_{\mathrm{low}} = (x * g) \downarrow 2$$
$$y_{\mathrm{high}} = (x * h) \downarrow 2$$

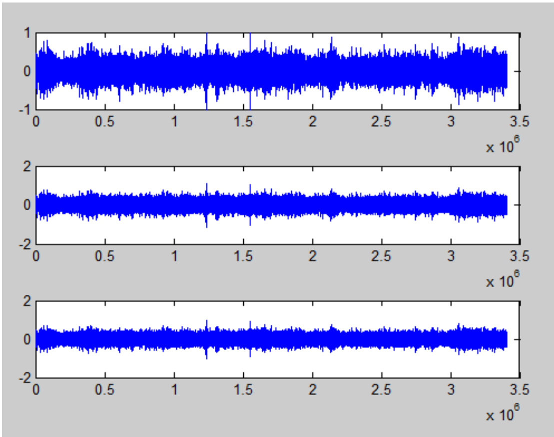
However computing a complete convolution $x * g$ with subsequent downsampling would waste computation time.

The Lifting scheme is an optimization where these two computations are interleaved.

IV. CONCLUSION:

Finally, by using WAVELET (DWT) transformation I have implemented the Speech enhancement effectively. By removing the external world noise or artefacts this can be improve and reject the background interference in the form of additive background noise which helps in the improvement of performance of communication system.

V. RESULT:



VI. REFERENCEs:

[1] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Processing, vol. 3, pp. 251–266, July 1995.

[2] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," IEEE Trans. Speech Audio Processing, vol. 8, pp. 159–167, Mar. 2000.

[3] J. Huang and Y. Zhao, "An energy-constrained signal subspace method for speech enhancement and recognition in colored noise," Speech Commun., vol. 1, pp. 165–181, 1998.

[4] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," IEEE Trans.

Speech Audio Processing, vol. 9, pp. 87–95, Feb. 2001.

[5] M. Dendrinos, S. Bakamidis, and G. Carayannis, “Speech enhancement from noise: A regenerative approach,” Speech Commun., vol. 10, pp. 45–57, Feb. 1991.

[6] P. S. K. Hansen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, “Experimental comparison of signal subspace based noise reduction methods, in Proc. ICASSP’99, vol. 1, 1999, pp. 101–104.

[7] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, “Reduction of broad-band noise in speech by truncated QSVD,” IEEE Trans. Speech Audio Processing, vol. 3, pp. 439–448, Nov. 1995.

[8] J. Huang and Y. Zhao, “A DCT-based fast signal subspace technique for robust speech recognition,” IEEE Trans. Speech Audio Processing, vol. 8, pp. 747–751, Nov. 2000.

[9] R. Vetter, “Single channel speech enhancement using MDL-based subspace approach in bark domain,” in Proc. ICASSP’01, vol. 1, 2001, pp. 641–644.

[10] M. Jeppesen, C. A. Rodbro, and S. H. Jensen, “Recursively updated eigenfilterbank for speech enhancement,” in Proc. ICASSP’01, vol. 1, 2001, pp. 653–656.

[11] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Processing, vol. 27, pp. 113–120, Apr. 1979.

[12] N. Virag, “Single channel speech enhancement based on masking properties of the human auditory system,” IEEE Trans. Speech Audio Processing, vol. 7, pp. 126–137, Mar. 1999.

[13] S. Gustafsson, P. Jax, and P. Vary, “A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics,” in Proc. ICASSP’98, 1998, pp. 397–400.

[14] A. Czyzewski and R. Krolikowski, “Noise reduction in audio signals based on the perceptual coding approach,” Proc. IEEE WASPAA, pp. 147–150, 1999.

[15] A. A. Azirani, R. J. Le Bouquin, and G. Faucon, “Optimizing speech enhancement by exploiting masking properties of the human ear,” in Proc. ICASSP’95, 1995, pp. 800–803.