# Wavelet Transform Domain Blind Equalization and Its Application to Speech Analysis

*Munehiro Namba and Yoshihisa Ishida*

Department of Electronics and Communication
School of Science and Technology, Meiji University
Kawasaki 214-8571 Japan

## ABSTRACT

In this paper, a wavelet transform domain realization of the blind equalization technique termed as EVA is applied to speech analysis. The conventional linear prediction problem can be viewed as a constrained blind equalization problem. Because the EVA does not impose any restriction to the probability distribution in the input (the glottal excitation), the principal features of speech can be effectively separated from a speech in a short duration. The computational complexity will be a problem, but the proposed implementation in a wavelet transform domain promotes the faster convergence in the analysis of speech signal.

## 1. INTRODUCTION

The feature of speech can be divided into two major elements, the glottal excitation and the vocal tract characteristics, as it is well known. The linear prediction, a traditional tool for speech analysis, has many advantages to extract the above features from a short-term segment of a speech [1]. However, the analysis result is erratic by nature; it is sensitive to the noise, the type of analysis window and even the magnitude of spikes corresponding to the pitch of speech [2]. This is chiefly caused by its Gaussian assumption in the excitation signal that assures the unique solution but inflicts the sensitivity to any change of the condition. Essentially, the prediction residual of a voiced speech is definitely not Gaussian because the glottal excitation signal includes not only the noises but also some large spikes corresponding to the pitch.

Recently, a theory called *blind equalization* or *blind deconvolution* has gained a lot of interests along with development of telecommunication networks. After a classical work of Stockham *et al*. [3] and a pioneer work by Benveniste *et al*. [4], several criteria and algorithms have been proposed and investigated in each specific application. We are interested in the fact that the linear prediction can be viewed as a constrained blind equalization @r deconvolution problem because almost works do not impose any restriction to the input sequence. The EigenVector Algorithm for blind equalization (EVA) proposed by Jelonnek *et al*. [5] is an extension of the Shalvi's criterion that is maximizing the magnitude of estimate's kurtosis [6]. The realization in the original work is based on the stochastic gradient method, hence the convergence is slow and tends to be fastened on a local minimum or maximum. However the EVA overcomes both the convergence assurance and uniqueness problem.

It remains a computational problem such that the EVA requests the maximum eigenvalue and its associate vector calculation of a matrix, repeatedly. We note that the matrix becomes approximately lower triangle in an appropriate orthogonal wavelet transform domain. By this transformation, the approximate maximum eigenvalue can be obtained by just picking up the maximum value of the diagonal elements. This utility promotes the faster convergence of well-known power method in the maximum eigenvector calculation.

The experiments on the analysis of Japanese vowels with the proposed wavelet transform domain EVA (WEVA) are presented in this paper. The results show that not only the computational advantage but also the essential superiority of WEVA.

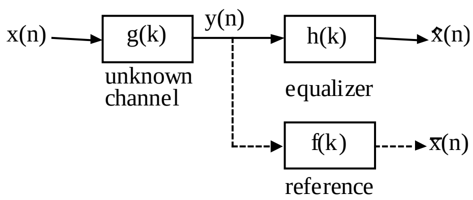



**Figure 1:** Blind Equalization

## 2. BLIND EQUALIZATION OF SPEECH

### 2.1. Recovering Glottal Excitation via Kurtosis Maximization

The blind equalization problem is to recover the input sequence $x(n)$ to an unknown system $g(k)$, from the observed sequence $y(n)$ with the equalizer $h(k)$. The procedure is illustrated in Fig.1. In speech analysis, in other words, the objective is to recover the glottal excitation signal $x(n)$ input to the vocal tract, from the observed speech signal $y(n)$ with the equalizer $h(k)$.

Of course we must have some preknowledge for proper equalization. In linear prediction, the probability distribution of $x(n)$ is supposed to be Gaussian and $g(k)$ be minimum phase. There are essentially inappropriate in voiced speech analysis. In Shalvi's work, $x(n)$ may consist of zero-mean i.i.d. real or complex variables with an arbitrary probability distribution, as long as the moments up to the fourth order exist. The reader should note, however, that $x(n)$ is assumed to be real throughout this paper. There is no restriction to $g(k)$, except that the equalized signal may have a constant delay. Under these assumptions, the criterion can be,

$$\text{maximize } |K[\hat{x}(n)]|$$
$$\text{subject to } E[\hat{x}^2(n)] = E[x^2(n)] \qquad (1)$$

where $K[\cdot]$ is the kurtosis of a sequence defined by

$$K[v] = E[v^4] - 3\{E[v^2]\}^2 \tag{2}$$

## 2.2 Eigenvector Algorithm

In EVA, Shalvi's criterion is expressed with vector forms for deriving the following quadratic formulas.

$$\begin{aligned} &\text{maximize } |\mathbf{h}^{\mathrm{T}}\mathrm{C}_4^{\bar{x},y}\mathbf{h}| \\ &\text{subject to } |\mathbf{h}^{\mathrm{T}}\mathrm{R}_{yy}\mathbf{h}| \end{aligned} \tag{3}$$

where

$$\mathrm{C}_4^{x,y} = E[\bar{x}^2(n)\mathbf{y}_n\mathbf{y}_n{}^{\mathrm{T}}] - 3E[\mathbf{y}_n\bar{x}(n)]E[\bar{x}(n)\mathbf{y}_n{}^{\mathrm{T}}] \tag{4}$$

$$\mathrm{R}_{yy} = E[\mathbf{y}_n\mathbf{y}_n{}^{\mathrm{T}}] \tag{5}$$

$$\bar{x}(n) = f(k) * y(n) = \mathbf{f}^{\mathrm{T}}\mathbf{y}_n \tag{6}$$

$$\hat{x}(n) = h(k) * y(n) = \mathbf{h}^{\mathrm{T}}\mathbf{y}_n \tag{7}$$

The signal $\bar{x}(n)$ is the output of a virtual reference model $f(k)$ that is used in the calculation only.

The EVA solution to the blind equalization problem is the eigenvector of

$$\mathrm{C}_4^{x,y}\mathbf{h} = \lambda\mathrm{R}_{yy}\mathbf{h} \tag{8}$$

corresponding to the eigenvalue with the maximum magnitude.

# 3. ORTHOGONAL DOMAIN EVA

Let the input sequence be finite, and the equalized sequence can be expressed as

$$\hat{\mathbf{x}} = \mathrm{Y}\mathbf{h} \tag{9}$$

where $\hat{\mathbf{x}}$ @is $N \times 1$ vector, $\mathrm{Y}$ is a $N \times M$ matrix of $[\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{N-1}]^{\mathrm{T}}$, and $\mathbf{h}$ is a $M \times 1$ vector. Considering this filtering procedure in an orthogonal domain that is formed by an $M \times M$ unitary matrix $\mathrm{W}$, we have

$$\begin{aligned} \hat{x} &= \mathrm{Y}\mathrm{W}^*\mathrm{W}\mathbf{h} \\ &= (\mathrm{W}\mathrm{Y}^{\mathrm{T}})^*\mathbf{h}_w \end{aligned} \tag{10}$$

where $\mathbf{h}_w = \mathrm{W}\mathbf{h}$ is the orthogonal domain expression of the equalizer. The above Eq.(10) indicates that the observed sequence $y(n)$ must be transformed into the orthgonal domain, then equalized by the filter $h_w(k)$. Therefore, the auto-correlation and the cross-cumulant matrices in Eq.(8) are also transformed into the same orthogonal domain.

$$\begin{aligned} W[C_4^{\bar{x},y}] &= E[\bar{x}^2(n)\mathrm{W}\mathbf{y}_n(\mathrm{W}\mathbf{y}_n)^*] \\ &\quad - 3E[\mathrm{W}\mathbf{y}_n\bar{x}(n)]E[\bar{x}(n)(\mathrm{W}\mathbf{y}_n)^*] \\ &= \mathrm{W}C_4^{\bar{x},y}\mathrm{W}^* \end{aligned} \tag{11}$$

$$\begin{aligned} W[\mathrm{R}_{yy}] &= E[\mathrm{W}\mathbf{y}_n(\mathrm{W}\mathbf{y}_n)^*] \\ &= \mathrm{W}\mathrm{R}_{yy}\mathrm{W}^* \end{aligned} \tag{12}$$

Substituting these matrices in Eqs.(11) and (12) into (8), we have

$$\mathrm{C}_w\mathbf{h}_w = \lambda\mathbf{h}_w \tag{13}$$

where

$$\mathrm{C}_w = \mathrm{W}\mathrm{R}_{yy}^{-1}\mathrm{C}_4^{\bar{x},y}\mathrm{W}^* \tag{14}$$

in which the eigenvector corresponding to the maximum eigenvalue becomes the EVA solution in the orthogonal domain.

In general, a square matrix $\mathrm{C}_w$ is not diagonalized but may be approximately triangular. According to the Schur triangularization theorem, we know that an arbitrary square matrix $\mathrm{A}$ can be transformed to be triangular matrix of $\mathrm{P}^*\mathrm{A}\mathrm{P}$, where $\mathrm{P}$ is an unitary matrix. It tells us that an arbitrary matrix can also be transformed to be triangular matrix of $\mathrm{W}\mathrm{A}\mathrm{W}^*$, where $\mathrm{W}$ is also an unitary matrix. The form of the transformed matrix depends on how the energy concentrates by the unitary matrix $\mathrm{W}$. The square matrix $\mathrm{W}\mathrm{A}\mathrm{W}^*$ may be upper, lower, even left or lower triangular.

The reason for why the orthogonal wavelet basis is preferred rather than the Fourier basis is its vanishing moment property. If a wavelet $w(t)$ has $p$ vanishing moments and the signal $s(t)$ has $p$ derivatives, its wavelet coefficients decay like $2^{-jp}$ [7]:

$$|b_{ij}| = \left|\int s(t)\, w_{jk}(t)\, dt\right| \le c\, 2^{-jp} \left\| s^{(p)}(t) \right\| \tag{15}$$

where $w_{jk}(t) = 2^{j/2} w(2^j t - k)$ and $c$ is a constant value. For explaining simply, consider the two-level orthogonal wavelet transformation such that the matrix $\mathrm{W}$ transforms a square matrix $\mathrm{A}$ into the following four-level resolution domain.

$$\mathrm{W}\mathrm{A}\mathrm{W}^{\mathrm{T}} = \begin{bmatrix} LL & LH \\ HL & HH \end{bmatrix} \tag{16}$$

In this transformation, the matrix $\mathrm{C}_w$ is expressed as follows:

$$\begin{aligned} \mathrm{C}_w &= (\mathrm{W}\mathrm{R}_{yy}\mathrm{W}^{\mathrm{T}})^{-1}\mathrm{W}\mathrm{C}_4^{\bar{x},y}\mathrm{W}^{\mathrm{T}} \\ &= \begin{bmatrix} LL_R & LH_R \\ HL_R & HH_R \end{bmatrix}^{-1} \begin{bmatrix} LL_{C_4} & LH_{C_4} \\ HL_{C_4} & HH_{C_4} \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} HH_R & -LH_R \\ -HL_R & LL_R \end{bmatrix} \begin{bmatrix} LL_{C_4} & LH_{C_4} \\ HL_{C_4} & HH_{C_4} \end{bmatrix} \end{aligned} \tag{17}$$

where $\Delta$ is the determinant of $\mathrm{R}_{yy}$. In both the transformed auto-correlation and cross-cumulant matrices,

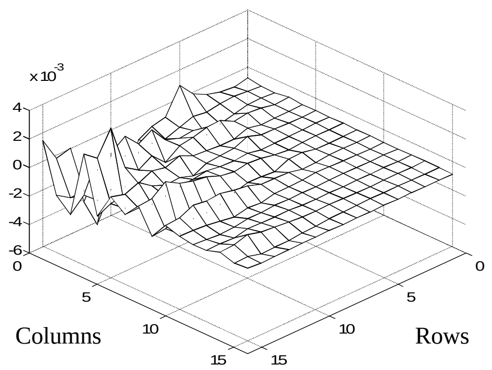
$$0 \approx |HH| << |HL| \le |LH| < |LL| \tag{18}$$

is tolerably satisfied if the matrices have $p$ derivatives and the wavelet with $p$ vanishing moments is applied. This results,

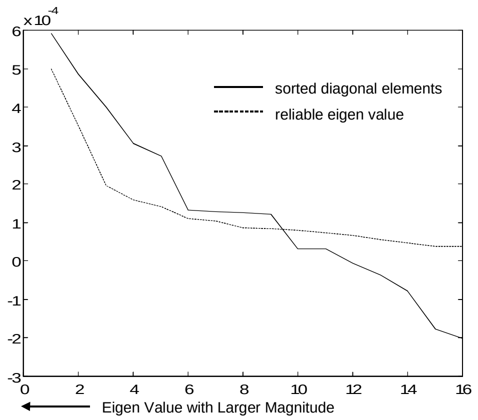
$$\mathrm{C}_w = \begin{bmatrix} HL_{C_w} & HH_{C_w} \\ LL_{C_w} & LH_{C_w} \end{bmatrix} \tag{19}$$

It should be noted that $LL$ or even $LH$ and $HL$ can be transformed into further four-level resolution domain in which the condition in Eq.(18) is also satisfied with tolerable degree, but the further transformation is not always needed.

Consequently, the matrix $C_w$ becomes nearly lower triangle. A typical example of the matrix is shown in Fig.2(a). The comparison of the reliable eigenvalues and the sorted diagonal elements of the matrix is shown in Fig.2(b). The analyzed signal is a partial segment of Japanese vowel /o/ sampled with 16[bits], 11.025[kHz]. The length of the segment, $N$, is 256 and the equalizer order $M$ is 16. The 8th order Daubechies wavelet is applied. The utility of this rough approximation promotes the faster convergence of the well-known power method.




(a) $C_w$ Appearance




(b) Comparison of Eigenvalues

**Figure 2:** Matrix property in wavelet domain.

# 4. CALCULATION PROCEDURE OF WEVA

The procedure of WEVA is the same as the original EVA, except for the wavelet transformation part and the eigenvector estimation method. In every EVA iteration, the eigenvector corresponding to the maximum eigenvalue is estimated by the shifted and inverted version of power method with the maximum diagonal element of $C_w$ as the initial estimate.

# 5. APPLICATION TO SPEECH ANALYSIS

## 5.1. Convergence

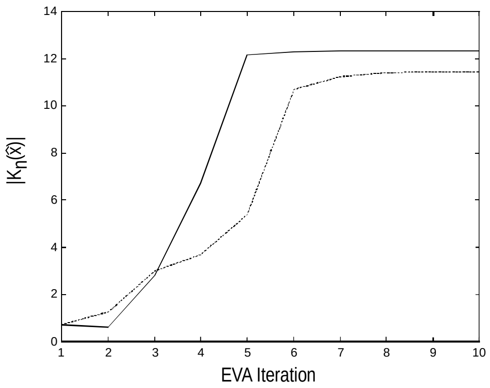
The analyzed speech is a partial segment of Japanese vowel /o/ that is sampled by 11.025[kHz] and the length of the segment is 256. We tested both EVA and WEVA in the condition that the equalizer order $M = 16$ and $M = 32$. WEVA exploits the Daubechies wavelet with the order 8 and 18, respectively. Table 1 shows the required iterations for the eigenvector estimation at each EVA iteration step $it$. The iteration continues until the preciseness of estimated maximum eigenvalue becomes less than $10^{-5}$. I should be noted that WEVA uses the shifted and inverted version of power method that is somewhat heavier than the original power method. However the WEVA reduces the total computational complexity and time because the iterations to estimate the eigenvalue corresponding to the maximum eigenvalue decreases remarkably. Fig.3 shows the normalized kurtosis of the equalized signal. The interesting point of this result is that the magnitude of kurtosis equalized by WEVA converges to a larger value than that by EVA. Such result is confirmed only when a suitable wavelet is applied, but indicates the possibility that the wavelet transform domain blind equalization essentially outperforms the conventionals. The waveforms of deconvoluted excitations are shown in Fig.4. We see that the glottal excitation is clearly presented in the equalized output from EVA and WEVA.

| | M=16 | | M=32 | |
|---|---|---|---|---|
| it | EVA | WEVA (Dau8) | EVA | WEVA (Dau18) |
| 1 | 31 | 11 | 197 | 19 |
| 2 | 69 | 18 | 12 | 19 |
| 3 | 22 | 9 | 9 | 11 |
| 4 | 16 | 7 | 8 | 7 |
| 5 | 28 | 7 | 8 | 5 |
| 6 | 16 | 7 | 7 | 4 |
| 7 | 12 | 7 | 6 | 2 |
| 8 | 9 | 7 | 6 | 2 |
| 9 | 8 | 7 | 5 | 1 |
| 10 | 7 | 7 | 5 | 1 |

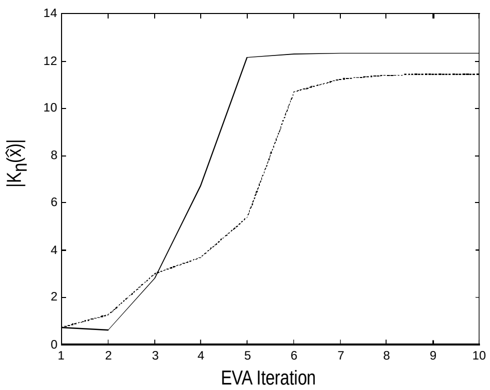
Table 1: Required Iterations

## 5.2. Spectral Analysis

The superiority of the blind equalization technique in the spectral analysis is that the spectrum of the equalizer holds the exact spectral envelope. The preciseness may be a match for or better than the homomorphic filtering common as cepstrum in speech researchers. Fig.5 shows the spectrum of the equalizer. The analyzed signal is the partial segment of a synthesized Japanese vowel /a/ with a pitch frequency 200[Hz]. The length of the segment is 256 with sampling frequency 11.025[kHz]. The parameters of equalizer estimated by WEVA are directly projected into discrete Fourier domain in which 512 points coefficients are calculated within Nyquist frequency.




(a) M=16




(b) M=32

**Figure 3:** Kurtosis Comparison

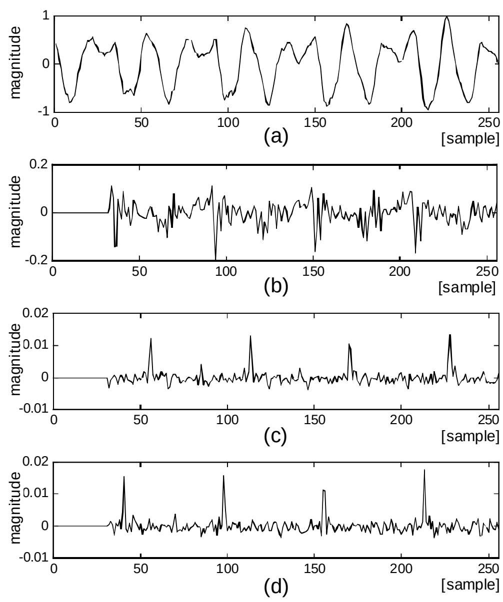



**Figure 4:** Waveform Comparison. (a) Speech. Equalized signal: (b) linear prediction, (c) EVA and (d) WEVA.

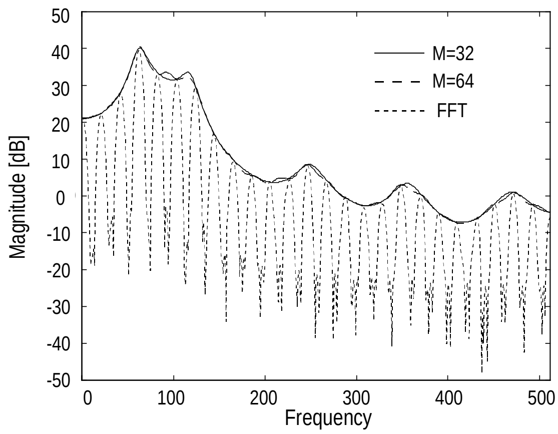



**Figure 5:** Spectrum of WEVA Equalizer

Although a certain order, experimentally more than 18 for EVA and 32 for WEVA, is required for good equalization, we can always have the spectral envelope that seems to be an interpolated FFT spectrum between the peaks at the pitch frequency and its harmonics.

## 6. REFERENCES

1. J. Makhoul, "Linear Prediction: A tutorial review", *Proc. IEEE*, vol. 63, pp. 561-580, 1975.
2. M. Namba and Y. Ishida, "Spectral Anaysis of Speech Signal with L1 norm linear prediction", Proc. *IEEE/ATHOS Workshop on HOS'95*, pp. 271-274, 1995.
3. T.G. Stockham, Jr., T.M. Cannon and R.B. Ingebretesen, "Blind Deconvolution Through Digital Signal Processing", Proc. IEEE, vol. 63, no. 4, pp. 678-692, 1975.
4. A. Benveniste, M. Goursat and G. Ruget, "Robust Identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communication", *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp.385-399, 1980.
5. B. Jelonnek, D. Boss and K.D. Kammeyer, "Generalized Eigenvector Algorithm for Blind Equalization", Elsevier Signal Processing, vol. 61, no. 3, pp. 237-264, 1997.
6. O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Sytems (Channels)", *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312-321, 1990.
7. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, 1996.