*Vol. 35* *No. 144* *Journal of the Radio Research Laboratory* *March 1988*
Printed in Tokyo, Japan *pp. 85–105*

# WEIGHTED LINEAR PREDICTION ANALYSIS OF SPEECH

By
Masuzo YANAGIDA and Osamu KAKUSHO*

(Received on December 16, 1987)


## ABSTRACT

Described here is an improvement of the linear prediction analysis, introducing a weight on each predictive equation, according to the amount of prediction residual of the corresponding data sample. The proposed method is a kind of weighted least-squares method and is called Weighted Linear Prediction (WLP). WLP is a generalization of the Sample-Selective Linear Prediction (SSLP) analysis method proposed by the authors, and SSLP is regarded as a special case of WLP with a binary weight 0 or 1. In the proposed WLP scheme, the weight is set close to 1 for those predictive equations which yield small prediction residuals, and close to 0 for those which yield large residuals, in order that the final estimation may not be seriously affected by the distribution curve outliers that correspond to source excitation.

In the present paper the WLP is first formulated in matrix form and then a practical computational procedure is presented, employing the Givens' reduction. The performance of the WLP on both synthetic and natural speech is shown compared with that obtained by the conventional linear prediction method.


## 1. Introduction

The authors have developed the Sample-Selective Linear Prediction (SSLP) analysis method(1) using the arbitrariness in selecting predictive equations to be evaluated in the least-squares scheme(2). It was an application of the Givens' reduction(3) to linear prediction analysis, making the most of the row-independency property of the process in obtaining the least-squares solution. Thus, the SSLP has improved accuracy in estimating the pole frequencies by removing outliers i.e., improper equations from the least-squares estimation.

Though the actual requirements for analysis accuracy are not clear at this moment, improvement in analysis accuracy is expected in practical applications. For example, methods that yield higher phoneme recognition rates are desirable for realizing practical speech recognition systems. Analysis accuracy can be studied only by using synthetic signals as test signals. It is uncertain, however, that an analysis method which gives high accuracy for synthesis signals also has high potential for natural speech, because the test signals might not adequately simulate the features of natural speech sounds. Practical effectiveness of a given analysis method can be confirmed by comparing the discrimination results replacing the analysis methods from one to another.

Described in this paper is a Weighted Linear Prediction (WLP) which is a generalized version of the SSLP in the sense that a weight is introduced as a function of the prediction equation in accordance with the corresponding residual value. The WLP would not have

* The Institute of Scientific and Industrial Research, Osaka University.

emerged if the Givens' reduction had not been introduced to the linear prediction analysis as a practical means for solving an over-determined set of linear equations. The principal idea of the WLP is that the preferable solution should be optimal in the least-squares sense, putting more weight on a large number of representative predictive equations while reducing the weights on a small number of counterrepresentative equations. Thus the WLP minimizes the sum of substantially weighted residuals, rather than minimizing the sum of equally weighted residuals as in the conventional linear prediction analysis method. The most substantial issue in the WLP is how to determine the weight for each predictive equation as a function of the residual value. The weight might best be determined from a statistical point of view, but in this paper its effects are investigated using computer simulation because the statistical nature of the prediction residual is not well known.

In this study, a matrix representation for the WLP is formulated, and its accuracy is, then, investigated experimentally, using synthetic vowels, and its effectiveness is demonstrated by discriminating voiced plosives with WLP as the analysis method in place of the conventional linear prediction method.

## 2. Formulation of the Weighted Linear Prediction Analysis

### 2.1 Linear Prediction by Using the Givens' Reduction

Linear prediction states that the $n$-th sample $y_n$ is approximated by a linear combination of the past $p$ samples, as in eq.(1), where $\alpha_i$ is the $i$-th predictive coefficient.

$$\sum_{i=1}^{p} \alpha_i y_{n-i} \doteqdot y_n \qquad (1)$$

Collecting these predictive equations for an appropriate period $N$, the analysis scheme is formulated so as to get the least-squares solution (lss) for $\alpha$ in eq.(2).

$$\underset{N\times p}{Y} \; \underset{p\times 1}{\alpha} \doteqdot \underset{N\times 1}{y} \qquad (2)$$

where

$$y = \begin{bmatrix} y_m \\ \vdots \\ y_{m+N-1} \end{bmatrix} \qquad \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}$$

$$Y = \begin{bmatrix} y_{m-1} & \cdots & y_{m-p} \\ \vdots & & \vdots \\ y_{m+N-2} & \cdots & y_{m+N-1-p} \end{bmatrix}.$$

The lss $\hat{\alpha}$ for $\alpha$ is obtained as eq.(3), where $Y^+$ is the generalized inverse of $Y$.

$$\hat{\alpha} = (Y^T Y)^{-1} Y^T y = Y^+ y, \qquad (3)$$

where T denotes the transpose of a matrix.

The Givens' reduction[3] is employed here as a practical procedure for eq.(3). It is summarized as a triangularization process of $Y$ by applying an orthogonal matrix $Q$ to eq.(2) as follows:

$$Q\,Y\alpha \doteqdot Qy \qquad (4)$$

Equaton (4) is then rewritten as eq.(5), where $R$ is an upper triangular matrix of size $p$x$p$, 0 is a ($N$-$p$ )x$p$ zero matrix, and $d$ and $e$ are column vectors of size $p$ and $N$-$p$, respectively.

$$\begin{bmatrix} R \\ 0 \end{bmatrix} \alpha \doteqdot \begin{bmatrix} d \\ e \end{bmatrix} \qquad (5)$$

The lss for $\alpha$ is obtained as follows:

$$\hat{\alpha} = R^{-1} d. \qquad (6)$$

In the actual computation, the process of applying $Q$ to eq.(2) is decomposed into a series of operations named plane rotations, applied each time a predictive equation in a form like that of eq.(1) is transcribed into the working area, as described later.

## 2.2 The Weighted Linear Prediction

### 2.2.1 Weighting Each Predictive Equation

Since each predictive equation expressed as eq.(1) can be manipulated independently in the Givens' reduction, a weighted linear prediction analysis is easily realized by putting weight on each predictive equation before the corresponding plane rotation described in 2.2.2. This type of linear prediction is formulated in matrix form as eq.(7), where $W$ is the weight matrix defined by eq.(8).

$$W\,Y\alpha \doteqdot W y, \qquad (7)$$

where

$$W = \begin{bmatrix} w_m & & & & 0 \\ & w_{m+1} & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & w_{m+N-1} \end{bmatrix}. \qquad (8)$$

The lss $\hat{\alpha}_w$ for this case is expressed as follows:

$$\hat{\alpha}_w = (W\,Y)^{+}\,W y\,. \qquad (9)$$

This analysis scheme is called Weighted Linear Prediction (WLP) analysis method because it is based on weighted predictive equations. Since eq.(7) is a weighted form of an over-determined set of linear equations, the weighted linear prediction analysis is regarded as a weighted least-squares estimation. Due to introducing the weight, the residual power (or the squared sum of the prediction errors) by this method is somewhat larger than that obtained by the ordinary least-squares method, or the conventional linear prediction analysis. This comes from the fact that the proposed method minimizes the squared sum of the weighted prediction equations rather than that of the raw or equally weighted prediction equations. An example for this situation will be shown in section 4.1.

### 2.2.2 Practical Computation

a) Basic Procedure

The practical computation procedure is easily realized by putting a weight $w_n$ on the data sequence corresponding to the predictive equation for the sample $y_n$ transcribed to the working matrix, depicted in Fig. 1, in which the Givens' reduction is performed. If the weight is 0 or 1, the proposed WLP degenerates to the SSLP, so the WLP can be regarded as a generalized version of the SSLP. In case the Givens' reduction is used for the conventional linear prediction analysis procedure, data transcription for the $n$-th prediction equation into the working matrix consists simply of writing the values of data sequence $y_{n-1}, y_{n-2}, \ldots,$ $y_{n-p}, y_n$ into the bottom row of the working matrix.

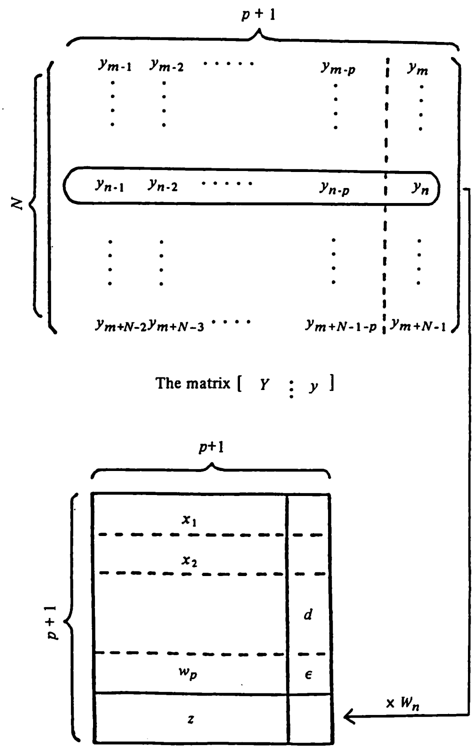


**Fig. 1. The working area used in the Givens' reduction.**
**The prediction equation for a sampled datum is transcribed onto the bottom row of the working matrix multiplied with its corresponding weight value. The plane rotation is performed on the working matrix. After a sequence of the plane rotations, the row vector $z$ in the bottom row becomes a zero vector and the value $\epsilon$ coincides with one element in vector $e$ in eq.(5) in case $w_n$=1.**

For the weighted linear prediction, only the sample values to be transcribed into the bottom row of the working matrix are to be modified as $w_n y_{n-1}$, $w_n y_{n-2}$, . . . . , $w_n y_{n-p}$, $w_n y_n$, instead of the sequence described above. The rest of the procedure is exactly the same as that for the conventional linear prediction analysis.

The following procedure, called plane rotation, is repeated for every existing row each time a new row is transcribed into the bottom row from eq.(7), until all the rows in eq.(7) are exhausted. The vertical size of matrix $Y$, or the number of rows in eq.(7) corresponds to the window length or the length of the analysis frame.

The plane rotation for row addition is represented as

$$\begin{bmatrix} r' \\ z' \end{bmatrix} \Leftarrow \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} r \\ z \end{bmatrix}, \quad \cdots\cdots (10)$$

where ′ denotes renewed value, and

$$\left.\begin{aligned} c &= r_{ii}/q \\ s &= z_i/q \\ q &= (r_{ii}^2 + z_i^2)^{1/2}. \end{aligned}\right\} \quad \cdots\cdots (11)$$

b) Frame Shifting

Frame shifting operation is realized by repeating two types of plane rotations, the above-mentioned one for row addition and the one described below for row deletion. The numbers of row addition and row deletion are arbitrary, but for realizing a fixed window-length frame shifting, they should be same. In order to maintain analysis stability it is preferable for the working matrix to be cleared every time a silence appears in the input speech.

The plane rotation for row deletion is represented as

$$\begin{bmatrix} r' \\ z' \end{bmatrix} \Leftarrow \begin{bmatrix} u & v \\ -v & u \end{bmatrix} \begin{bmatrix} r \\ z \end{bmatrix}, \quad \cdots\cdots (12)$$

where

$$\left.\begin{aligned} u &= r_{ii}/q' \\ v &= z_i/q' \\ q' &= (r_{ii}^2 - z_i^2)^{1/2}. \end{aligned}\right\} \quad \cdots\cdots (13)$$

The actual computation is performed with the Gentleman's algorithm[4] since it speeds up the computation process by avoiding square root calculations. Some useful properties of the Givens' reduction when it is applied to the linear prediction are discussed in reference (5).

## 3. Determining the Weight

Since the performance of the WLP can be controlled by adjusting the weight, the WLP is more flexible than the conventional linear prediction method and even than the SSLP.

The main problem here is how to determine the weight on which the performance of the WLP depends. Unfortunately, no practical way to improve the accuracy of formant estimation by putting weights on predictive equations has been developed so far. It would be in principle, reasonable, however, to determine the weight for each predictive equation according to the prediction error i.e., the prediction residual for the corresponding sample.

### 3.1 Statistical Considerations

The conventional linear prediction model assumes that the excitation signal to the vocal tract is either a pulse train or white Gaussian noise. However, this is not true for the real speech production system, particularly for voiced speech. In fact, according to our offhand investigation, the gross features of the distribution of prediction residuals are fairly well approximated by a single normal distribution curve if the input speech is differentiated with an appropriate pre-emphasis factor. However, the temporal locations of a small number of outliers on the distribution plot often concentrate around the excitation points. This indicates the non-Gaussian nature of the excitation signal of the vocal tract, even if its distribution is not "heavy-tailed".

For heavy-tailed distribution or mixed distribution cases a statistical robustness theory is proposed by Huber(6), and a robust estimation of the system parameters of AR models are also proposed by Martin(7), where a loss function is designed as the integral of a minimax type of Huber's psi-function. Since the present objective, however, is just to reduce ill-effects of the outliers related to source excitation, a class of empirical weight functions is discussed in the following section from a practical point of view.

### 3.2 Empirical Approach

Considering the desirable nature of the weight function, values for $w_n$ ought to be a function of the absolute value of the prediction residual $\epsilon_n$, and they should satisfy the following conditions.

(a) $w(-\epsilon) = w(\epsilon)$, monotone with $w(0) = 1$ and $w(\infty) = 0$.

(b) Its steepness is controllable by a shaping parameter.

(c) Drastic changes such as threshold characteristics should be avoided on the weight function in order to keep the analysis not so sensitive to the shaping parameter.

Taking these factors in account, some heuristic functions have been investigated(8), and the normal distribution curve is suggested(9) as in eq.(14) as the most proper candidate for the weight function, where the standard deviation $\sigma$ is assigned as the shaping parameter.

$$w_n = w(\epsilon_n) = \exp\left[-\frac{1}{2}\left(\frac{\epsilon_n}{\sigma}\right)^2\right] \qquad (14)$$

where $\epsilon_n$: prediction residual for sample $y_n$,

and $\sigma$: shaping parameter to control the weight function.

Some examples of the weight functions are depicted in Fig. 2, where the prediction residual is normalized by its possible maximal value. An appropriate value for $\sigma$ will be chosen according to simulation results.

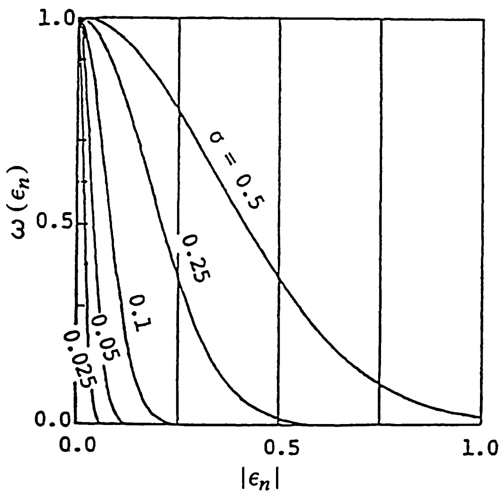


**Fig. 2. Examples of the weight function using the normal distribution curve.**

## 4. Performance of the WLP

### 4.1 Estimation Accuracy for Synthetic Vowel-like Sounds

Synthetic vowels are used to determine the appropriate value for the shaping parameter $\sigma$ and to reconfirm the performance of the proposed WLP. Table 1 shows the conditions for synthesizing test sounds.

The test sounds are analyzed by both the conventional LP method and the proposed WLP method under the conditions shown in Table 2. In the simulation, pre-emphasis is

**Table 1. Synthesis conditions of the test sounds.**

| | |
|---|---|
| Synthesis method | LP of order 10 |
| Excitation | One pitch residual wave |
| Pitch | 125 and 263 Hz |
| Spectral envelope | 50 patterns |

**Table 2. Analysis conditions**

| | |
|---|---|
| Analysis order | 12 |
| Window length (ms) | 20 |
| Pre-emphasis | non |

not employed because the excitation wave here is one pitch residual signal, as described in Table 1, obtained by the LP method with pre-emphasis, while pre-emphasis is needed for analysis of synthetic sounds excited by the Rosenberg wave or the like.

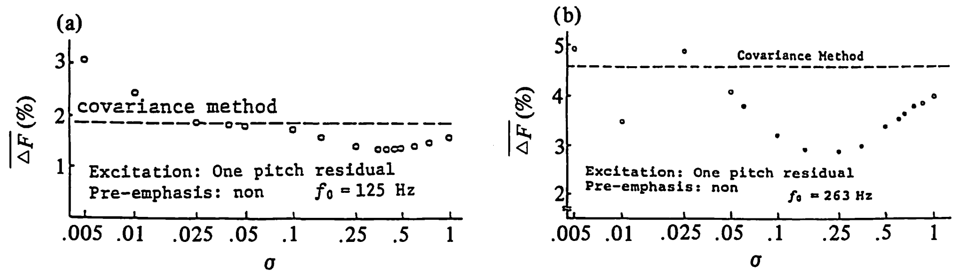


**Fig. 3.** Average errors of pole frequency estimation by using the WLP with the normal distribution curves as the weight function. The broken lines indicate the average error by the conventional covariance method.
(a) Average estimation error for low pitch sounds.
(b) Average estimation error for high pitch sounds.

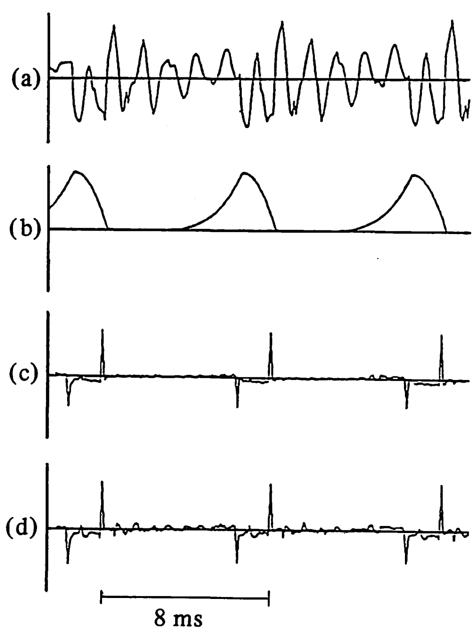


**Fig. 4.** Comparison of the residual signals for a synthetic vowel.
(a) The synthetic vowel for performance evaluation.
(b) The excitation wave.
(c) The residual signal obtained by the WLP.
(d) The residual signal obtained by the conventional LP.

The performance is evaluated by the average errors in estimating the five formant frequencies normalized by the correct values. Figure 3 shows the results for (a) low pitch sounds and (b) high pitch sounds, where the broken lines represent the estimation error by the conventional LP (covariance) method, and the abscissa corresponds to the shaping parameter $\sigma$ normalized by the maximum value of $\epsilon_n$.

Figure 3 says that the proposed WLP shows higher accuracy than the conventional LP method in estimating the pole frequencies of synthetic stationary vowel-like sounds, especially for high pitch sounds. The results state that the appropriate value for $\sigma$ is 0.3~0.6 for low-pitch voices and 0.15~0.3 for high-pitch voices. The value 0.5 will be employed as the default value for $\sigma$ hereafter.

The performance of the WLP is shown in Fig. 4 from a different point of view. A synthetic vowel (a), excited by the Rosenberg wave (b), is used as a test signal here. The corresponding residual signals estimated by the WLP and the conventional LP are compared in (c) and (d), respectively, where the ordinate is normalized by the maximal value of each sequence. From Fig. 4 it can be clearly recognized that the amplitude of the glottal closure duration in the residual signal obtained by the WLP is smaller than that obtained by the conventional LP method. Of course, the residual power is larger for (c) than for (d), since the WLP employs weights to reduce the residual power of the closure period rather than the excitation period.

## 4.2 Effects on Discrimination of Voiced Plosives

Since analysis accuracy for natural speech cannot be evaluated directly, a phoneme discrimination test was employed to confirm the effectiveness of the proposed WLP. The test adopted here is the discrimination of three Japanese voiced plosive sounds in isolated CV syllables composed of a voiced plosive (/b/, /d/ or /g/) followed by one of the five Japanese vowels. Speech samples uttered by 38 male adults were recorded on magnetic tape with an electret condenser microphone in an anechoic chamber and were sampled at the rate of 10 kilo samples/sec with 12 bit accuracy after low-pass filtering (−260 dB/oct).

The discrimination process is depicted in Fig. 5. First, the burst point is detected using the spectral difference between a pair of unequal-length frames whose starting points coincide. Then, based on the detected burst point, the following vowel is identified using decision by majorities in five-nearest neighbors (5-NN) in the four-dimensional Fisher space. The phoneme discrimination of the preceding plosives is performed as decision by majorities in three-nearest neighbors (3-NN) in the two-dimensional Fisher space obtained from the temporal values and the gradients of acoustic parameters. Short-term power and LPC cepstrum coefficients are employed here as the acoustic parameters. The window length and the frame shifting interval for a series of analyses are determined according to the identification results of the following vowel. The phoneme templates were prepared for each following vowel. Figure 6 shows an example of the phoneme templates in the two-dimensional Fisher space, where the ordinate and the abscissa correspond to the first and the second axes in the multi-dimensional Fisher space.

In order to obtain an open discrimination score for speakers, a single discrimination test was repeated 38 times, assuming one set of samples uttered by a speaker as the test sample and the rest as training samples.

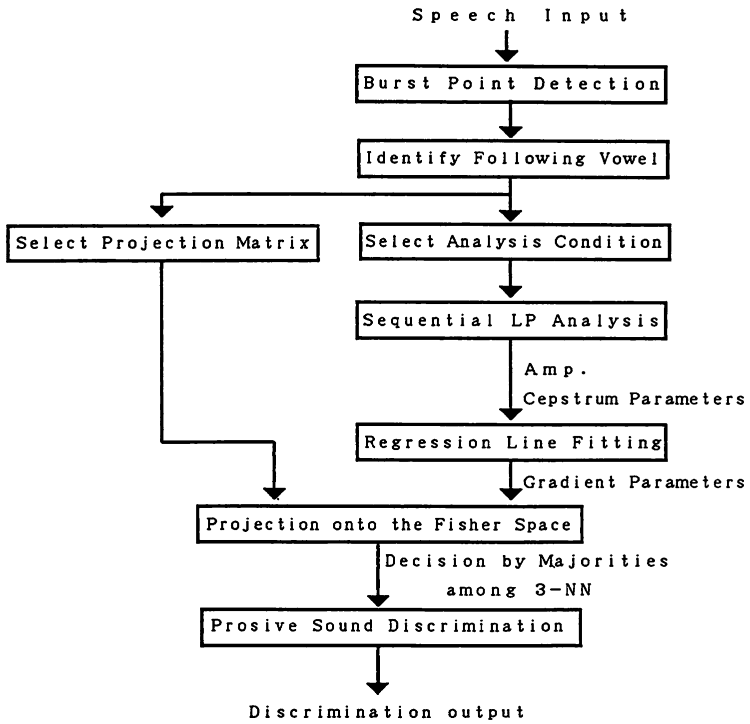


**Fig. 5.** The schematic flow for discriminating voiced plosives.

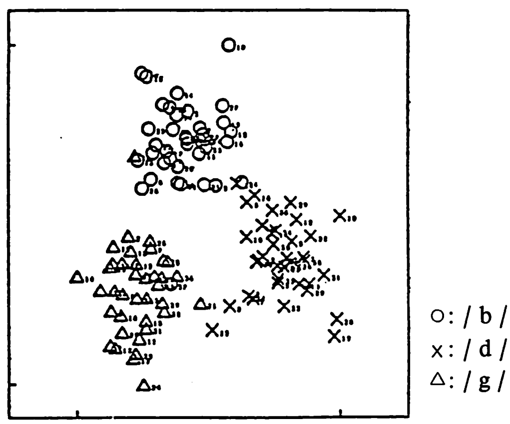


**Fig. 6.** An example of the phoneme templates in the Fisher space for discrimination of voiced plosives followed by vowel /u/. The ordinate and the abscissa correspond to the most and the second significant axes, respectively, in the Fisher space.

Table 3 shows the comparison of the discrimination score by the WLP and the conventional LP (covariance) method for voiced plosives /b/, /d/ and /g/. The figures show the remarkable effects of introducing both the dynamics of the feature parameters and following-vowel dependency into the system, while they show only small improvement with the WLP. The final score for the WLP is expected to improve a little bit more because the total analysis period was optimally adjusted for the conventional LP method, but it was fixed for the WLP in the present experiment. It becomes harder to improve the discrimination score as the results approach a perfect score which is unrealizable unless semantics is employed. Though the score improvement by introducing the WLP is small, it will help somewhat to reduce the ambiguity among phoneme candidates.

**Table 3. The comparison of the discrimination score among /b/, /d/ and /g/ uttered by 38 male adults in isolated CV context.**
**(* indicates employment of the corresponding feature)**

| | | | Discrimination Score (%) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Following Vowel | | | | | |
| Dynamics | Vowel-dependency | Analysis | a | e | i | o | u | Average |
| – | – | LP | 87 | 62 | 70 | 75 | 69 | 72.6 |
| – | – | WLP | 82 | 65 | 76 | 83 | 75 | 76.1 |
| – | * | LP | 90 | 75 | 90 | 84 | 76 | 83.0 |
| * | – | LP | 85 | 80 | 84 | 84 | 88 | 84.2 |
| * | * | LP | 95 | 89 | 89 | 95 | 90 | 91.6 |
| * | * | WLP | 95 | 91 | 93 | 95 | 92 | 93.2 |

## 5. Further Applications

Some further applications of the weighted linear prediction are discussed in this section, namely introducing the weight to a non-stationary linear prediction model, to the ARMA model and to pitch synchronous analysis.

### 5.1 Application of the WLP to Non-stationary LP Model

Several non-stationary analysis models(10-12) have been proposed so far in order to correctly represent the nonstationary nature of natural speech. The introduction of the weight to a non-stationary linear prediction model proposed by Grenier(12) which approximates the time variations of the AR parameters within the analysis frame by means of an expansion of the AR parameters with an orthogonal set of functions.

#### 5.1.1 The Grenier's Model(12)

The time varying AR model proposed by Grenier is summarized in the following expression for the time-varying AR parameters.

$$\alpha_i(n) \doteqdot \sum_{j=0}^{m} a_{ij} f_j(n) \quad , \qquad (15)$$

where $m$ represents the expansion order and $\{f_i(n)\}$ is a set of orthogonal functions which satisfies the following orthogonality condition:

$$\int f_j(n) f_k(n) dn = \begin{cases} 1, & j=k \\ 0, & j \neq k \end{cases} \qquad (16)$$

The linear prediction equations is then expressed as

$$\sum_{i=1}^{p} \sum_{j=0}^{m} a_{ij} f_j(n) y_{n-i} \doteq y_n, \qquad (17)$$

and is rewritten as

$$[\, Y_{n-1}^{T} \cdots Y_{n-p}^{T} \,]\, \theta = y, \qquad (18)$$

where

$$\underset{(m+1)\times 1}{Y_n} = [\, f_0(n) y_n \cdots f_m(n) y_n \,]^{T} \qquad (19)$$

and

$$\underset{p(m+1)\times 1}{\theta} = [\, a_{10} \cdots a_{1m} a_{20} \cdots a_{2m} \cdots a_{p0} \cdots a_{pm} \,]. \qquad (20)$$

Collecting these predictive equations for $n = n, n+1, \ldots\ldots, n+N-1$, a matrix representation is obtained as follows:

$$\underset{N\times p(m+1)}{\mathbf{Y}} \; \underset{p(m+1)\times 1}{\theta} \doteq \underset{N\times 1}{y}, \qquad (21)$$

where

$$\mathbf{Y} = \underset{N\times p(m+1)}{\begin{bmatrix} f_0(n-p) y_{n-p} & \cdots & f_m(n-p) y_{n-p} & f_0(n-p+1) y_{n-p+1} & \cdots & f_m(n-p+1) y_{n-p+1} & \cdots & f_0(n-1) y_{n-1} \cdots f_m(n-1) y_{n-1} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \quad \vdots \\ f_0(n-p-N+1) y_{n-p-N+1} & \cdots & f_m(n-p-N+1) y_{n-p-N+1} & f_0(n-p-N+2) y_{n-p-N+2} & \cdots & f_m(n-p-N+2) y_{n-p-N+2} & \cdots & f_0(n-N) y_{n-N} \cdots f_m(n-N) y_{n-N} \end{bmatrix}} \qquad (22)$$

Then the Yule-Walker equation for this case is expressed as

$$\mathbf{Y}^{T} \mathbf{Y} \theta = \mathbf{Y}^{T} y. \qquad (23)$$

However, as the structure of this equation is similar to the ordinary Yule-Walker equation, the same computational procedure described in the previous section can be employed again. The number of columns in this extended Yule-Walker equation is $(m+1)$ times larger than that of the conventional one.

5.1.2 Formulation of the Weighted Time-varying LP Model or the Non-stationary WLP

Weight is easily introduced to the type of prediction expressed in eq.(21) as

$$W \mathrm{Y} \theta \doteqdot W y \quad (24)$$

and the weighted extended Yule-Walker equation is expressed as follows:

$$(W \mathrm{Y})^{\mathrm{T}} (W \mathrm{Y}) \hat{\theta} = (W \mathrm{Y})^{\mathrm{T}} y. \quad (25)$$

The computational procedure is realized as a two-stage analysis depicted in Fig. 7.

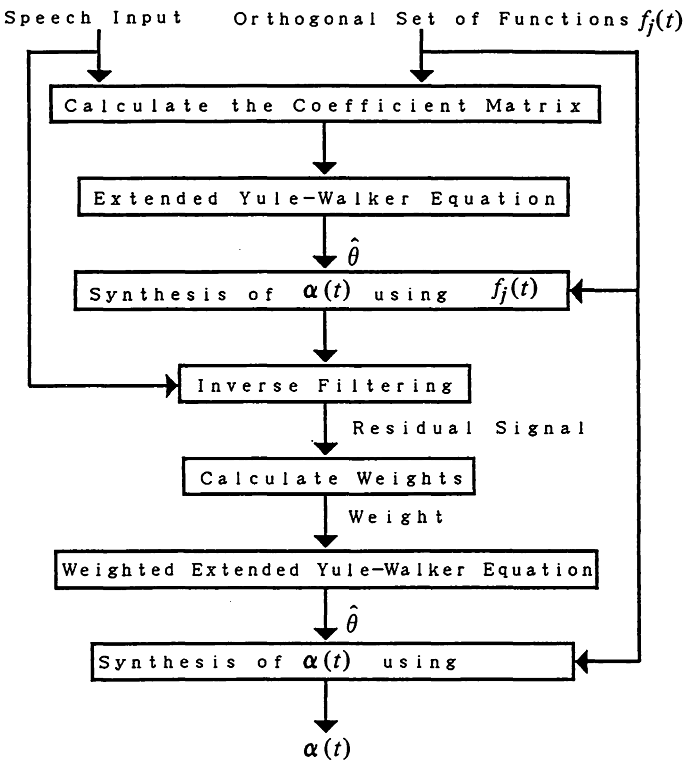


**Fig. 7. The schematic flow of the non-stationary WLP analysis method.**

1. Choose a set of orthogonal functions and decide the expansion order $m$.
2. Calculate the coefficient matrix $\mathbf{Y}$ based on the set of orthogonal functions.
3. Get the preliminary least-squares solution $\hat{\theta}$ for the expansion coefficient vector $\boldsymbol{\theta}$ by solving the extended Yule-Walker equation.
4. Get the preliminary estimate for the AR parameter vector $\alpha(t)$ as a function of time.

5. Calculate the prediction residual by inverse filtering.
6. Calculate the weight for each extended prediction equation.
7. Set up the weighted extended Yule-Walker equation.
8. Get the least-squares solution for $\boldsymbol{\theta}$ by solving the weighted extended Yule-Walker equation.
9. Get the final estimate for $\boldsymbol{\alpha}(t)$ as a function of time.

5.1.3 Performance of the Non-stationary WLP

In order to investigate the performance of the non-stationary WLP, time varying synthetic sounds are used. An example of the test sounds is depicted in Fig. 8, which is a connected sequence of three vowels. The time-varying part of the test sound is analyzed by the conventional LP with frame shifting, non-stationary LP and the proposed non-stationary WLP, according to the analysis conditions shown in Table 4.

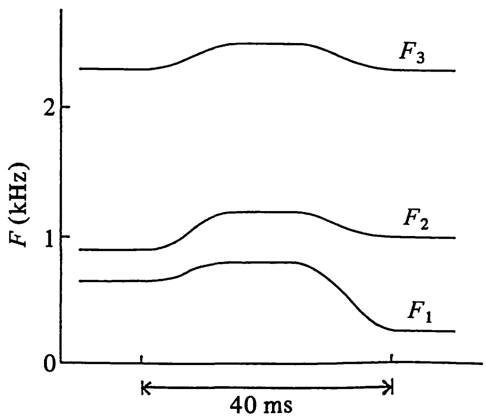


**Fig. 8. Formant trajectories of the synthetic sound for performance evaluation of the proposed non-stationary WLP method.**

**Table 4. Analysis conditions for the time-varying sound.**

| | Conventional LP | Non-stationary LP | Non-stationary WLP |
|---|---|---|---|
| Pre-emphasis factor | 0.93 | 0.93 | 0.93 |
| Analysis order | 12 | 12 | 12 |
| Frame length | 20 ms | 40 ms | 40 ms |
| Frame shifting | Every 20 ms | – | – |
| Evaluation point | Window center | Every 20 ms | Every 20 ms |
| $\sigma$ | – | 0.5 | 0.5 |
| Orthogonal function | – | Legendre | Legendre |
| Expansion order | – | 2 | 2 |

The normalized estimation errors for the first three formants for the middle 40 ms-long part are shown in Fig. 9, where the weighted version (denoted by ○) shows better accuracy than the non-weighted version (△) and better than the conventional stationary linear prediction method (x).

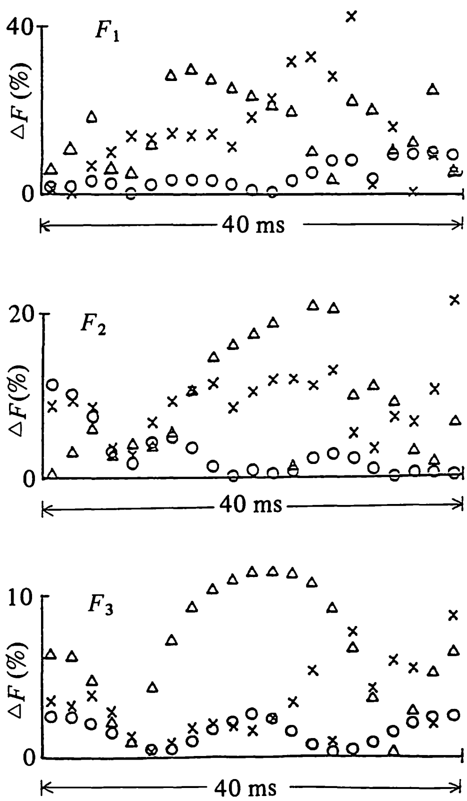


**Fig. 9. Normalized estimation errors for the first three formants.**
**○: Non-stationary WLP**
**△: WLP**
**x: Conventional LP**

## 5.2 Application of the WLP to the ARMA Model

For analysis of nasal sounds and certain other consonants the all-pole model or the AR model is not sufficient to describe the nature of those sounds. The pole-zero model i.e., the ARMA model has been developed to meet that demand. Discussed in this section is introduction of weight into the estimation of the ARMA parameters based on the ARMA model(13).

### 5.2.1 Introduction of Weight to ARMA parameter Estimation

The WLP is expected to reduce the errors in estimating the MA parameters from the residual signal obtained through the inverse filter corresponding to the AR part. The analysis scheme is depicted in Fig. 10, where the weighted high-order linear prediction method is employed to get more accurate estimates for the AR parameters, and the MA parameters are obtained as the coefficients of the inverse polynomial expression based on the AR parameters obtained from the residual signal by the conventional LP method. The pole and zero frequencies are calculated based on the AR and the MA parameters, respectively.

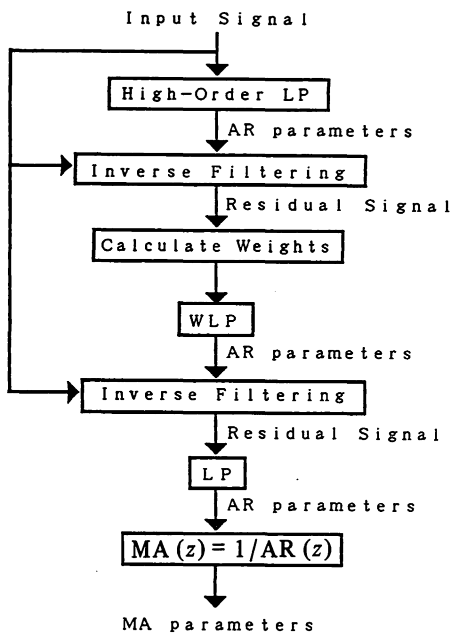


**Fig. 10. The schematic flow of the MA parameter estimation employing the WLP in the first stage analysis.**

### 5.2.2 Performance Evaluation

Performance of the proposed method on nasalized vowels is compared with the Cadzow's method[13] in Fig. 11, where the rigid line represents the test signal, the dotted broken line, the estimated spectra by the proposed method, and the broken line, by the conventional method without weighting. The test sounds are stationary ones synthesized with an ARMA filter ($p = 8$, $q = 2$) excited by the Rosenberg wave. Additional Gaussian noise is added to the test sounds to show the effectiveness of the proposed method even for slightly noisy cases. The test sounds are prefiltered by a pre-emphasis factor of 0.93. The analysis is carried out by the corresponding methods with $p = 8$, $q = 2$ through a 20 ms window. The estimation errors of pole and zero frequencies are compared in Table 5. Table 5 also shows that further improvement was obtained by introducing both the high-order property* and the over-determination property** together with weight.

The effect of weighting is most evident in estimation of zero frequencies. It is because the estimation accuracy of zero frequencies depends on the estimation accuracy of the residual signal waveform which is obtained more precisely by introducing weight in the first stage analysis.

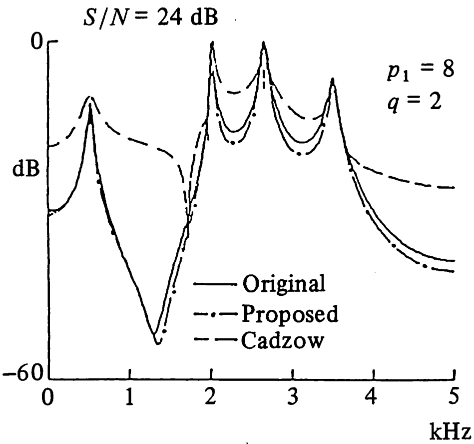


**Fig. 11. Performance comparison of estimating frequency spectra by the weighted 2-stage method and the conventional 2-stage method on a synthetic sound with poles and zeros.**

**———: test signal**
**— · —: the proposed weighted 2-stage method**
**- - - - -: the 2-stage method by Cadzow(13)**
**$p_1$ :analysis order in the first stage analysis**
**$q$: assumed order for the MA process**

**Table 5. Estimation errors (%) for pole and zero frequencies of nasalized synthetic vowels.**

| | Pole Frequency | Zero Frequency |
|---|---|---|
| Cadzow's method | 1.4 | 12.5 |
| Weighted LP | 1.5 | 5.8 |
| Hi-order Over-det. WLP | 0.4 | 2.9 |

* The high-order property is shifting a definite number of rows ($d = 2$, here) in the Yule-Walker equation.

** The over-determination property is adding a definite number of excess equations ($e = 8$, here) to the Yule-Walker equation.

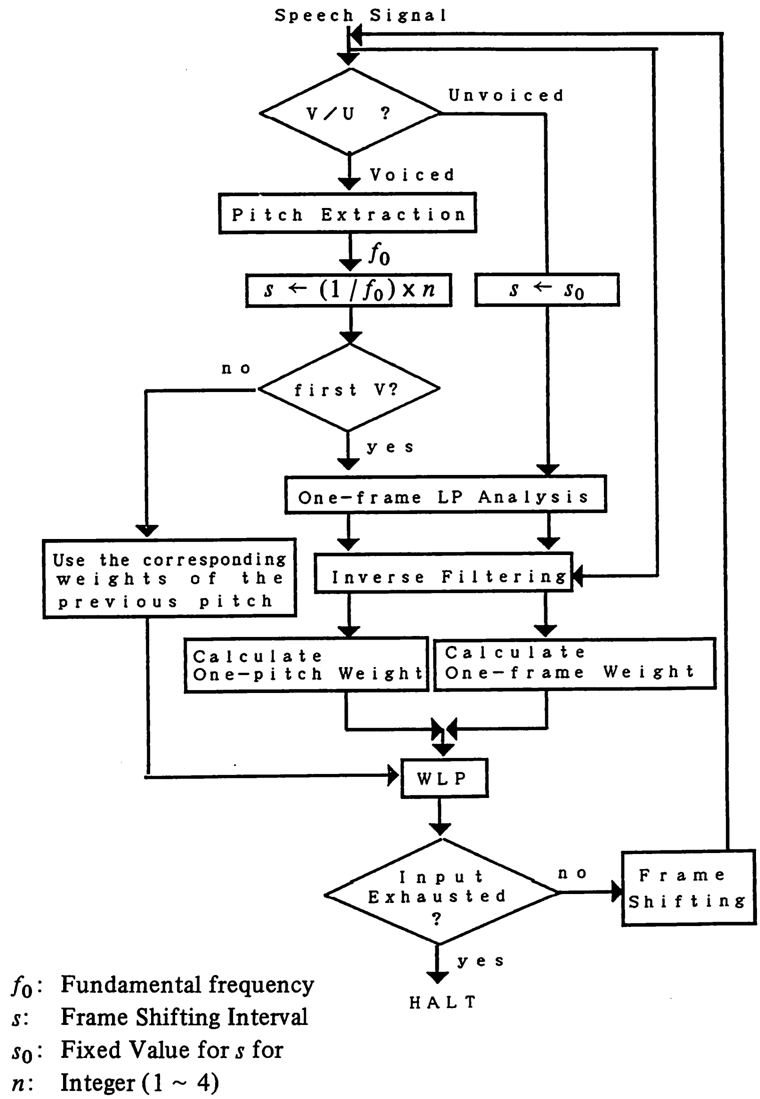


**Fig. 12. The schematic flow of the pitch synchronous WLP.**

## 5.3 Pitch Synchronous WLP[14]

Described in this section is a means to speed up the computation process of the WLP. The basic idea is to use the same weight sequence for a series of analysis frames on voiced i.e., periodic portions of speech. The validity of this idea comes from the fact that the prediction residual seems to repeat almost the same patterns pitch-synchronously.

### 5.3.1 Pitch Synchronous Weighted Linear Prediction

The WLP offers the difficulty that preliminary analysis and inverse filtering are required before the final analysis for each analysis frame in order to calculate the weight for prediction equatios in the analysis frame concerned. Most of these preparatory procedures, however, can be omitted by assuming that the residual signal repeats similar patterns in every pitch period for a speaker even if its fundamental frequency varies.

If the residual signal can be assumed to repeat the same pattern in every pitch period, a common weight pattern can be employed for each analysis frame shifted synchronously to pitch. Therefore, the calculation for determining the weight is required only for the first analysis frame. Thus the calculation process of the WLP is drastically reduced. This method is called Pitch Synchronous Weighted Linear Prediction analysis (PSWLP). The window length for the PSWLP is preferably made one- or integer-pitch period long so that the analysis is independent of the mutual position to pitch excitation. In the PSWLP the correspondence between sample points on the speech signal and those on the weight pattern has great importance, so the shifting interval should be strictly synchronized to the pitch interval.

The processing scheme of the PSWLP is depicted in Fig. 12. In the present system pitch synchronization is realized by fundamental frequency extraction based on the auto-correlation of the residual signal. The shifting interval for analysis of natural speech is synchronized to local maximal points on voiced portions of the input speech in order to avoid mis-matching in correspondence between residual values and weight values.

### 5.3.2 Performance Evaluation

In order to confirm the performance of the PSWLP, accuracy comparison in estimating formant frequencies was made among the following methods:

COV($N$): Covariance method ($N$: window length).
SSLP($\theta$): Sample Selective LP, $\theta$: The threshold value for $|\epsilon_n|$.
WLP($\sigma$): Weighted LP, $\sigma$: The shaping factor.
PSWLP($\sigma$): Pitch Synchronous WLP.

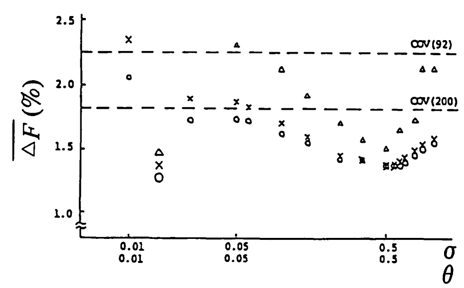


**Fig. 13. Performance comparison among estimation errors among PSWLP, WLP, SSLP and the conventional LP on synthetic stationary sounds.**
**Broken lines represent the estimation error by the conventional covariance method of window length noted in ( ).**

An all-pole filter of order 10 was employed to synthesize vowel-like test sounds excited with one-pitch residual sequence of 80 samples long. Fig. 13 shows the results of the comparison with a common analysis order 12 on a non-pre-emphasized data sequence. The results are evaluated for 10 different window locations and for five formants over 50 synthetic sounds of different formant patterns. The abscissa represents the shaping parameter $\sigma$ for the weight function for WLP and PSWLP, and corresponds to the threshold value $\theta$ for the SSLP. From Fig. 13, it can be seen that the PSWLP retains the same accuracy as the WLP over a broad range of the shaping parameter, although the window length is as short as one pitch period.

The processing times required for the above-mentioned methods are compared in Table 6, where the average times required to process one frame are shown. The processing time required for PSWLP is reduced to one third of that required for WLP.

**Table 6. Comparison of the time required for one frame analysis on DG MV/8000 II.**

| Analysis method | Time required (sec) |
|---|---|
| Covariance method(15) | 0.04 |
| Givens' Reduction(4) | 0.09 |
| WLP(8) | 1.22 |
| PSWLP(14) | 0.36 |

## 6. Conclusions

The concept of the weighted linear prediction is described from the view point of weighted least-squares estimation. The Givens' reduction with the Gentleman's algorithm is employed as a practical computational procedure for solving the weighted least-sqaures problem concerned.

The weight is introduced to improve estimation accuracy of the linear prediction method for speech signal analysis. Although the improvement of error in formant frequency estimation remain only about 30%, the proposed weighting schemes show stable contribution to error reduction. The effect of improvement of formant frequency estimation on phoneme discrimination is confirmed by employing the proposed method as the analysis tool for the plosive sound discrimination system developed by the authors, resulting in a definite improvement in discrimination rate.

Applications of the proposed method to a non-stationary model and to pitch synchronous analysis are discussed from the view point of constructing a practical model and/or seeking computational efficiency with satisfactory results.

A multi-stage WLP, or a multi-stage weighting, is a possible means to refine the analysis accuracy. It is under investigation as an extension of the WLP, promising higher accuracy in estimating pole frequencies.

Computational difficulties and practical effects of the proposed method are left for future investigation.

**Acknowledgment**

The authors express their thanks to Messrs. Y. Yamashita, K. Yokota, Y. Takahashi, Y. Miyazaki, M. Tsunoda, Y. Okada, T. Mori and S. Tsukada, Osaka University, for their help in this research project. The authors also thank Dr. T. Igarashi, Chief of the Publication Committee, Radio Research Laboratory, for giving us the opportunity to summarize the project.

This work has been partly supported by Grant-in-Aid for Scientific Research for fiscal 1984–85 (No. 59550039), 1986–87 (No. 61850072) and 1987 (No. 62608008), and by Special Coordination Funds for Promoting Science and Technology, Science and Technology Agency, for 1986–88.

*References*

(1) Kakusho, O., Yanagida, M., and Mizoguchi, R.: "A sample-Selective Linear Prediction of speech", Proc. 9th ICASSP, San Diego, 2.1, Mar. 1984.

(2) Yanagida, M., Kakusho, O., and Graham-Stuart, D.; "Speech Analysis by Generalized Inverse of Matrices", Proc. 10th ICA, Sydney, A1–10.3, Jul. 1980.

(3) Nash, J.C.; "Compact Numerical Methods for Computers: Linear Algebra and Function Minimization", Adam-Hilger, Bristol, pp. 40–43, 1979.

(4) Gentleman, W.M.; "Least-Squares Computations by Givens' Transformations without Square-Root", J. Inst. Maths. Applc., **12**, pp. 329–336, 1973.

(5) Yanagida, M., Yamashita, Y., Kakusho, O., and Graham-Stuart D.; "Some Properties of the Givens' Reduction Applied to Linear Prediction Analysis of Speech", J. Acoust. Soc. Jpn. (E), **7**, 1, pp. 47–56, Jan. 1986.

(6) Huber, P.J.; "Rubust Statistics", Wiley, New York, 1981.

(7) Martin, D.R.; "Rubust Methods for Time Series", in Applied Time Series II, D.F. Findley ed., pp. 683–759, Academic, New York, 1981.

(8) Yanagida, M. and Kakusho, O.; "A Weighted Linear Prediction of Speech Signals by Using Givens' Reduction", IASTED International Symposium on Applied Signal Processing and Digital Filtering, Paris, pp. 129–131, June 1985.

(9) Yanagida, M., Tsukada, S., and Kakusho, O.; "High-Order Over-Determined Weighted Linear Prediction Analysis Method" Reports of Inst. Elec., Inf. & Com. Engnrs. Jpn., SP87–14, May 1987.

(10) Bohlin, T.; "Four Cases of Identification of Changing Systems", in "System Identification: Advances and Case Studies", ed. R.K. Mehra, and D.G. Lainiotis, Academic 1976.

(11) Miyanaga, K., Miki, N., Nagai, N., and Hatori, K.; "Adaptive Simultaneous Estimation of Time Variable Autoregressive and Moving-Average Parameters", Trans. IECE, Jpn., **J64-D**, 4, pp. 308–315, 1981.

(12) Grenier, Y.; "Time-dependent ARMA modeling of nonstationary signals", IEEE Trans. Acoust., Speech Signal Process., **ASSP-31**, 4, pp. 899–911, Aug. 1983.

(13) Cadzow, J.A.; "High Performance Spectral Estimation – A New ARMA Method", IEEE Trans. Acoust., Speech Signal Process., **ASSP-28**, 5, pp. 524–529, 1980.

(14) Yanagida, M. and Kakusho, O.; "A Pitch Synchronous Linear Prediction Analysis of Speech", Proc. of 12th Intern. Congress on Acoust., Montreal, A1–6, 1986.

(15) Markel, J.D. and Gray Jr., A.H.; "Linear Prediction of Speech", Springer, Berlin, p. 221, 1976.