Electronics and Communications in Japan, Part 1, Vol. 70, No. 9, 1987
Translated from Denshi Tsushin Gakkai Ronbunshi, Vol. 69-A, No. 4, April 1986, pp. 525-534

Vowel Recognition Using an LPC Deviation Model

Itsuo Kumazawa and Taizo Iijima, Members

Faculty of Engineering, Tokyo Institute of Technology, Tokyo, Japan 152

SUMMARY

When the linear prediction coefficients (LPC) are used as a feature parameter of speech, a problem in speech recognition is the variation of LPC due to the difference of the speaker and the effect of the preceding or succeeding utterances. Because of this variation, LPC is distributed in a certain region in the LPC space. This paper proposes a method which approximates the region of distribution by a linear manifold model described in [1], and recognizes the category permitting the variation of LPC. The method is applied to the vowel recognition of CV syllable, and its usefulness is verified experimentally. It is shown first by spectrum analysis experiment that the linear manifold model is useful in the approximation of the variation of the vowel spectrum due to the preceding consonant. Second, a recognition experiment is carried out based on the model, and it is shown that the recognition rate can be improved by a simple model of linear manifold which is suited to computation.

1. Introduction

The feature parameter of speech exhibits a great variation by individuals or by the effect of the preceding and succeeding utterances, which is one of the most serious reasons for difficulty in speech recognition. One possible approach to such variations of feature parameters is to prepare a model for the variation and to perform the matching within the model, thereby permitting the variation from the template. The method proposed in this paper is based on such an idea.

In the following, the speech to be considered is restricted to the vowel, and the linear prediction coefficients (LPC) are used as the feature parameter which is known to be adequate in representing the vowel. The model for the variation used in the following is the expression

$$\bar{a} + \sum_{l=1}^{L} c_l b_l \qquad (1)$$

which is equivalent to the one used in POLPEC (Polarized linear Predictive Error Coding) [2]. The variation from the standard LPC $\bar{a}$ is approximated by a linear combination of vectors $b_1, \cdots, b_L$ representing the major variation components. In [2] - [4], this model is considered for the prediction error signal, and the recognition is performed by the processing in the time-domain. On the other hand, [1] presented the determination of $\bar{a}$, $b_1, \cdots, b_L$ and the parametric spectrum analysis by using Eq. (1) as a model of speech variations. Based on those results, this paper derives a method of recognition in the frequency-domain. By using such a simple and linear model as Eq. (1) the computational cost of spectrum analysis is made quite low. Thus, using the adaptive filter which is an extension of the lattice filter, the analysis can be implemented by simple hardware.

Section 3 experimentally shows that the variation of the vowel spectrum due to the preceding consonant can easily be approximated by such a simple model. Section 4 presents a recognition experiment, and it is shown that the recognition rate can be improved, compared with the case using the fixed template LPC. It may be obvious that a higher recognition rate can be achieved by using a more complex model. The purpose of this paper is to show that the recognition performance can be improved by such a simple model suited for computation as in Eq. (1), by adequately setting $\bar{a}$ and $\{b_l\}$.

2. LPC Variation Model

The method of determining $\bar{a}$ and b_l in the model of Eq. (1) is described in [1], but is summarized in the following. Let the speech signal be $x(n)$ and the prediction

ISSN8756-6621/87/0009-0030$7.50/0

error be $\varepsilon(n)$. Consider the linear prediction model

$$\varepsilon(n)=x(n)+a_1 x(n-1)+\cdots+a_p x(n-P) \quad (2)$$

The set of prediction coefficients is represented as a vector as

$$\boldsymbol{a}=(1,\ a_1,\ \cdots,\ a_p)^t \quad (3)$$

In the following, this vector is simply called LPC. The space composed of all such vectors is called LPC space.

The set of signal pieces $\{x(n)\}$ belonging to the specified vowel category is represented as follows using a marker ω.

$$\{x(n,\omega),\ \omega\in\Omega\} \quad (4)$$

The LPC of the signal $x(n,\omega)$ is denoted by $\boldsymbol{a}(\omega)$. The autocorrelation matrix of the signal is denoted by $R(\omega)$. LPC belonging to Ω always corresponds to the same vowel, but exhibits a variation by individuals and by the influence of neighboring utterances. Consequently, the above set is distributed in the LPC space due to the variation.

The following model is employed in this paper in order to approximate the distribution of $\boldsymbol{a}(\omega)$ in the LPC space and the model is called LPC deviation model:

$$\hat{\boldsymbol{a}}_L(\omega)=\bar{\boldsymbol{a}}+\sum_{l=1}^{L} c_l(\omega)\boldsymbol{b}_l \quad (5)$$

$$\bar{\boldsymbol{a}}=(1,\ \bar{a}_1,\ \cdots,\ \bar{a}_p)^t$$

$$\boldsymbol{b}_l=(0,\ b_{l1},\ \cdots,\ b_{lp})^t$$

which corresponds geometrically to a linear manifold in LPC space.

$\bar{\boldsymbol{a}}$ is the standard LPC, and $\{\boldsymbol{b}_l\}$ is the vector indicating the major direction of LPC variation; $c_l(\omega)$ is a parameter which can vary depending on ω, which is estimated by a parametric signal analysis technique. The number L of such varying parameters $c_l(\omega)$ is called the degree of freedom technique of the model. By contrast, $\bar{\boldsymbol{a}}$ and $\{\boldsymbol{b}_l\}$ are fixed vectors for all $\omega\in\Omega$, and can be considered as the standard pattern representing the properties of the vowel, including its variation. $\bar{\boldsymbol{a}}$ and $\{\boldsymbol{b}_l\}$ are determined for Ω so that all $\boldsymbol{a}(\omega)$, $\omega\in\Omega$ belonging to Ω are best approximated by the model. For the purpose of the speech recognition, it is desirable that all $\boldsymbol{a}(\omega)$ belonging to a vowel category can be approximated with as small degrees of freedom as possible.

The notation $\underset{\omega\in\Omega}{E}$ used in the above discussion denotes the averaging over ω. Actually, it is calculated as follows using the samples of Ω:

$$\underset{\omega\in\Omega}{E}[\boldsymbol{a}(\omega)]=\frac{1}{M}\sum_{i=1}^{M}\boldsymbol{a}^{(i)} \quad (6)$$

where $\boldsymbol{a}^{(i)}$ is the sample of LPC belonging to ω, and M is the number of samples. Using this notation, the approximation problem for the LPC distribution by the model can be formulated as follows, denoting the distance in LPC space by $d(\boldsymbol{a},\boldsymbol{b})$.

The mean error, when all LPC's in Ω are approximated by the model with m degrees of freedom, is represented as

$$\Delta_m=\underset{\omega\in\Omega}{E}\left[\min_{c_1(\omega),\cdots,c_m(\omega)} d\left(\boldsymbol{a}(\omega),\ \bar{\boldsymbol{a}}+\sum_{l=1}^{m}c_l(\omega)\boldsymbol{b}_l\right)\right] \quad (7)$$

$\bar{\boldsymbol{a}}$, $\boldsymbol{b}_1$, $\cdots$, $\boldsymbol{b}_L$ are successively determined to minimize Δ_0, Δ_1, $\cdots$, Δ_L, respectively. The degrees of freedom L should be determined so that Δ_L is sufficiently small.

In the following, the log likelihood ratio [5] is defined by

$$d(\boldsymbol{a}(\omega),\ \hat{\boldsymbol{a}}_L(\omega))=\log\frac{\hat{\boldsymbol{a}}_L(\omega)^t R(\omega)\hat{\boldsymbol{a}}_L(\omega)}{\boldsymbol{a}(\omega)^t R(\omega)\boldsymbol{a}(\omega)} \quad (8)$$

and is used as the measure of distance in the LPC space. As is shown in [1], $\bar{\boldsymbol{a}}$, $\boldsymbol{b}_1$, $\cdots$, $\boldsymbol{b}_L$ can be determined approximately by solving the following series of eigenvalue problems. In the following, it is assumed that $R(\omega)$ is normalized as follows by the prediction error power:

$$R(\omega)/\boldsymbol{a}(\omega)^t R(\omega)\boldsymbol{a}(\omega) \quad (9)$$

The autocorrelation matrix which is already normalized is written as $\hat{R}(\omega)$.

(i) Determination of $\bar{\boldsymbol{a}}$

Let the matrix obtained by averaging $\hat{R}(\omega)$ be

$$A=\underset{\omega\in\Omega}{E}[\hat{R}(\omega)] \quad (10)$$

Then $\bar{\boldsymbol{a}}$ is determined as the solution of

$$A\bar{\boldsymbol{a}}=\lambda_0(1,\ 0,\ \cdots,\ 0)^t \quad (11)$$

This is the LPC corresponding to the mean autocorrelation.

(ii) Determination of b_1

Let the vector obtained by deleting the first component of

$$b_l = (0,\ b_{l1}, \cdots,\ b_{lP}) \tag{12}$$

be

$$\hat{b}_l = (b_{l1}, \cdots,\ b_{lP})^t \tag{13}$$

Calculate the matrix

$$B_0 = \underset{\omega \in \Omega}{E} [\hat{R}(\omega)\, \bar{a}\, \bar{a}^t\, \hat{R}(\omega)] \tag{14}$$

Let the matrix obtained by deleting the first row and the first column of this matrix be $\hat{B}_0$, and let the matrix obtained by deleting the first row and the first column of A be $\hat{A}$. Then $\hat{b}_1$ is obtained as the eigenvector corresponding to the maximum eigenvalue of the characteristic equation

$$\hat{B}_0 \hat{b}_1 = \lambda_1 \hat{A} \hat{b}_1 \tag{15}$$

Assuming that $\bar{a}$, b_1, $\cdots$, b_{m-1} are already determined, b_m is calculated inductively as follows.

(iii) Determination of b_m

Calculate the matrix

$$B_{m-1} = \underset{\omega \in \Omega}{E} \left[\hat{R}(\omega) \left(\bar{a} + \sum_{l=1}^{m-1} c_l(\omega)\, b_l \right) \cdot \left(\bar{a} + \sum_{l=1}^{m-1} c_l(\omega)\, b_l \right)^t \hat{R}(\omega) \right] \tag{16}$$

where $c_l(\omega)$ is obtained by solving the following system of equation for each ω, by setting that $L = m - 1$:

$$\sum_{l=1}^{L} c_l(\omega)\, b_l^t R(\omega)\, b_k = -\bar{a}^t R(\omega)\, b_k \qquad k = 1, \cdots,\ L \tag{17}$$

Let the matrix obtained by deleting the first row and the first column of B_{m-1} be $\hat{B}_{m-1}$. Then $\hat{b}_m$ is obtained as the eigenvector corresponding to the maximum eigenvalue of the characteristic equation

$$\hat{B}_{m-1} \hat{b}_m = \lambda_m \hat{A} \hat{b}_m \tag{18}$$

b_m thus obtained is not unique and has the freedom within the scalar multiplier in performing the analysis using the adaptive filter described later, the following normalization is convenient,

$$b_m^t A\, b_m = \bar{a}^t A \bar{a}$$

3. Speech Analysis Based on the LPC Deviation Model

We use the LPC deviation model defined above for the parametric spectrum analysis. In the analysis the LPC closest to the input LPC, among the LPC's represented by the model, is found. In this way, the constraint represented by the model is introduced to the analysis. Then, the recognition is performed as follows.

By the method described in Sect. 2, a and b_m representing a vowel category are determined for each category. The LPC deviation model is constructed for each category. The distance between the LPC $a(\omega)$ of the input speech and the model

$$\hat{a}_L(\omega) = \bar{a} + \sum_{l=1}^{L} c_l(\omega)\, b_l \tag{19}$$

is defined as follows:

$$D_L = \min_{c_1(\omega), \cdots, c_L(\omega)} d(a(\omega),\ \hat{a}_L(\omega)) \tag{20}$$

The distance is calculated for each model of the vowel categories and the category corresponding to the model providing the minimum distance is defined as the result of recognition.

The determination of $\hat{a}_L(\omega)$ minimizing the right-hand side of Eq. (20) is equivalent to the LPC analysis with the constraint that the determined LPC must be contained in the linear manifold of Eq. (19). In other words, the parametric analysis is made based on the model of Eq. (19). The constraint is stronger as L is smaller. The desirable model representing a vowel is such that all $a(\omega)$ belonging to that vowel category can well be approximated by $\hat{a}_L(\omega)$ with the smallest L. In the following, the ability of the model to approximate the variation of LPC is examined by comparing the spectrum calculated from $a(\omega)$ of the input speech and $\hat{a}_L(\omega)$ of the model closest to $a(\omega)$.

3.1 Accuracy of analysis by adaptive filter

The log likelihood ratio is used as the measure of distance in the LPC space;

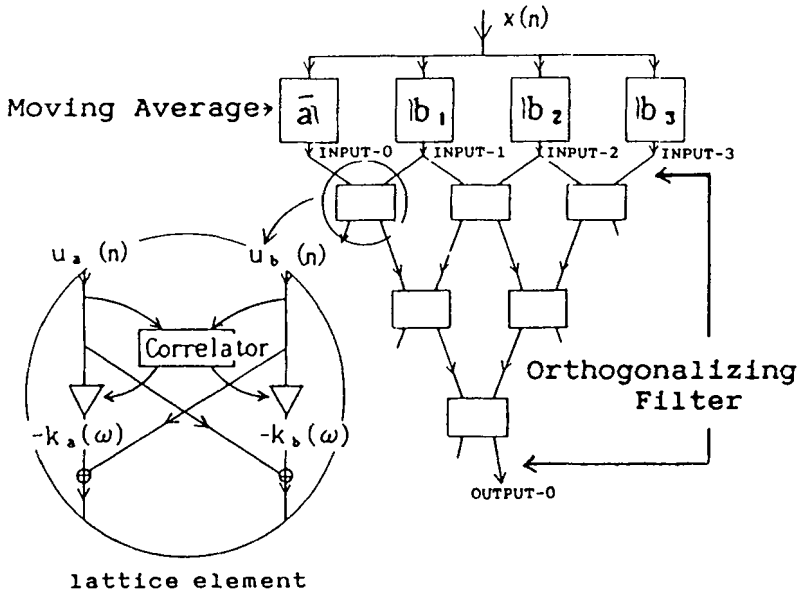


Fig. 1. Adaptive analysis based on the LPC deviation model.

$\hat{a}_L(\omega)$ minimizing the right-hand side of Eq. (20) can either be obtained by substituting the solution $c_l(\omega)$ of the system of Eqs. (17) into Eq. (19), or adaptively calculated by the filter of Fig. 1 [1]. Such a simple method of analysis results from the simplicity of the model of Eq. (19).

In this section we demonstrate the accuracy of the spectrum analysis by the adaptive filter.

The analysis procedure is shown in the following for the filter of Fig. 1 with the case of $L = 3$ as an example. Using $\bar{a}$ and b_l as the MA coefficients, the moving average is applied to the input speech $x(n)$, and the obtained signals are given to the orthogonalizing filter, which is the lower part of the filter of Fig. 1, as the inputs [6]. The parameters in the individual lattice elements are calculated from each input to the element by

$$\left.\begin{aligned} k_a(\omega) &= \frac{E_n[u_a(n)u_b(n)]}{E_n[u_a(n)^2]} \\ k_b(\omega) &= \frac{E_n[u_a(n)u_b(n)]}{E_n[u_b(n)^2]} \end{aligned}\right\} \quad (21)$$

where $u_a(n)$, $u_b(n)$ are shown in Fig. 1. The thus obtained parameters are fixed for the successive determination of $\hat{a}_3(\omega)$. The ith component of the vector $\hat{a}_3(\omega)$ is obtained from the terminal 'OUTPUT-0' in Fig. 1 by inputting the ith components of $\bar{a}$, b_1, $\cdots$, b_3 to the terminals 'INPUT-0', 'INPUT-1', $\cdots$, 'INPUT-3,' respectively.

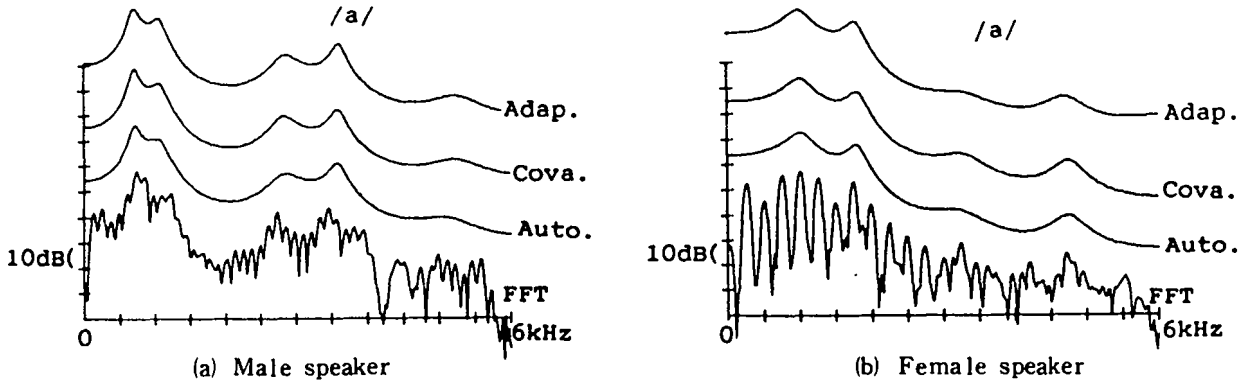
The time average $E_n[\ ]$ in Eq. (21) is adaptively calculated by Makhoul's three-pole window [7]. The parameter β in the window is set as $\beta = 0.975$ so that the width of the window is 20 ms [7].

To examine the accuracy of such an adaptive analysis, set $P = 12$ and $L = 12$ and considers the case where

$$\left.\begin{aligned} \bar{a} &= (1, 0, \cdots, \cdots, 0)^t \\ b_l &= (\underset{0}{0}, \cdots, 0, \underset{l}{1}, 0, \cdots, \underset{P}{0})^t \end{aligned}\right\} \quad (22)$$

Under these conditions, Fig. 2 shows the spectrum calculated from $\hat{a}_L(\omega)$ obtained by the above method. In this case the degree of freedom (L) is equal to the prediction order (P) as was described in [1], no constraint is imposed on the analysis and the analysis is equivalent to the lattice analysis of 12th order. Consequently, $\hat{a}_L(\omega) = a(\omega)$ is ensured as long as the filter operates in a stable way. The accuracy and the stability of the above method are checked by comparing the spectrum obtained by $\hat{a}_L(\omega)$ with the one obtained by $a(\omega)$.

For comparison, the spectra obtained by FFT, autocorrelation method of 12th order, and the covariance method, are shown in the figure with shifts. It is seen from these results that the adaptive analysis by the



(a) Male speaker (b) Female speaker

Adap.: Adaptive analysis Cova.: Covariance Method

Auto.: Autocorrelation Method FFT : Short time Fourier Spectrum

Fig. 2. Spectrum obtained by the adaptive analysis. Vowel /a/, 12 kHz sampling, 256 samples per frame.

filter of Fig. 1 has a sufficiently high accuracy. Now that the accuracy of the adaptive analysis is verified, the spectrum is calculated in the following experiment by solving the system of Eqs. (17), which is suited to the calculation by the computer.

3.2 Approximation of vowel variation by preceding consonant

From 100 CV syllables uttered by a speaker, the vowel part is extracted by observation. The model is constructed for each vowel category by the method described in Sect. 2. The ability of the model to approximate the vowel variation by the preceding consonant is examined by comparing the spectrum of LPC $\alpha(\omega)$ of the varied vowel part and the spectrum of the LPC $\hat{a}_L(\omega)$ which is closest to $\alpha(\omega)$ among the models.

The speech data used are 100 isolated vowels and CV syllables uttered by a male speaker. The data are sampled by 12 kHz with 12 bits. The number of /a/, /i/, /u/, /e/ and /o/ in 100 data are 26, 12, 24, 13 and 25, respectively. Using these speech data, the model is made for each vowel category. Then the experiment is performed for the data used to construct the model. The degree of prediction P is set as 12, and the frame length of the analysis is set as 256 samples.

To quantitatively evaluate the approximation ability of the model, the decrease of Δ_m in Eq. (7), i.e., accumulated log likelihood ratio, with the increase of m was examined. The result is shown in Fig. 3. For comparison, Δ_m for the case, where $\bar{a}, b_1, \cdots, b_L$ are determined by Eq. (22), i.e., the case of ordinary LPC analysis, is also shown. It is seen from the result that the error can be decreased efficiently by using the model of Eq. (19).

Based on this result, the degree of freedom of the model L is set as 6 in the experiment. Figure 4(b) shows the obtained spectrum. For comparison, the spectrum estimated by 12th-order autocorrelation method is shown in Fig. 4(a). The figures show five typical cases which have remarkable variations among the 100 pronunciations for each vowel. By comparing (a) and (b), it is seen that the model with the 6 degrees of freedom has a very high approximating ability.

The above result indicates the following. In the ordinary 12th-order LPC analysis, there are twelve free parameters for the representation of the spectrum. But for a particular vowel category of a particular speaker, six free parameters are sufficient for the spectrum representation by appropriately modeling the variation.

4. Recognition of Vowel with Variation Preceding Consonant

The spectrum, which is the result of the approximating experiment in section 3.2, indicated that the variation of the vowel of a particular speaker due to the preceding consonant can well be approximated by the model of Eq. (19) with 6 degrees of freedom. As the next step, the model is applied to the recognition of the vowel.

4.1 Result of recognition by model without restriction

First, the coefficient $c_l(\omega)$ representing the variation along the direction of b_l in Eq. (19) is permitted to take any large value, and the experiment is performed. In this case, as far as the variation of the input LPC from the standard LPC $\bar{a}$ is repre-

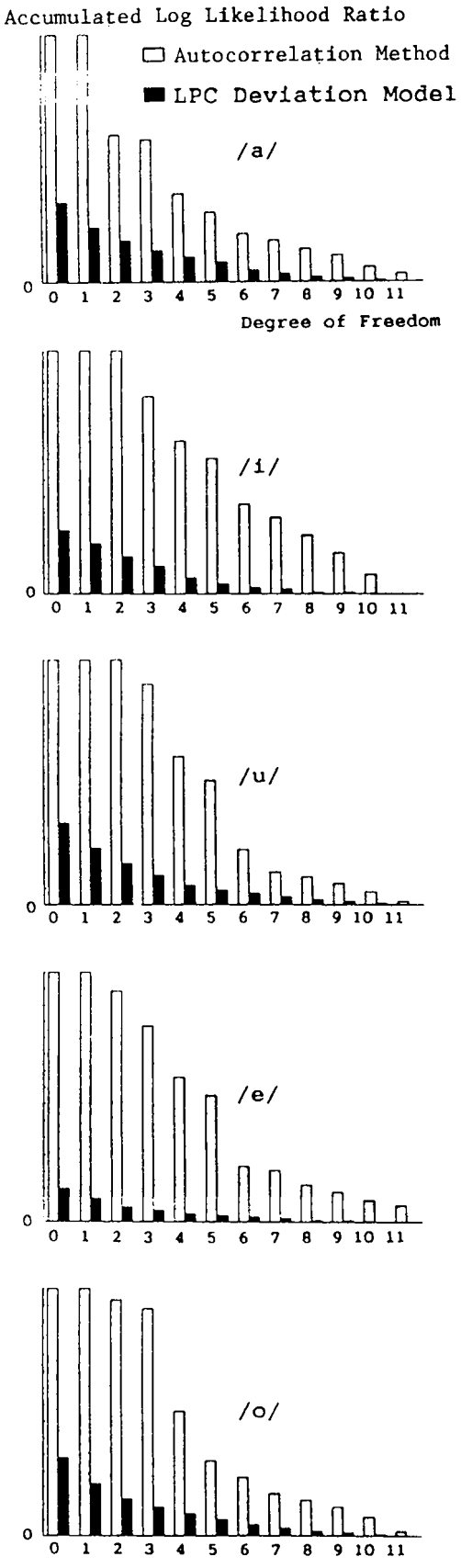


Fig. 3. Freedom of the model and residual (accumulated log likelihood ratio) in the approximation.

Table 1. Recognition rate for each speaker

Power method speaker	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M10	M11	Total
Residual	93.23	93.93	95.99	80.63	85.58	96.44	88.78	87.94	77.90	85.77	94.17	88.96
LPC variation	93.87	90.73	93.65	83.81	81.09	95.79	90.38	95.24	84.05	83.63	96.12	89.45
Restricted LPC model	95.48	96.49	97.32	85.71	88.78	98.06	92.31	94.29	86.23	88.61	97.41	92.54

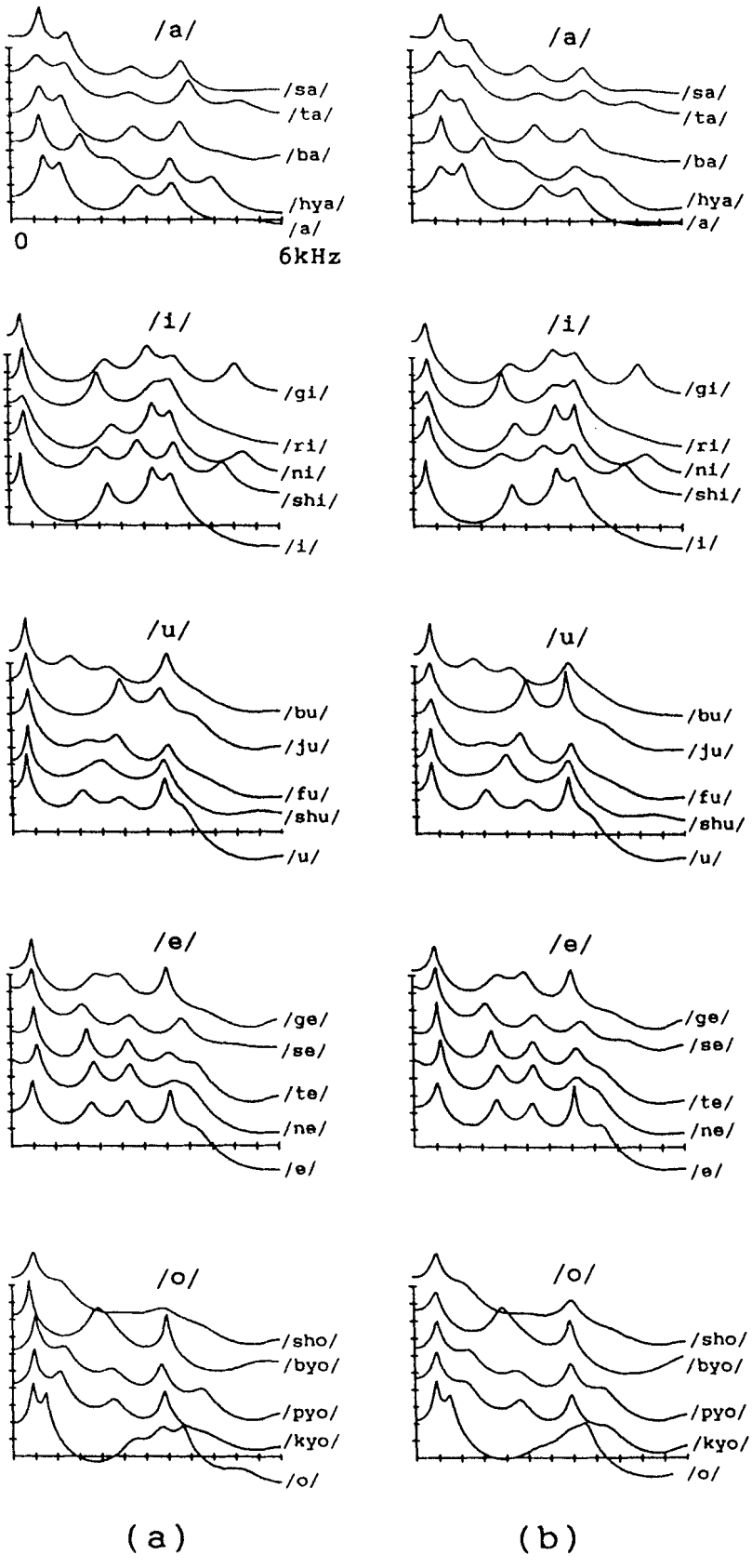


Fig. 4. Spectrum of the varied vowel part in CV syllable (a) and the approximation of it by the LPC deviation model (b).

Table 2. Recognition rate for each vowel

Vowel category	/a/	/i/	/u/	/e/	/o/	Total
Residual power method	88.59	96.29	82.70	93.51	89.15	88.96
LPC variation model	83.10	94.98	89.62	98.38	88.14	89.45
Restricted LPC variation model	88.91	96.72	91.81	98.38	91.64	92.54

Table 3. Distance (logarithmic likelihood ratio) between vowel model and input /myu/

utterance /myu/

Frame number	1	2	3
Model /a/	1.51	1.71	1.57
Model /i/	0.95	1.35	1.43
Model /u/	0.69	0.85	0.77
Model /e/	0.96	1.25	1.22
Model /o/	1.34	1.56	1.54
Result of recog.	/u/	/u/	/u/

(a) Residual power method

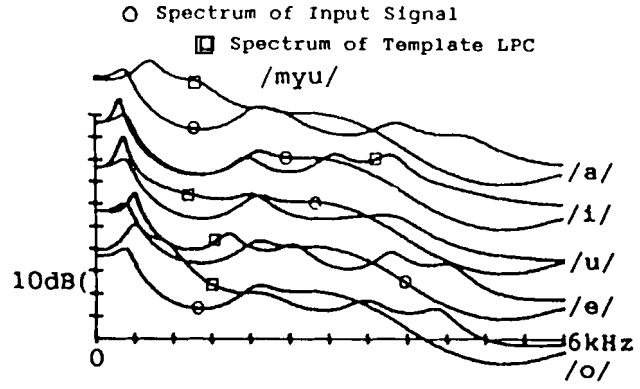
Frame number	1	2	3
Model /a/	0.77	0.80	0.80
Model /i/	0.20	0.17	0.21
Model /u/	0.06	0.08	0.05
Model /e/	0.21	0.19	0.14
Model /o/	0.67	0.50	0.48
Result of recog.	/u/	/u/	/u/

(b) LPC variation model 6 degrees of freedom

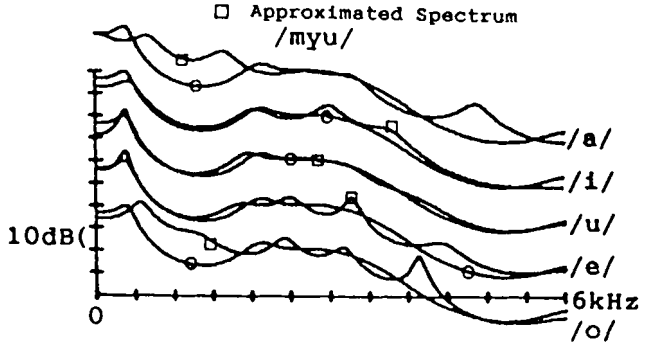
sented by a linear combination of $\{b_1, b_2, \cdots, b_L\}$, the model approximates that LPC, however large the variation.

The log likelihood ratio is used as the LPC distance measure, and Eq. (20) is used to define the distance between the model and the LPC of the input speech. The vowel category of the model minimizing the distance is adopted as the result of recognition. If the experiment is made only for a particular speaker, it may happen that the method applies well only to that speaker. To avoid this problem, 11 male speakers are employed, and the model was constructed for each of the speakers. The recognition experiment was made for the data used in the construction of the model.

The frame length for the analysis is set as 256 samples. By shifting the frame by 128 samples, $\alpha(\omega)$ and $\hat{\alpha}_6(\omega)$ are calculated, and the result of recognition is obtained. Thus the recognition result is obtained for each frame, where the recognition rate is defined as the ratio of the number of correctly recognized frames to the total number of frames. For some kinds of vowels, the utterance period is relatively short and only a small number of frames could be obtained. Consequently, it should be noted

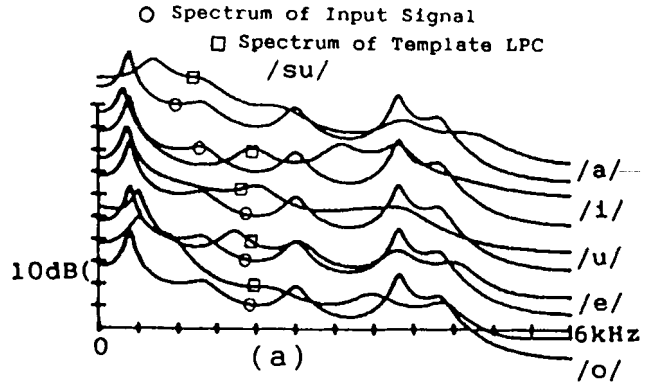


(a) Template LPC (Fixed reference LPC)

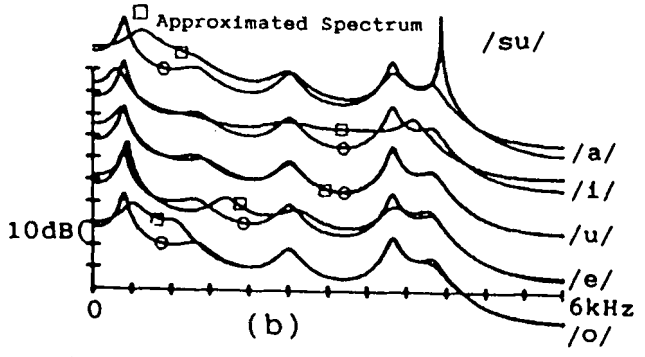


(b) LPC deviation model (Variable reference LPC)

Fig. 5. Spectrum when the both method success.



(a) Template LPC (Fixed reference LPC)



(b) LPC deviation model (Variable reference LPC)

Fig. 6. Spectrum when the fixed LPC gives a wrong result.

Table 4. Distance (logarithmic likelihood ratio) between vowel model and input /su/ utterance /su/

Frame number	1	2	3
Model /a/	0.94	1.09	1.01
Model /i/	1.06	1.29	1.17
Model /u/	0.66	0.76	0.68
Model /e/	0.45	0.61	0.54
Model /o/	1.01	1.25	1.14
Result of recog.	/e/	/e/	/e/

(a) Residual power method

Frame number	1	2	3
Model /a/	0.62	0.66	0.68
Model /i/	0.30	0.38	0.34
Model /u/	0.10	0.10	0.13
Model /e/	0.25	0.34	0.35
Model /o/	0.26	0.24	0.28
Result of recog.	/u/	/u/	/u/

(b) LPC variation model 6 degrees of freedom

that all utterances are not evaluated uniformly by the above definition of the recognition rate. For comparison, the result of the ordinary residual power method (i.e., recognition result by the model with 0 degree of freedom) using the LPC $\bar{\alpha}$ as the standard is also shown.

Table 1 is the recognition rate for all vowels for each speaker; M1 to M11 are the identification codes for the male speakers. In the table, the top row shows the rate by the ordinal residual power method and the middle row shows the rate by the modeling method. Depending on the speaker, the recognition rate is improved or degraded compared with the residual power method. In sum, a slight improvement is observed. Table 2 is the total recognition rate for each vowel. It is observed again that the recognition rate is sometimes degraded depending on the vowel.

4.2 Reason for misrecognition

The reason for the unsatisfactory recognition rate in Sect. 4.1 is discussed in the following based on the spectrum comparison by the method described in Sect. 3. As is shown in Fig. 4, the vowel variation model used in the experiment can approximate with a high accuracy the variation of the vowel spectrum due to the preceding consonant. The misrecognition is still produced despite this accuracy. The spectrum was analyzed for the vowel /u/ of the speaker M10, where the recognition rate was low.

Figure 5 shows the spectra for the case correctly recognized by both the residual power method and the vowel variation model (the pronunciation is /myu/). Figure 6 is the spectra for the case where the former is in error (the pronunciation is /su/) and Fig. 7 is the spectra for the case where the latter is in error (the pronunciation is /ju/).

In (b), the true spectrum of the input speech and the approximate spectra by each vowel model are superposed. The spectra are shown being shifted from each other for the vowels /a/, /i/, /u/, /e/ and /o/. The spectrum of the input speech is the one obtained from $\alpha(\omega)$ of the LPC by the 12th-order autocorrelation method. The approximate

Table 5. Distance (logarithmic likelihood ratio) between vowel and input /ju/

Frame number	1	2	3
Model /a/	1.37	1.40	1.52
Model /i/	1.02	1.37	1.40
Model /u/	0.50	0.50	0.48
Model /e/	0.85	0.91	0.98
Model /o/	1.07	1.13	1.03
Results of recog.	/u/	/u/	/u/

(a) Residual power method

Frame number	1	2	3
Model /a/	0.48	0.58	0.53
Model /i/	0.45	0.61	0.64
Model /u/	0.06	0.12	0.17
Model /e/	0.03	0.04	0.06
Model /o/	0.72	0.55	0.42
Results of recog.	/e/	/e/	/e/

(b) LPC variation model 6 degrees of freedom

spectra by the vowel variation model is obtained by the method shown in Sect. 3. For comparison, (a) shows the spectrum of the fixed template LPC. The spectrum obtained from the standard LPC $\underline{a}$ is shown for each vowel, being superposed with the true spectrum.

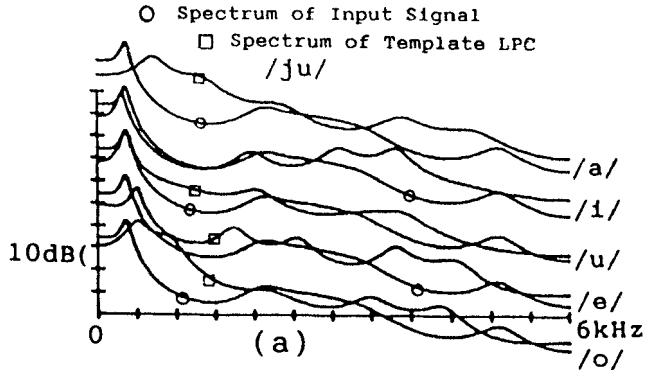
Tables 3, 4 and 5 (b) show the distance (log likelihood ratios) between the LPC's of 3 frames of the input waveforms and the corresponding vowel models. Tables 3, 4 and 5 show the distance to the standard LPC of each vowel category.

An immediate observation from those results is that the model for the vowel /u/ approximates well the spectrum of the input speech. Figure 5, for example, is the case where both methods give the correct result. In (a), however, it is difficult to tell by observation which of the standard LPC is closer to the input spectrum. On the other hand, in (b), it becomes clear by using the variation model that the model for /u/ approximates best the true spectrum.

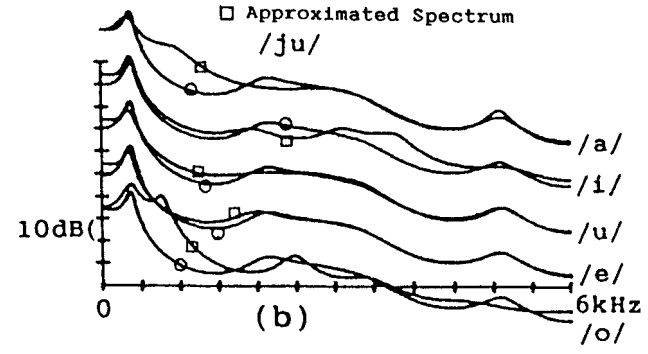
Figure 7 shows the case where the misrecognition is increased by using the model. In this case again, the model for /u/ approximates well the input spectrum. The reason for the misrecognition is not that the model is lacking in approximating ability, but that another vowel model (/e/) gives a better approximation. To circumvent such a problem, where the input is better approximated by other vowel models, restrict the range of values of $c_l(\omega)$.

4.3 Recognition by restricted allowance

The following method is considered to cope with the problem pointed out in the preceding section. The improvement of the recognition rate is then examined. Using each vowel model, the LPC of the signal bandwidth to that vowel category is approximated, and the distribution of $c_l(\omega)$ is examined. The threshold θ_l is set so that the most of the distribution is included in the range



(a) Template LPC (Fixed reference LPC)



(b) LPC deviation model (Variable reference LPC)

Fig. 7. Spectrum when the LPC deviation model gives a wrong result.

$$|c_l(\omega)| \leqq \theta_l. \tag{23}$$

Let the LPC closest to the LPC of the input speech among the model be

$$\hat{\boldsymbol{a}}_6(\omega) = \bar{\boldsymbol{a}} + \sum_{l=1}^{6} c_l(\omega)\, \boldsymbol{b}_l \tag{24}$$

where $c_l(\omega)$ is obtained by solving Eq. (17).

However, when the value of $c_l(\omega)$ above is extraordinarily large compared with the threshold θ the value is restricted as follows:

$$\tilde{\boldsymbol{a}}_6(\omega) = \bar{\boldsymbol{a}} + \sum_{l=1}^{6} \tilde{c}_l(\omega)\, \boldsymbol{b}_l \tag{25}$$

where $\tilde{c}_l(\omega)$ is determined as follows:

$$\tilde{c}_l(\omega) = \begin{cases} \theta_l & \theta_l < c_l(\omega) \\ c_l(\omega) & -\theta_l < c_l(\omega) \leqq \theta_l \\ -\theta_l & c_l(\omega) \leqq -\theta_l \end{cases} \tag{26}$$

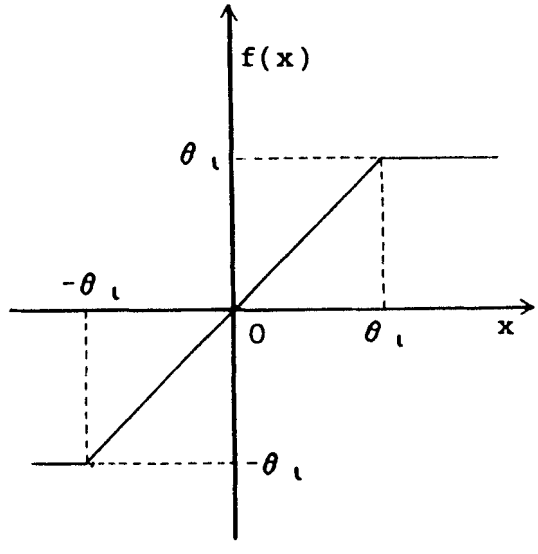


Fig. 8. Function for the restriction of free parameters in the model.

This can also be represented as follows using the function $f(x)$ of Fig. 8:

$$\tilde{c}_l(\omega) = f(c_l(\omega)) \tag{27}$$

Using $\tilde{a}_6(\omega)$ whose parameter range is restricted as above, the distance between the model and the input LPC is calculated as follows:

$$\log \frac{\tilde{\boldsymbol{a}}_6(\omega)^t R(\omega) \tilde{\boldsymbol{a}}_6(\omega)}{\boldsymbol{a}(\omega)^t R(\omega) \boldsymbol{a}(\omega)} \tag{28}$$

The forementioned variable is calculated for each vowel model and the vowel category corresponding to the model exhibiting the minimum distance is defined as the result of recognition.

The bottom row of Tables 1 and 2 shows the result of recognition by this method; θ_l is determined intuitively as follows from the histogram of $c_l(\omega)$:

$$\left.\begin{aligned} &\theta_1=\theta_2=0.6,\ \theta_3=0.5,\ \theta_4=0.4 \\ &\theta_5=0.3,\quad \theta_6=0.2 \end{aligned}\right\} \tag{29}$$

Other conditions of experiment are the same as in Sect. 4.1. It is seen from the result that the recognition rate is improved for all speakers and for all vowels.

5. Conclusions

The method proposed in this paper is advantageous in that the model is simple and the processing is theoretically clear. In this paper, the experiment was made for the most simple case since the aim is to indicate the characteristics of the basic principle. The following points were observed.

(i) By a simple model of a linear manifold, the vowel spectrum variation within a particular speaker can well be approximated. By the experiment, a model with six parameters provided a nearly equal spectrum as the 12th-order LPC analysis.

(ii) Even though the model can well approximate the vowel variation, it cannot be applied directly to the recognition. A problem is that the input vowel is better approximated by another vowel model.

(iii) Considering the distribution of the parameters and limiting the range of parameters in the model, the recognition rate can be improved compared with the case where the fixed template is used.

The following points should be considered in the future.

(i) In the recognition procedure in this paper, the LPC closest to the input is determined for each vowel model independently. Consequently, the models of other categories, as well as input category, also approach the input LPC, which is a reason for the misrecognition. This problem will be improved by utilizing the relative relations among vowels [8].

(ii) In this experiment, the model is constructed for each speaker, but it will be required to realize the recognition method for the unspecified speaker, using the technique of learning.

(iii) In the experiment of Sect. 4.3, the threshold θ_l is determined intuitively. The threshold should be generalized using the function $f(x)$ of Fig. 8, and the automatic determination of the function $f(x)$ should be established.

This work was supported by a Sci. Grant, Min. Education (no. 59460112).

Acknowledgement. The authors acknowledge the assistance of the Toshiba Co. in providing the experimental speech data. They thank Assoc. Prof. H. Ogawa, Dr. M. Sato, Tokyo Inst. Tech., and Dr. M. Akagi, NTT Basic Res. Lab., for their advice, as well as Mr. K. Takahashi, student in master's program and Mr. Y. Kato, senior student, in this Lab., for their assistance.

REFERENCES

1. I. Kumazawa and T. Iijima. Linear prediction analysis based on statistical model of variation of linear prediction coefficient, Trans. (A) I.E.C.E., Japan, J69-A, 2, pp. 224-231 (Feb. 1986).
2. M. Akagi and T. Iijima. Speech recognition by polarity error discrimination --POLPEC method, Trans. (A) I.E.C.E., Japan, J65-A, 8, pp. 759-766 (Aug. 1982).
3. M. Akagi and T. Iijima. Recognition of Japanese phonemes by polarity error discrimination method--POLPEC method II, Trans. (A) I.E.C.E., Japan, J67-A, 5, pp. 439-446 (May 1984).
4. M. Akagi and T. Iijima. Recognition of Japanese burst by polarity error discrimination method, Trans. (A), I.E.C.E., Japan, J67-A, 11, pp. 1013-1019 (Nov. 1984).
5. F. Itakura. Minimum prediction residual principle applied to speech recognition, I.E.E.E. Trans. Acoust., Speech, and Signal Process, ASSP-23, 1, pp. 67-72 (Feb. 1975).
6. I. Kumazawa and T. Iijima. A design technique of adaptive filter using lattice elements as adaptation module,

Trans. (A), I.E.C.E., Japan, 68-A, 12, pp. 1341-1349 (Dec. 1985).
7. J.I. Makhoul and L.K. Cosell. Adaptive lattice analysis of speech, I.E.E.E. Trans. Acoust. Speech, and Signal Process. ASSP-29, 3, pp. 654-659 (June 1981).
8. Sugiyama and Kano. Learning without teacher of vowel standard pattern, Tech. Rep. Speech, Acoust. Soc. Jap., S83-48 (Dec. 1983).

AUTHORS (from left to right)

Itsuo Kumazawa graduated 1981 Dept. Electrical and Electronics Eng., Fac. Eng., Tokyo Inst. Tech. Completed Master's program, 1983 Grad. School. Presently, student in doctoral proram. Engaged in research in signal processing and pattern recognition.

Taizo Iijima graduated 1948 Dept. Electrical Eng., Tokyo Inst. Tech., and affiliated with Electrotechnical Lab. Engaged in researches in electromagnetic field analysis and pattern recognition and development of OCR. Professor 1972, Tokyo Inst. Tech. Editorial secretary, secretary of planning and secretary for general affairs, I.E.C.E., Japan. Chairman, Study Group for Pattern Recognition. Contributions Award 1976. Paper award (4 times) and Author Award. Doctor of Eng.