Spectral estimation of speech corrupted by colored noise

Estimation du spectre du signal de parole en présence d'un bruit de fond coloré

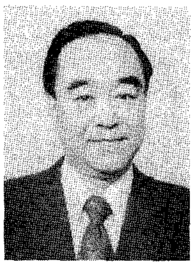
Hiroyoshi MORIKAWA

Department of Electronic Engineering, the Faculty of Engineering, the University of Tokyo, Bunkyo-ku, TOKYO, 113 JAPAN.

Hiroyoshi Morikawa was born in Japan on February 7, 1949. He received the BS degree in management engineering from Osaka Electrocommunications University in 1971, and the MS degree in radiocommunications from University of Electrocommunications, Tokyo, in 1973.

Since 1973 he has been with the Faculty of Engineering, the University of Tokyo, as a Research Associate. His research interests range from stochastic system modeling and system identification to speech processing and application of adaptive filtering to communication systems.

Mr. Morikawa is a senior member of the Institute of Electrical and Electronics Engineers, INC, and a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Acoustical Society of Japan.



Hiroya FUJISAKI

Department of Electronic Engineering, the Faculty of Engineering, the University of Tokyo, Bunkyo-ku, TOKYO, 113 JAPAN.

Hiroya Fujisaki was born in Japan on October 18, 1930. He received the BS, MS, and Dr. Eng. degrees in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1954, 1956, and 1962, respectively.

In 1962, he joined the Faculty of Engineering of the University of Tokyo as Assistant Professor. He has held the position of Professor of Electronic Engineering since 1973. He has been engaged research on speech communication and information processing and published a number of technical papers and several books.

Dr. Fujisaki is Vice-President of the Acoustical Society of Japan. He has been Chairman of the Tokyo Chapter of the IEEE Acoustics, Speech, and Signal Processing Society. He is a fellow of the Acoustical Society of America, and a member of the Institute of Electronics, Information and Communication Engineers of Japan, the Information Processing Society of Japan, the Linguistic Society of Japan, Tau Beta Pi, and Sigma Xi.

SUMMARY

A modified SEARMA method is proposed for estimating the speech spectrum in the presence of colored background noise. The following assumptions are used in developing the analysis. The speech production process is represented by an autoregressive moving-average (ARMA) model. The background noise is represented by an MA process. The noise process is locally stationary during speech activity. Following these assumptions, the process during speech activity can be represented by an extended ARMA model. In this formulation, unique estimation of AR parameters of the vocal tract transfer function is always possible if the MA parameters of the noise process can be estimated separately, but the estimation of MA parameters of the speech production process requires further assumption of a high SNR. The validity of the proposed method is demonstrated by spectral estimation of both synthetic and natural speech sounds in the presence of additive colored noise, and by comparing the results with those obtained by the LPC method.

KEY WORDS

Speech analysis, parameter estimation, spectral analysis, ARMA model, modified SEARMA method.

RÉSUMÉ

Une version modifiée de l'algorithme SEARMA est proposée pour l'estimation du spectre du signal de parole en présence d'un bruit de fond coloré. Les hypothèses suivantes sont utilisées pour la mise en œuvre de l'algorithme. Le processus de génération du signal de parole est modélisé par un modèle ARMA. Le bruit est supposé localement stationnaire devant la production du signal de parole. Compte tenu de ces hypothèses le signal de parole peut alors être représenté par un modèle ARMA étendu. Dans cette formation, une estimation unique des paramètres AR de la fonction de transfert du conduit vocal peut être effectuée séparément si les paramètres MA du modèle du bruit peuvent être estimés séparément, mais l'estimation des paramètres MA du modèle de signal de parole nécessite l'hypothèse supplémentaire d'un rapport signal sur bruit élevé. La validité de la méthode proposée est illustrée par l'estimation spectrale à la fois de signaux de parole synthétique et naturelle en présence de bruit additif coloré et en comparant les résultats avec ceux obtenus par la technique LPC.

MOTS CLÉS

Analyse parole, estimation paramètres, analyse spectrale, modèle ARMA, méthode SEARMA modifiée..

1. Introduction

In practical speech transmission systems, the speech signal is always accompanied by some noise. The performance from a voice system is dependent mainly on the estimation accuracy of the speech spectrum. Therefore, accurate estimation of speech spectrum in the presence of background noise is an important problem in speech communication.

The autoregressive power spectral density estimate, which is known as the linear prediction (LPC) spectral estimate in speech research (e. g. [1, 10]), has been shown to possess excellent performance for noise-free input speech. The addition of white noise to an AR process, however, results in an incorrect power spectral density estimate (e. g. [3, 20]). The reason for the degradation of LPC analysis in the presence of noise is that the diagonal elements of the Toeplitz matrix are modified by the variance of the additive noise [4, 14, 16]. Therefore, the lower the signal-to-noise ratio (SNR), the poorer the spectral estimate is obtained from LPC analysis. One previous approach to enhance speech spectrum in the presence of white noise is based on an AR model of speech [4] while another approach is based on an ARMA model [14, 16]. Since the overall observation process does not include spectral information of the background noise, however, neither of these approaches are applicable to the estimation of the speech spectrum in the presence of colored noise.

In this paper we propose a method for the accurate estimation of speech spectrum in the presence of colored background noise. The estimation is accomplished in two stages. The first stage is the detection of the speech activity interval. The second stage is the estimation of the vocal tract transfer function in the extended ARMA model. For the first stage, we can use the previous methods (e. g. [2, 6, 18]. The validity of the proposed method for the second stage is demonstrated by spectral estimation of both synthetic and natural speech sounds in the presence of additive colored noise, and by comparing the results with those obtained by the LPC method.

2. Statement of the problem

The block diagram of Figure 1 summarizes the whole procedure [17]. At the first step, a decision is made on the presence/absence of speech activity within the current frame of about 30 ms duration, which is the standard length of one analysis frame in speech analysis. In the case of a decision for "silence" where speech is absent but the background noise is present, we estimate noise spectrum. In the case of a decision for "speech" where speech activity and background noise are both present, we estimate the speech spectrum using the noise spectrum estimated from the "silence" intervals. In this paper, we assume that the speech activity interval is detected by the previous methods [6, 18].

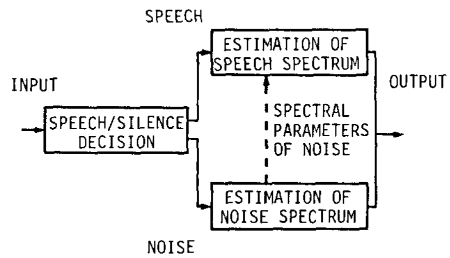


Fig. 1. — Schematic diagram of spectral estimation of speech corrupted by colored noise.

A. MODEL OF ADDITIVE NOISE

The following assumptions are used with regard to the background noise. The noise process v_i is represented by an MA process

$$v_i = w_i + \gamma_1 w_{i-1} + \ldots + \gamma_l w_{i-l} \quad (1)$$

and is always present regardless of presence or absence of the speech activity. The noise process is locally stationary during speech activity. Hence, the values of γ_i's are constant and w_i's are represented by a zero-mean white Gaussian noise with constant variance R during speech activity. Even if the background noise process changes to a new state, there exists enough time (more than 300 ms) to estimate new noise spectral parameters. The background noise is added to the speech signal s_i, and the observed process can be expressed as:

$$y_i = s_i + v_i. \tag{2}$$

In all practical situations, conversational speech is accompanied by silent intervals. Some of background noise such as room noise, traffic noise and speech-like noise may be approximated by an MA process. Therefore, above assumptions with regard to the background noise have a generality for the practical noise environments.

B. MODEL OF THE SPEECH PRODUCTION PROCESS

A model of the vocal tract can be constructed by representing it as a discrete time-varying linear system. The entire process of conversion from the source to the speech signal can be somewhat idealized and represented by the response of a linear system whose input is either a train of volume velocity impulses or random pressure fluctuations, if we include all other characteristics into the system characteristics. The process of speech production is represented by an ARMA process that converts the source signal u_i into the speech signal s_i [13]:

$$s_i + \alpha_1 s_{i-1} + \ldots + \alpha_n s_{i-n} = u_i + \beta_1 u_{i-1} + \ldots + \beta_m u_{i-m}. \tag{3}$$

The order n of the AR scheme and the order m of the MA scheme vary with the vocal tract configuration and the source location, and m is smaller than n. It is assumed that the source signal u_i is either a zero-mean white Gaussian noise or a train of randomly spaced impulses with variance Q. The transfer function of (3) can be represented by a z-transformation:

$$G(z) = \frac{1 + \beta_1 z^{-1} + \ldots + \beta_m z^{-m}}{1 + \alpha_1 z^{-1} + \ldots + \alpha_n z^{-n}}. \tag{4}$$

Corresponding to the transfer function $G(z)$ of the vocal tract, the power spectrum $f(\omega)$ at the angular frequency ω is given by

$$f(\omega) = \frac{Q}{2\pi} \cdot \frac{B_0 + 2B_1 \cos\omega + \ldots + 2B_m \cos m\omega}{A_0 + 2A_1 \cos\omega + \ldots + 2A_n \cos n\omega}, \tag{5}$$

where

$$A_i = \sum_{j=0}^{n-i} \alpha_j \alpha_{i+j}, \qquad \alpha_0 = 1,$$

$$B_i = \sum_{j=0}^{m-i} \beta_j \beta_{i+j}, \qquad \beta_0 = 1.$$

Therefore, the power spectrum of the speech signal can be obtained from the AR and MA parameters.

3. Estimation of spectral parameters of speech

After detecting the onset of speech activity, the algorithm estimates the speech spectrum by using the estimated MA parameters of the noise process. From (1) and (2), we have

$$y_i = s_i + w_i + \gamma_1 w_{i-1} + \ldots + \gamma_l w_{i-l}. \tag{6}$$

Substituting (6) into (3) and rearranging with respect to the regression of y_i, we have

$$\sum_{j=0}^{n} \alpha_j y_{i-j} = \sum_{j=0}^{m} \beta_j u_{i-j} + \sum_{j=0}^{n+l} \delta_j w_{i-j}, \tag{7}$$

where

$$\delta_j = \sum_{k=0}^{j} \alpha_k \gamma_{j-k}, \qquad \alpha_0 = \beta_0 = \gamma_0 = 1,$$

$$\alpha_k = 0 \quad (k > n), \qquad \gamma_k = 0 \quad (k > l).$$

The whole model of the observation process is represented as an extended ARMA model with two sources. In this formulation, unique estimation of AR parameters of the vocal tract transfer function is always possible if the MA parameters of the noise process can be estimated separately, but the estimation of MA parameters of the speech production process requires further assumption of a high SNR. From a practical point of view, an accurate estimation of the AR parameters of the vocal tract transfer function is important at a low SNR condition. We have shown that using zeros in the spectral model does not improve the spectral estimation at low SNR conditions (e. g., less than 5 dB in the case of white noise) [14, 16].

We have proposed a method for simultaneous estimation of ARMA parameters (i. e., the SEARMA Method) which estimates the orders as well as the parameter values of the transfer function of the vocal tract [12, 13]. The SEARMA method is capable of controlling the location and the length of the analysis interval to adapt to speech sounds with rapid spectral changes and a high fundamental frequency [15]. Similar methods for estimating ARMA parameters in an ARMA process have been developed (e. g. [7, 5, 8, 9, 19, 11]. However, these methods cannot be directly applied to the spectral estimation of speech corrupted by colored noise. We propose two algorithms for estimating the vocal tract transfer function. The Algorithm 1 estimates only AR parameters of the vocal tract transfer function under a low SNR condition, while the Algorithm 2 estimates both AR and MA parameters of the vocal tract transfer function under a high SNR condition. We also propose a method for selecting one of these algorithms to adapt the changing of SNR of the observed speech signal.

A. ALGORITHM 1

Under a low SNR condition, we may assume that

$$\sum_{j=0}^{m} \beta_j u_{i-j} \ll \sum_{j=0}^{n+l} \delta_j w_{i-j}. \tag{8}$$

We then assume that the MA scheme of the speech production process has a flat spectrum and is replaced by a single compensation factor. Thus a value of ξ

exists such that it satisfies the following relation:

$$(9) \qquad \sum_{j=0}^{m} \beta_j u_{i-j} \approx \xi w_{i-1}.$$

Substituting (9) into (7), (7) reduces to

$$(10) \qquad y_i = Y_{i-1}^T \Phi + W_{i-1}^T \Omega + w_i,$$

where

$$Y_{i-1}^T = [y_{i-n}, \ldots, y_{i-1}, w_{i-1}],$$
$$\Phi^T = [-\alpha_n, \ldots, -\alpha_1, \xi],$$
$$W_{i-1}^T = [w_{i-n-l}, \ldots, w_{i-1}],$$
$$\Omega^T = [\delta_{n+l}, \ldots, \delta_1].$$

The problem is to obtain an estimate $\hat{\Phi}$ for Φ which minimizes the following performance index:

$$(11) \quad I_k(\Phi) = \sum_{i=1}^{k} (y_i - Y_{i-1}^T \Phi - W_{i-1}^T \Omega)^2 + \|\Phi - \Phi_0\|^2 P_0^{-1}.$$

where Φ_0 is an *a priori* estimate of Φ, P_0 is the initial value of the symmetric and positive weighting matrix P_k. For a given set of n, the parameter values can be estimated by setting $\partial I_k(\Phi)/\partial\Phi$ equal to zero and by solving for Φ, and then replacing Y_k by $\hat{Y}_k$.

$$(12) \qquad \hat{\Phi}_{k+1} = \hat{\Phi}_k + K_k (y_{k+1} - \hat{Y}_k \hat{\Phi}_k)$$

$$(13) \qquad K_k = P_k \hat{Y}_k (\hat{Y}_k^T P_k \hat{Y}_k + \hat{R})^{-1}$$

$$(14) \quad P_k = P_{k-1} - P_{k-1} \hat{Y}_{k-1} \times (\hat{Y}_{k-1}^T P_{k-1} \hat{Y}_{k-1} + \hat{R})^{-1} \hat{Y}_{k-1}^T P_{k-1},$$

where

$$\hat{Y}_{k-1}^T = [y_{k-n}, \ldots, y_{k-1}, \hat{w}_{k-1}],$$
$$\hat{\Phi}^T = [-\hat{\alpha}_n, \ldots, -\hat{\alpha}_1, \hat{\xi}],$$
$$\hat{W}_{k-1}^T = [\hat{w}_{k-n-l}, \ldots, \hat{w}_{k-1}],$$
$$\hat{\Omega}^T = [\hat{\delta}_{n+l}, \ldots, \hat{\delta}_1],$$
$$\hat{w}_k = y_k - \hat{Y}_{k-1}^T \hat{\Phi}_k - \hat{W}_{k-1}^T \hat{\Omega}_k,$$
$$\hat{\delta}_j = \sum_{k=0}^{j} \hat{\alpha}_k \hat{\gamma}_{j-k}.$$

The initial values Φ_0 and P_0 for the formulas above are arbitrary as long as P_0 is positive definite. $\hat{R}$ and $\hat{\gamma}$ are obtained from the analysis of background noise in the silence interval. We initiate the computation process after the first $n+1$ measurements by assigning the initial values to Φ_k and P_k (e. g., $\Phi_0 = 0$ or estimate at the immediately preceding frame, and $P_0 = 100\,I$). Starting from $k = n+1$, the above formulas can be applied recursively to renew the estimate $\hat{\Phi}_k$ at every sampling point. The convergence of estimation is detected by the same procedure of the SEARMA method and the estimated parameters are obtained at the end of an analysis time window including N samples. The orders of the noise process and of the speech production process are estimated by the previously proposed method which minimize the variance of residuals.

B. Algorithm 2

Under a high SNR condition, we can write

$$(15) \qquad \sum_{j=0}^{m} \beta_j u_{i-j} \gg \sum_{j=0}^{n+l} \delta_j w_{i-j}.$$

Then, we have the same formulas as described in (12)-(14), where

$$\hat{Y}_{k-1}^T = [y_{k-n}, \ldots, y_{k-1}, \hat{u}_{k-m}, \ldots, \hat{u}_{k-1}],$$
$$\hat{u}_k = y_k - \hat{Y}_{k-1}^T \hat{\Phi}_k,$$
$$\hat{\Phi}^T = [-\hat{\alpha}_n, \ldots, -\hat{\alpha}_1, \hat{\beta}_m, \ldots, \hat{\beta}_1].$$

The difference between the Algorithm 2 and the SEARMA method is only $\hat{R}$ in (13) and (14).

C. Selection of Algorithm 1 and Algorithm 2

During the analysis of the first frame of speech after detecting the onset of speech activity, the selection of the Algorithm 1 and the Algorithm 2 can be performed by the following procedure:

(16) If SNR $<\theta$, then the algorithm 1 is selected.
$\geqq\theta$, then the algorithm 2 is selected.

In (16), we approximate SNR by

$$\mathrm{SNR} = 10\log_{10}\left(\frac{O-R}{R}\right)$$

where O denotes the variance of the observed speech. In the next frame, Φ is estimated by the algorithm which is selected in the previous frame. The Algorithm (16) is adopted at every analysis frame if the SNR condition of the observed speech is time-variable. The threshold value θ is an experimental value which is determined from the comparative evaluation of the spectral estimation accuracy of the Algorithm 1 and the Algorithm 2 at various SNR conditions. A suitable value of θ is 20 dB. We will give a simulation to determine the threshold value θ in Section 4.

We call the method including Algorithm 1, Algorithm 2 and selection of these algorithms as the Modified SEARMA method (the MSEARMA method).

4. Experimental results

In order to assess the validity of the theory, the speech analysis system described above was tested by using both synthetic and natural speech sounds sampled at 10 kHz with a quantization accuracy of 11 bits. The parameters of the speech production process were then determined for each analysis frame according to the procedure described in above section.

To quantitatively evaluate the estimation accuracy of the speech spectrum obtained by the proposed method, we use the following spectral distortion measures:

$$(17)\quad D_{AR}=\sqrt{\frac{1}{L}\sum_{i=1}^{L}(10\log_{10} f(\omega_i)-10\log_{10}\hat{f}_{AR}(\omega_i))^2}$$

and

$$(18)\quad D_{ARMA}=\sqrt{\frac{1}{L}\sum_{i=1}^{L}(10\log_{10} f(\omega_i)-10\log_{10}\hat{f}_{ARMA}(\omega_i))^2}$$

where $\hat{f}_{AR}(\omega_i)$ and $\hat{f}_{ARMA}(\omega_i)$ denote estimated spectral densities obtained by the AR parameters and ARMA parameters, respectively, and the frequency range corresponds to half of the sampling frequency (i. e., 10 kHz) is divided into L (i. e., 80) equal frequency portions. The spectral distortion measure D_{AR} is used to evaluate the spectral estimation accuracy for the Algorithm 1 and the LPC method, while the spectral distortion measure D_{ARMA} is used to evaluate that for the Algorithm 2.

The parameters of the noise process are shown in Table I. The estimated parameters in Table I are given by the Algorithm 2 with $n=0$ and $\hat{R}=1$. These estimated parameters and the estimated variance $\hat{R}$ of the noise are then used in the estimation of speech spectrum.

TABLE I

Parameters of additive noise.

	γ_1	γ_2	γ_3
Actual value	1.654	0.882	0.152
Estimated value	1.361	0.480	0.137

A. EVALUATION OF SPECTRAL ESTIMATION FOR SYNTHETIC SPEECH

We show here the results of the spectral estimation for the synthetic speech sounds $|i|$, $|m|$ and $|\int|$. The parameters of the synthetic speech sounds are shown in Table II.

Figure 2 shows the estimated spectral envelope obtained by the Algorithm 1 for the synthetic vowel $|i|$ with an SNR varying from 0 to 30 dB. The estimated speech spectrum is slightly degraded by the additive noise at every SNR condition as shown in Figure 2.

Figure 3 shows the comparison of the Algorithm 1 and the LPC method in the spectral estimation. In the LPC analysis, the order n is set at 11 which is equal to the order 8 of the speech process plus the order 3 of the noise process. The estimation accuracy of the Algorithm 1 is superior to that of the LPC method at all SNR conditions. The Algorithm 1 with

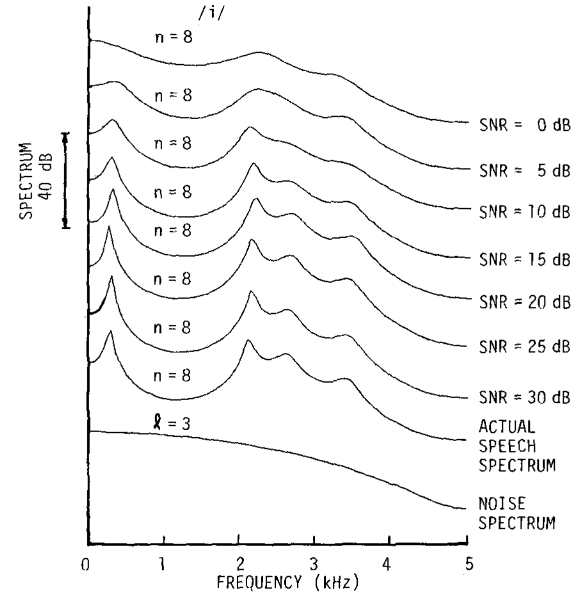


Fig. 2. — Spectral estimation of synthetic vowel $|i|$ at various SNR conditions.

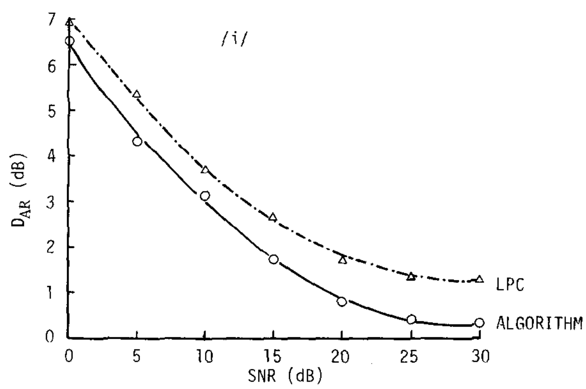


Fig. 3. — Comparison of the Algorithm 1 with $n=8$ (○) and the LPC method with $n=11$ (△) in the values of spectral distortion, for the synthetic vowel $|i|$ at various SNR conditions.

lower order than that of the LPC method yields more accurate estimated spectrum in the analysis of speech sound.

As example of speech sounds in which the role of zeros are more conspicuous, Figure 4 shows the estimated spectral envelope for the synthetic nasal $|m|$. In this figure, the spectral envelopes at SNR over 20 dB are obtained by the Algorithm 2 and the others are obtained by Algorithm 1. Figure 5 shows the comparison of the MSEARMA method and the LPC method in the spectral estimation. The difference between D_{ARMA}(●) and D_{AR}(○) indicates the contribution of the MA scheme in the ARMA model to

TABLE II

Parameters of synthetic speech sounds.

	First pole frequency	Second pole frequency	Third pole frequency	Fourth pole frequency	First zero frequency	Second zero frequency
$\|i\|$	290	2,130	2,640	3,450	–	–
$\|m\|$	220	1,050	2,380	4,100	1,600	3,320
$\|\int\|$	2,680	3,970	–	–	1,250	–

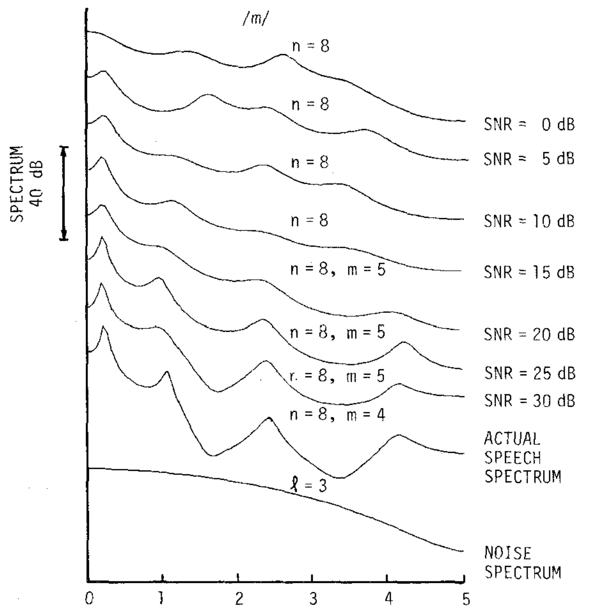


Fig. 4. — Spectral estimation of synthetic nasal |m| at various SNR conditions. The spectral envelopes at SNR over 20 dB are obtained by the Algorithm 2 and the others are obtained by the Algorithm 1.

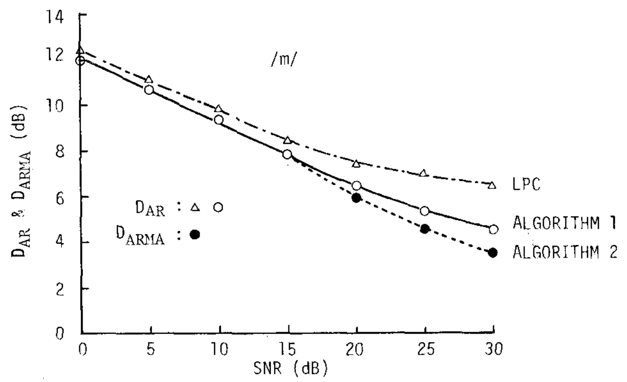


Fig. 5. — Comparison of the MSEARMA method with $n=8$(○, ●) and the LPC method with $n=11$(△) in the values of spectral distortion, for the synthetic nasal |m| at various SNR conditions.

the modeling of the spectrum of noisy speech. This difference vanishes at an SNR below 15 dB as shown in Figure 5. This fact suggests that using zeros in the spectral model does not improve the spectral estimation at low SNR conditions. However, the AR scheme in the ARMA model approximates the spectrum of speech with less distortion than the AR model of the LPC analysis.

Figure 6 shows, as well as Figure 4, a result of spectral estimation for the synthetic voiceless fricative |ʃ| with an SNR varying from 0 to 30 dB. In this figure, the spectral envelopes at SNR over 20 dB are obtained by the Algorithm 2 and the others are obtained by the Algorithm 1. The difference between D_{ARMA}(●) and D_{AR}(○) in Figure 7, as well as Figure 5, vanishes at an SNR of 15 dB. Therefore, the threshold value of $\theta = 20$ dB for selecting one of the Algorithm 1 and the Algorithm 2 is suitable. The threshold setting is not so critical since the difference between D_{ARMA}(●) and D_{AR}(○) is always less than 0.5 dB at SNR from 10 to 20 dB.

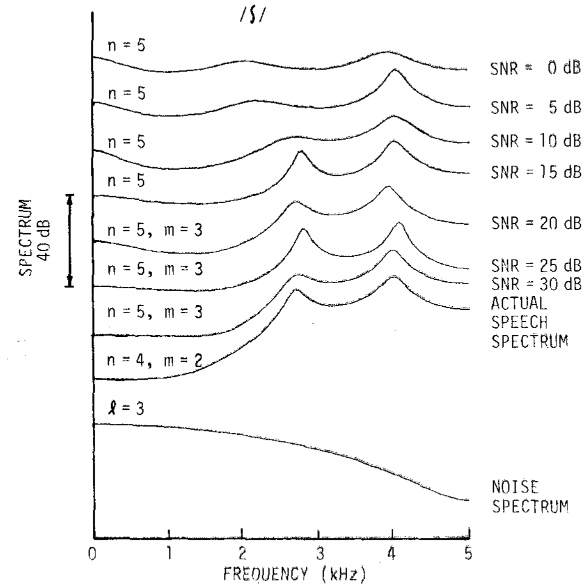


Fig. 6. — Spectral estimation of synthetic voiceless fricative |ʃ| at various SNR conditions. The spectral envelopes at SNR over 20 dB are obtained by the Algorithm 2 and the others are obtained by the Algorithm 1.

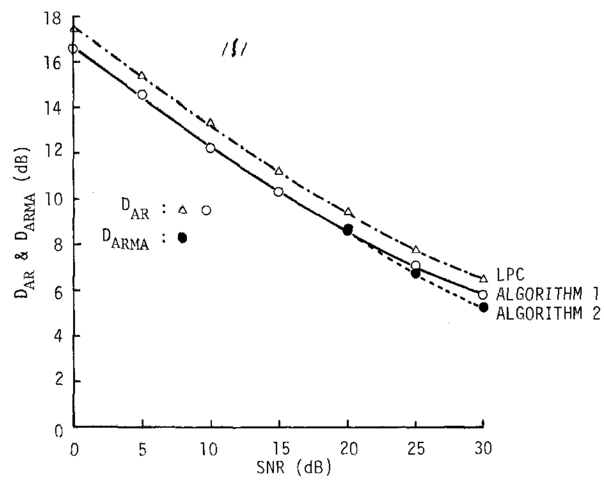


Fig. 7. — Comparison of the MSEARMA method with $n=5$(○, ●) and the LPC method with $n=7$(△) in the values of spectral distortion for the synthetic voiceless fricative |ʃ| at various SNR conditions.

B. Evaluation of spectral estimation for natural speech

As a last example, we show a result of running spectral analysis using natural connected speech uttered by a male speaker. Figure 8 illustrate the result of estimated running spectra for the Japanese utterance "korekara" (...from now to...) at a SNR of 10 dB. As seen from Figure 8, the running spectra of the natural connected speech are successfully analyzed.

As examplified using both synthetic and natural speech sounds, the proposed method has an effect of the enhancement of speech spectrum corrupted by the colored noise.

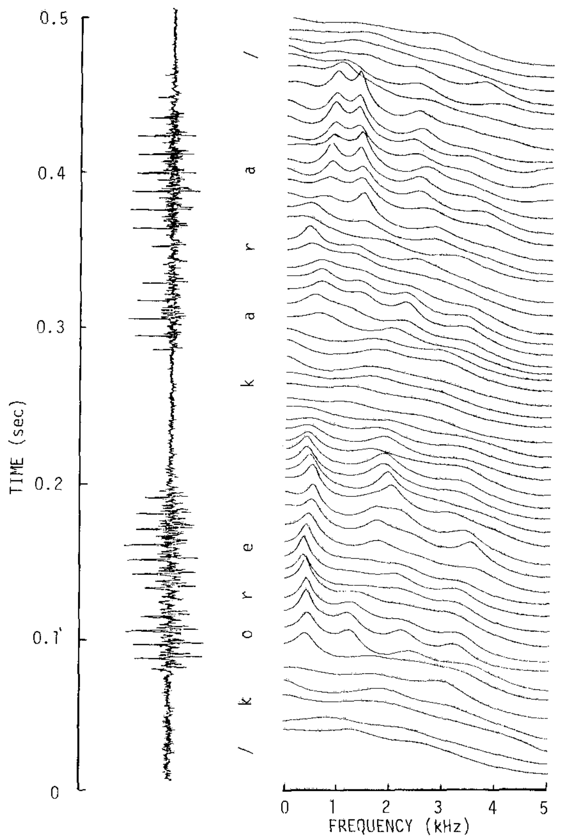


Fig. 8. — Estimation of running spectrum of the natural connected speech at an SNR of 10 dB. The speech is the Japanese utterance "korekara" (... from now to...) spoken by a male speaker.

5. Conclusions

We have proposed a new method for estimating the speech spectrum corrupted by colored noise. The speech production process is represented by an ARMA model. The background noise is represented by an MA model. The whole model of the observation process is then represented as an extended ARMA model with two sources. After detecting the onset of speech activity and estimating the noise spectrum, the method can estimate AR or ARMA parameters of the vocal tract transfer function according to the SNR condition of the observed speech.

The validity of the proposed method was demonstrated by spectral estimation of both synthetic and natural speech sounds. Since the method can estimate the parameters required for speech synthesis, it can be applied to noise-reduction of speech.

Manuscrit reçu le 7 octobre 1986.

REFERENCES

[1] B. S. Atal and S. L. Hanauer, Speech analysis and synthesis by linear prediction of the speech wave, *J. Acoust., Soc. Amer.*, 50, 1971, pp. 637-655.

[2] B. S. Atal and L. R. Rabiner, A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-24, 1976, pp. 201-212.

[3] S. M. Kay, The effects of noise on the autoregressive spectral estimator, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-27, 1979, pp. 478-485.

[4] S. M. Kay, Noise compensation for autoregressive spectral estimates, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-28, 1980, pp. 292-303.

[5] R. L. Kashyap, Estimation of parameters in partially whitened representation of a stochastic process, *IEEE Trans. Automat. Contr.*, AC-19, 1974, pp. 13-21.

[6] H. Kobatake, Optimization of voiced/unvoiced decisions in nonstationary noise environments, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-35, 1987, pp. 9-18.

[7] R. C. K. Lee, *Optimal Estimation, Identification, and Control*, MIT Press, 1964, pp. 79-126.

[8] I. D. Landau, Unviased recursive identification using model reference adaptive techniques, *IEEE Trans. Automat. Contr.*, AC-21, 1976, pp. 194-202.

[9] L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Trans. Automat. Contr.*, AC-22, 1977, pp. 551-575.

[10] J. Makhoul, Linear prediction: A tutorial review, *Proc. IEEE*, 63, 1975, pp. 561-580.

[11] Y. Miyanaga, N. Miki, N. Nagai and K. Hatori, A speech analysis algorithm which eliminates the influence of pitch using the model reference adaptive system, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-30, 1982, pp. 88-96.

[12] H. Morikawa and H. Fujisaki, Speech analysis based on simultanious estimation of autoregressive moving-average parameters, *Trans. IECE Japan*, J60-A, 1978, pp. 195-202.

[13] H. Morikawa and H. Fujisaki, Adaptive analysis of speech based on a pole-zero representation, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-30, 1982, pp. 77-88.

[14] H. Morikawa and H. Fujisaki, Spectral estimation of speech in noisy environments, *Trans. IECE Japan*, J65-A, 1982, pp. 987-994.

[15] H. Morikawa and H. Fujisaki, Adaptive control of analysis interval in the SEARMA speech analysis methods, *J. Acoust. Soc., Japan*, 39, 1983, pp. 512-520.

[16] H. Morikawa and H. Fujisaki, System identification of the speech production process based on a state-space representation, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-32, 1984, pp. 252-262.

[17] H. Morikawa and H. Fujisaki, Spectral estimation of speech corrupted by colored noise, Reprints of *IFAC/IFORS Conference on Control Science and Technology for Development*, Beijing, China, 3, 1985, pp. 1401-1412.

[18] H. Morikawa and H. Fujisaki, Noise reduction of speech signal by adaptive Kalman filtering, *Traitement du signal*, 4, special No. 1, 1987, pp.

[19] T. Soderstrom, L. Ljung and I. Gustavsson, A theoretical analysis of recursive identification methods, *Automatica*, 14, 1978, pp. 231-244.

[20] J. Tiereny, A study of LPC analysis of speech in additive noise, *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-28, 1980, pp. 389-397.