Linear Statistical Modeling of Speech and its Applications
-- Over 36 year history of LPC --

Fumitada ITAKURA

Graduate School of Engineering, Nagoya University
Furo-cho 1, Chikusa-ku Nagoya 464-8603, Japan
itakura@nuee.nagoya-u.ac.jp

Abstract

Linear prediction of speech signal is now a common knowledge for students, engineers and researchers who are interested in speech science and technology. The first introduction of the concept to the international speech community was two presentations [6],[7] at the 6th ICA held in Tokyo in 1968. While Atal and Schroeder used the concept of linear prediction to encode speech waveform efficiently, Itakura and Saito used the concept of maximum likelihood estimation of speech spectrum to implement a new type of vocoder. Since this epoch, these concepts are investigated widely in order to achieve accurate speech analysis, efficient transmission of speech and speech recognition. In this paper, the history of LPC will be surveyed mainly based on my personal experience.

1. Introduction

Homer Dudley is a foremost pioneer of analysis and synthesis of speech. He invented the "Channel Vocoder", as early as 1939 at Bell Telephone Laboratories [1]. He established the principles of the carrier nature of speech. Following his footstep, various speech analysis and synthesis methods have been proposed for the purpose of transmitting speech signal as efficient as possible. Most research has been concentrated on finding feature parameters which express speech spectral characteristics efficiently. Among these methods, the most effective and successful one is based on the autoregressive or all-pole model of speech signal. From 1966 through 1968 [2],[3],[4] we proposed the maximum likelihood estimation of speech signal (MLES) and its application to vocoder. However the MLES approach itself was not been very efficient due to its poor quantization characteristics. The PARCOR scheme was proposed in 1969 by F. Itakura, in which all-pole model is specified by a set of parameters called PARCOR coefficients or the reflection coefficients. It is shown that a significant information reduction is accomplished through optimal non-linear quantization and bit allocation of parameters. In 1975, the line spectral pair (LSP) representation of all-pole model was proposed by F.Itakura, which is now widely used in many speech compression systems.[14] This paper describes the speech analysis and synthesis methods developed in 1960's through 1980's at the Electrical Communications Laboratories of the Nippon Telegraph and Telephone Public Corporation. These methods involve MLES, PARCOR, LSP and CSM speech analysis/synthesis.

2. All-pole speech production model based autoregressive (AR) stochastic process

Speech is a continuously time-varying process. By making reasonable assumptions, it is possible to develop linear models which are locally time-invariant for describing important speech events. A time series obtained by sampling a speech signal shows a significant autocorrelation between adjacent samples. The short time autocorrelation function is related to the running spectrum, which plays the most important role in speech perception. Let $\{x[n], n = \text{integer}\}$ be a discrete time speech waveform sampled at every ΔT seconds. Assume that $\{x[n-i], i = 0,\cdots, p\}$ are (p+1) dimensional random variables taken from a stochastic process which is stationary over a short interval such as 30 milliseconds. When $x[n]$ is a real sampled value, the next autoregressive (AR) relation is assumed:

$$\sum_{i=0}^{p} \alpha[i] x[n-i] = \sigma\varepsilon[n] \qquad (1)$$

where $\alpha[i]'$s (with $\alpha[0]=1$) are the AR or linear predictive coefficients, and $\sigma\varepsilon[n]$ is the residual and σ is its rms value. An explicit representation of Eq. (1) in a digital filter form is shown in Fig. 1, where D is the unit time delay ΔT . The transfer function H(z) equivalent to Eq.(1) is

$$H(z) = \frac{X(z)}{E(z)} = \frac{\sigma}{1 + \sum_{i=1}^{p} \alpha[i]\, z^{-i}} \qquad (2)$$

where $z = \exp(j\omega\Delta T)$, $(-\pi < \omega\Delta T < \pi)$.

3. Maximum likelihood estimation of speech spectrum

An approach to estimating parameters in the frequency domain has been first reported [2]. It is assumed that the speech signal has the following characteristics.

(1)The speech production system can be represented by a transfer function with poles only.

(2) The speech signal is generated by adding a random white Gaussian signal to the system mentioned above. The averaged value of the random signal is zero, and its variance is assumed to be σ^2 .

Based on these assumptions, parameters such as $\Theta = (\sigma^2, \alpha[1], \alpha[2], \cdots, \alpha[p])$ can be estimated from $\mathbf{X} = (x[1],\ x[2],\ ...,\ x[N])$.

The all-pole speech envelope is assumed to

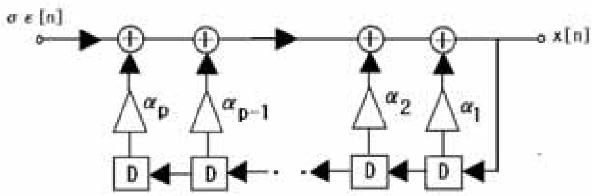


Fig.1 All-pole (AR) model of speech production

$$
\begin{aligned}
S_{AR}(\omega) &= \frac{\sigma^2}{2\pi}\left|H(\exp(-j\omega\Delta T)\right|^2 \\
&= \frac{\sigma^2}{2\pi}\frac{1}{\left|\prod_{i=1}^{p}\left(1-z_i z^{-1}\right)\right|^2} \\
&= \frac{\sigma^2}{2\pi}\frac{1}{\left|\sum_{i=1}^{p}\alpha[i]\exp(-j\omega i\Delta T)\right|^2} \\
&= \frac{\sigma^2}{2\pi}\frac{1}{\sum_{i=-p}^{p}A[i]\cos(\omega i\Delta T)}
\end{aligned}
\qquad (3)
$$

where z_i represents roots of a polynomial $1+\alpha[1]z^{-1}+\alpha[2]z^{-2}+\cdots+\alpha[p]z^{-p}=0$ and $A[i]$ is given by

$$A[i]=\sum_{j=0}^{p-|i|}\alpha[j]\alpha[j+|i|] \qquad (4)$$

A logarithmic likelihood function

$$
\begin{aligned}
L(\mathbf{X}\,|\,\Theta) &= \frac{-N}{2}\left[\log(2\pi\sigma^2)+\frac{1}{\sigma^2}\sum_{k=-p}^{p}A_k V[k]\right] \\
&= \frac{-N}{2}\left[2\log(2\pi)+\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\log S_{AR}(\omega)+\frac{S_{DFT}(\omega)}{S_{AR}(\omega)}\right]d\omega\right]
\end{aligned}
\qquad (5)
$$

where $V[k]$ is the short time autocorrelation function of samples $\mathbf{X}$:

$$V[i]=\frac{1}{N}\sum_{j=1}^{N-|i|}x[j]\,x[j+|i|]$$

(6)

and $S_{DFT}(\omega)$ is the short-time power spectrum which is obtained by the discrete Fourier transformation of $\mathbf{X}$:

$$S_{DFT}(\omega)=\frac{1}{2\pi N}\left|\sum_{k=1}^{N}x[k]\exp(-j\omega\Delta T k)\right|^2 . \qquad (7)$$

On the basis of these equations, when $\mathbf{X}$ is given, the logarithmic likelihood ratio is represented by $V[k]$ alone. The approach which determines Θ , by maximizing $L(\mathbf{X}\,|\,\Theta)$, is called the maximum likelihood method. $L(\mathbf{X}\,|\,\Theta)$ is maximized with respect to σ^2 ; by computing $\tilde{\sigma}^2=\arg\max_{\sigma^2} L(\mathbf{X}|\Theta)$. Consequently, the maximization of $L(\mathbf{X}\,|\,\Theta)$ is equivalent to the minimization of

$$\tilde{\sigma}^2 = J(\alpha[1],\ \alpha[2],\ ...,\ \alpha[p])=\sum_{i=0}^{p}\sum_{j=0}^{p}\alpha[i]\hat{V}[i-j]\alpha[j] \qquad (8)$$

and the corresponding likelihood function is given by

$$L(\mathbf{X}\,|\,\Theta)\Big|_{\sigma^2=\tilde{\sigma}^2}=\frac{-N}{2}\log\left[2\pi e\left\{\sum_{i=0}^{p}\sum_{j=0}^{p}\alpha[i]V[i-j]\alpha[j]\right\}\right] \qquad (9)$$

By maximizing the above expression with respect to $\{\alpha[i], i=1,2,\cdots,p\}$, we get the following Yule-Walker (or normal) equation:

$$\sum_{j=0}^{p}V[i-j]\,\alpha[j]=0,\quad (i=1,2,\cdots,p) \qquad (10)$$

It has been shown that this maximization of the likelihood function is equivalent to the minimization of the following spectral distance measure defined by

$$
\begin{aligned}
E\{d(\omega)\} &= \int_{-\pi}^{\pi}2\left[d(\omega)+\exp(-d(\omega))-1\right]d\omega \\
&= \int_{-\pi}^{\pi}\left[d^2(\omega)-\frac{2}{3!}d^3(\omega)+\frac{2}{4!}d^4(\omega)-\cdots\right]d\omega
\end{aligned}
\qquad (11)
$$

where $d(\omega)=\log\{S_{AR}(\omega)/S_{DFT}(\omega)\}$ is the log spectral difference between the AR spectrum and observed DFT spectrum. It is non-negative and zero if and only if $S_{AR}(\omega)\equiv S_{DFT}(\omega)$ for all ω ; it is a measure of the goodness of fit between two spectra $S_{AR}(\omega)$ and $S_{DFT}(\omega)$. The integrant of Eq.(11) is the penalty function against the mismatch between $S_{AR}(\omega)$ and $S_{DFT}(\omega)$ as shown in Fig.2. Due to its asymmetricity, it is very sensitive to the underestimation $S_{AR}(\omega) \ll S_{DFT}(\omega)$, but is tolerant for the overestimation $S_{AR}(\omega) \gg S_{DFT}(\omega)$. For this reason, MLES is well suited for extracting spectral peaks such as the formants

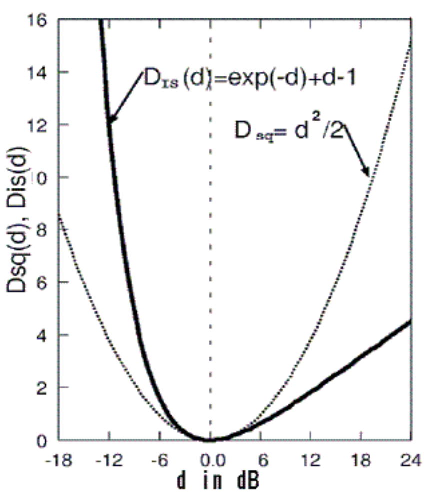
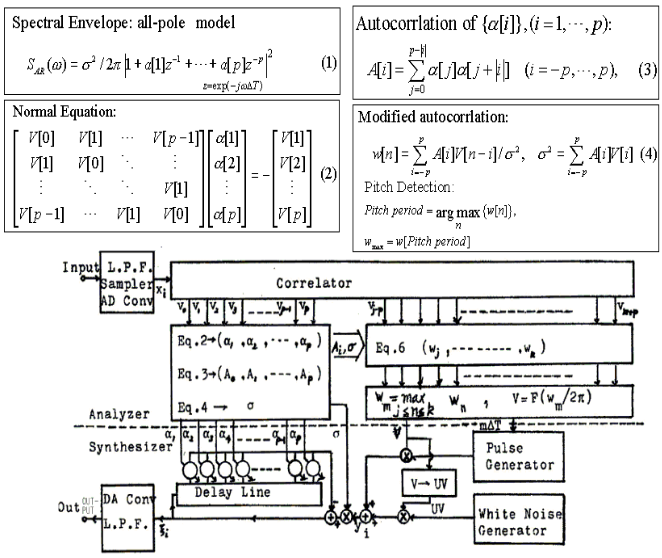


Fig.2 Distance measure



Fi.3 Maximum Likelihood Vocoder

of speech.

4. PARCOR speech analysis and synthesis method

The direct form speech synthesis filter $H(z)$ with parameters $\{\alpha[i]\}$ requires a higher quantization accuracy to maintain the stability of $H(z)$. To overcome this problem, the partial autocorrelation (PARCOR) lattice filter was invented in 1969 [4]. Since then, the optimum coding of the PARCOR coefficients have been extensively studied in order to improve the synthetic speech quality. [13],[14]

4.1 Definition of PARCOR coefficients

The conventional autocorrelation coefficients $V[k]$ can be regarded as a measure of linear time shift dependency, but the set of these parameters is still redundant since there is significant dependency among them. The notion of partial autocorrelation was introduced to reduce the redundancy using successive linear prediction techniques. Suppose that $\hat{x}[n]$ and $\breve{x}[n-p]$ are predicted values of $x[n]$ and $x[n-p]$ from the intermediate samples $x[n-p+1],\cdots,x[n-2],x[n-1]$ in the forms,

$$\begin{aligned} \hat{x}^{(p-1)}[n] &= -\sum_{i=1}^{p-1}\alpha^{(p-1)}[i]x[n-i], \\ \breve{x}^{(p-1)}[n-p] &= -\sum_{i=1}^{p-1}\beta^{(p-1)}[i]x[n-i] \end{aligned} \quad (12)$$

where the prediction coefficients $\alpha^{(p-1)}[i]$ and $\beta^{(p-1)}[i]$ are determined to minimize the residuals

$$E\{(x[n]-\hat{x}^{(p-1)}[n])^2\}$$

and (13)

$$E\{(x[n-p]-\breve{x}^{(p-1)}[n-p])^2\}$$

of the forward and backward prediction.

The PARCOR coefficient k_p between $x[n]$ and $x[n-p]$ is defined as the correlation coefficient between two residuals $x[n]-\hat{x}^{(p-1)}[n]$ and $x[n-p]-\breve{x}^{(p-1)}[n-p]$, as shown in Fig. 3.

$$\varepsilon_f^{(p-1)}[n] = \sum_{i=0}^{p-1}\alpha^{(p-1)}[i]x[n-i]$$

$$\varepsilon_b^{(p-1)}[n-p] = \sum_{i=1}^{p}\beta^{(p-1)}[i]x[n-i]$$

$$k_p = \frac{E\left\{\varepsilon_f^{(p-1)}[n]\varepsilon_b^{(p-1)}[n-p]\right\}}{\sqrt{E\{\varepsilon^2{}_f^{(p-1)}[n]\}E\{\varepsilon^2{}_b^{(p-1)}[n-p]\}}}$$

(14)

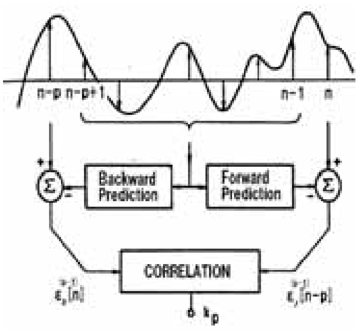


Fig.3 Definition of PARCOR, Partial Correlation coefficients

Fig.3 Definition of PARCOR coefficient

Physically, these k parameters correspond to reflection coefficients in the acoustic tube model of the vocal tract.

The following relation can be obtained by substituting Eqs. (12) , (13) and (14).

$$k_m = \frac{\sum_{i=0}^{m-1}\alpha[i]V[m-i]}{\sum_{i=0}^{m-1}\alpha[i]V[i]} = \frac{w[m-1]}{u[m-1]} \quad (15)$$

$$\begin{aligned} u[m] &= u[m-1](1-k_m{}^2) \\ \alpha^{(m)}[i] &= \alpha^{(m-1)}[i] - k_m\beta^{(m-1)}[i], \\ \beta^{(m)}[i] &= \beta^{(m-1)}[i-1] - k_m\alpha^{(m-1)}[i-1], \\ \beta^{(m-1)}[i] &= \alpha^{(m-1)}[m-i] \end{aligned} \quad (16)$$

4.2 Direct PARCOR coefficient derivation

The PARCOR coefficients are also derived directly from the speech signal. This is called the PARCOR lattice analysis method. The PARCOR coefficients have already been defined as the correlation coefficient between the residuals of the forward and of the backward predictions. The forward and backward residual linear operators are introduced:

$$\begin{aligned} &\varepsilon_f^{(p-1)}[n] = \sum_{i=0}^{p-1}\alpha^{(p-1)}[i]D^i x[n] = A_{p-1}(D)x[n] \\ &\text{where } A_{p-1}(D) = \sum_{i=0}^{p-1}\alpha^{(p-1)}[i]D^i, \\ &\varepsilon_b^{(p-1)}[n-p] = \sum_{i=1}^{p}\beta^{(p)}[i]D^i x[n] = B_{p-1}(D)x[n] \\ &\text{where } B_{p-1}(D) = \sum_{i=1}^{p}\beta^{(p-1)}[i]D^i, \end{aligned} \quad (17)$$

where D is the delay operator for unit time. From Eqs. (14) and (16), the following recursions are obtained:

$$\begin{aligned} A_m(D) &= A_{m-1}(D) - k_m B_{m-1}(D), \quad A_0(D) = 1 \\ B_m(D) &= D[B_{m-1}(D) - k_m A_{m-1}(D)], \quad B_0(D) = D \end{aligned} \quad (18)$$

From this recursion we can implement a PARCOR lattice analyzis/synthesis system as shown in Fig.4

4.3 Extraction of source parameters using the modified autocorrelation method

As the input signal passes through the PARCOR lattice analyzer, the autocorrelation between the adjacent samples are gradually removed. If the number of sections *p*, is sufficiently

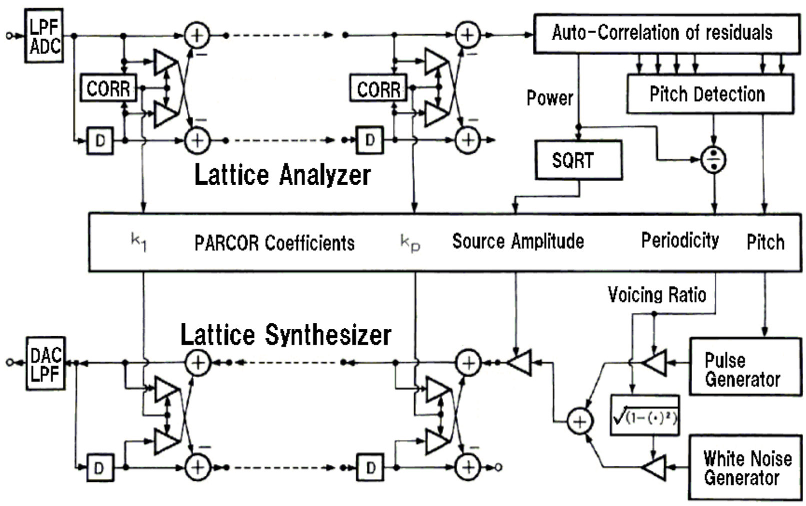


Fig.4 Lattice form PARCOR vocoder system

large, for example, from 10 to 16, the spectral envelop of the input signal will be extracted almost completely, and the spectrum envelope of the residual will be flattened. Thus, only excitation source characteristics such as signal amplitude, voicing, and pitch period are contained in it. The signal amplitude is the root mean square value of the residual variance. The autocorrelation coefficients of the residual are computed to detect periodicity and to determine pitch period. Pitch period is determined by searching the maximum autocorrelation.

4.4 PARCOR speech synthesis method

Speech synthesis from the PARCOR coefficients and excitation source parameters is the inverse process of speech analysis. The excitation source is generated by controlling the pulse and white noise generators according to pitch period, voicing and amplitude excites a time-varying filter composed of lattice sections to produce synthesized speech.

5. Line spectrum pair (LSP) speech analysis and synthesis method

PARCOR method *does have* limitations with regard to data compression. PARCOR-synthesized speech quality rapidly deteriorates at bit rates lower than 2.4 kbps. There are two main reasons for this. Firstly, the required quantization bits per parameter are four to eight bits for the PARCOR coefficients [7,8,9,10]. Secondly, the spectral distortion due to parameter interpolation increases rapidly as the frame (parameter refreshing) period is increased. This chapter describes an approach to speech analysis and synthesis, which compensates for the problems. This approach, called the "line spectrum pair" (LSP) method [16], again exploits the all-pole modeling of speech. The LSP parameters are one of LPC parameters in the frequency domain.

Theorem: All the roots of $P(z) = Q(z) = 0$ lie on the unit circle, and alternate one another. In this case, the all-pole filter $1/A_p(z^{-1})$ is guaranteed to be stable. For even *p*, $P(z)$ and $Q(z)$ are factored as follows:

$$\begin{aligned} P(z) &= (1-z^{-1})\prod_{i=1}^{p/2}(1\text{-}2\cos\omega_{2i}z^{-1}+z^{-2}) \\ Q(z) &= (1+z^{-1})\prod_{i=1}^{p/2}(1\text{-}2\cos\omega_{2i\text{-}1}z^{-1}+z^{-2}) \end{aligned} \quad (20)$$

It can be readily seen that a set of parameters $(\omega_1,\omega_2,\omega_3,\cdots,\omega_{p-1},\omega_p)$ is derived from $A_p(z^{-1})$, and conversely, $A_p(z^{-1})$ is also reconstructed in terms of $(\omega_1,\omega_2,\omega_3,\cdots,\omega_{p-1},\omega_p)$. Therefore the set of $\{\alpha_i\}^p{}_{i=1}$ is also equivalent to set $(\omega_1,\omega_2,\omega_3,\cdots,\omega_{p-1},\omega_p)$, similar to $\{k_i\}^p{}_{i=1}$. A considerable amount of computation is eliminated by an elegant algorithm used to obtain the roots of $P(z)$ and $Q(z)$ by exploiting the fact that both have complex conjugate roots on the unit circle.

5.2 LSP speech synthesis method

All pole LSP filter is a digital filter whose transfer function is identical to $1/A_p(z^{-1})$. The all-pole digital filter $1/A_p(z^{-1}) = 1/(1-(1-A_p(z^{-1})))$ is realized by a digital filter with a negative feedback loop, whose transfer function is $(1-A_p(z^{-1}))$ .By using this relation, the LSP all-pole filter can be implemented by the signal flow graph shown in Fig.5. Using this configuration, generation of one sample output requires *p* multiplications and (3*p*+1) additions or subtractions.

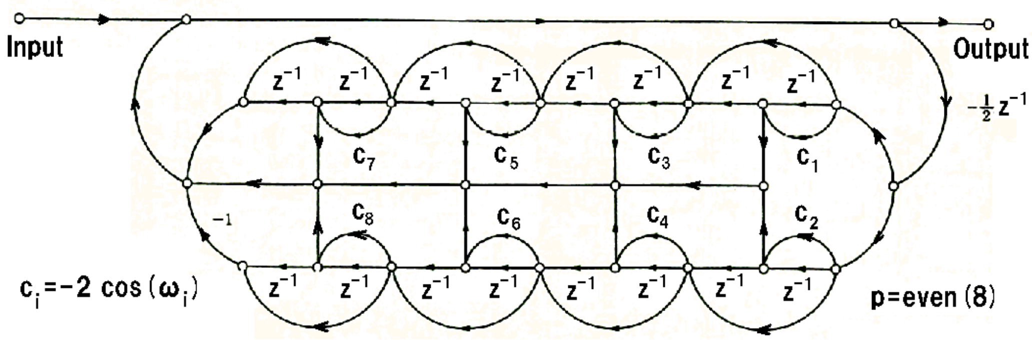


Fig.5 LSP all-pole filter for speech synthesis

5.1 LSP speech analysis method

Let $1/A_p(z^{-1})$ be a stable all-pole digital filter in Eq. (4), and define the following $(p+1)$ th order polynomials. $A_{p+1}(z)$ corresponds to relations Eq.(18),

$$\begin{aligned} P(z) &= A_p(z^{-1}) - z^{-(p+1)}A_p(z) \\ Q(z) &= A_p(z^{-1}) + z^{-(p+1)}A_p(z) \end{aligned} \quad (19)$$

5.3 Physical meaning of LSP

The LSP parameters have an interesting physical interpretation. If vocal tract characteristics can be expressed by $1/A_p(z^{-1})$, the vocal tract is modeled as a non-uniform section acoustic tube consisting of *p* short sections of equal length. The acoustic tube is open at the lips terminal, and each section is numbered from the lips. Differences of vocal tract areas between the adjacent sections n and n+1 causes reflection of traveling wave. The reflection coefficients are equal to the n^{th} PARCOR coefficients k_n. Section p+1, which corresponds to the glottis terminal, is terminated by a matched resistance. The excitation signal applied to the glottis drives the acoustic tube. Now consider a pair of artificial boundary conditions, where the acoustic tube is completely closed or open at the glottis. These conditions correspond to $k_{p+1}=1$ and $k_{p+1}=-1$. Under these conditions, $A_{p+1}(z^{-1})$ should be identical to either $P(z)$ or $Q(z)$. The acoustic tube is now loss-less, and therefore the transfer function displays a line spectrum structure.

6. Composite Sinusoidal Modeling(CSM)

In 1975, Itakura proposed the line spectrum representation (LSR) concept and its algorithm to obtain a set of parameters for new speech spectrum representation. Independently from this, Sagayama developed a composite sinusoidal modeling (CSM) concept which is equivalent to LSR but give a quite different formulation, solving algorithm and synthesis scheme. Sagayama clarified the duality of LPC and CSM and provided the unified view covering LPC, PARCOR, LSR, LSP and CSM, CSM is not only an new concept of speech spectrum analysis but also a key idea to understand the linear prediction from a unified point of view. Consider the composite sinusoidal model which is a sum of n sinusoids of arbitrary amplitudes $\{\sqrt{2m_i}\}$, frequencies $\{\omega_i\}$ and phases $\{\phi_i\}$.

$$x[n]=\sum_{i=1}^{q}\sqrt{2m_i}\sin(\omega_i n+\phi_i) \qquad (21)$$

Its autocorrelation function is given in terms of amplitudes and frequencies as follows:

$$V[n]=\sum_{i=1}^{q}m_i\cos(\omega_i n) \qquad (22)$$

The above formula includes $2q$ parameters $\{\omega_i\}, i=1,\cdots,q$ and $\{\phi_i\}, i=1,\cdots,q$. The problem is to find a set of parameters $\{\omega_i\}$ and $\{m_i\}$, which makes $\{V[n]\}$ equal to the sample autocorrelation function for $n=0,1,\cdots,2q-1$; a kind of correlation matching problem. The amplitude $\{m_i\}$ is determined by solving a system of linear equation fo Hankel form, and the frequencies $\{\omega_i\}$ is obtained by solving a Van der Monde matrix equation.

Conclusion

Several speech analysis and synthesis methods based on linear prediction, PARCOR , LSP and CSM have been described. Their principles and physical interpretations have also been presented. The characteristics of these methods have been clarified by means of a number of objective and subjective experiments. As a result, the LSP speech analysis and synthesis method has been found to be superior to other methods. Today active research research efforts in narrow-band speech coding is focused on extremely low bit rate transmission, under 1000 bps.[30],[31] A very interesting problem remains on how to eliminate speech signal redundancy. In most of these methods, redundancies in feature parameters in both space and time domains, are eliminated as far as possible.

Reference

[1] H. Dudley, “The Vocoder”, Bell Labs. Rec., vol. I8, pp. 122-126, Dec. 1939 (1939) 169-177.

[2] S. Saito and F. Itakura, “A theoretical consideration of statistically optimum methods for speech spectral density”, Report No. 3107, Electrical Communication Laboratory, NTT, Tokyo (1966) (in Japanese).

[3] F. Itakura and S. Saito, “Analysis-synthesis transmission system based on maximum likelihood spectrum estimation”, Proc. Conv. Acoust. Soc. Japan, 2-3-1 (Nov., 1967) 231-232 (in Japanese).

[4] F. Itakura and S. Saito, “Speech analysis-synthesis system based on the partial autocorrelation coefficients”, Proc. Conv. Acoust. Soc. Japan, 2-2-6 (Oct. 1969) 199-200 (in Japanese).

[5] B. S. Atal and M. R. Schroeder, “Predictive coding of speech signals”, Proc. 1967 Conf. Comm. and Process., (1967) 360-361.

[6] B. S. Atal and M. R. Schroeder, “Predictive coding of speech signals”, Proc. Int. Cong. Acoust. C-5-4,1968(Tokyo)

[7] F. Itakura and S. Saito, “Analysis-synthesis telephony based on maximum likelihood method”, Proc. Int. Cong. Acoust. C-5-5,1968(Tokyo)

[8] B. S. Atal, “Speech analysis and synthesis by linear prediction of the speech wave”, J. Acoust. Soc. Am., 47 (1970) 65(A).

[9] J. D. Markel, “Digital inverse filtering, a new tool for formant trajectory estimation”, IEEE Trans., AU-20 (1972) 129-137.

[10] J. Makhoul and J. Wolf, “Linear prediction and the spectral analysis of speech”, NTIS No. AD-749066, BBN Report No. 2304 Bolt Beraneck and Newman, Inc., Cambridge, Massachusetts (1972).

[11] I. Makhoul, “Spectral analysis of speech by linear prediction”, IEEE Trans., AU-21 (1973) 140-148.

[12] J. D. Markel and A. H. Gray, Jr., “On autocorrelation with application to speech analysis”, IEEE Trans., AU-21 (1973) 69-79.

[13] N.Kitawaki and F.Itakura, ”Efficient coding 0f speech by nonlinear quantization and nonuniformity of PARCOR coefficients”, IECE Trans., Vol. J61-A, No.6,1978,pp.543-550

[14] Y.Tohkura, ”Improvement of voice quality in PARCOR bandwidth compression system”, IECE Trans. Vol.J61-A, No.3,1978

[15] M. J. Narasimha, K. Shenoi, and A. M. Peterson, “A Hilbert space approach to linear predictive analysis of speech signals”, Tech. Report 3606-10, Radio Science Lab., Stanford Electronics Lab., Stanford University, California (1974).
[16] F. Itakura, “Line spectrum representation of linear predictor coefficients of speech signals”, J. Acoust. Soc. Am., 157 (Suppl.) (1975) R9, 535.
[17] F. Itakura, “Line spectrum representation of linear predictors”, Trans. of the Committee on Speech Research, S75-34 (1975) (in Japanese).
[18] S. Sagayama and F. Itakura, “Composite sinusoidal modeling applied to spectrum analysis of speech”, Trans. of the Committee on Speech Research, 579-06 (1979) (in Japanese).
[19] S. Sagayama, “Symmetrical Formulations of linear predictive coding and composite sinusoidal modeling”, Proc. Annual Convention of Acoust. Soc. Japan, 1-1-1, (Oct., 1980) 359-360 (in Japanese).
[20] U. Grenander and G. Szego, Toeplitz Forms and their Applications, University of California Press (1958).
[21] F. Itakura and S. Saito, “Digital filtering techniques for speech analysis and synthesis”, 7th Int. Cong. Acoust, Budapest, 25 C 1 (1971).
[22] N. Levinson, “The Wiener RMS (root Mean Square) error criterion in filter design and prediction”, J Math. Phys., 25 (1947) 261-278.
[23] N. Sugamura and F. Itakura, “Speech data compression by LSP speech analysis-synthesis technique”, Trans. IECE 81/8, J64-A, 8, (1981) 599-606 (in Japanese).
[24] S. Sagayama, “Stability condition of LSP speech synthesis digital filters”, Proc. Annual Conv. Acoust. Soc. Japan, 3-4-12 (March, 1972) 153-154 (in Japanese).
[25] T. Kobayashi and F. Itakura, “An algorithm to extract line spectrum pair parameters”, 2-7-9 (May, 1980) 599-600 (in Japanese).
[26] F. Itakura and N. Sugamura, “Principle and organization of the LSP speech synthesizer”, Trans. of the Committee on Speech Research, 579-46 (1979) (in Japanese).
[27] S. Sagayama, “Theoretical study in CSM, LSP, and their properties”, Trans. of the Committee on Speech Research, The Acoustical Society of Japan, 582-14 (May, 1982) (in Japanese).
[28] S. Sagayama and F. Itakura, “Composite sinusoidal modeling applied to spectral analysis of speech”, Trans. IECE81/2, J64-A, No. 2 (1981) 105-112 (in Japanese).
[29] F. Itakura and S. Saito, “PARCOR digital speech synthesizer”, Proc. Annual Convention of Acoust. Soc. Japan, 2-1-2 (Oct. 1970) 121-122 (in Japanese).
[30] Y.Shiraki and M.Honda,”Very low bit rate speech coding based on joint segment quantization and variable length segment quantizer”, J. Acoust. Soc. Amer. Suppl. 1, vl.79, 1986
[31] Y.Shiraki and M.Honda,”LPC speech coding based on variable-length segment quantization”,IEEE Trans. Acoust. Signal Processing, ASSP-36, 9, pp.1437-1444, Sep. 1988