THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.20, No.4E 2001. 12. pp. 39~44

Stabilizing Linear Prediction for Discrete Harmonic Spectra of Audio Signals

Seung Hyon Nam*, Kyeongok Kang**, Jin Woo Hong**

*Department of Electronic Engineering, Paichai University, Taejon
**Broadcasting Media Technology Department Electronics and Telecommunications Research Institute, Taejon
(Received 2 June 2001; accepted 17 October 2001)

Abstract

We investigate the numerical instability of linear prediction for discrete harmonic spectra of audio signals. It is identified that the eigenvalue spread is very large when discrete harmonic spectra are confined only in a lower part of the entire signal bandwidth. A simple method that redefines the sampling frequency and associate harmonic frequencies is proposed to improve the numerical stability. Simulation results using real audio signals indicate its superior stabilizing ability and improved accuracy in the discrete spectral estimation for both LP and DAP.

Keywords: *LP, DAP, Autocorrelation matrix, Discrete harmonic spectra, Eigenvalue spread, Stability*

I. Introduction

Audio signals are often described as mixtures of harmonics and residual signals. An efficient representation of the harmonic part is very crucial in audio processing. All-pole modeling is one of efficient representations for harmonic spectra, where the spectral envelope of the all-pole filter approximates discrete harmonic spectra at given harmonic frequencies.

Linear prediction (LP) is a popular way to obtain all-pole filters that minimize the spectral distance between the signal spectrum and the all-pole spectrum[1]. When the signal spectrum is continuous (or smooth enough), LP provides stable all-pole filters which are accurate enough. When the signal spectrum is available at only a set of harmonic frequencies, however, LP suffers from several shortcomings. One of such shortcomings is the poor estimation accuracy evident in the spectral valley. This problem is addressed by El-Jaroudi and Makhoul[2]. They proposed DAP (Discrete All-Pole) modeling that employes the Itakura-Saito error measure, to solve the problem. However, both LP and DAP may suffer from numerical instability due to ill-conditioned autocorrelation matrices.

In this paper, we investigate the numerical instability due to ill-conditioned autocorrelation matrices for discrete harmonic spectra of audio signals. Specifically, the eigenvalue spread of the autocorrelation matrix for discrete harmonic spectra is investigated as a function of the largest harmonic frequency and filter order. A simple technique proposed to reduce the eigenvalue spread is to redefine the sampling frequency and associate harmonic frequencies. Simulation results on real audio signals demonstrate its stabilizing ability and improved estimation accuracy.

Corresponding author: Seung Hyon Nam (shnam@pcu.ac.kr)
Paichai University, Taejon 302-735, Korea

II. LP and DAP for Discrete Harmonic Spectra

The signal composed of L harmonics can be expressed as

$$s[n] = \sum_{k=1}^{L} A_k \cos(k\omega_0 n + \phi_k) \tag{1}$$

where $\omega_0 = 2\pi f_0$ is the fundamental (or pitch) frequency. The power spectrum for this signal is expressed by

$$P(\omega) = \sum_{k=1}^{L} \frac{|A_k|^2}{2} \{\delta(\omega + k\omega_0) + \delta(\omega - k\omega_0)\}, \tag{2}$$

and the autocorrelation coefficient of lag m is given by

$$r_m = \sum_{k=1}^{L} \frac{|A_k|^2}{2} \cos(k\omega_0 m). \tag{3}$$

Most often, information on discrete harmonic spectra $P(k\omega_o)$ are unknown and to be estimated from the signal spectrum which is a mixture of harmonics and noise as shown in Figure 1.

Given an p^{th} order all-pole filter

$$H(z) = \frac{1}{\sum_{k=0}^{p} a_k z^{-k}}, \tag{4}$$

the all-pole envelope is defined as

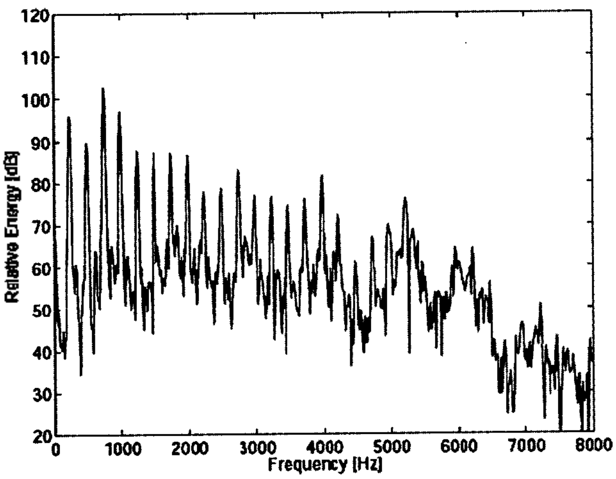

Figure 1. Typical audio spectrum.

$$\hat{P}(\omega) = \frac{G^2}{\left|\sum_{k=0}^{p} a_k e^{-j\omega k}\right|^2} \tag{5}$$

where G is the gain factor. Then both LP and DAP try to find the best all-pole envelope which matches best to the signal spectrum $P(\omega_k)$.

In LP, the all-pole filter is obtained by minimizing the MSE criterion

$$E_{LP} = \frac{1}{L} \sum_{k=1}^{L} \frac{P(\omega_k)}{\hat{P}(\omega_k)} \tag{6}$$

with respect to a_k. This is accomplished by solving the normal equation

$$\boldsymbol{R}_p \ \boldsymbol{a}_p = - \ \boldsymbol{r}_p \tag{7}$$

where

$$\boldsymbol{R}_p = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix}$$
$$\boldsymbol{a}_p = [a_1 \ a_2 \ a_3 \cdots a_p]^T \tag{8}$$
$$\boldsymbol{r}_p = [r_1 \ r_2 \ r_3 \cdots r_p]^T$$

Since the error criterion (6) emphasizes the region where $P(\omega) > \hat{P}(\omega)$, spectral error is larger for smaller discrete harmonic spectra. To overcome this shortcoming, DAP employs the Itakura-Saito (I-S) error criterion defined as

$$E_{DAP} = \frac{1}{L} \sum_{k=1}^{L} \left\{ \frac{P(\omega_k)}{\hat{P}(\omega_k)} - \log \frac{P(\omega_k)}{\hat{P}(\omega_k)} - 1 \right\}. \tag{9}$$

Minimization of (9) with respect to a_k yields

$$\boldsymbol{R}_p \ \boldsymbol{a}_p = \ \boldsymbol{h}_p \tag{10}$$

where $\boldsymbol{h}_p = [h(-1) \ h(-2) \ h(-3) \cdots - h(-p)]^T$. Here, $h(-i)$ is the time-reversed impulse response of the all-pole filter defined as

$$h(-i) = \frac{1}{L} \sum_{k=1}^{L} H(\omega_k) e^{-j\omega_k i}. \tag{11}$$

Equation (10) cannot be solved directly since $h(-i)$ is also a function of a_k. In[2], the following iterative algorithm is proposed:

$$a_p^{m+1} = (1-\alpha)\ a_p^{m} + \alpha\ R^{-1}\ h_p^{m}, \quad (12)$$

where m is the iteration number and α is the control parameter for convergence speed. The starting value a_p^0 is obtained from LP. Normally, α is chosen to be about 0.5 for fast stable convergence of the algorithm. DAP is known to improve the spectral matching significantly over LP[2].

III. The Eigenvalue Spread of the Auto-correlation Matrix

Both LP and DAP require inversion of the auto-correlation matrix. Whether the inversion is performed in a recursive manner or not, conditioning of the auto-correlation matrix affects numerical stability of a solution critically.

In general, conditioning of a matrix is defined by the eigenvalue spread of a matrix, which is given by

$$\chi_p = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad (13)$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and the minimum eigenvalues of the matrix, respectively.

3.1. The Eigenvalue Spread as a Function of the Largest Harmonic Frequency

First of all, we investigate the behavior of the eigenvalue spread as a function of the largest harmonic frequency $\omega_{\max} = (L+1)\omega_0$. In fact, we vary $\omega_{\max}$ by varying ω_0 with L fixed.

For $p=2$, the eigenvalue spread of R_2 is given by

$$\begin{aligned} \chi_2 &= \frac{r_0 + r_1}{r_0 - r_1} \\ &= \frac{\sum_{k=1}^{L}\left[\frac{|A_k|^2}{2}(1+\cos(k\omega_0))\right]}{\sum_{k=1}^{L}\left[\frac{|A_k|^2}{2}(1-\cos(k\omega_0))\right]} \end{aligned} \quad (14)$$

assuming that $r_1 > 0$. In this case, $r_1 = \sum_{k=1}^{L}\frac{|A_k|^2}{2}\cos(k\omega_0)$ is positive since it is dominated by first few largest spectral lines. Moreover, $r_1 \to r_0$ as $\omega_0 \to 0$. Thus, as $\omega_0 \to 0$, $\lambda_{\max} = (r_0 + r_1) \to 2r_0$ while $\lambda_{\min} = (r_0 - r_1) \to 0$. As a result, χ_2 increases as $\omega_0 \to 0$.

For $p=3$, the eigenvalues of R_3 are found to be[3]

$$\begin{aligned} \lambda_1 &= r_0 - r_2 \\ \lambda_2 &= r_0 + \frac{r_2}{2} + \sqrt{\frac{r_2^2}{4} + 2r_1^2} \\ \lambda_3 &= r_0 + \frac{r_2}{2} - \sqrt{\frac{r_2^2}{4} + 2r_1^2}. \end{aligned} \quad (15)$$

In practice, $r_2 > 0$ for most audio signals. Therefore, $\lambda_{\max} = \lambda_2$ and $\lambda_{\min} = \lambda_3$. It is easy to see that

$$\begin{aligned} \lambda_{\max} &= r_0 + \frac{r_2}{2} + \sqrt{\frac{r_2^2}{4} + 2r_1^2} \\ &> r_0 + \frac{r_2}{2} + \sqrt{\frac{r_2^2}{4} + 2r_2^2} \\ &= r_0 + 2r_2 \\ &= \sum_{k=1}^{L}\left[\frac{|A_k|^2}{2}(1+2\cos(2k\omega_0))\right] \end{aligned} \quad (16)$$

and

$$\begin{aligned} \lambda_{\min} &= r_0 + \frac{r_2}{2} - \sqrt{\frac{r_2^2}{4} + 2r_1^2} \\ &< r_0 + \frac{r_2}{2} - \sqrt{\frac{r_2^2}{4} + 2r_2^2} \\ &= r_0 - r_2 \\ &= \sum_{k=1}^{L}\left[\frac{|A_k|^2}{2}(1-\cos(2k\omega_0))\right] \end{aligned} \quad (17)$$

As $\omega_0 \to 0$, $\lambda_{\max} = (r_0 + 2r_2) \to 3r_0$ while $\lambda_{\min} = (r_0 - r_2) \to 0$. Again, χ_3 increases as $\omega_0 \to 0$.

To demonstrate these analytical expectations, the magnitude values of first 19 discrete harmonic spectra have been extracted from the spectrum shown in Figure 1. The eigenvalue spreads for $p=2$ and $p=3$ are then calculated by varying $\omega_{\max} = (L+1)\omega_0$, where $L=19$. As shown in Figure 2, the eigenvalue spreads are increasing sharply as $\omega_0 \to 0$. It is very important to notice that the eigenvalue spread is not a function of ω_0 but $\omega_{\max} = (L+1)\omega_0$. That is, for a fixed ω_0, the eigenvalue spread is still increasing as L increases.

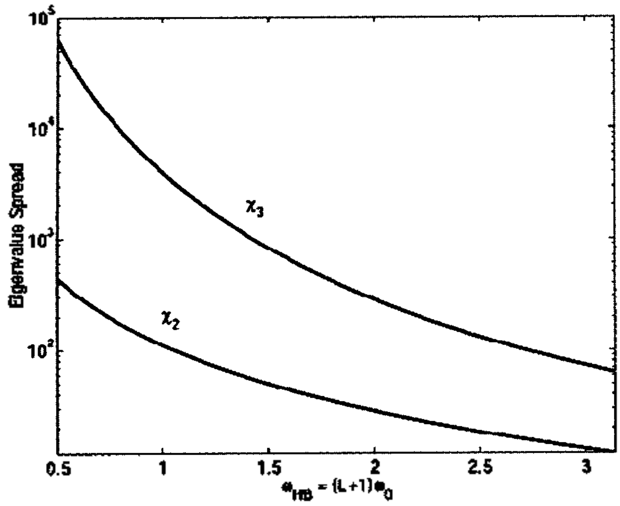

Figure 2. The eigenvalue spread with respect to $(L+1)\omega_0$ when $p=2$ and $p=3$.

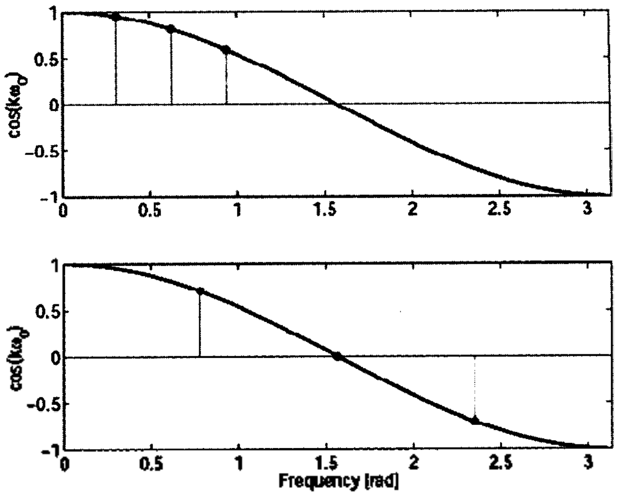

Figure 3. The effect of frequency redefinition: before (upper) after (lower) redefinition.

3.2. The Eigenvalue Spread as a Function of Filter Order

As shown in Figure 2, $\chi_2 < \chi_3$ for all $\omega_{\max}$. In fact, $\chi_p < \chi_{p+1}$ in general as we are going to prove. For filters of order higher than $p>3$, it is not possible to find any analytical expression for χ_p. However, we can predict its behavior from the Bordering theorem which states an interlacing property of the eigenvalues of the Hermitian matrix[4]: if $\{\lambda_i\}_{i=1}^{p}$ and $\{\tilde{\lambda}_i\}_{i=1}^{p+1}$ are the ordered eigenvalues of R_p and R_{p+1}, respectively, then those eigenvalues are interlaced each other as

$$0 < \tilde{\lambda}_1 \leq \lambda_1 \leq \tilde{\lambda}_2 \leq \lambda_2 \leq \cdots \tilde{\lambda}_p \leq \lambda_p \leq \tilde{\lambda}_{p+1}. \tag{18}$$

As a matter of fact, it is conjectured that, as p increases, the minimum eigenvalue approaches zero rapidly while and the maximum eigenvalue is bounded by pr_0. This makes the eigenvalues spread χ_p increases sharply as p increases for a given $\omega_{\max} = (L+1)\omega_0$.

IV. A Stabilization Method

From the investigation in the previous section, it is shown that the autocorrelation matrix may become ill-conditioned when $\omega_{\max} = (L+1)\omega_0$ is much less than the maximum signal frequency π. Furthermore, the situation becomes worse as filter order p increases. Stability monitoring and checking methods, normally employed, are insufficient for obtaining adequate spectral estimates.

One solution to this problem is to redefine the maximum signal frequency correspond to the maximum harmonic frequency $\omega_{\max}$. Specifically, a set of discrete frequencies $\{\omega_k,\ 1 \leq k \leq L\}$ is mapped into $\{\tilde{\omega}_k, 1 \leq k \leq L\}$ by

$$\tilde{\omega}_k = \pi \frac{\omega_k}{\omega_{\max}}. \tag{19}$$

Such frequency redefinition has been proposed by Makhoul in terms of "selective LP" in order to improve estimation accuracy by applying all-pole filters selectively to various parts of the signal spectrum[1]. In our study, however, its conditioning property that resolves the numerical instability of linear prediction is illuminated. The effect of frequency redefinition above is depicted in Figure 3 which shows $\cos(k\omega_0)$ terms in (14). Clearly, $\sum_{k=1}^{L} \cos(k\omega_0)$ in (14) is reduced greatly after frequency redefinition.

In order to illustrate the advantage of the proposed stabilization method, the eigenvalue spread of R_p has been computed as a function of p for the set of 19 harmonics obtained from the real audio signal shown in Figure 1. The fundamental frequency is fixed to $f_0 = 250Hz$.

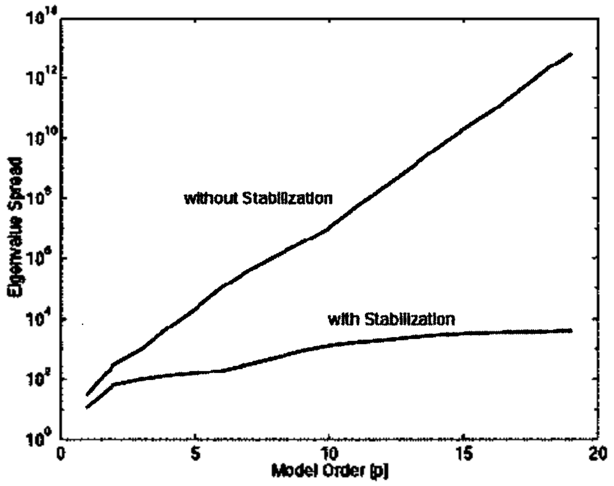

Figure 4. The eigenvalue spread of the autocorrelation matrix with respect to filter order p for LP with and without stabilization.

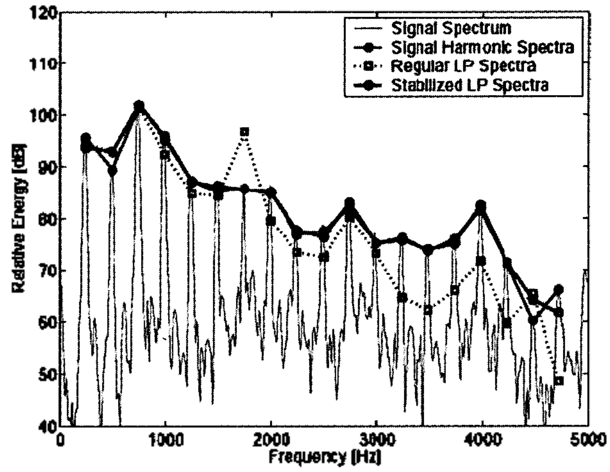

Figure 5. Estimation of discrete harmonic spectra using LP with and without stabilization.

As expected from the Bordering theorem, the eigenvalue spread is increasing exponentially as p increases as shown in Figure 4. If order p is chosen to be the number of harmonics, numerical instability would be a serious problem in regular LP without stabilization. With the help of the proposed stabilization method, however, the eigenvalue spread can be maintained low enough.

V. Simulation Results

Results of spectral estimation for a real audio signal using LP with and without stabilization of order p=19 are shown in Figure 5. Large spectral error for regular LP without stabilization is clear. In fact, numerical instability due to the ill-conditioned autocorrelation matrix prevents the Levinson-Durbin algorithm from computing a_k. Therefore, filter order is limited to some $p_0 < p$ to guarantee numerical stability. On the other hand, LP with stabilization is free from any numerical instability and provides good spectral matching.

Next, the stabilization method has been applied to DAP modeling. To measure performance of the proposed method numerically, we define the total spectral error by

$$SD = \sum_{k=1}^{L} \left| 10\log_{10}(P(k\omega_0)) - 10\log_{10}(\hat{P}(k\omega_0)) \right|.$$

In simulation, DAP is devised to iterate until the decreasing rate of spectral error is smaller than some fixed threshold. Although DAP is devised to minimize spectral error recursively, regular DAP without stabilization fails to minimizes the spectral error sufficiently due to numerical instability caused by the ill-conditioned autocorrelation matrix as shown in Figure 6. On the other hand, stabilized DAP minimizes the spectral error further efficiently as shown in Figure 7.

VI. Conclusions

The eigenvalue spread of the autocorrelation matrix for discrete harmonic spectra of audio signals is investigated. It is shown that the autocorrelation matrix is seriously ill-conditioned when the largest harmonic frequency is relatively small compared to the half of the sampling frequency, which is very common in audio signals. Numerical instability caused by the ill-conditioned autocorrelation matrix affects implementation of the algorithm as well as estimation accuracy. To improve the eigenvalue spread, we proposed a simple and efficient stabilization method that makes harmonic frequencies virtually larger by redefining the maximum frequency. It is demonstrated that the proposed stabilization method eliminates the

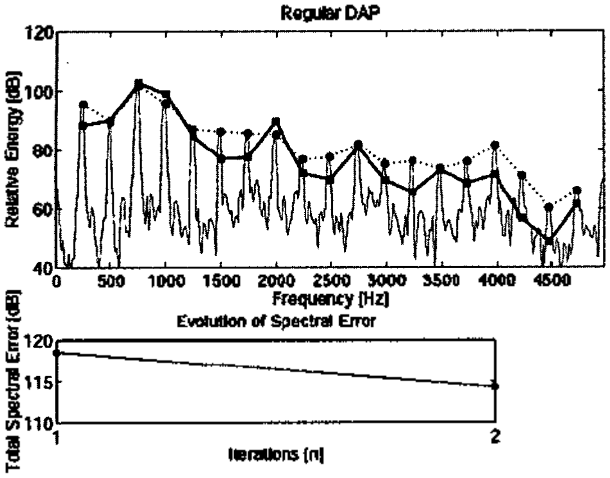

Figure 6. Estimation of discrete harmonic spectra (dashed line with circles) using regular DAP without stabilization (solid line with squares).

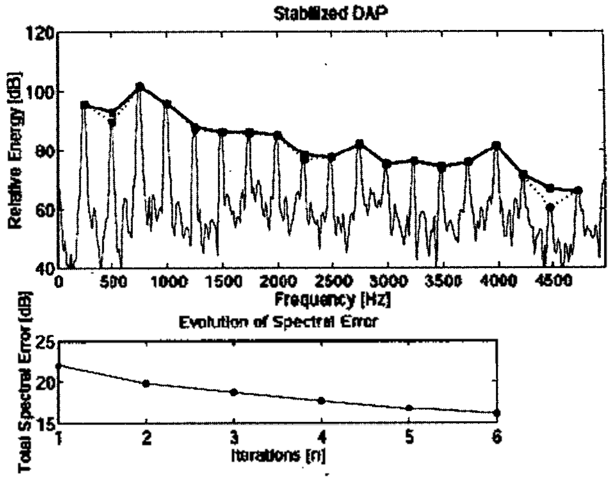

Figure 7. Estimation of discrete harmonic spectra (dashed line with circles) using stabilied DAP (solid line with squares).

numerical instability and improves spectral estimates for both LP and DAP.

Acknowledgements

This study was supported by ETRI in 2000.

References

1. J. Makhoul, "Spectral linear prediction: Properties and applications," *IEEE Trans. Signal Processing,* vol. ASSP-23, pp. 283-296, 1975.
2. A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Processing,* vol. 39, pp. 411-423, 1991.
3. J. Makhoul, "On the eigenvectors of symmetric teoplitz matrices," *IEEE Trans. Signal Processing,* vol. ASSP-29, pp. 868-872, 1981.
4. Monson H. Hayes, *Statistical Digital Signal Processing and Modeling,* John Wiley & Sons, Inc, 1996.

[Profile]

• Seung Hyon Nam

Seung Hyon Nam received the B.S. degree from Sogang University, Korea, in 1980, the M.S. degree from the University of Alabama, Huntsville, in 1987, and the Ph.D. degree from Texas A&M University, College Station TX, in 1992. From 1979 and 1985, he was with Agency for Defense Development in Korea as a Research Engineer. Since 1993, he has been with Paichai University, Taejon, Korea, where he is currently an Associate Professor in the Department of Electronic Engineering. His research interests include speech and audio processing, data compression, adaptive filter theory.

• Kyeongok Kang

Kyeongok Kang was born in Namhae in 1962. He received B.S. and M.S. degrees both in physics from Pusan National University in 1985 and 1988, respectively. He is now a graduate student for Ph. D. degree in electrical engineering at Hankuk Aviation University. He has joined in ETRI since 1991, and he is now a senior member of engineering staff at and a leader of Interactive Broadcasting Research Team. His major interests are in low-bitrate audio coding, such as sine transform coder, audio signal processing, and MPEG-7 related issues.

• Jin Woo Hong

Jin Woo Hong was born on April 15, 1959 in Yujoo, Korea. He received the B.S. and M.S. degree in Electronic Engineering from Kwangwoon University, Seoul, Korea, in 1982 and 1984, respectively. He also received the Ph.D. degree in Computer Engineering from the same university in 1993.

Since 1984, he has been with ETRI (Electronics and Telecommunications Research Institute) in Taejon, Korea, as a principal member of engineering staff, where he is currently a team manager of the Broadcasting Contents Research Team. From 1998 to 1999, he researched at Fraunhofer Institute in Erlangen, Germany, as a visiting researcher. His research interests include audio & speech coding, acoustic signal processing, audio watermarking, and IPMP.