

論文93-30B-10-4

적응디지털필터를 사용한 음질향상 방법

(A New Speech Enhancement Method Using Adaptive Digital Filter)

林龍勳*, 金元九*, 車日煥*, 尹大熙*

(Yong Hoon Lim, Weon Goo Kim, Il Whan Cha and Dae Hee Youn)

要約

본 연구에서는 적응디지털필터를 사용하여 주변 잡음에 의해 손상된 음성의 음질을 향상시키는 새로운 방법을 제안하였다. 두 센서신호는 마이크로폰으로부터 얻은 음성신호와 목에 부착한 진동센서로부터 얻은 신호를 말한다. 두 센서는 동일한 근원신호(source signal)를 갖기 때문에 상관관계가 큰 신호이며, 진동센서신호는 주변의 잡음에 영향을 거의 받지 않는 특성이 있다. 제안한 음질향상 방법에서는 이러한 두센서신호 사이의 선형의존특성을 이용하여 두신호 사이의 최적선형시스템을 추정하며, 추정된 시스템에 진동센서신호를 통과시켜 잡음의 영향이 제거된 음성신호를 얻는다. 시스템 추정 과정에서는 음성의 시변 특성을 반영하기 위해 적응디지털필터를 사용하였다. 시뮬레이션 결과 백색잡음이 첨가되어 0 dB의 SNR을 갖는 음성이 7 dB 정도로 향상되었다.

Abstract

In this paper, a new speech enhancement method for speech signal corrupted by environmental noise is proposed. Two signals are obtained from the microphone and from the accelerometer attached to the neck, respectively. Since two signals are generated from the same source signal, both signals are closely correlated. And environmental noise has no effect on the accelerometer signal. The speech enhancement system identifies the optimum linear system between two signals on the basis of the dependence between the signals. The enhanced speech can be obtained by filtering the noise-free accelerometer signal. Since the characteristics of the speech signal and environmental noise are changing with time, adaptive filtering system has to be used for characterizing the time-varying system. Simulation results show 7dB enhancement with 0dB speech signal level relative to the white noise.

1. 서론

* 正會員, 延世大學校 電子工學科
(Dept. of Elec. Eng., Yonsei Univ.)

※본연구는 연세대학교 학술 연구비 지원에 의해 이루어진 것임.

接受日字: 1993年 9月 14日

음질향상시스템은 주변 잡음의 영향으로 인하여 음성신호처리시스템의 성능이 저하하는 현상을 방지하는 목적으로 사용되는 경우가 많기 때문에 시스템이 사용되는 잡음환경에 적합하도록 설계되어야 한다.

기존의 음질향상시스템의 대부분은 하나의 센서(마이크로폰)만을 사용하는 시스템으로서, 신호의 통계적 특성을 가정하거나 잡음에 의해 손상된 음성으로부터 음성 또는 잡음만의 통계량 추출과정이 별도로 필요하다.^[1] 그러나 신호의 특성을 가정하거나 별도의 통계량 추출과정을 사용하는 것은 음질향상시스템을 실제의 잡음환경에 최적화시키는 문제에 많은 제한을 준다.

다수의 센서를 사용하는 음질향상방법은 잡음 또는 음성의 특징을 분리된 센서를 통하여 얻을 수 있으므로, 하나의 센서만을 사용하는 방법^[1]에 비하여 좋은 성능을 갖는 음질향상 시스템을 구성할 수 있다. 이러한 시스템의 하나로 잡음이 발생하여 음성에 첨가되기까지의 경로를 적응디지털필터로 모델링하여 음성에 포함된 잡음을 제거하는 적응잡음제거방법(adaptive noise cancelling)^{[2] [8]}이 연구되고 있다. 적응잡음제거방법은 자동차나 비행기와 같은 제한된 공간에서 좋은 성능을 나타내지만 높은 차수의 디지털필터를 사용해야 하므로 많은 계산량을 요구한다는 단점이 있다.^[2] 최근에 진동센서를 목에 부착하여 얻은 신호가 음질향상에 유용한 정보라는 연구결과가 보고되었다.^[3] 진동센서에 수신되는 신호는 음의 발생원으로부터 피부를 통하여 전달되기 때문에 배경잡음의 영향을 거의 받지 않는 특성이 있다. 이 신호를 잡음에 의해 손상된 음성에 적당한 크기로 첨가하면, 잡음의 영향을 상대적으로 억압하여 음성의 이해도 향상에 상당한 효과를 나타낸다. 이러한 음질향상방법은 전투기 소음 환경에서의 음질향상시스템에 응용되어 좋은 성능을 나타내었다.^{[3] [4] [5]}

본 연구에서는 마이크로폰만을 사용하는 음질향상방법을 적용하기 어려운 잡음환경, 예를들면 전투기 기내와 같은 제한된 공간이나 잡음에 대한 정보 추출이 어려운 환경에서 성대의 진동에 관한 신호를 음질향상에 부가적인 정보로 이용하는 음질향상방법을 제안하였다. 제안된 방법에서는 진동센서신호로부터 마이크로폰에 수신된 음성 신호의 파형(waveform)을 추정하여 잡음의 영향을 제거한다. 이는 두 신호 사이의 시스템인식(system identification) 문제로 생각할 수 있다. 또한 두 신호 사이의 시스템은 시변(time varying)이므로 적응디지털필터를 사용하여 시변 특성을 반영하였다. 제안된 방법은 적응잡음제거방법과 유사한 구조를 갖지만, 두 센서에 감지되는 신호의 발생경로를 모델링하기 때문에 잡음의 발생경로를 모델링하는 적응잡음제거방법 보다 적은 차수의 디지털필터로 구성이 가능하며, 진동센서에 수신되는 신호에서 잡음의 양은 거의 무시할 수 있기 때문에

잡음의 특성에 영향을 적게 받는 장점이 있다.

II. 신호의 발생모델

음질향상시스템에서 이용하는 신호는 그림 1에 나타낸 바와 같이 마이크로폰으로 수집되는 음성신호와 성대 부근의 목에서 진동센서를 사용하여 수집한 진동센서 신호이며, 앞으로 두센서신호(two-sensor signal)라 부른다. 두센서신호의 발생과정을 선형음원필터(linear source-filter) 모델^[6]을 사용하여 그림 2에 나타내었다. 여기신호(excitation signal) $ex(k)$ 는 무성음의 경우에는 백색잡음의 형태를 갖고, 유성음의 경우에는 주기적인 임펄스 형태로 모델링된다. 음원신호 $ex(k)$ 가 성도모델(vocal tract model) $V(z)$ 를 통과하여 음성신호 $s(k)$ 로 출력되고, 진동센서에 감지되는 신호 $x(k)$ 는 성대의 진동에 관한 신호 $g_h(k)$ 와 음성의 발생과정중에 피부를 통해서 전달되는 음성신호 성분 $g_s(k)$ 로 구성된다.^[3]

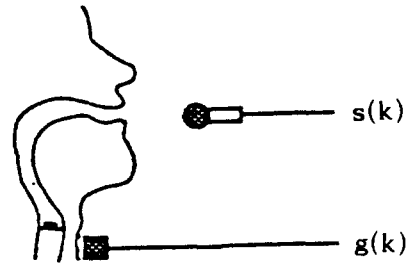


그림 1. 두센서신호
Fig. 1. Two-sensor signal.

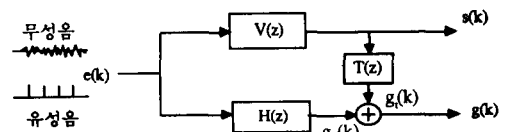


그림 2. 신호의 발생모델
Fig. 2. Signal model.

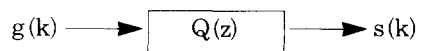


그림 3. 두센서 신호 사이의 선형시스템
Fig. 3. Linear system between the signals.

이와같이, 두신호가 동일한 근원에서 발생하여 선형시스템을 통과한다고 가정하면, 그림 3과 같이 두신호 사이의 관계를 선형시스템 $A(z)$ 로 나타낼 수 있다.

$$Q(z) = \frac{V(z)}{T(z) + H(z)} \quad (1)$$

제안한 음질향상방법은 A(z)를 선형시스템으로 인식(identification)하는 과정을 기초로 하므로 두 신호 사이의 선형성 정도는 시스템 성능에 크게 영향을 미친다. 선형성의 척도로서 선형의존도를 나타내는 MSC(Magnitude Squared Coherence) 함수^{[7] [10]}를 측정하였다. 그림 4는 유성음 /아/에 대한 두 신호 사이의 MSC함수 측정값이다. 유성음 /아/에 대한 두센서신호의 파형(waveform)과 전극모델스펙트럼(all-pole model spectrum)^[6]은 그림 6에 나타내었다. MSC함수가 1에 가까운 값을 갖는 것은 음성신호와 진동센서신호 사이에 선형관계가 크다는 것을 의미하므로, 두신호 사이의 시스템을 그림 3과 같이 선형시스템으로 근사화할 수 있다.

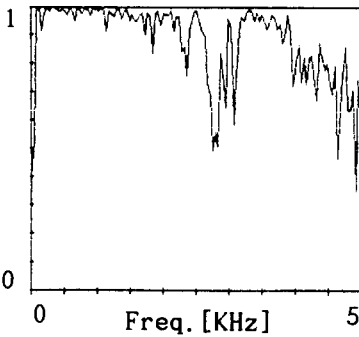


그림 4. MSC함수 (유성음 /아/)
Fig. 4. MSC function. (voiced /a/)

Ⅲ. 음질 향상 방법

1. 두센서신호 사이의 시스템인식(system identification)

그림 3에 나타낸 바와 같이 진동센서신호 g(k)가 선형 시스템 Q(z)를 통과하여 음성신호 s(k)가 발생하며, 여기에 잡음 n(k)가 첨가되어 x(k)가 얻어진다.

$$s(k) = \sum_{i=0}^{N-1} c_i g(k-i) \quad (2)$$

$$x(k) = s(k) + n(k) \quad (3)$$

여기에서 c_i는 Q(z)의 시스템계수(system parameter)를 나타낸다.

$$Q(z) = \sum_{i=0}^{N-1} c_i z^{-i} \quad (4)$$

음질향상의 문제는 손상되어 관측된 신호 x(k)로부터 음성신호의 추정치 ŝ(k)를 구하는 문제로 생각할 수 있다. 2-1절에서 고찰한 바와 같이 s(k)와 g(k)는 선형적으로 종속관계에 있는 신호이므로, 두신호 사이의 최적선형시스템 Q̂(z)을 구하고, g(k)를 통과시켜 그 출력으로 s(k)를 얻을 수 있다.

$$\hat{s}(z) = \sum_{i=0}^{N-1} \hat{c}_i g(k) \quad (5)$$

여기에서, {c_i, i = 0, ..., N-1}는 최적선형필터 Q̂(z)의 계수이며, N은 필터 차수를 나타낸다. 스테이션너리(stationary) 환경에서 g(k)와 n(k)사이 상관관계가 없다고 가정하면, 추정오차 e(k) = s(k) - ŝ(k)의 파워를 최소화하는 최적선형 시스템의 필터계수는 Wiener 필터계수^[9]로 주어진다.

성도 시스템 Q(z)는 음성의 변화와 함께 특성이 변하는 시변시스템이므로, 이를 모델링하는 시스템 Q̂(z)도 시변시스템이어야 한다. 시변시스템을 추정하는 효율적인 방법으로는 적응필터링방법이 있으며, LMS(Least Mean Square) 알고리즘^[8]은 간단하며 널리 이용되는 적응알고리즘으로 계수 추정방법은 다음과 같다.

$$e(k) = x(k) - \sum_{i=0}^{N-1} \hat{c}_i(k)g(k-i) \quad (8)$$

$$\hat{c}_i(k+1) = \hat{c}_i(k) + \mu e(k)g(k-i) \quad (9)$$

여기에서, ĉ_i(k)는 시변필터계수이며, μ는 알고리즘의 수렴상태를 조정하는 수렴상수(convergence parameter)이다.

2. 음질향상시스템

그림 5에 적응디지털필터를 사용하는 제안된 음질향상시스템의 블록도를 나타내었다. 잡음에 의해 손상된 음성 신호 x(k)와 진동센서신호 g(k)는 각각 적응디지털필터의 원시입력신호(primary signal)와 참고입력신호(reference signal)로 작용한다. 적응디

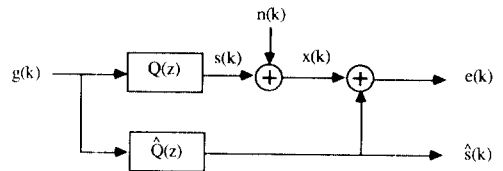


그림 5. 제안된 음질 향상 시스템
Fig. 5. The proposed system.

디지털필터 $A(z)$ 는 시스템 $A(z)$ 를 추정하며, 그 출력 $s(k)$ 는 잡음이 제거된 음성신호가 된다. 제안된 시스템은 잡음에 손상되지 않은 진동센서신호를 필터링하여 향상된 음성신호를 얻기 때문에 별도로 잡음의 특성을 측정할 필요가 없으며, 잡음의 특성이 시변인 환경에도 적용이 가능하다.

실제 응용에 있어서 성능에 영향을 미치는 요소의 하나는 적응디지털필터의 수렴속도이다. 적응 디지털 필터의 수렴속도는 음성의 시변특성을 반영하기에 충분히 빨라야 한다. 수렴속도가 늦은 경우에는 천이구간에서 음성이 왜곡되기 때문에 음성의 명료도가 저하된다.

IV. 실험 및 결과 고찰

1. 실험환경

두센서신호를 그림 1에 나타낸 바와 같이 마이크로폰과 진동센서를 사용하여 녹음한 후 표본화주파수 10KHz의 A/D 변환하여 컴퓨터 시뮬레이션을 수행하였다. 진동센서는 0Hz에서 5KHz 까지 평탄한 주파수 응답 특성을 갖는 B&K Type 4366을 사용하였다.

시뮬레이션 과정에서 사용한 적응 디지털필터의 차수는 실험 데이터 유성음 /아/에 대해 FPE(Fininal Prediction Error)^[12], AIC(Akaike Information Criterion)^[12], 필터출력신호의 SNR을 구하여 결정하였다. 유성음 /아/에 대해 50차 정도가 적당하였다.

2. 두센서신호 사이의 시스템인식(system identification) 실험

음질향상시스템은 두센서신호 사이의 시스템인식에 기초하므로, 잡음이 없는 환경에서 두센서신호 사이의 시스템인식에 관한 실험을 하였다. 유성음 /아/에서 통계적 특성이 일정한 20ms 시간구간의 음성신호와 진동센서신호로부터 최소자승법(least squares method)^[12]을 적용하여 50차의 최적필터의 계수를 구하였으며, 진동센서 신호를 통과시켜 음성신호의 추정치 $\hat{s}(k)$ 를 얻었다.

그림 6에서 (a)는 음성신호를 나타내고, (c)는 진동센서신호를 나타낸다. (b)는 진동센서신호를 최적필터에 통과시켜 얻은 추정된 음성신호이다. (d) (e) (f)는 시스템인식의 결과를 주파수영역에서 해석으로 10차 전극(all pole)모델의 전력밀도 스펙트럼^[6]을 구하였다. (e)와 같은 스펙트럼을 갖는 진동센서신호

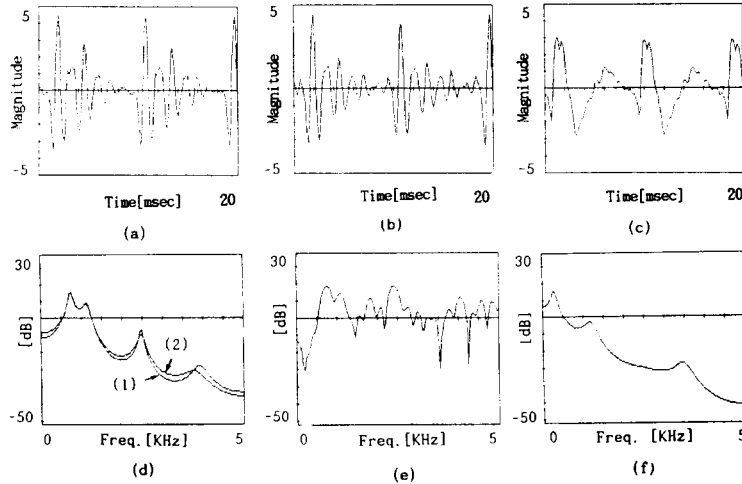


그림 6. 시스템 인식에 관한 실험

- (a) 음성 신호 /아/ (b) 추정된 음성 신호 (c) 진동센서 신호
- (d) (1)음성 신호 /아/의 전력 밀도 스펙트럼 (2) 추정된 음성신호의 전력 밀도 스펙트럼
- (e) 진동센서 신호의 전력 밀도 스펙트럼 (f) 추정된 시스템의 주파수 크기 응답

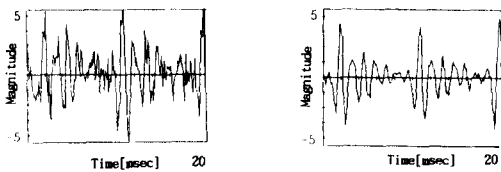
Fig. 6. Experimental results for system identification.

- (a) Speech signal /a/, (b) Estimated speech signal, (c) Accelerometer signal,
- (d) (1)Power density spectrum of the speech signal /a/, (2) Power density spectrum of the estimated speech signal,
- (e) Power density spectrum of the accelerometer signal,
- (f) Frequency magnitude response of the estimated system.

가 (f)와 같은 주파수 크기응답을 갖는 필터에 통과되어 (a)의 (1)과 같은 스펙트럼을 갖는 추정된 음성 신호를 얻었다. 그림 6에서 (a)와 (b)를 비교하고 (d)에서 (1)과 (2)를 비교하면, 추정된 시스템의 출력이 원음과 매우 유사하게 나타난 것을 관찰할 수 있다. 원래의 음성신호에 대한 추정된 음성신호의 SNR은 12.27dB 이었다. 추정된 선형 시스템의 주파수 전달함수(e)에서 음성의 포먼트(formant)에 해당하는 부분의 이득이 커져서 추정된 신호에 포먼트가 형성된 것을 관찰할 수 있다.

3. 음질향상 실험

음성신호 /아/에 인위적으로 백색잡음을 첨가하여 SNR에 따른 성능을 평가하였다. 그림 7에서 (a)는 백색잡음이 첨가되어 0dB SNR을 갖는 음성신호이고, (b)는 적응디지털필터를 사용하여 처리한 향상된 음성신호이다. 잡음이 섞인 음성(a)와 비교할 때 잡음의 영향이 많이 제거된 것을 관찰할 수 있으며, SNR이 6.88dB 개선되었다. 그림 8은 그림 7에 나타난 신호에 대한 주파수 영역의 해석으로 (a)는 음성신호, (b)는 잡음이 섞인 음성신호, (c)는 개선된 음성의 스펙트럼을 나타낸다. 개선된 음성의 스펙트럼이 낮은 주파수 영역에서는 원음과 유사하지만 높은 주파수 영역에서는 원음과 많은 차이가 있음을 관찰할 수 있다. 이러한 이유는 고주파 영역에서 두 센서신호 사이에 선형의존도가 낮은 값을 보이며, 적응디지털필터의 전달함수가 신호의 에너지가 집중된 저주파 영역에서의 오차를 줄이도록 조정되기 때문이다. 표 1은 실험에 사용된 유성음 /아/에 대한 SNR에 따른 음질개선정도를 나타낸다.



(a) (b)

그림 7. 백색 잡음에 대한 시뮬레이션 결과(시간 영역)

- (a) 백색 잡음에 의해 손상된 음성 신호 (SNR 0dB)
- (b) 향상된 음성 신호 (SNR 6.88dB)

Fig. 7. Simulation results for additive white noise (time domain).

- (a) Speech signal corrupted by additive white noise (SNR 0dB),
- (b) Enhanced speech signal (SNR 6.88dB).

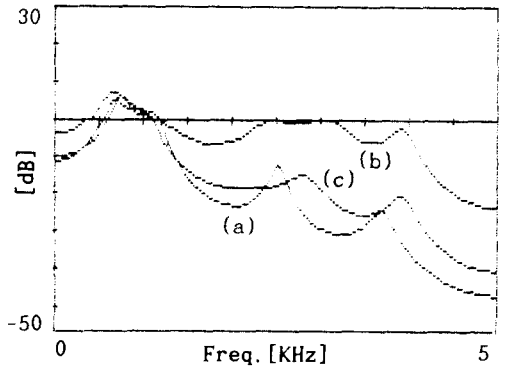


그림 8. 백색 잡음에 대한 시뮬레이션 결과(주파수 영역)

- (a) 음성신호의 스펙트럼
- (b) 백색잡음에 의해 손상된 음성의 스펙트럼

Fig. 8. Simulation results for additive white noise (frequency domain).

- (a) Spectrum of the speech signal,
- (b) Spectrum of the speech signal corrupted by additive white noise.
- (c) Spectrum of the enhanced speech signal.

표 1. SNR에 따른 음질 개선 정도

Table 1. Enhanced speech signal level relative to the SNR.

입력 신호의 SNR(dB)	-10	0	10	∞
처리된 신호의 SNR(dB)	-1.95	6.88	11.14	11.75

그림 9는 실제의 잡음 환경에서 수집한 데이터에 대한 음질 향상 실험의 결과이다. (a)는 백색 잡음 환경에서 수집한 음성 신호 /그림과 이쉬움에 가슴 조이던/의 일부이고, 10dB 정도의 segmental SNR을 갖는 것으로 추정되었다. (b)는 적응 디지털 필터를 사용하여 향상시킨 음성으로 청취 실험 결과 잡음의 영향이 많이 감소되었으나, 신호의 특성이 변화하는 부분에서 적응 디지털 필터의 수렴 특성에 의한 음질 왜곡 현상이 관찰되었다.

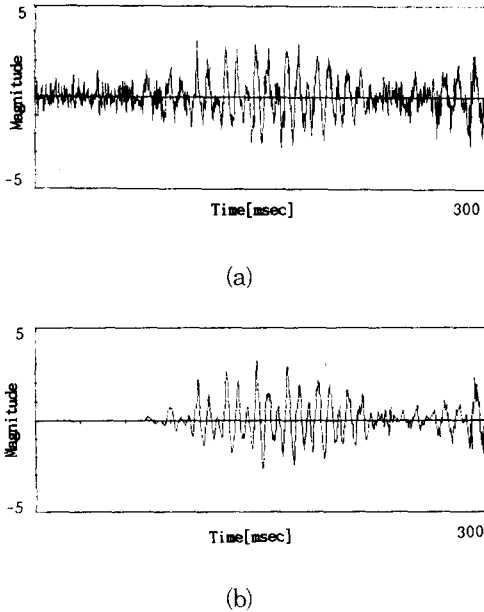


그림 9. 실제 잡음환경에서 수집한 데이터에 대한 실험 결과

Fig. 9. simulation result for data obtained in noise environment.

V. 결론

본 연구에서는 진동 센서를 사용하여 얻은 신호로부터 음성 신호에 포함된 배경 잡음의 영향을 줄이는 음질 향상 방법을 제안하였다. 제안한 시스템의 특징은 잡음에 손상되지 않은 신호를 필터링하여 향상된 음성 신호를 얻기 때문에 별도로 잡음의 특성을 측정할 필요가 없으며, 잡음의 특성이 시변인 환경에도 적용이 가능하다. 시뮬레이션 결과 백색 잡음이 첨가되어 0 dB의 SNR을 갖는 음성이 7 dB 정도로 향상되었으나, 고주파 영역에서는 저주파 영역에 비하여 성능이 저하되었다.

앞으로의 연구 과제로서 시스템의 성능에 대한 주관적 평가가 필요하며, 높은 주파수 영역에서의 성능 향상을 위한 대책과 빠른 변화를 갖는 음성 구간에 대한 음질 개선 방법이 필요하다.

參考文獻

[1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth com-

pression of noisy speech," *Proc. IEEE*, vol. 67, no.12, Dec. 1979.

- [2] S. F. Boll and D. C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation," *IEEE Trans., Speech, Signal Processing*, vol. ASSP-28, no. 6, pp. 752-754, Dec. 1980.
- [3] V. R. Viswanathan, K. F. Karnofsky, K. N. Stevens and M. N. Alakel, "Multisensor speech input for enhanced immunity to acoustic background noise," *ICASSP-84*, pp. 15A.3-1-18A.3.4, 1984.
- [4] V. R. Viswanathan, C. M. Henry and A. G. Derr, "Noise-immune speech transduction using multiple sensors," *ICASSP-85*, pp. 1911 - 1914, 1985.
- [5] V. R. Viswanathan and C. Henry, "Noise-immune multisensor speech input: formal subjective testing in operational conditions," *ICASSP-89*, pp. 373-376, 1989.
- [6] J. D. Markel, *Linear Prediction of Speech*, Springer-Verlag Berlin Heidelberg New York 1976.
- [7] D. H. Youn, A Class of adaptive methods for estimating coherence and time delay function, Ph.D. Thesis, Kansas State University, Manhattan KS., 1982.
- [8] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Inc., 1987.
- [9] H. L. Van Trees, *Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons, 1968.
- [10] J. S. Bendat, A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, John Wiley & Sons, 1980.
- [11] L. Ljung, *System Identification; Theory for user*, Englewood Cliffs, 1987
- [12] S. L. Marple, Jr., *Digital Spectral Analysis with Application*, Prentice-Hall, Inc., 1987.

著 者 紹 介



林 龍 勳(正會員)

1963年 9月 26日生. 1989年 2月 연세대학교 전자공학과 졸업. 1991年 2月 연세대학교전자공학과 석사학위 취득. 1991年 3月 ~ 현재 연세대학교 전자공학과 박사과정재학중. 주관심 분야는 적응신

호처리, 비선형적응디지털필터 및 음성신호처리 등임.



金 元 九(正會員)

1964年 4月 3日生. 1987年 2月 연세대학교 전자공학과 졸업. 1989年 8月 연세대학교전자공학과 석사학위 취득. 1989年 9月 ~ 현재 연세대학교 전자공학과 박사과정재학중. 주관심 분야는 디지

탈 신호처리, 음성신호처리 및 음성인식 등임.

尹 大 熙(正會員) 第 28卷 第 3號 參照

현재 연세대학교 전자공학과 교수

車 日 煥(正會員) 第 28卷 第 3號 參照

현재 연세대학교 전자공학과 교수