

프리엠퍼시스를 이용한 CPSP 기반의 도달시간차이 추정 성능 개선

Performance Improvement of CPSP Based TDOA Estimation Using the Preemphasis

권 흥 석*, 배 건 성**
(Hongseok Kwon*, Keunsung Bae**)

*삼성전자 무선사업부, **경북대학교 전자전기컴퓨터학부
(접수일자: 2009년 4월 7일; 채택일자: 2009년 6월 1일)

본 연구에서는 CPSP (Cross Power Spectrum Phase) 함수를 이용한 프레임 기반의 TDOA (Time Difference of Arrival) 추정시 나타나는 문제점들을 분석하고 연구하였다. 구형 윈도우함수를 이용해서 음성신호의 프레임을 추출할 때 나타나는 스펙트럼 누설현상은 CPSP 스펙트럼의 추정을 부정확하게 한다. 또한 스펙트럼 누설을 줄이기 위하여 구형이 아닌 다른 윈도우함수를 사용하여 프레임을 추출하면 프레임의 끝부분에서 발생하는 윈도우함수 가중치의 불일치 때문에 신호를 왜곡시킨다. 이 문제점들은 CPSP 기반의 TDOA 추정성능을 감소시킨다. 본 논문에서는 음성신호의 프리엠퍼시스를 이용하여 이러한 문제점들을 완화시키는 방법을 제안한다. 프리엠퍼시스된 음성신호의 다이내믹 레인지를 줄여줌으로써 스펙트럼 누설을 감소 시킨다. 제안한 프리엠퍼시스 방법을 검증하기 위하여, 다양한 잡음 및 잔향환경에서 TDOA 추정실험을 수행하였다. 실험결과 프리엠퍼시스된 마이크 출력에 구형 윈도우함수를 적용시켜 CPSP를 구할 경우 프리엠퍼시스를 하지 않거나 다른 윈도우를 사용하는 경우에 비해 TDOA 추정성능이 향상됨을 확인하였다.

핵심용어: 음원 위치 추정, TDOA 추정, CPSP 함수

투고분야: 음향 신호처리 분야 (1,1)

We investigate and analyze the problems encountered in frame-based estimation of TDOA (Time Difference of Arrival) using CPSP function. Spectral leakage occurring in framing of a speech signal by a rectangular window could make estimation of CPSP spectrum inaccurate. Framing with other windows to reduce the spectral leakage distorts the signal due to the asynchronous weighting around the frame specifically both ends of the frame. These problems degrade the performance of the CPSP-based TDOA estimation. In this paper, we propose a method to alleviate those problems by pre-emphasis of the speech signal. It reduces the influence of the spectral leakage by reducing dynamic range of the spectrum of a speech signal with pre-emphasis. To validate the proposed method of pre-emphasis, we carry out TDOA estimation experiments in various noise and reverberation conditions. Experimental results have shown that the framing of pre-emphasized microphone output by a rectangular window achieves higher success rate of TDOA estimation than any other framing methods.

Keywords: Sound Source Localization, TDOA Estimation, CPSP Function

ASK subject classification: Acoustic Signal Processing (1,1)

I. 서론

현재 음성신호를 이용한 대부분의 응용분야에서는 헤드셋 마이크로폰 또는 데스크탑 마이크로폰과 같이 단일

채널 마이크로폰이 주로 사용되고 있다. 그러나 디지털 신호처리 기술의 발달과 더불어 마이크로폰 배열을 이용한 음원의 방향 추정에 대한 연구가 꾸준히 수행되어 왔다 [1]. 마이크로폰 배열을 이용하면 주변 잡음과 간섭신호를 줄이는 동시에 고품질의 음성신호를 수집할 수 있다. 그리고 화자의 위치에 상관없이 음성신호를 수집하는 것이 용이해지며, 화자의 위치를 추정, 추적할 수도

책임저자: 권 흥 석 (ksbae@mirbbs.knu.ac.kr)
경북 구미시 임수동 94-1
삼성전자 무선사업부 개발 6그룹 S/W 파트
(전화: 053-940-8627; 053-940-8827)

있다. 이러한 특징으로 인하여 마이크로폰 배열은 원격 회의 시스템, 음성인식, 화자식별, 음성수집, 방법 시스템 등에 광범위하게 응용될 수 있다.

마이크로폰 배열을 이용할 경우 음원의 방향이나 신호의 도달시간차이 (TDOA)의 정확한 추정을 필요로 한다. 이를 위한 방법은 크게 마이크로폰 배열의 빔조향 (beam-steering)에 기반한 방법 [2], 고해상도 스펙트럼 추정에 기반한 방법 [3], GCC (Generalized Cross Correlation) 함수에 기반한 방법 등으로 구분할 수 있다. 빔조향에 기반한 방법은 근거리 (near-field) 화자인 경우에 가능한 모든 화자의 위치에 대해서 조사해야 되므로 많은 연산량을 필요로 하며, 고해상도 스펙트럼 추정에 기반한 방법은 협대역 신호에 대해서 고안된 방법으로서 주파수 분해와 고유치 분해로 인한 많은 연산량이 요구되므로 광대역 신호인 음성신호에 적용하기에는 부적합하다. GCC 기반의 도달시간차이 추정 방법은 마이크로폰 쌍의 출력신호에서 상호전력스펙트럼을 구하고 적당한 프리필터 (prefilter)를 곱하여 얻은 GCC 함수의 최대값을 찾아서 도달시간차이로 간주하는 방법이다 [4][5]. 이 방법은 알고리즘이 간단하고, 적은 연산량으로 만족할 만한 성능을 갖는다는 장점을 가지기 때문에 널리 사용된다.

GCC 함수를 이용하여 도달시간차이를 추정할 때 여러 가지 프리필터가 제안되었는데, 잔향이 존재하는 환경에서는 일반적으로 CPSP (Cross Power Spectrum Phase) 함수가 가장 우수한 성능을 갖는다고 알려져 있다. CPSP 함수는 상호상관 스펙트럼의 위상만을 이용하여 도달시간차이를 추정하는 방법이다. 실제로 CPSP 함수를 이용하여 도달시간차이를 추정하기 위해서는 이산 신호 형태의 마이크로폰 출력신호를 DFT하여 이산 주파수 영역에서 수행된다. 특히 유성음 구간에서 프레임 단위의 CPSP 함수로 도달시간차이를 추정하면 잡음과 잔향이 존재하지 않음에도 불구하고 도달시간차이의 추정이 실패하는 경우가 빈번히 발생한다. 이러한 이유는 유성음 구간에서 구형 윈도우로 프레임을 구성할 때 스펙트럼 누설로 인하여 GCC 스펙트럼이 부정확하게 되기 때문이다. 해밍 윈도우 (Hamming window)와 같은 다른 윈도우를 이용하면 이러한 스펙트럼 누설의 영향은 줄일 수 있지만, 시간적으로 동일한 신호에 다른 윈도우 가중치를 곱하기 때문에 신호의 왜곡이 발생되어 추정 성능이 저하된다. 본 논문에서는 프레임 단위의 CPSP 함수를 이용한 도달시간차이 추정시에 생길 수 있는 문제점을 제시하고, 이를 해결하기 위하여 구형 윈도우를 사용하면서 프리엠퍼시스로 마이크로폰 출력신호의 스펙트럼 동적범위를

줄여 줌으로써 추정 성능을 개선할 수 있음을 보인다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 이산 주파수 영역에서 CPSP 함수를 이용하여 도달시간차이를 추정하는 방법을 설명하고, 구형 윈도우로 프레임을 구성할 때 생길 수 있는 CPSP 방법의 문제점을 제시한다. 3장에서는 제시된 문제점의 원인을 분석하고 이를 해결하기 위하여 프리엠퍼시드된 마이크로폰 출력신호에 구형 윈도우로 프레임을 구성하는 방법을 제안한다. 4장에서는 제안한 방법의 성능을 평가하기 위하여 수행된 모의 실험 결과를 비교, 분석하며, 마지막으로 5장에서 결론을 맺는다.

II. GCC 함수를 이용한 도달시간차이 추정

2.1. GCC 함수를 이용한 도달시간차이 추정

마이크로폰 쌍의 출력신호 $x_i(n)$ 은 식 (1)과 같이 음성 신호 $s(n)$ 이 실내공간 임펄스 응답 (room impulse response) $h_i(n)$ 을 통과하고 주변 잡음 $v_i(n)$ 이 첨가된 형태로 표현할 수 있다. 그리고 N -포인트 DFT로 구한 이산 스펙트럼 $X_i(k)$ 는 식 (2)와 같이 표현된다. 마이크로폰 쌍의 GCC 함수를 이용한 도달시간차이 추정을 위해서는 먼저 각 마이크로폰 출력신호의 스펙트럼을 곱하여 상호 전력 스펙트럼을 구한다. 그리고 상호 전력 스펙트럼에 적당한 프리필터 $\Psi(k)$ 를 곱하여 GCC 스펙트럼 $C(k)$ 를 구하고, IDFT로 GCC 함수 $\hat{c}(m)$ 을 얻는다. 이때 식 (3)의 *는 공액 복소수를 의미한다.

$$x_i(n) = s(n) * h_i(n) + v_i(n), \quad i=1,2 \quad (1)$$

$$X_i(k) = S(k) \cdot H_i(k) + V_i(k), \quad i=1,2 \quad (2)$$

$$\begin{aligned} \hat{c}(m) &= \sum_{k=0}^{N-1} \Psi(k) \cdot X_1(k) \cdot X_2^*(k) e^{j\frac{2\pi mk}{N}} \\ &= \sum_{k=0}^{N-1} C(k) e^{j\frac{2\pi mk}{N}} \end{aligned} \quad (3)$$

CPSP 함수는 식 (4)와 같은 프리필터를 가지는 GCC 함수로서 모든 상호 전력 스펙트럼의 크기를 1로 만든다. 최종적인 도달시간차이는 식 (5)와 같이 GCC 함수가 피크를 나타내는 샘플 위치로 결정된다.

$$\Psi(k) = \frac{1}{|X_1(k) \cdot X_2^*(k)|} \quad (4)$$

$$\hat{m} = \begin{cases} \arg \max_m [\hat{c}(m)], & \arg \max_m [\hat{c}(m)] < \frac{N}{2} \\ \arg \max_m [\hat{c}(m)] - N, & \frac{N}{2} \leq \arg \max_m [\hat{c}(m)] \end{cases} \quad (5)$$

2.2. 구형 윈도우로 구성된 프레임의 문제점

그림 1은 마이크로폰 출력신호가 유성음이고 구형 윈도우로 프레임을 구성하였을 때, 각 마이크로폰 출력신호와 GCC의 전력 스펙트럼, 각 마이크로폰 출력신호와 GCC의 위상 스펙트럼, 정규화된 CPSP 함수를 각각 나타

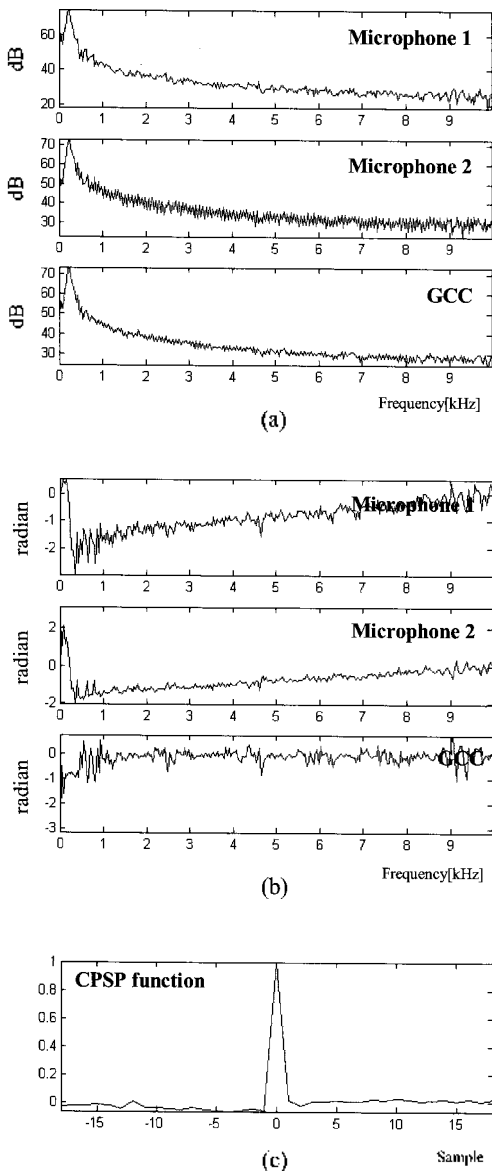


그림 1. 구형 윈도우로 구성된 유성음 프레임의 예 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수
Figure 1. An example of voiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.

낸 것이다. 이때 샘플링율과 프레임의 길이는 20 kHz와 256 샘플로 두고, 512-포인트 FFT를 사용하였다. 그리고 첫번째 마이크로폰 출력신호는 식 (6)과 같이 두번째 마이크로폰 출력신호를 12 샘플 지연시켜서 사용하였다. 따라서 도달시간차이를 정확히 검출한 경우에는 CPSP 함수가 12 샘플에서 피크를 보인다.

$$x_1(n) = x_2(n-12) \quad (6)$$

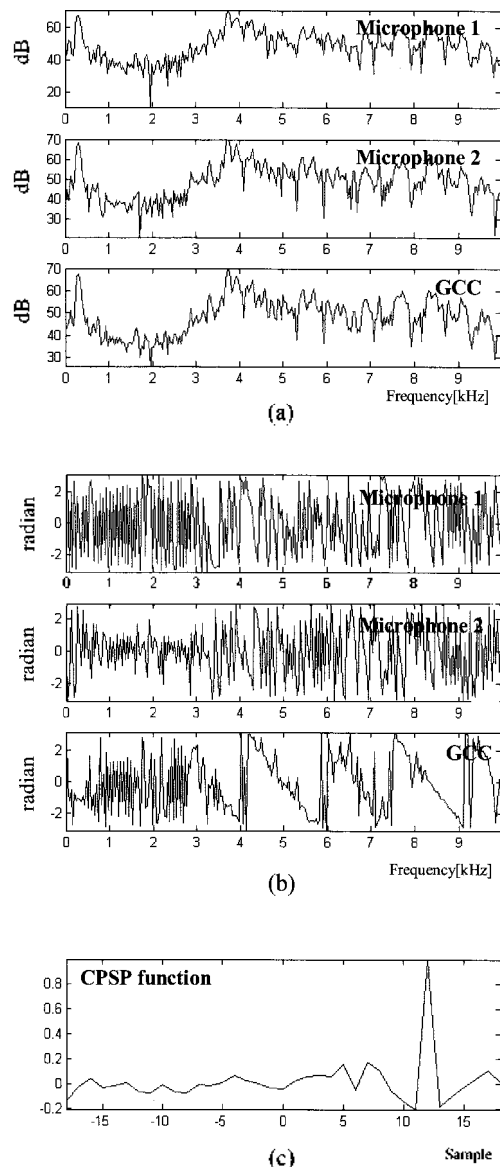


그림 2. 구형 윈도우로 구성된 무성음 프레임의 예 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수
Figure 2. An example of unvoiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.

CPSP 함수는 GCC 전력 스펙트럼을 모두 1로 두고 오직 위상만을 이용하여 도달시간차이를 추정한다. 그림 1 (b)와 1(c)를 보면 알 수 있듯이 잔향과 잡음이 전혀 없음에도 불구하고 GCC 위상 스펙트럼이 거의 0도 부근에 존재하고, CPSP 함수의 피크가 0 샘플에서 나타나기 때문에 도달시간차이 추정에 실패한다.

그림 2는 무성음 구간에서 구형 윈도우로 프레임을 구성한 경우를 나타낸 것이다. 그림 2 (b)의 GCC 위상 스펙트럼을 보면 12 샘플에 해당하는 선형위상이 비교적 잘 나타남을 알 수 있다. 따라서 그림 2 (c)의 CPSP 함수가 12 샘플에서 피크를 나타냄으로써 도달시간차이를 정확히 추정한다. 그림 1과 그림 2로부터 알 수 있듯이 CPSP 함수로 도달시간차이를 추정할 때 구형 윈도우로 프레임을 구성하면 무성음 구간에서는 도달시간차이를 잘 추정하지만, 유성음 구간에서는 잡음과 잔향이 없음에도 불구하고 도달시간차이를 정확히 추정하지 못하는 경우가 발생한다.

III. 프리엠퍼시스를 이용한 CPSP 함수의 성능 개선

3.1. 구형 윈도우로 구성된 프레임의 분석

유성음 구간에서 구형 윈도우로 프레임을 구성하는 경우에 발생하는 CPSP 함수의 도달시간차이 추정 실패 원인을 분석하기 위하여 20 kHz 샘플링을 가지며 기본 주파수가 160 Hz인 음성신호를 합성하였다. 이 신호를 두번째 마이크로폰 출력신호로 두고, 첫번째 마이크로폰 출력신호는 두번째 마이크로폰 출력신호를 12 샘플 지연시켜서 사용하였다. 따라서 도달시간차이를 정확히 추정하는 경우에는 12 샘플에서 피크를 가진다. 한 프레임의 길이는 256 샘플로 두었으며, 512-포인트 FFT를 사용하였다. 이때 신호의 스펙트럼 포락선은 앞 절에서 제시한 유성음과 무성음의 스펙트럼 포락선을 갖도록 하였으며, 각 하모닉의 위상은 균일 분포를 갖는 랜덤 위상으로 모델링하였다.

그림 3은 합성한 유성음에 대하여 구형 윈도우로 프레임을 구성한 경우를 나타낸 것이다. 그림 3 (a)는 각 마이크로폰 출력신호와 GCC의 전력 스펙트럼을 나타낸 것이다. 이때 굵은 선은 전력 스펙트럼이 가장 큰 주파수 성분에 해당하는 구형 윈도우의 스펙트럼 포락선을 보인 것이다. 그림 3 (b)는 각 마이크로폰 출력신호와 GCC의 위상 스펙트럼을 나타낸 것이고, 그림 3 (c)는 정규화된 CPSP

함수를 그린 것이다. 그림을 보면 구형 윈도우의 스펙트럼 누설과 다른 주파수 성분의 전력 스펙트럼이 비슷한 크기를 가짐을 알 수 있다. 그리고 큰 전력 스펙트럼을 갖는 주파수 성분의 스펙트럼 누설이 대부분 주파수의 스펙트럼에 더해짐으로써 정확한 스펙트럼을 표현하지 못한다. 따라서 그림 3 (c)와 같이 CPSP 함수의 피크가 0 샘플에서 발생하여 도달시간차이 추정에 실패한다.

구형 윈도우는 부엽 (sidelobe)의 감쇄가 작아서 스펙트럼 누설이 크다. 따라서 큰 전력을 갖는 주파수 성분의 스펙트럼이 다른 주파수의 스펙트럼에 영향을 많이 끼친

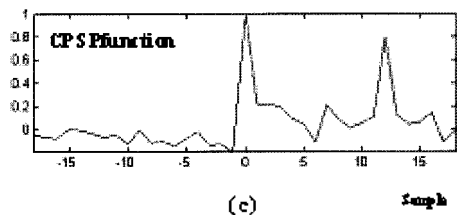
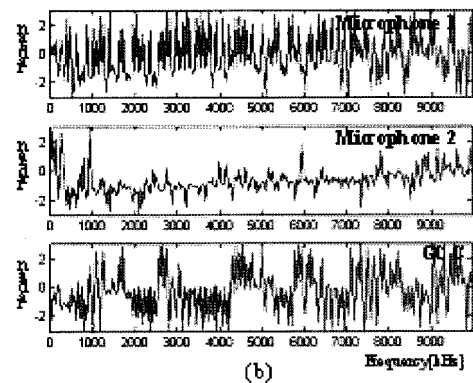
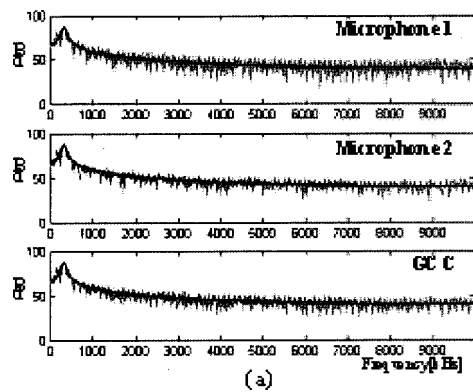


그림 3. 구형 윈도우로 구성된 유성음 프레임의 문제점 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (굵은 선: 구형 윈도우의 스펙트럼 포락선) (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수

Figure 3. Problem of voiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (thick lines: spectral envelope of rectangular window) (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.

다 [6]. 특히 유성음은 저주파 영역의 전력 스펙트럼이 상대적으로 아주 크기 때문에 구형 윈도우의 스펙트럼 누설이 다른 주파수의 스펙트럼에 영향을 많이 끼친다. 즉, 어떤 주파수의 스펙트럼은 실제 이 주파수의 스펙트럼과 큰 전력 스펙트럼을 갖는 주파수의 스펙트럼 누설의 합으로 표현되므로 정확한 GCC 스펙트럼의 추정을 어렵게 만든다. 특히 이런 부정확한 GCC 스펙트럼은 위상 정보만을 이용하는 CPSP 함수의 성능 저하를 초래한다.

그림 4는 합성된 무성음에 대하여 구형 윈도우로 프레임임을 구성한 경우를 나타낸 것이다. 그림 4를 보면 큰 전력 스펙트럼을 갖는 주파수 성분이 몇 군데 존재 하지만 유성음에 비해서 그 동적범위가 작다. 따라서 큰 전력을 갖는 주파수 성분의 전력 스펙트럼이 유성음에 비해서 상대적으로 작으므로 구형 윈도우의 스펙트럼 누설이 다른 주파수의 스펙트럼에 큰 영향을 미치지 못한다. 그림 4 (b)의 GCC 위상 스펙트럼을 보면 선형성이 잘 나타나고, 그림 4 (c)의 CPSP 함수가 12 샘플에서 피크를 보임으로써 도달시간차이를 정확히 추정한다.

구형 윈도우의 스펙트럼 누설로 인한 CPSP 함수의 도달시간차이 추정 실패는 스펙트럼 누설이 상대적으로 작은 해밍 윈도우를 사용하여 해결할 수 있다. 해밍 윈도우는 구형 윈도우에 비해서 스펙트럼 누설이 작기 때문에 큰 전력 스펙트럼을 갖는 주파수 성분이 존재하더라도 다른 주파수의 스펙트럼에 영향을 많이 끼치지 않는다. 그러나 해밍 윈도우는 그림 5와 같이 화자가 발성한 시간을 기준으로 하였을 때 동일한 샘플에 다른 윈도우 가중치가 곱해지므로 신호의 왜곡을 발생시킨다. 이러한 왜곡은 전향과 잡음이 없는 경우에는 성능에 큰 영향을 끼치지 않으나 4절의 실험 결과에서 제시하듯이 잡음이나 잔향이 존재하는 환경에서는 도달시간차이 추정 성능을 저하시킨다. 이는 시간적으로 동일한 샘플이 잡음과 잔향으로 인하여 완전히 일치하지 않으므로 해밍 윈도우에 의한 신호의 왜곡이 더 커지기 때문이다.

3.2. 프리엠퍼시스를 이용한 구형 윈도우의 문제점 해결

본 논문에서는 식 (7)과 같이 정의되는 프리엠퍼시스 전처리 과정을 이용하여 음성신호 스펙트럼의 동적범위를 줄여주어 유성음과 같이 큰 전력을 갖는 저주파 성분의 스펙트럼 누설이 다른 주파수의 스펙트럼에 영향을 끼치지 않도록 한다. 프리엠퍼시스는 일종의 고역통과필터로서 낮은 주파수의 전력 스펙트럼을 감쇄시킨다. 그리고 동일한 필터로 각 마이크로폰 출력신호를 필터링하

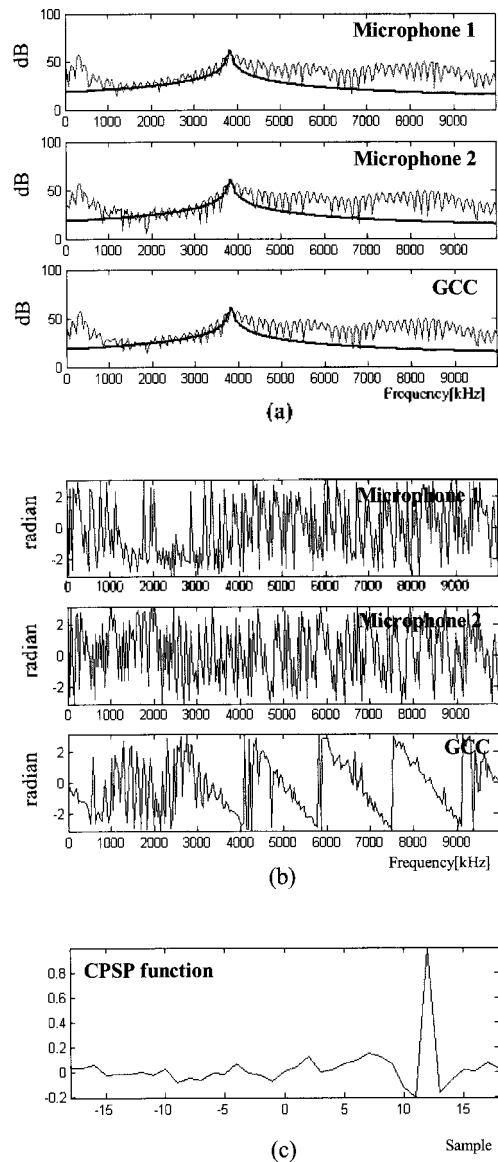


그림 4. 구형 윈도우로 구성된 무성음 프레임의 문제점 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (굵은 선: 구형 윈도우의 스펙트럼 포락선) (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수

Figure 4. Problem of unvoiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (thick lines: spectral envelope of rectangular window) (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.

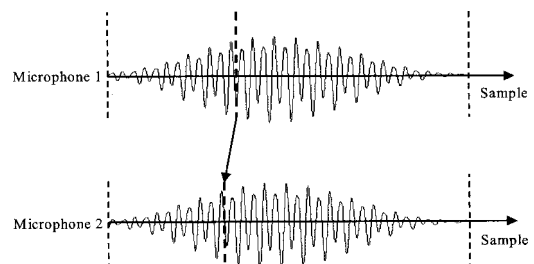


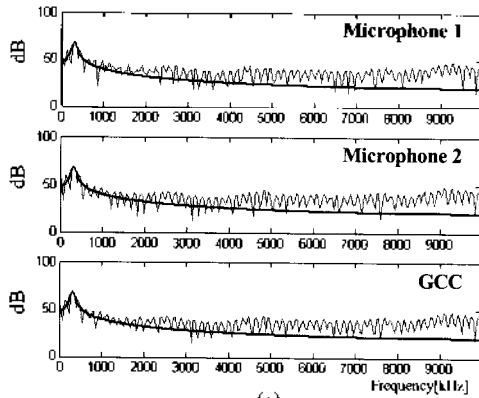
그림 5. 해밍 윈도우에 의한 신호의 왜곡
Figure 5. Signal distortion occurred by hamming window.

기 때문에 프리엠퍼시스의 위상 스펙트럼이 GCC 위상 스펙트럼에 전혀 영향을 미치지 않는다. 본 논문에서는 식 (7)의 α 를 0.95로 두고 사용하였다.

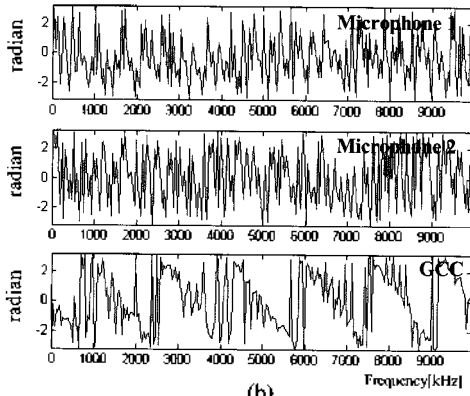
$$y_i(n) = x_i(n) - \alpha \cdot x_i(n-1), \quad i=1,2 \quad (7)$$

유성음과 무성음에 대하여 각 마이크로폰 출력신호를 프리엠퍼시스한 다음 구형 윈도우로 프레임을 구성한 경

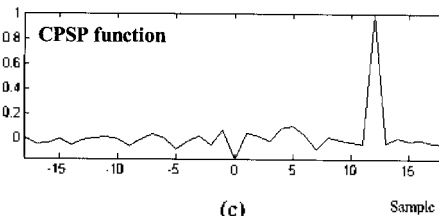
우를 그림 6과 그림 7에 각각 나타내었다. 유성음에 해당 되는 그림 6 (a)의 전력 스펙트럼을 보면 큰 전력 스펙트럼을 갖는 저주파 성분의 스펙트럼 누설이 다른 주파수의 전력 스펙트럼에 비해서 작다는 것을 알 수 있다. 따라서 구형 윈도우의 스펙트럼 누설이 다른 주파수의 스펙트럼에 크게 영향을 끼치지 못하므로 그림 6 (b)의 GCC 위상 스펙트럼에 선형성이 잘 나타난다. 결과적으로 그림 6 (c)와 같이 CPSP 함수의 피크가 12 샘플에 나타나게 되어



(a)

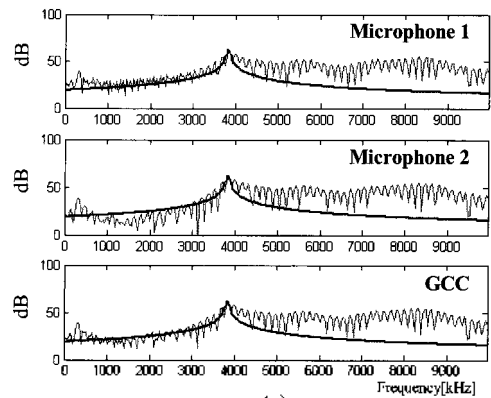


(b)

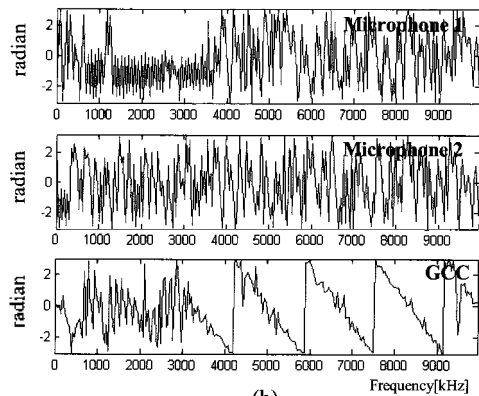


(c)

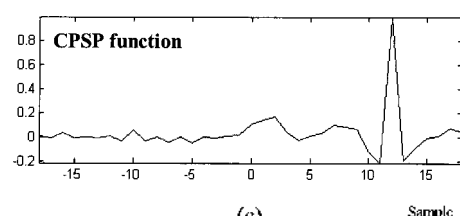
그림 6. 프리엠퍼시스된 유성음을 구형 윈도우로 구성한 프레임 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (굵은 선: 구형 윈도우의 스펙트럼 포락선) (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수
Figure 6. Pre-emphasized voiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (thick lines: spectral envelope of rectangular window) (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.



(a)



(b)



(c)

그림 7. 프리엠퍼시스된 무성음을 구형 윈도우로 구성한 프레임 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (굵은 선: 구형 윈도우의 스펙트럼 포락선) (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수
Figure 7. Pre-emphasized unvoiced frame by a rectangular window (a) Power spectra of microphone outputs and GCC (thick lines: spectral envelope of rectangular window) (b) Phase spectra of microphone outputs and GCC (c) Normalized CPSP function.

도달시간차이를 정확히 추정한다.

프리엠퍼시스를 무성음에 대하여 구형 원도우로 프레임 을 구성한 경우인 그림 7도 마찬가지로 큰 전력을 갖는 주파 수 성분의 스펙트럼 누설이 다른 주파수의 전력 스펙트럼 보다 작다. 따라서 그림 7 (b)의 GCC 위상스펙트럼에 선형 성이 잘 나타나며, 그림 7 (c)에서 CPSP 함수의 피크가 12 샘플에 나타남으로써 도달시간차이를 정확히 추정한다.

지금까지의 설명으로부터 알 수 있듯이 유성음에 대하 여 구형 원도우로 프레임을 구성하고 CPSP 함수로 도달 시간차이를 추정하면, 저주파의 스펙트럼 누설로 인하여 추정 실패가 빈번히 발생한다. 그러나 본 논문에서 제안 한 방법과 같이 프리엠퍼시스를 유성신호에 대하여 구형 원도우로 프레임을 구성하면 신호의 왜곡을 발생시키지 않으면서 스펙트럼의 동적범위를 줄여주어 스펙트럼 누 설로 인한 CPSP 함수의 도달시간차이 추정 실패를 상당 히 줄일 수 있다.

IV. 실험결과 및 분석

제안한 방법의 도달시간차이 추정 성능을 평가하기 위 하여 그림 8과 같은 실내공간에서 모의실험을 수행하였 다. 화자는 마이크로폰 쌍의 중심에서 3 m 떨어진 위치에 10도, 40도, 70도의 방향을 가지도록 위치시켰다. 실내공 간 임펄스 응답은 Allen에 의해 제안된 image 방법을 이 용하여 0초, 0.1초, 0.2초, 0.3초를 갖는 잔향시간 T_{60} 에 대하여 구하였으며, 20 kHz 샘플링율에 임펄스 응답의 길이는 8192 샘플로 두었다 [7]. 그리고 마이크로폰은 등 방향 빔패턴 (omni-directional beampattern)을 가지고, 마이크로폰이 벽면에 붙어 있으므로 후면에서 반사되는 반사파는 없다고 가정하였다. 각 잔향시간에 해당되는 반사율 β 는 식 (8)로 주어지는 Eyring 공식 [8]으로 계산 하였다. 이때 c 는 음속으로서 340 m/s로 두었고, θ 는

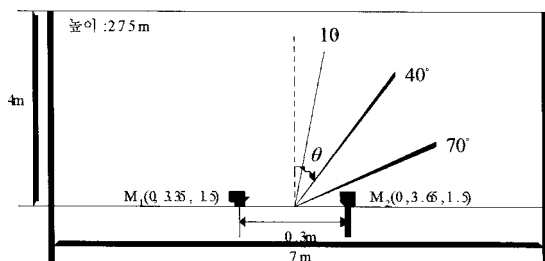


그림 8. 도달시간차이 추정 성능 평가를 위한 모의실험 환경
Figure 8. Simulation environment for performance evaluation of TDOA estimation.

마이크로폰 쌍의 수직방향을 0도로 정의하였다.

$$\beta = \exp\left(\frac{-13.82}{c \cdot (L_x^{-1} + L_y^{-1} + L_z^{-1}) \cdot T_{60}}\right) \quad (8)$$

부잔향실에서 수집된 3개의 음성신호를 각 잔향시간 에 해당되는 실내공간 임펄스 응답으로 필터링하여 잔향 신호를 만들고, 주변 잡음으로서 백색잡음을 첨가하여 신호대잡음비가 0 dB에서 40 dB까지 5 dB 간격을 갖도록 하였다. 이때 신호대잡음비는 식 (9)와 같이 잔향신호와 잡음의 전력비로 정의하였고, $r_i(n)$, $v_i(n)$, N_i 은 각각 번 째 마이크로폰의 잔향신호, 백색잡음, 실험에 사용된 신 호의 샘플 개수를 의미한다. 실험에서 구한 GCC 함수에 cubic-spline 보간을 수행하여 샘플 단위의 해상도를 0.1 샘플 단위로 높였다 [9]. 그리고 도달시간차이를 추정할 때 수행되는 피크 검출은 최대 도달시간차이 구간 내에서 만 수행하였다.

$$SNR[dB] = 10 \log_{10} \left(\frac{\sum_{n=0}^{N_i} |r_i(n)|^2}{\sum_{n=0}^{N_i} |v_i(n)|^2} \right), \quad i = 1, 2 \quad (9)$$

그림 9는 SNR 25 dB, 잔향시간 0.2초, 70도 방향의

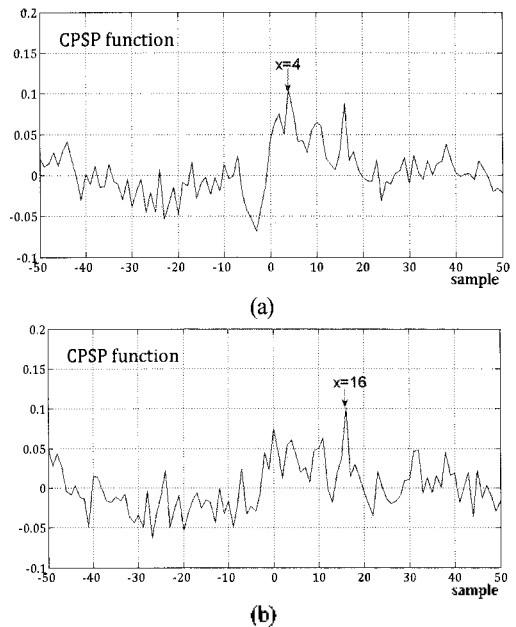


그림 9. 실제 음성신호에 대한 CPSP 함수 예 (a) 프리엠퍼시스를 하지 않은 경우, (b) 프리엠퍼시스를 한 경우
Figure 9. Example of CPSP function for real speech signals (a) Without pre-emphasis, (b) with pre-emphasis.

환경에서 생성된 음성신호로부터 구한 CPSP 함수의 예를 보인 것이다. 추정성공으로 간주되는 도달시간 차이는 16~17샘플인데, 그림 9 (a)와 같이 프리엠퍼시스를 하지 않은 경우에는 탐색범위 내에서 피크가 4 샘플로 바르게 추정되지 않았지만 그림 9 (b)의 경우와 같이 프리엠퍼시스를 한 경우에는 16샘플에서 피크가 바르게 추정됨을 볼 수 있다.

실험은 아래와 같은 3 가지 방법을 정의하여 각 방법의 도달시간차이 추정 성공율을 비교하였다. 이때 실제 화자방향 θ 의 ± 5 도에서 계산된 이론적 도달시간차이 내에 추정된 도달시간차이 \hat{m} 이 존재하는 경우를 도달시간차이 추정 성공으로 간주하였다. 그리고 백색잡음을 20회 첨가하여 구한 평균 도달시간차이 추정 성공율을 퍼센트로 나타내었다. 한 프레임은 구형 윈도우를 이용하여 2048 샘플로 두고 1024 샘플씩 프레임을 이동시켰으며, 4096-포인트 FFT를 사용하였다.

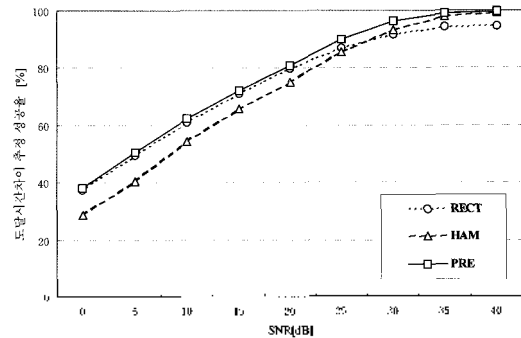
- RECT: 구형 윈도우로 프레임을 구성
- HAM : 해밍 윈도우로 프레임을 구성
- PRE : 프리엠퍼시스된 마이크로폰 출력신호에 구형 윈도우로 프레임을 구성

그림 10은 화자가 70도에 위치한 경우에 대하여 신호대잡음비와 잔향시간에 따른 CPSP 함수의 도달시간차이 추정 성공율을 나타낸 것이다. RECT는 잔향시간과 무관하게 높은 신호대잡음비에서의 성능이 상대적으로 낮은 신호대잡음비에서의 성능보다 저하되는 경우를 보인다. 이것은 구형 윈도우의 스펙트럼 누설로 인하여 도달시간차이를 정확히 추정하지 못하는 경우가 발생하기 때문이다. 화자가 10도와 40도에 위치한 경우에도 70도 경우와 마찬가지로 프리엠퍼시스를 함으로써 향상된 추정 성능을 나타냄을 확인하였다.

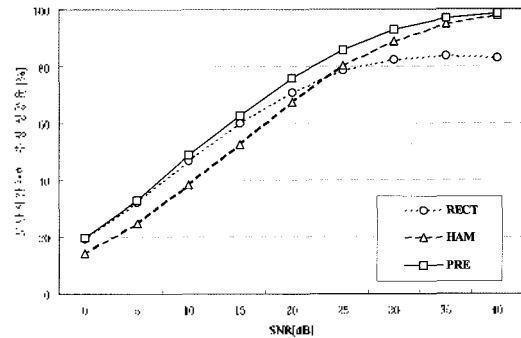
HAM은 RECT와는 달리 스펙트럼 누설이 작기 때문에 낮은 신호대잡음비보다 높은 신호대잡음비에서의 도달시간차이 추정 성공율이 높다. 그러나 낮은 신호대잡음비에서는 잔향에 상관없이 RECT보다 도달시간차이 추정 성공율이 저조하다. 이는 잡음이 많은 경우에 각 마이크로폰 출력신호의 파형이 달라지고, 시간적으로 동일한 샘플에 다른 윈도우 가중치를 곱하기 때문에 더 많은 왜곡이 발생하기 때문이다.

PRE는 저주파의 전력 스펙트럼을 감쇄시켜서 스펙트럼 누설을 줄였기 때문에 높은 신호대잡음비에서의 도달시간차이 추정 성공율이 높다. 그리고 HAM과는 달리 신

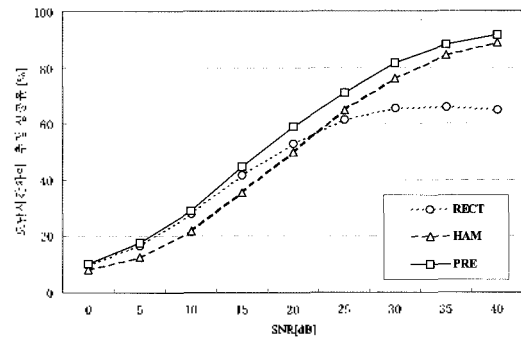
호대잡음비가 낮더라도 RECT보다 도달시간차이 추정 성공율이 저하되지 않는다. 이것은 구형 윈도우를 사용하



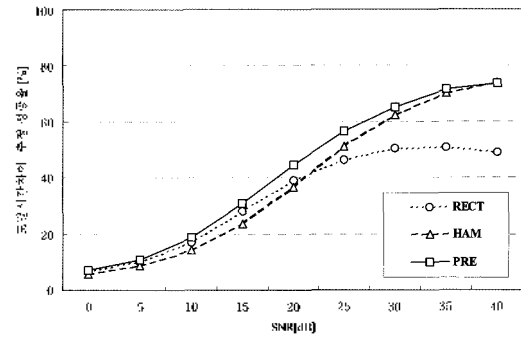
(a)



(b)



(c)



(d)

그림 10. 70도 화자에 대한 도달시간차이 추정 성공율 (a) 잔향시간=0초 (b) 잔향시간=0.1초 (c) 잔향시간=0.2초 (d) 잔향시간=0.3초

Figure 10. Success ratio of TDOA estimation for the source in 70 degree (cont.) (a) $T_{60} = 0s$ (b) $T_{60} = 0.1s$ (c) $T_{60} = 0.2s$ (d) $T_{60} = 0.3s$

면 시간적으로 동일한 샘플에 같은 윈도우 가중치를 곱하므로 신호의 왜곡을 발생시키지 않기 때문이다.

그림 10 (b)에서 그림 10 (d)를 보면 PRE의 도달시간차이 추정 성공율이 RECT에 비해서 잔향시간이 길수록 현저히 높게 나타남을 알 수 있다. 일반적으로 유성음 파형은 무성음 파형에 비해서 큰 값을 가지므로 잔향시간이 길고 유성음 뒤에 무성음이 존재하면 무성음의 작은 진폭이 유성음의 큰 진폭에 의해서 영향을 받게 된다. 따라서 이러한 경우에는 무성음 구간이지만 유성음과 같이 저주파에서 큰 전력 스펙트럼을 가지므로 스펙트럼 누설로 인한 도달시간차이 추정 성능이 저하될 수 있다. 그러나 PRE는 이러한 경우에도 저주파 영역의 큰 전력 스펙트럼을 감쇄시키기 때문에 스펙트럼 누설로 인한 성능 저하를 줄일 수 있다.

따라서 프리엠퍼시스된 마이크로폰 출력신호에 구형 윈도우로 프레임 구성하면 구형 윈도우의 문제점인 스펙트럼 누설과 해밍 윈도우의 문제점인 신호의 왜곡으로 발생하는 도달시간차이 추정 성능의 저하를 동시에 줄일 수 있다.

V. 결론

본 논문에서는 프레임 단위의 CPSP 함수로 도달시간 차이를 추정할 때 발생하는 문제점을 제시하고 분석하였다. 구형 윈도우는 스펙트럼 누설로 인하여 CPSP 스펙트럼을 부정확하게 추정하는 문제점이 있고, 해밍 윈도우를 비롯한 다른 윈도우 함수는 프레임의 양쪽 끝부분에서 발생하는 윈도우 가중치의 불일치 때문에 신호의 왜곡을 발생시키는 단점이 있다. 이를 해결하기 위하여 프리엠퍼시스된 마이크로폰 출력신호에 구형 윈도우로 프레임을 구성하는 방법을 제안하였다. 프리엠퍼시스와 구형 윈도우의 사용은 유성음 구간에서 저주파 영역의 큰 전력 스펙트럼을 감쇄시키고 시간적으로 동일한 샘플에 같은 윈도우 가중치를 곱하기 때문에 구형 윈도우의 문제점인 스펙트럼 누설과 해밍 윈도우의 문제점인 신호의 왜곡으로 발생하는 도달시간차이 추정 성능의 저하를 동시에 줄일 수 있다. 제안한 방법의 도달시간차이 추정 성능을 평가하기 위하여 도달시간차이 추정 성공율을 모의실험 하였다. 실험 결과 프리엠퍼시스된 마이크로폰 출력신호에 구형 윈도우를 적용시켜 CPSP를 구할 경우 프리엠퍼시스를 하지 않거나 다른 윈도우를

사용하는 경우에 비해 도달시간차이 추정 성능이 크게 향상될 수 있음을 보였다.

감사의 글

본 연구는 경북대학교 2009학년도 학술연구지원으로 수행되었습니다.

참고 문헌

1. M. Branstein and D. Ward, *Microphone Arrays—Signal Processing Technique and Applications*, Springer, New York, 2001.
2. B.D. Van Veen, and K.M. Buckley, "Beamforming: a Versatile Approach to Spatial Filtering," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol.5, no.2, pp.4–24, 1988.
3. E.D. Di Claudio, R. Parisi, and G. Orlandi, "Multi-source Localization in Reverberant Environment by ROOT-MUSIC and Clustering," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.921–924, June, 2000.
4. C.H.Knapp and G.C.Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol.24, no.4, pp.320–327, 1976.
5. D.V. Rabinikin, R.J. Renomeron, J.C. French, and J.L. Flanagan, "Estimation of Wavefront Arrival Delay Using The Cross-power Spectrum Phase Technique," *132nd Meeting of the Acoustical Society of America, Honolulu, HI, USA*, pp.1–10, December 4, 1996.
6. A.V. Oppenheim and R.W. Schacter, *Discrete-Time Signal Processing*, Prentice Hall, New Jersey, 1989.
7. J.B. Allen and D.A. Berkey, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am*, vol. 65, no.4, pp.943–950, 1979.
8. H. Kultrull, *Room Acoustics*, Elsevier, London, 1991.
9. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: the art of scientific computing*, press syndicate of the university of Cambridge, NewYork, 1992.

저자 약력

• 권 홍 석 (Hongseok Kwon)



1997년 2월: 경북대학교 전자공학과 졸업
 1999년 2월: 경북대학교 전자공학과 석사
 2004년 8월: 경북대학교 전자공학과 박사
 2004년 8월~ 현재: 삼성전자 무선사업부

•배 건 상 (Keunsung Bae)



1977년 2월: 서울대학교 전자공학과 졸업
1979년 2월: 한국과학기술원 전기 및 전자공학과
(공학석사)
1989년 5월: University of Florida (공학박사)
1979년 3월 ~ 현재: 경북대학교 전자전기컴퓨터학부
교수