마이크로폰 쌍을 이용한 음원의 도달시간차이 추정에서 음성신호의 프리엠퍼시스 영향 분석

권 홍 석, 김 시 호, 배 건 성
경북대학교 전자공학과

Preemphasis of Speech Signals in the Estimation of Time Difference of Arrival with Two Microphones

Hongseok Kwon, Siho Kim, Keunsung Bae
Department of Electronics, Kyungpook National University
hskwon@mir.knu.ac.kr, ksbae@mir.knu.ac.kr

요약

In this paper, 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 makes estimation of CPSP spectrum inaccurate. Framing with a Hamming window to reduce the spectral leakage effect distorts the signal due to the different weighting at temporally same sample, which make the TDOA estimation using CPSP function inaccurate. In this paper, we solve this problem by reducing the dynamic range of the spectrum of a speech signal with preemphasis. Experimental results confirm that the framing of pre-emphasized microphone output with a rectangular window shows higher success ratio of TDOA estimation than any other framing methods.

1. 서론

마이크로폰 배열은 단일 채널 마이크로폰에 비해 주변 잡음과 간섭신호를 크게 줄이며, 화자의 위치를 추정/추적할 수 있다는 장점을 가진다. 이러한 특징들로 인하여 마이크로폰 배열은 원격회의 시스템, 음성 인식, 화자 식별, 음성 수집, 방범 시스템 등에 광범위하게 사용되고 있다.

마이크로폰 배열을 이용하는 많은 기법들은 기본적으로 음원의 방향이나 신호의 도달시간차이의 정확한 추정을 필요로 한다. 도달시간차이 추정 기법은 크게 마이크로폰 배열의 빔조향(beamsteering)에 기반한 방법[1], 고해상도 스펙트럼 추정에 기반한 방법[2], GCC(Generalized Cross Correlation) 함수에 기반한 방법 등으로 구분할 수 있다. 그 중에서 GCC 함수에 기반한 도달시간차이 추정 방법은 마이크로폰 쌍의 출력신호에서 상호 전력 스펙트럼을 구하고 적당한 프리필터(prefilter)를 곱하여 얻은 GCC 함수의 최대값을 찾아서 도달시간차이로 간주하는 방법으로 알고리즘이 간단하고, 적은 연산량으로 만족할 만한 성능을 얻을 수 있기 때문에 널리 사용된다[3,4].

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 상호 상관 스펙트럼의 위상만을 이용하여 도달시간차이를 추정하는 방법이다. CPSP 함수를 이용하여 도달시간차이를 추정하기 위해서는 이산 신호 형태의 마이크로폰 출력신호를 DFT하여 이산 주파수 영역에서 수행된다. 그러나 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)를 이용하면 이러한 왜곡은 줄일 수 있지만, 해밍 윈도우는 시간적으로 동일한 신호에 다른 윈도우 가중치를 곱하는 결과가 되어 신호의 왜곡이 발생된다. 따라서 본 논문에서는 프레임 단위의 CPSP 함수를 이용한 도달시간차이 추정에 대한 문제점을 제시하고, 이를 해결하기 위하여 프리엠퍼시스로 마이크로폰 출력신호의 동적범위를 줄이는 방법을 제안한다.

2. GCC 함수를 이용한 TDOA 추정

2.1 GCC 함수를 이용한 TDOA 추정

마이크로폰 쌍의 GCC 함수를 이용하여 도달시간차이를 추정하기 위해서는 먼저 식 (1)과 같이 각 마이크로폰 출력신호의 스펙트럼($X_1(k)$, $X_2(k)$)을 곱한 상호 전력 스펙트럼에 적당한 프리필터 $\Psi(k)$ 를 곱하여 IDFT 함으로써 GCC 함수 $\hat{c}(m)$ 을 얻는다.

$$\hat{c}(m) = \sum_{k=0}^{N-1} \Psi(k) \cdot X_1(k) \cdot X_2^*(k) e^{j\frac{2\pi mk}{N}} \qquad (1)$$

CPSP 함수는 식 (2)와 같은 프리필터를 가지는 GCC 함수로서 모든 상호 전력 스펙트럼의 크기를 1로 만든다. 최종적인 도달시간차이 추정은 GCC 함수 $\hat{c}(m)$ 에서 피크를 나타내는 샘플 위치로 결정된다.

$$\Psi(k) = \frac{1}{\left| X_1(k) \cdot X_2^*(k) \right|} \qquad (2)$$

2.2 구형 윈도우를 사용하는 CPSP 함수의 문제점

그림 1은 유성음 신호에 대해서 CPSP를 이용한 도달시간차이 추정 결과의 한 예를 보이고 있다. 두 마이크로폰의 출력신호는 12 샘플 지연이 있기 때문에 CPSP 함수는 샘플 12에서 피크가 나타나야 한다. 그러나 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 해당되는 선형위상이 비교적 잘 나타남을 알 수 있다.

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.

3. 프리엠퍼시스를 이용한 CPSP 함수의 성능 개선

3.1 구형 윈도우로 구성된 프레임의 분석

유성음 구간에서 구형 윈도우로 프레임을 구성하는 경우에 CPSP 함수가 도달시간차이 추정에 실패하는 원인은 구형 윈도우에 의한 신호 누설에 기인한다.

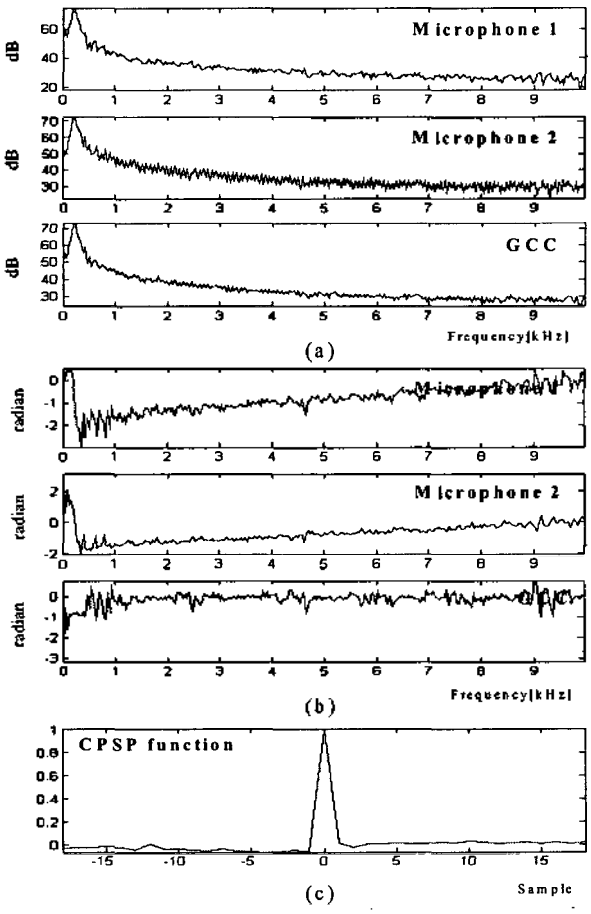

그림 1. 구형 윈도우로 구성된 유성음 프레임의 예
(a) 마이크로폰 출력신호와 GCC 전력 스펙트럼
(b) 마이크로폰 출력신호와 GCC 위상 스펙트럼
(c) 정규화된 CPSP 함수

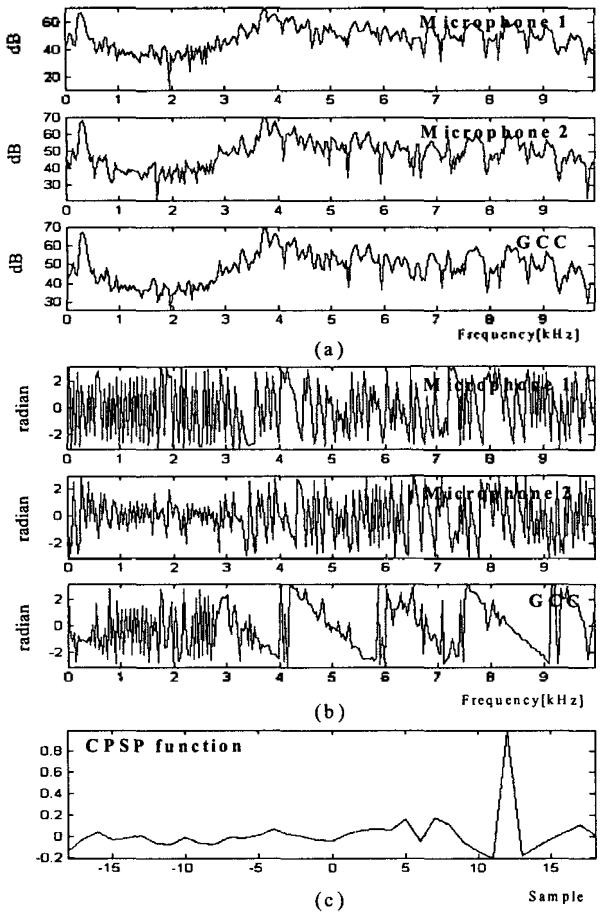

그림 2. 구형 윈도우로 구성된 무성음 프레임의 예

(a) 마이크로폰 출력신호와 GCC 전력 스펙트럼

(b) 마이크로폰 출력신호와 GCC 위상 스펙트럼

(c) 정규화된 CPSP 함수

구형 윈도우는 부엽(sidelobe)의 감쇄가 작아서 스펙트럼 누설이 크기 때문에 큰 전력을 갖는 주파수 성분이 다른 주파수에 영향을 많이 끼친다[5]. 음성의 유성음은 저주파에 에너지가 몰려있고 이러한 에너지가 구형 윈도우에 의해 누설되어 전 주파수 대역에 영향을 미침으로써 음성신호에 의한 위상 특성이 제대로 나타나지 못한다. 그림 3을 보면 음성의 1차 포만트를 중심으로 하는 구형 윈도우의 스펙트럼(굵은선)이 음성신호의 스펙트럼과 비슷한 형태를 보인다. 따라서 큰 전력 스펙트럼을 갖는 주파수 성분의 스펙트럼 누설이 다른 주파수의 스펙트럼에 더해짐으로써 음성신호의 스펙트럼이 정확히 표현되지 못하게 된다.

구형 윈도우의 스펙트럼 누설로 인한 CPSP 함수의 도달시간차이 추정 실패는 스펙트럼 누설이 상대적으로

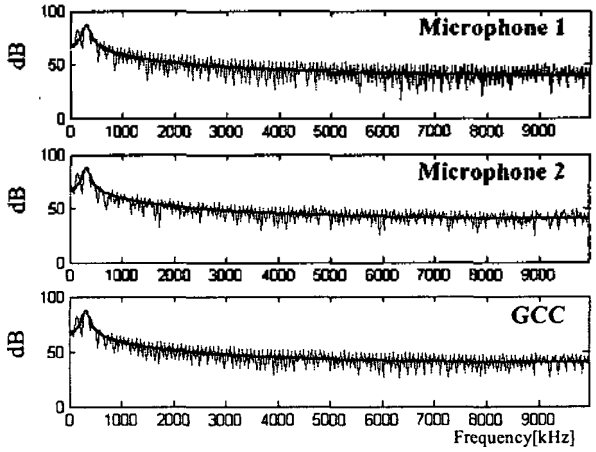

그림 3. 유성음에서 구형 윈도우의 스펙트럼 누설

작은 해밍 윈도우를 사용하여 해결할 수 있다. 그러나 해밍 윈도우는 두 마이크로폰 신호에서 동일한 샘플에 다른 윈도우 가중치가 곱해지므로 신호의 왜곡을 발생시킨다. 이러한 왜곡은 잔향과 잡음이 없는 경우에는 성능에 큰 영향을 끼치지 않으나 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 성능 개선

앞 절에서 언급했듯이 유성음에서 구형 윈도우를 사용한 CPSP에서 검출 오류가 발생하는 원인은 유성음의 저주파 성분이 고주파에 비해 상대적으로 크기 때문에 나타나는 현상이다. 따라서 음성신호 스펙트럼의 고주파 성분을 boosting하여 스펙트럼의 동적 범위를 줄이면 구형 윈도우에 의한 영향을 줄일 수가 있다. 이것은 적절한 고역통과필터로 구현 가능하지만 식 (3)과 같이 간단한 프리엠퍼시스를 사용함으로써 해결할 수 있다. 프리엠퍼시스는 일종의 고역통과필터로서 낮은 주파수의 전력 스펙트럼을 감쇄시키고, 동일한 필터로 각 마이크로폰 출력신호를 필터링하기 때문에 프리엠퍼시스의 위상 스펙트럼이 GCC 위상 스펙트럼에 전혀 영향을 미치지 않는다. 실험에서 프리엠퍼시스에 사용되는 α 의 값은 0.95로 두었다.

$$y_i(n) = x_i(n) - \alpha \cdot x_i(n-1), \quad i = 1,2 \tag{3}$$

4. 실험 결과 및 분석

유성음과 무성음에 대하여 각 마이크로폰 출력신호를 프리엠퍼시스한 후 구형 윈도우로 프레임을 구성하고 CPSP 검출 실험을 수행하였다. 그림 4는 유성음에 대한 결과를 보이고 있다. 그림 4(a)의 전력 스펙트럼을 보면 큰 전력 스펙트럼을 갖는 저주파 성분의 누설 전력이 다른 주파수의 전력 스펙트럼에 비해서 작다는 것을 알 수 있다. 따라서 구형 윈도우의 스펙트럼 누설이 다른 주파수의 스펙트럼에 크게 영향을 끼치지 못하므로 그림 4(b)의 GCC 위상 스펙트럼에 선형성이 잘 나타난다. 결과적으로 그림 4(c)와 같이 CPSP 함수의 피크가 12 샘플에 나타나게 되어 도달시간차이를 정확히 추정한다.

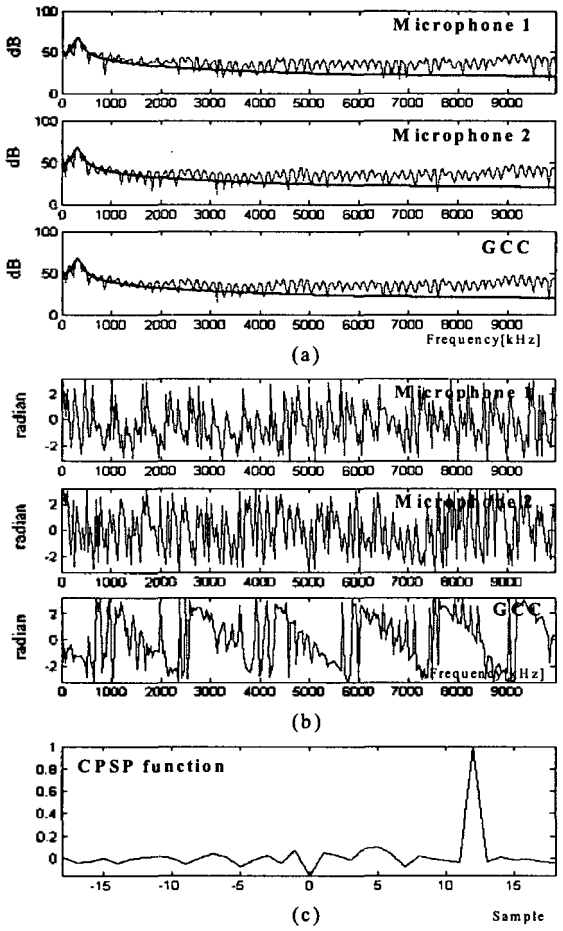

그림 4. 프리엠퍼시스된 유성음을 구형 윈도우로 구성한 프레임 (a) 마이크로폰 출력신호와 GCC 전력 스펙트럼 (굵은 선: 구형 윈도우의 스펙트럼 포락선) (b) 마이크로폰 출력신호와 GCC 위상 스펙트럼 (c) 정규화된 CPSP 함수

5. 결론

본 논문에서는 프레임 단위의 CPSP 함수로 도달시간차이를 추정할 때 구형 윈도우에 의해 발생되는 문제점을 제시하고 분석하였다. 구형 윈도우는 스펙트럼 누설로 인하여 CPSP 스펙트럼을 부정확하게 추정하는 문제점이 있고, 해밍 윈도우는 시간적으로 동일한 신호에 다른 윈도우 값을 곱하므로 신호의 왜곡을 발생시키는 단점이 있다. 이를 해결하기 위하여 프리엠퍼시스된 마이크로폰 출력신호에 구형 윈도우로 프레임을 구성하는 방법을 제안하였다. 제안한 방법은 유성음 구간에서 저주파의 큰 전력 스펙트럼을 감쇄시키고 시간적으로 동일한 샘플에 같은 윈도우 가중치를 곱하기 때문에 구형 윈도우의 문제점인 스펙트럼 누설과 해밍 윈도우의 문제점인 신호의 왜곡으로 발생되는 도달시간차이 추정 성능의 저하를 동시에 줄일 수 있었다.

본 연구는 한국과학재단 목적기초연구(R01-2003-000-10242-0)지원으로 수행되었음.

참고문헌

1. B.D. Van Veen and K.M. Buckley, "Beamforming: a Versatile Approach to Spatial Filtering," IEEE Trans. on Acoust., Speech, Signal Processing, Vol.5, No.2, pp.4-24, 1988
2. E.D. Di Claudio, R. Parisi, and G. Orlandi, "Multi-source Localization in Reverberant Environment by ROOT-MUSIC and Clustering," ICASSP, pp.921-924, 2000
3. C.H. Knapp and G.C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," IEEE Trans. on Acoust., Speech, Signal Processing, Vol.24, No.4, pp.320-327, 1976
4. D.V. Rabinkin, 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, pp.1-10, 1996
5. A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1989