말소리와 음성과학 제3권 제3호 (2011)
pp. 109~114

CPSP 문턱값 설정을 통한 음원도달 방향 추정 성능 개선

Performance Improvement of Sound Direction of Arrival Estimation by Applying Threshold to CPSP

전 성 일[1] · 배 건 성[2]
Quan, Xingri · Bae, Keunsung..

ABSTRACT

To estimate sound direction of arrival with a pair of microphones, a method based on Time Difference of Arrival (TDOA) estimation using the Cross Power Spectrum Phase (CPSP) function is largely used due to its simplicity and good performance. In this paper, we investigate CPSP maximum values for various SNRs and adverse environments, and propose a novel method to improve the estimation performance of sound direction of arrival. The proposed method applies a threshold to the CPSP values and increases the reliability of the estimated sound direction. Through computer simulation for various SNRs, we validate the effectiveness of the proposed method. When the threshold was set to 0.1, more than 90% of success rate of sound direction of arrival estimation has been achieved for directions of 10°, 40°, 70° from the source location even with reverberation times of 0.1s.

Keywords: sound localization, CPSP, TDOA

1. 서론

음원도달방향 추정은 마이크로폰 어레이를 이용하여 화자의 위치를 찾아내기 위한 것으로 로봇, 화상회의, 무인방범 시스템, 음성개선 등의 분야에 널리 응용된다. 음원도달방향 추정 방법은 크게 마이크로폰 배열의 빔조향(beam steering)에 기반한 방법[1], 고해상도 스펙트럼 추정에 기반한 방법[2], GCC (Generalized Cross Correlation) 기반의 도달시간차이, 즉, TDOA(Time Difference of Arrival) 추정에 의한 방법[3] 등이 있다. 빔조향에 기반한 방법은 근거리 화자인 경우 모든 가능한 화자의 위치에 대해서 빔형성(beam forming)을 해야 하므로 많은 연산량을 필요로 하며, 고해상도 스펙트럼 추정에 기반한 방법은 음원이 협대역 신호인 경우에 대해 주파수 분해와 고유치 분해 등을 적용하므로 광대역 신호인 음성신호에 적용하기에는 부적합하다. GCC 기반의 도달시간차이 추정에 의한 방법은 마이크로폰 쌍의 출력신호에서 상호전력스펙트럼을 구하고 적당한 프리필터를 곱하여 얻은 GCC 함수의 최대값을 찾는 것으로, 알고리즘이 간단하고 적은 연산량으로도 만족할 만한 성능을 얻을 수 있는 장점이 있다.

GCC 함수를 이용하여 도달시간차이를 추정할 때 여러 가지 프리필터가 제안되었는데, 두 마이크로폰에 수신된 신호의 상호전력스펙트럼의 위상만을 이용하여 도달시간차이를 추정하는 CPSP(Cross Power Spectrum Phase) 함수가 알고리즘이 간단하면서 우수한 성능을 갖는다고 알려져 있다[3]. 본 논문에서는 2개 마이크로폰을 이용하여 CPSP 방법으로 음원도달방향을 추정할 때 잡음과 잔향이 존재하는 환경에서 추정 신뢰도를 향상시키기기 위하여 CPSP의 최대값에 적절한 문턱값을 적용하는 방법을 제안한다. 다양한 SNR을 갖는 모의실험 환경에서 CPSP를 이용하여 도달시간차이를 추정할 때 얻어지는 CPSP 최대값의 분포를 분석하고, 적절한 문턱값을 설정하여 신뢰도가 높은 프레임만으로 음원도달방향을 추정함으로써 일반적인 사무실 환경에서 음원도달방향 추정 시스템의 안정성을 높일 수 있음을 보인다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 CPSP 함수를 이용한 음원도달방향 추정 방법에 대해 간략히 설명하고, 3장에서는 본 논문에서 제안한 CPSP 함수에 문턱값을 설정하여 도달시간차이 추정의 신뢰도를 향상 시킬 수 있는 방법에 대해

1) 경북대학교, comet263@hotmail.com
2) 경북대학교, ksbae@ee.knu.ac.kr, 교신저자

접수일자: 2011년 6월 5일
게재결정: 2011년 6월 28일

설명한다. 4장에서는 모의실험 조건 및 문턱값 설정을 이용한 음원도달방향 추정 실험결과를 제시하며, 마지막으로 5장에서 결론을 맺는다.

2. CPSP 함수를 이용한 음원도달방향 추정

마이크로폰과 음원의 거리가 멀고 평면파라고 가정하면 두 마이크로폰의 출력신호 $x_1(n), x_2(n)$은 식 (1)과 같이 표현된다. 여기에서 $s(n)$은 마이크로폰 1에 입력되는 음성신호이고 $v_1(n), v_2(n)$은 마이크로폰에 입력되는 잡음신호이며 α및 D는 마이크로폰 1을 기준으로 한 마이크로폰 2까지의 감쇠인자와 시간지연을 의미한다. 이때, 음성신호 $s(n)$과 잡음신호 $v_1(n), v_2(n)$은 상관성이 없다고 가정한다.

$$\begin{aligned} x_1(n) &= s(n)+v_1(n) \\ x_2(n) &= \alpha s(n-D)+v_2(n) \end{aligned} \quad (1)$$

두 마이크로폰 신호 사이의 상호상관함수는 식 (2)와 같고, 상호전력스펙트럼과 상호상관함수의 관계식에 의하여 식 (3)과 같이 쓸 수 있다.

$$R_{x_1x_2}(\tau) = E[x_1(n+\tau)x_2(n)] \quad (2)$$

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} G_{x_1x_2}(f)e^{j2\pi f\tau}df \quad (3)$$

여기에서 $G_{x_1x_2}(f)$는 $x_1(n), x_2(n)$ 두 신호의 상호전력스펙트럼이다. 식 (3)을 이용하여 상호상관함수를 구할 때 잔향과 잡음의 영향을 줄이고 신호 성분을 강조하기 위해 상호전력스펙트럼에 프리필터 $\psi(f)$를 곱하여 얻어지는 식 (4)를 GCC (Generalized Cross Correlation) 함수라고 한다.

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} \psi(f)G_{x_1x_2}(f)e^{j2\pi f\tau}df \quad (4)$$

식 (4)에서 $\psi(f)=1$이면 CC(Cross Correlation) 함수가 되고, $\psi(f)=1/|G_{x_1x_2}(f)|$ 일 때를 CPSP 함수라고 한다. CPSP를 이용한 최종적인 도달시간차이 추정은 식 (5)에서와 같이 $R_{x_1x_2}(\tau)$ 함수가 최대값을 나타내는 샘플 위치로 결정된다.

$$D = \arg max[R_{x_1x_2}(\tau)] \quad (5)$$

3. CPSP 문턱값 설정을 통한 음원도달방향 추정 성능 향상

<그림 1>과 같은 모의실험 환경에서 음원이 10° 방향에 위치할 때 SNR 40dB 및 SNR 0dB에서 잔향이 없는 경우와, SNR 0dB 잔향이 0.2초인 경우에 대해 구한 CPSP 함수의 값을 <그림 2>에 나타내었다. <그림 2(a)>와 같이 SNR이 높고 잔향이 없는 경우에는 CPSP 최대값이 1에 가까운 큰 값을 가지지만 <그림 2(b), (c)>의 경우처럼 SNR이 낮거나 잔향이 클 경우에는 CPSP 최대값이 0에 가까운 쪽으로 작은 값을 가지게 되는 것을 볼 수 있다.

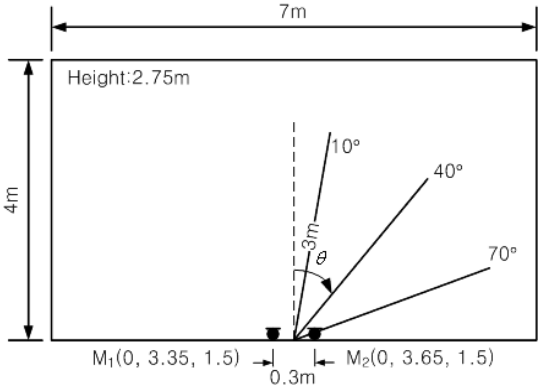
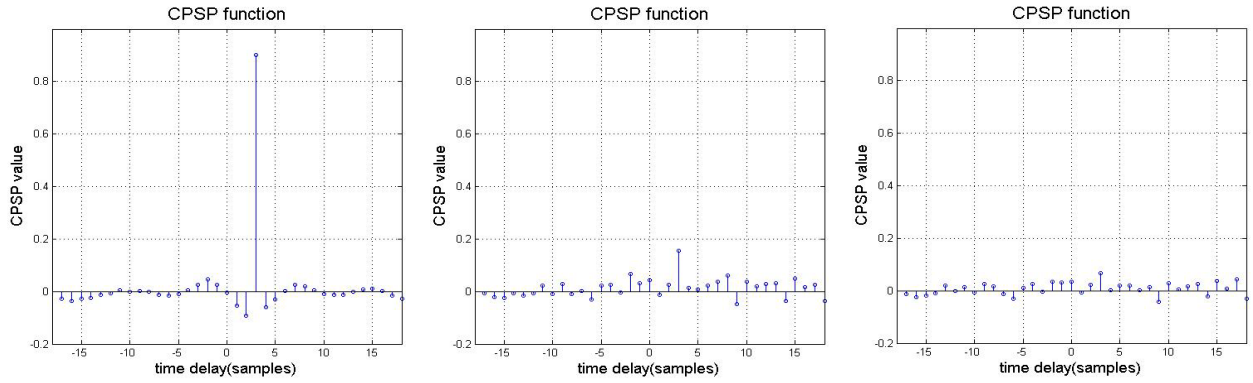


그림 1. 모의실험 환경
Figure 1. Simulation environment

<그림 3>은 음원의 위치가 10°, 40°, 70° 일 때, 잡음 및 잔향 환경에서 음원도달방향 추정 시 CPSP 함수의 최대값을 구하고, 5° margin을 허용한 범위에서 음원도달방향이 바르게 추정된 경우와 잘못 추정된 경우에 대한 CPSP 최대값의 분포를 히스토그램으로 나타낸 것이다. 실선은 음원도달방향이 바르게 추정되었을 경우의 CPSP 최대값의 분포를 타나낸 것이고, 점선은 잘못 추정되었을 경우의 CPSP 최대값의 분포를 나타낸 것이다. 음원도달방향의 추정 성공 여부는 원래의 음원방향에서 ±5° margin을 허용하였을 때의 도달시간차이를 기준으로 하였다. 즉, 추정된 도달시간차이 값이 ±5°에 해당되는 시간지연 값 범위 내에 들어오면 바르게 추정된 것으로 간주하였다. 음성신호의 샘플링주파수가 20kHz, 마이크로폰 사이의 간격이 30cm 일 때 10°, 40°, 70° 방향에 대한 ±5° margin에 대응되는 시간지연 허용 범위는 <표 1>과 같다.

<그림 3>의 히스토그램에서, 음원의 위치가 10°인 경우에는 CPSP의 최대값 크기가 작은 경우에도 상대적으로 다른 음원 위치에 비해 음원도달방향이 잘못 추정되는 프레임 수가 훨씬 적음을 볼 수 있는데 이것은 <표 1>에서 보듯이 10° 에서 허용되는 샘플 margin이 제일 큰 영향이라고 생각된다. 하지만 <그림 3>에서 보면 음원의 위치에 관계없이 음원도달방향의 추정에 실패하는 경우는 CPSP의 최대값이 0.2에 미치지 못하는 경우가 대부분이라는 것을 알 수 있다. 따라서 CPSP 함수 최대값에 적절한 문턱값을 설정하여 문턱값보다 큰 CPSP 최대값에 대해서만 음원도달방향을 추정한다면 추정 신뢰도를 크게 향상시킬 수 있을 것이다.



(a) (b) (c)

그림 2. 음원의 위치가 10° 방향에 있을 때 SNR과 잔향에 따른 CPSP 함수 값의 예
(a) SNR=40dB, T_{60}=0s (b) SNR=0dB, T_{60}=0s (c) SNR=0dB, T_{60}=0.2s

Figure 2. Examples of CPSP values depending on SNR and reverberation when sound source is at 10 degree position
(a) SNR=40dB, T_{60}=0s (b) SNR=0dB, T_{60}=0s (c) SNR=0dB, T_{60}=0.2s

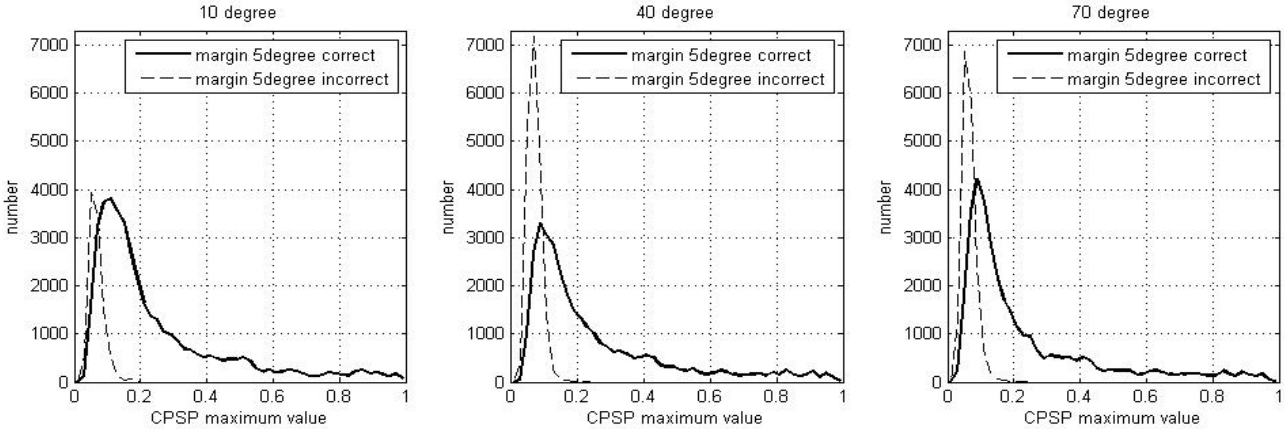


그림 3. 음원 위치에 따른 CPSP 함수 최대값의 분포

Figure 3. Histogram of maximum values of CPSP depending on sound source position

<그림 3>에서 음원도달방향을 바르게 추정한 경우를 ω_1, 잘못 추정한 경우를 ω_2라고 하면, 전체에 대한 히스토그램은 <그림 4>와 같다. 이에 대해 식 (6)의 Bayes' decision rule을 적용하여 적절한 문턱값을 계산하면 다음과 같다.

$$P(\omega_j|x) = \frac{p(x|\omega_j)\,P(\omega_j)}{p(x)},\ j=1,2 \tag{6}$$

$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)\,P(\omega_j) \tag{7}$$

식 (6)에서의 사전확률(a priori probability)은 식 (8)과 같다.

$$P(\omega_j) = \frac{N_j}{N},\ j=1,2 \tag{8}$$

여기에서, N은 히스토그램의 총 합이고, N_1은 ω_1 경우에 해당되는 히스토그램의 총 합, N_2는 ω_2 경우에 해당되는 히스토그램의 총 합이다. 식 (6)에서의 likelihood는 식 (9)로 얻어지며,

$$p(x|\omega_j) = \frac{H(x|\omega_j)}{N_j},\ j=1,2 \tag{9}$$

여기에서, $H(x|\omega_j)$는 ω_j 경우의 히스토그램을 의미한다. Bayes' decision rule에 의하여 $p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$이면 ω_1의 경우로, $p(x|\omega_1)p(\omega_1) < p(x|\omega_2)p(\omega_2)$이면 ω_2의 경우로 판정하게 된다. <그림 4> 히스토그램에 Bayes' decision rule를 적용하여 계산한 최적의 문턱값은 0.1이다.

표 1. 5° margin에 해당되는 시간지연 값
Table 1. Delayed samples corresponding to 5° margin

Position / Margin	10° (samples)	40° (samples)	70° (samples)
-5°	1.5	10.1	16.0
0°	3.1	11.3	16.6
+5°	4.6	12.5	17.0

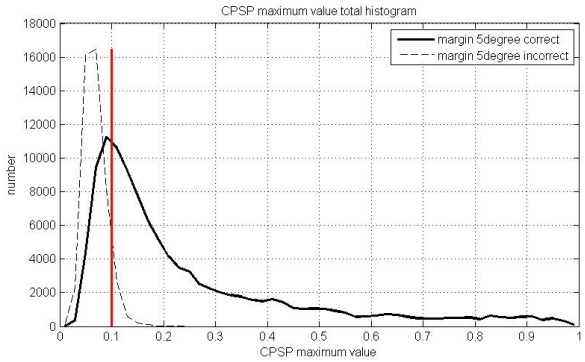


그림 4. 전체 음원에 대한 히스토그램 및 문턱값
Figure 4. Histogram for all data and its threshold

4. 실험 및 검토

제안한 방법의 음원도달방향 추정 성능을 분석하기 위하여 <그림 1>과 같은 실내공간을 가정하고 모의실험을 수행하였다. 두 마이크로폰은 0.3m 간격으로 두었으며 등방성 빔패턴을 가지고 마이크로폰 후면에서 반사되는 반사파는 없다고 가정하였다. 음원의 위치는 두 마이크로폰의 중심에서 3m 떨어진 위치에 10°, 40°, 70° 방향을 가지도록 위치시켰는데, 이때 두 마이크로폰의 중심에서 수직방향을 0°로 정의하였다.

Image method[4]를 이용하여 음원의 위치에서 각 마이크로폰까지의 실내공간의 임펄스 응답을 20kHz 샘플링주파수에 8192 샘플의 길이를 갖도록 생성하였으며, 이때 반사계수는 식 (10)

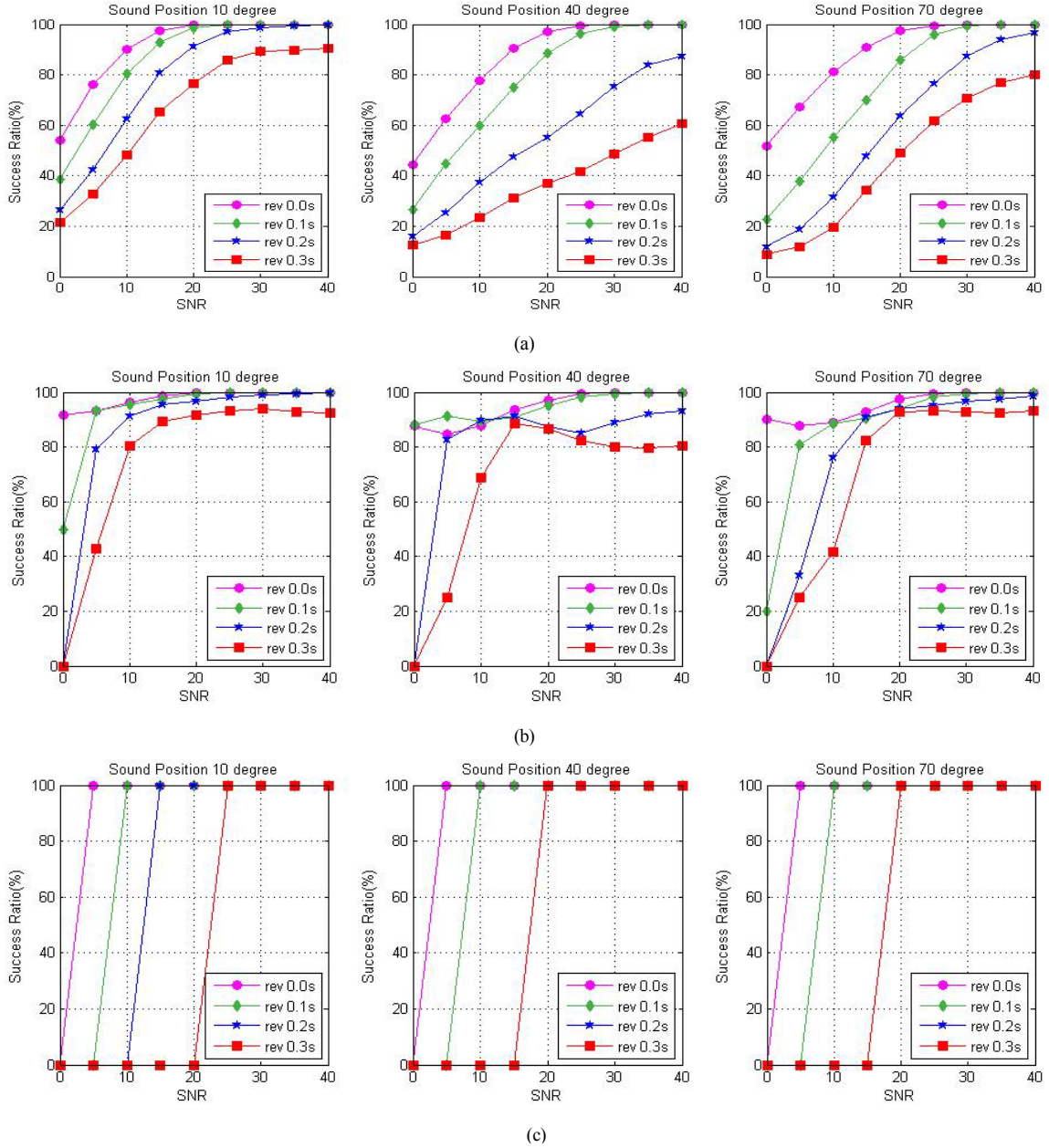


그림 5. 음원도달방향 추정 성공률 (a) 문턱값을 설정하지 않은 경우 (b) 문턱값 0.1 (c) 문턱값 0.2
Figure 5. Success rate of sound direction of arrival estimation (a) without threshold (b) threshold 0.1 (c) threshold 0.2

의 Eyring 공식으로 계산된다. 식 (10)에서 L_x, L_y, L_z 는 실내 공간의 가로, 세로, 높이 길이를 의미하고, T_{60}은 잔향시간을 나타내는 파라미터이며, c 는 공기 중에서의 음속으로 $340m/s$ 로 가정하였다.

$$\beta = e^{-13.82/[c(L_x^{-1}+L_y^{-1}+L_z^{-1})T_{60}]} \quad (10)$$

<표 2>에 음원도달방향 추정에 사용된 실험조건을 나타내었다. CPSP 추정을 위한 상호전력스펙트럼을 구할 때 프리엠퍼시스와 구형윈도우를 적용하였으며[5], 잡음환경은 SNR 0dB 에서 40dB 까지 5dB 간격으로 백색잡음을 첨가하고 다양한 잔향시간을 갖는 환경에 대해 실험하였다. 실험결과에 대한 음원도달방향 추정성능 평가 척도로는 ±5° margin을 둔 추정성공률로 분석하였다.

표 2. 실험조건
Table 2. Experimental condition

Frame size	2048 samples, 1024 moving
Window	Rectangular with pre-emphasis
Reverberation	T_{60}= 0s, 0.1s, 0.2s, 0.3s
Noise	White noise, SNR 0~40dB
Success rate	TDOA with ±5° margin

<그림 5>는 음원의 위치가 10°, 40°, 70°에 있을 때의 CPSP를 이용한 음원도달방향 추정성공률을 보인 것이다. <그림 5(a)>는 문턱값을 설정 하지 않은 경우 각 음원 위치에서의 SNR과 0초, 0.1초, 0.2초, 0.3초 잔향에 따른 추정성공률을 보인 것이다. <그림 5(b)> 및 <그림 5(c)>는 CPSP 최대값의 문턱값을 0.1 및 0.2로 설정하였을 경우의 음원도달방향 추정성공률을 보인 것인데, <그림 5(a)>의 결과와 비교하면 전반적으로 추정성공률이 많이 개선된 것을 볼 수 있다. <그림 5(b)>의 문턱값을 0.1로 설정한 경우를 보면, SNR 15dB 이상인 환경에서는 잔향시간이 0.1초 이내인 경우 모든 음원 위치에서 약 90% 이상의 높은 추정성공률을 얻을 수 있음을 알 수 있다. 그러나 <그림 5(c)>의 경우처럼 CPSP의 문턱값을 0.2로 높게 설정할 경우 도달시간차이 값을 잘못 추정한 프레임이 대부분 제거되어 음원도달방향 추정성공률이 크게 개선될 수 있지만, SNR이 낮거나 잔향시간이 클 경우 도달시간차이 값을 바르게 추정한 프레임의 수가 급격히 줄어들게 되는 문제점이 발생하게 된다. <표 3> 및 <표 4>에 음원의 방향이 10° 인 경우에 대해서, 각각 CPSP 문턱값을 0.1 및 0.2로 설정하였을 때 CPSP 최대값이 문턱값 이상인 유효프레임 수와 도달시간차이 값이 바르게 추정된 프레임 수를 함께 표시하였다. <표 3>에서 음원 파일은 총 1440프레임인데, SNR 15dB, 잔향이 0.1초인 경우를 예로 들면, CPSP 최대값이 문턱값 0.1 이상인 프레임은 1184프레임이고 그 중 1155프레임은 정확하게 추정하였으며 추정 성공률은 98%이다. <표 3> 및 <표 4>에서 보면, SNR이 낮아지고 잔향시간이 증가할수록 CPSP 최대값이 문턱값을 넘는 유효프레임 수가 많이 감소함을 볼 수 있으며, 문턱값을 0.2로 설정하였을 경우에는 그 수가 급격히 줄어듦을 알 수 있다.

표 3. CPSP 최대값이 문턱값 0.1 이상인 프레임 수
(TDOA가 바르게 추정된 프레임 수/ 전체 유효 프레임 수)
Table 3. The number of frames whose CPSP maximum value is greater than 0.1
(number of frames with correct TDOA estimation/number of total frames whose CPSP maximum value is greater than 0.1)

SNR \ T_{60}	0s	0.1s	0.2s	0.3s
0dB	241/263	5/10	0/1	0/0
5dB	682/731	249/267	19/24	3/7
10dB	1122/1167	710/743	236/259	58/72
15dB	1357/1375	1155/1184	646/675	297/333
20dB	1434/1439	1365/1372	1058/1093	649/708
25dB	1438/1440	1428/1430	1284/1306	995/1068
30dB	1440/1440	1439/1440	1378/1393	1157/1233
35dB	1440/1440	1440/1440	1422/1429	1226/1320
40dB	1440/1440	1440/1440	1437/1440	1244/1347

표 4. CPSP 최대값이 문턱값 0.2 이상인 프레임 수
(TDOA가 바르게 추정된 프레임 수/ 전체 유효 프레임 수)
Table 4. The number of frames whose CPSP maximum value is greater than 0.2
(number of frames with correct TDOA estimation/number of total frames whose CPSP maximum value is greater than 0.2)

SNR \ T_{60}	0.0s	0.1s	0.2s	0.3s
0dB	0/0	0/0	0/0	0/0
5dB	181/181	0/0	0/0	0/0
10dB	418/418	72/72	0/0	0/0
15dB	730/730	358/358	1/1	0/0
20dB	1152/1152	589/589	111/111	0/0
25dB	1408/1408	1021/1021	309/309	67/67
30dB	1439/1439	1357/1357	475/475	130/130
35dB	1440/1440	1423/1423	667/667	148/148
40dB	1440/1440	1439/1439	825/825	170/170

5. 결론

본 논문에서는 마이크로폰 쌍을 이용하여 잡음과 잔향이 존재하는 환경에서 CPSP 함수를 이용하여 음원도달방향을 추정할 때, CPSP의 최대값에 문턱값을 설정하여 음원도달방향 추정 성능을 향상시킬 수 있는 방법을 제안하였다. 잡음 및 잔향 환경에서 CPSP 최대값의 분포에 대한 분석을 통하여 문턱값을 0.1 ~ 0.2로 설정 시 추정 신뢰도를 크게 향상 시킬 수 있음을 보이고, 모의실험을 통해 제안한 방법의 타당성을 검증하였다.

참고문헌

Veen, B. V., and Buckley, K. (1988). “Beamforming: A versatile approach to spatial filtering”, *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4-24, Apr.

Claudio, E. D., Parisi, R., and Orlandi, G. (2000). “Multi-source localization in reverberant environments by root-music and clustering”, *ICASSP*, vol. 2, pp. 921-924, Jun.

Knapp, C., and Carter, G. (1976). “The generalized correlation method for estimation of time delay”, *Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320-327, Aug

Allen, J., and Berkley, D. (1979). “Image method for efficiently simulating small-room acoustics”, J. Acoust. Soc. Am, vol. 65, no. 4, pp. 943-950, Apr.

Kwon, HongSeok, Bae Keunsung (2009). “*Performance Improvement of CPSP Based TDOA Estimation Using the Preemphasis*”, Acoustical Society of Korea, vol. 28, no. 5, pp. 461-470 July.
(권홍석, 배건성, (2009). “프리엠퍼시스를 이용한 CPSP 기반의 도달시간차이 추정 성능 개선”, 한국음향학회지, 제28권 제5호, pp. 461-470.)

• **전성일 (Quan, Xingri)**
경북대학교 전자전기컴퓨터학부
대구 북구 산격동 1370번지
Tel: 053-940-8627
Email: comet263@hotmail.com
관심분야: 음성신호처리, PCG 심음신호처리
현재 전자전기컴퓨터 대학원 박사과정 재학 중

• **배건성 (Bae, Keunsung)**, 교신저자
경북대학교 IT대학 전자공학부
대구 북구 산격동 1370번지
Tel: 053-950-5527
Email: ksbae@ee.knu.ac.kr
관심분야: 음성신호처리, 디지털신호처리, 수중음향신호처리
1979~현재 전자공학부 교수