

## Modified Subband CPSP를 이용한 음원 추적 시스템에 관한 연구

오상현\*, 박규식\*

\*상명대학교 컴퓨터·정보·통신학부

E-mail : sanghun@smuc.sangmyung.ac.kr

kspark@smuc.sangmyung.ac.kr

### 요약

본 연구는 폐쇄된 임의의 공간상에서 2개의 마이크로폰 어레이를 이용 수신된 2개 신호의 도착 시간차를 계산하여 음원의 방향 각을 추정하는 새로운 알고리즘을 제안한다. 제안된 MSCPSP 알고리즘은 기존의 CPSP 알고리즘을 개선한 것으로, 마이크로폰에 수신된 2개의 입력 신호에 대해 서브밴드 필터 뱅크를 이용하여 대역 분할하고 각 서브밴드 대역에서 구해지는 신호 대 잡음비(SNR)를 대역별 CPSP 결과에 가중치로 제공한다. 이러한 대역 분할 가중방식은 잡음의 영향을 각 대역으로 한정 분산시켜 보다 정확한 지연 시간 추정을 가능하게 한다. 제안된 알고리즘의 성능을 입증하기 위해 기존의 CPSP와 MSCPSP 알고리즘의 컴퓨터 모의 실험을 수행하였으며, 실험 결과 제안된 MSCPSP의 우수함을 볼 수 있었다.

### 1. 서론

임의의 공간상에서 인위적으로 발생하는 음성이나 음향으로부터 음원의 위치를 추적하는 sound localization 기술은 고품질의 음성/음향 획득을 위한 sound acquisition, 원격회의 시스템, 음성인식 시스템, hand-free audio 통신 시스템, video 감시 시스템 등 다양한 분야에서 응용 가능한 기초 기반 기술이다. 폐쇄된 공간에서 발생된 음원으로부터 음원의 위치를 추적하는 방법은 2개나 혹은 그 이상의 마이크로폰 어레이를 사용하여 수신된 2개의 마이크로폰 입력 신호간의 시간 지연(Time Delay)을 측정함으로써 음원의 방위각을 계산하고 마이크로폰의 배치 기하학

을 이용하여 음원의 위치를 결정하게 된다.

마이크로폰에 수신된 2개의 신호간에 시간 지연을 추정하는 알고리즘은 다양한 응용분야에 적용 될 수 있는 기반 기술로서 최근 많은 관련 알고리즘들이 개발되고 있는 실정이다. 이러한 알고리즘에서는 폐쇄 공간을 작은 사무실이나 회의실 등으로 제한하고 이러한 사무 공간 내에서 발생할 수 있는 주변 잡음(컴퓨터, 에어컨 등)과 음향 반향 등 장애 요소들의 영향을 최소화하여 가급적 추정된 시간 지연 차 분산을 최소화 할 수 있는 알고리즘을 제안하고 있다. 최근에 제안된 대부분의 시간 지연 추정 알고리즘은 주로 Cross-Correlation[1]이나 Generalized Cross-Correlation 방법(GCC)[2]을 사용하고 있는데 이러한 방법들로서는 Roth Processor[3], SCOT(Smoothed Coherence Transform)[4], CPSP(Cross-Power Spectrum Phase)[5], Eckart Processor[6], ML Estimation[2] 등이 있다.

본 연구는 지연 시간 추정시 암펄스형태의 비교적 정확한 시간차 계산이 가능한 아정과 적은 계산량으로 인하여 최근 많이 사용되고 있는 CPSP[5][7] 알고리즘에서 단점으로 알려져 있는 저주파 대역 잡음에 대한 성능저하를 향상시킬 수 있는 Modified Subband CPSP(MSCPSP) 알고리즘을 제안한다. 제안된 MSCPSP 알고리즘은 마이크로폰에 수신된 2개의 입력 신호를 서브밴드 필터 뱅크를 이용하여 대역 분할하고 각 서브밴드 대역에서 구해지는 신호 대 잡음비(SNR)를 대역별 CPSP 결과에 가중치로 제공함으로써, 잡음에 의한 시간 지연 추정시 잡음의 영향을 각 대역으로 한정하여 최소화하는 방법을 제공한다.

## II. 지연 시간 추정 알고리즘

2개의 microphone에 수신된 신호로부터 음원의 지연시간차 추정은 크게 두 신호간의 상관성 관계를 이용한 Cross-Correlation 방법과 신호의 상관성에 가중치를 부가하는 Generalized cross-correlation (GCC) 2가지로 나누어진다.

먼저 2개의 마이크로폰에 수신되는 신호를  $x_1(t)$ ,  $x_2(t)$ 이라고 할 때, 수신 신호는 다음과 같이 표현 될 수 있다.

$$\begin{aligned} x_1(t) &= s_1(t) + n_1(t) \\ x_2(t) &= s_1(t+D) + n_2(t) \end{aligned} \quad (1)$$

수식 (1)에서 D 는 음원신호의 도착 지연 시간,  $s_1(t)$  는 음원 신호,  $n_1(t), n_2(t)$ 는 공간 내에 존재하는 잡음신호와 반향 신호등의 음향 잡음을 의미한다. 여기서 음원 신호  $s_1(t)$  는 잡음 신호와 비 상관 관계에 있고, 잡음신호  $n_1(t), n_2(t)$  는 AWGN(Additive White Gaussian Noise)로 가정되며 지연 시간 추정 알고리즘은 음원의 도착 지연 시간차인 D를 계산하게 된다.

Cross-Correlation 방식은 마이크로폰에 수신된 2개 신호의 Cross-Correlation 함수를 구하여 이 함수를 최대화하는 지연 시간을 계산하는 방식으로, Cross-Correlation은 다음과 같이 정의된다.[1]

$$\begin{aligned} R_{x_1x_2}(\tau) &= E[x_1(t) x_2(t-\tau)] \\ &= \int_{-\infty}^{\infty} S_{x_1x_2}(f) e^{j2\pi f\tau} df \end{aligned} \quad (2)$$

수식 (2)에서  $S_{x_1x_2}(f)$ 는 Cross Power Spectrum을 나타내며, 이러한 Cross-Correlation 방식은 공간내의 주변잡음과 반향에 상당히 민감하게 반응하는 단점을 가지고있다.

Generalized Cross-Correlation(GCC) 방법은 수식 (2)의 Cross Power Spectrum에 가중함수를 주어 기존의 Cross-Correlation 방법의 단점을 보완하는 것으로서 다음과 같이 정의된다.[2]

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

여기서  $\phi_g(f)$ 는 가중함수로서 실제 구하고자 하는 지연 시간값은 가중 Cross Power Spectrum 인  $S_{x_1x_2}^{(g)}(f) = \phi_g(f) S_{x_1x_2}(f)$ 의 역 푸리에 변환값인 GCC 함수를 최대화하는 값으로 구해진다. 참고로 수식 (3)에서 가중 함수를  $\phi_g(f) = 1$ 로 하면 Cross-Correlation 방법과 동일함을 볼 수 있다.

## III. CPSP(Cross Power Spectrum Phase) 알고리즘

Cross-Power Spectrum Phase 알고리즘은 수식 (3)의 GCC 에서 가중 함수를 다음과 같이 입력신호  $x_1(t), x_2(t)$  의 cross-power spectrum의 절대값을 취함으로서 정의된다.

$$\phi_c(f) = \frac{1}{|S_{x_1x_2}(f)|} \quad (4)$$

여 경우 GCC 함수는 다음과 같이 표현될 수 있으며

$$R_{x_1x_2}^{(c)}(\tau) = \int_{-\infty}^{\infty} \frac{S_{x_1x_2}(f)}{|S_{x_1x_2}(f)|} e^{j2\pi f\tau} df \quad (5)$$

여기서 잡음 및 음원신호의 비상관성을 가정하면 일반화된 Cross Power Spectrum 과 Cross-Correlation은 다음과 같이 구해진다.

$$\begin{aligned} S_{x_1x_2}^{(c)}(f) &= \frac{S_{x_1x_2}(f)}{|S_{x_1x_2}(f)|} = e^{j2\pi fD} \\ R_{x_1x_2}^{(c)}(\tau) &= \delta(t-D) \end{aligned} \quad (6)$$

따라서 CPSP의 GCC 함수는 임펄스 함수 형태로 구해질 수 있으며 이러한 GCC의 특성은 Cross-Correlation의 최대 값을 구할 때 날카로운 피크 값을 용이하게 찾을 수 있기 때문에 성능이 뛰어난 것으로 알려져 있다. 하지만 이러한 임펄스 특성은 실제 실험에서 잡음 및 음원의 상관성 정도로 인하여 피크 값이 널리 퍼지게 된다. 또한 CPSP 알고리즘은 입력신호의 스펙트럼을 스펙트럼의 절대값으로 나누어줌으로서 가중함수 특성상 신호의 에너지가 작은 대역에서는 CPSP의 spectral 크기가 커지게 되어 잡음에 의한 심각한 오차가 발생하게 된다.

따라서 CPSP의 이러한 단점을 극복하기 위해

서는 신호의 에너지가 높은 대역과 낮은 대역 (혹은 잡음 에너지가 작거나 높은 대역)을 구분하여 따로 가중치를 줌으로서 알고리즘의 정확성을 높일 수 있다. 예를 들면, 일반적인 작은 사무실 환경에서는 대부분의 잡음(컴퓨터, 소형 냉장고, 에어컨등)이 500Hz 미만에 놓이게 됨으로 잡음 에너지가 높은 대역과 낮은 대역을 구분하여 가중치를 줌으로서 CPSP 알고리즘의 성능을 향상시킬 수 있다.

#### V. Modified Subband CPSP (MSCPSP) 알고리즘

위에서 언급한 바와 같이 단일 CPSP 알고리즘은 가중 함수의 특성상 잡음 에너지가 많은 대역에서는 상대적인 오차가 많이 발생함으로써 전체 CPSP 알고리즘의 성능을 저하시키게 된다. 따라서 본 연구에서는 그림 1과 같이 마이크로폰에 입력되는 신호를 N개의 서브밴드 대역으로 나누어서 대역별 SNR 가중함수를 대역별 CPSP 알고리즘에 적용하는 Modified Subband CPSP를 제안한다.

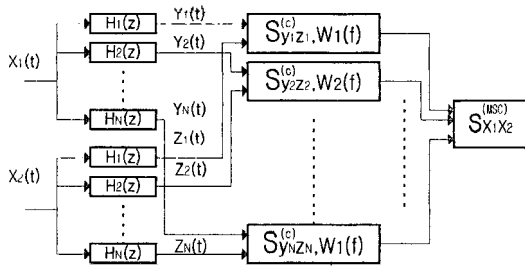


그림 1. Modified Subband CPSP

그림 1에서  $H_1(z), H_2(z), \dots, H_N(z)$ 는 N개의 Subband Bandpass filter 이며, 각 Bandpass filter를 거친 신호들은 각 대역별로 SNR 가중치  $W_i(f)$ 를 부여한 CPSP를 계산하여 다음과 같이 Modified Subband CPSP(MSCPSP)를 구하게 된다.

$$S_{x_1x_2}^{(MSC)}(f) = \sum_{i=0}^{N-1} W_i(f) S_{y_{zi}}^{(C)}(f) \quad (7)$$

수식 (7)에서  $S_{y_{zi}}^{(C)}(f)$ 는  $i$  번째 대역의 CPSP 그리고  $W_i(f)$ 는  $i$  번째 대역의 SNR 가중치를 나타내며 다음과 같은 유사도 함수로 구해진다.

$$W_i(f) = \frac{r^2(f)}{1-r^2(f)} \quad (8)$$

$$r^2(f) = \frac{|S_{y_{zi}}|^2}{S_{y_{zi}} S_{z_{zi}}}$$

여기서  $S_{y_{zi}}, S_{y_{zi}}, S_{z_{zi}}$ 는  $i$  번째 대역 신호에 대한 각 밴드별 Cross-Power Spectrum과 Auto-Power Spectrum을 나타낸다. 한편 수식 (7)의 MSCPSP는 역푸리에 변환 과정을 거쳐 다음과 같은 MSCPSP의 GCC 함수를 구하게 되고

$$R_{x_1x_2}^{(MC)}(\tau) = \sum_{i=0}^{N-1} [w_i(\tau) * R_{y_{zi}}^{(C)}(\tau)] \quad (9)$$

여기서 구하고자 하는 지연 시간차는 수식(9)의 일반화된 cross correlation을 최대화하는 D를 구하면 된다.

#### V. 컴퓨터 모의 실험

본 연구의 컴퓨터 모의 실험 환경은 다음과 같다. 먼저 수식 (1)에 따라 인공적으로 음원 신호  $s_1(t)$ , 잡음신호  $n_1(t), n_2(t)$ 를 발생시키기 위해 샘플링 주파수를 16Khz로 하여 2초 분량의 3개 white gaussian sequence를 MATLAB rplid() 함수를 이용하여 발생하였다. 이 중에서 하나의 white gaussian sequence는 AR(6) 필터를 통과시켜 음성과 유사한 특징을 갖는 신호를 만들었으며 나머지 2개의 white sequence는 잡음신호로 사용하였다. 이러한 잡음신호는 모의 실험 환경이 작은 사무실이라는 조건하에 잡음신호를 500Hz cutoff 주파수를 갖는 저 대역(LPF) 통과 필터를 통과시켜 실제 환경과 유사하게 잡음신호를 발생하였다. 또한 CPSP와 제안된 MSCPSP 알고리즘의 성능을 사무실내 잡음뿐만 아니라 음향 반향 환경 하에서도 고려하기 위해, 음원신호의 크기를 감쇠 하여 인위적으로 지연시켜 반향 신호를 만들었다. 이때 반향신호의 감쇠정도는 Eyring의 image model을 사용하여 계산하였다.

위에서 만들어진 음원신호, 잡음신호, 반향신호는 서로 더해져서 2개의 마이크로폰에 수신되는 입력 신호들을 만들었으며 이때 두 개의 입력 신호간에 상대적인 시간 지연을 갖도록 인위적으로 2 sample 지연 시간을 갖도록 하였다.

한편 CPSP와 MSCPSP 알고리즘 모의 실험시

Cross Power spectrum 계산은 2초 분량의 입력 데이터 sequence를 20개의 구간으로 나누어서 Periodogram Spectrum 추정 방법을 사용하였다. 모의 실험에 사용된 MSCPSP의 Subband 수는 2개로서 이는 모의 실험의 계산 량을 줄이기 위함이며 사용된 LPF 와 HPF는 Binomial 특성을 갖는 7 tap 필터를 사용하였다.

CPSP 와 제안된 MSCPSP 알고리즘의 모의 실험은 다양한 잡음환경을 고려하여 각각 SNR이 20dB, 15dB, 10dB, 5dB, 0dB, -5dB 6가지 신호 대 잡음비를 적용하였으며 각 SNR에서 20번의 반복실험을 하여 총 240번의 독립 실험을 수행하였으며, 각각의 알고리즘 성능을 비교하기 위해 최종 Cross-Correlation 함수로부터 추정된 지연 시간과 PFR(Peak to Floor Ratio)을 계산하였다. PFR은 최종 GCC 결과로부터 최대 피크치와 이 값을 제외한 나머지 값들의 합과의 비교를 통해 최대 피크치 근처에서 발생하는 사이드 피크 값의 크기를 측정하는 것으로서 지연 시간 추정의 정확성을 평가할 수 있는 중요한 척도가 되며 다음과 같이 정의된다.

$$PFR = \frac{\max_k R_{x_1 x_2}^{(GCC)}(k)}{\sum_{j \neq k} R_{x_1 x_2}^{(GCC)}(j)} \quad (10)$$

그림 2는 각 주어진 SNR에서의 PFR 값의 변화를 보이고 있다. 그림에서 보듯이 SNR이 비교적 높은 상황에서는 2개의 알고리즘이 차이가 없지만 SNR이 낮아질수록 제안된 MSCPSP 알고리즘의 성능이 우수함을 알 수 있다.

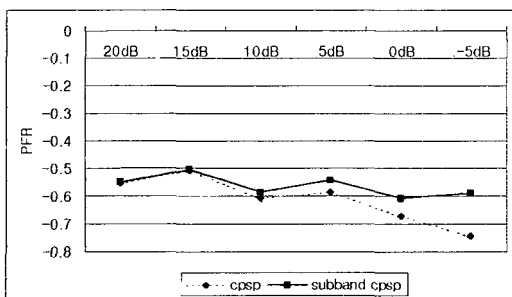


그림 2. SNR에 따른 PFR 비교

본 모의 실험에서는 비교적 단순한 계산 량을 위해서 MSCPSP의 subband를 2개의 주파수 대역을 분할하였으나 subband 수를 늘릴수록 제안된 MSCPSP 알고리즘의 성능이 우수함을 보이고 있다.

## V. 결론

본 연구는 임의의 공간상에서 2개의 마이크로폰 어레이를 이용 수신된 2개 신호의 도착 시간차를 계산하여 음원의 방향 각을 추정하는 새로운 알고리즘을 제안하였다. 제안된 MSCPSP 알고리즘은 서브밴드 필터 뱅크를 이용하여 각 대역별 SNR을 CPSP 에 가중치로 제공하는 방법으로 이러한 대역 분할 가중방식은 잡음의 영향을 각 대역으로 한정, 분산시켜 보다 정확한 지연 시간 추정을 수행할 수 있다. 컴퓨터 모의 실험결과 제안된 MSCPSP 가 PFR 측면에서 더욱 좋은 성능을 보이고 있으며, 특히 잡음이 없을 수록 MSCPSP의 성능이 우수함을 알 수 있었다.

본 연구에서 제안된 MSCPSP는 Subband filter를 이용한 대역 분할 방법으로서, bandpass filter의 최적 설계 방식에 따라 지연 시간 추정 성능을 향상시킬 수가 있을 것으로 기대된다.

## 참고문헌

- [1] A. Papoulis, Probability, Random Variables and Stochastic Process, New York, McGraw-Hill, 1965
- [2] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans., Acoust., Speech, Signal Processing, vol. ASSP-24, Aug., 1976
- [3] P. R. Roth, "Effective measurements using digital signal analysis", IEEE Spectrum, vol. 8, Apr. 1971
- [4] G. C. Carter, A. H. Nuttall and P. G. Cable, "The smoothed coherence Transform", Proc. IEEE, vol. 6, Oct. 1973
- [5] M. Omologo and P. Svaizer, "Acoustic event localization using a cross-power spectrum phase based technique", Proceedings of ICASSP'94, Adelaide, Australia, 1994
- [6] C. Eckhart, "Optimal rectifier systems for detection of steady signals", Scripps. Inst Oceanography, Marine Physical Lab., Univ. California, Rep. SIO 12692, 1952
- [7] M. Omologo and P. Svaizer, "Acoustic source location in noisy and reverberant environment", Proceedings of ICASSP'96, Atlanta, GA, 1996