

공간좌표로 사상된 GCC 함수의 다 음원에 대한 해석과 음원 위치 추정 방법

Spatially Mapped GCC Function Analysis for Multiple Source and Source Localization Method

권 병 호*, 박 영 진, 박 윤 식
(Byoung-ho Kwon, Youngjin Park, and Youn-sik Park)

Abstract: A variety of methods for sound source localization have been developed and applied to several applications such as noise detection system, surveillance system, teleconference system, robot auditory system and so on. In the previous work, we proposed the sound source localization using the spatially mapped GCC functions based on TDOA for robot auditory system. Performance of the proposed one for the noise effect and estimation resolution was verified with the real environmental experiment under the single source assumption. However, since multi-talker case is general in human-robot interaction, multiple source localization approaches are necessary. In this paper, the proposed localization method under the single source assumption is modified to be suitable for multiple source localization. When there are two sources which are correlated, the spatially mapped GCC function for localization has three peaks at the real source locations and imaginary source location. However if two sources are uncorrelated, that has only two peaks at the real source positions. Using these characteristics, we modify the proposed localization method for the multiple source cases. Experiments with human speeches in the real environment are carried out to evaluate the performance of the proposed method for multiple source localization. In the experiments, mean value of estimation error is about 1.4° and percentage of multiple source localization is about 62% on average.

Keywords: multiple source localization, TDOA (Time Delay of Arrival), GCC (Generalized Cross Correlation)

I. 서론

마이크로폰 어레이에서 측정된 음향 신호를 이용하여 음원의 방향 및 위치를 알아내는 방법을 음원 위치 추정(sound source localization)이라고 한다[1]. 음원 위치 추정 방법은 기계장치에서 소음의 위치를 알아낼 목적으로 사용되기도 하며, 화상회의에 화자의 위치로 카메라의 방향을 제어하기 위해서 사용되기도 한다[2,3]. 뿐만 아니라, 사람과 로봇의 원활한 상호작용을 위한 로봇의 청각 시스템에 활용되기도 한다[4]. 로봇의 청각 시스템에 적용된 음원 위치 추정 방법에는 로봇의 머리전달함수를 이용한 방법, 다수의 마이크로폰 어레이를 이용하여 빔포밍을 적용한 방법, 인공귀를 이용한 방법 등이 있다[5-9]. 또한 두 마이크로폰 사이의 도달시간지연을 이용한 음원 위치 추정 방법도 널리 연구되고 있으며, 이는 적은 계산량과 넓고 정확한 방향 감지 특성을 갖고 있다[10,11]. 도달시간지연을 이용한 음원 위치 추정 방법은 3차원 공간상에 존재하는 음원의 위치를 추정하거나, 음원 위치 추정 성능을 향상시키기 위해서 다수의 마이크로폰이 사용되는데, 이 경우 다수의 마이크로폰 쌍에서 추정된 도달시간지연의 조합으로부터 음원의 위치를 추정하는 방법이

필요하게 된다. 일반적으로 음원의 위치는 추정된 도달시간지연에 해당하는 음원의 위치들 사이에서 오차를 최소화하는 위치로 결정하게 된다. 하지만 기존의 연구에서 소수의 마이크로폰만을 이용하는 경우에도 높은 음원 위치 추정 분해 성능을 가지며, 낮은 신호대잡음비에서도 좋은 추정 성능을 갖는 음원 위치 추정 방법을 제안하였다[12]. 제안된 방법은 도달시간지연에 기반을 둔 음원 위치 추정 방법으로, 공간좌표로 사상된 GCC 함수를 이용하여 하나의 음원의 위치를 추정할 수 있는 방법이었다. 즉, 로봇과 한 사람과의 상호작용만을 가정하여 화자의 위치를 추정할 수 있도록 하였다. 하지만 로봇과 다수의 사람이 동시에 상호작용을 할 수 있기 때문에 다 음원이 존재하는 경우에도 적용할 수 있는 음원 위치 추정 방법이 필요하다.

따라서 본 논문에서는 제안된 공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법을 다 음원이 존재하는 경우에도 적용할 수 있도록 수정·보완하며, 그 성능에 대해서 검증하고자 한다.

II. 다 음원 위치 추정 방법

본 논문에서 제안하는 음원 위치 추정 방법은 시간영역에서 기존의 도달시간지연을 이용한 음원 위치 추정 방법과는 달리 상호상관함수에서 시간지연을 추정하지 않고, 상호상관함수를 적절한 사상함수를 이용하여 공간좌표로 사상시킨 후 음원의 위치를 추정하는 방법이다[12].

먼저 식 (1)와 같이 각 마이크로폰 쌍에서 측정된 신호를 이용하여 일반화된 상호상관 방법(Generalized Cross Correlation method, PHAT 가중치 함수 적용)을 이용해 상호상관함수를 구한다[13].

* 책임저자(Corresponding Author)

논문접수: 2009. 10. 21., 수정: 2009. 12. 24., 채택확정: 2010. 1. 20.

권병호, 박영진, 박윤식: KAIST, 기계공학부

(bhkwon@kaist.ac.kr/yjpark@kaist.ac.kr/yspark0117@kaist.ac.kr)

* 이 논문은 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업[2008-F-044-01, 전자파, 음향 및 건물 환경을 개선하는 지능형 건설 IT 융합 신기술 개발], 한국과학재단을 통해 교육과학기술부의 국가지정연구실 사업(R0A-2005-000-10112-0) 일환으로 수행하였음.

* 상기 논문은 KACC2009 학술대회에서 초안이 발표되었습니다.

$$R_i(\tau) = \int_{-\infty}^{\infty} \frac{G_i(f)}{|G_i(f)|} e^{j2\pi f\tau} df \quad (1)$$

G_i 는 i 번째 마이크로폰 쌍에서 측정된 신호의 상호스펙트럼밀도함수이며, R_i 는 상호상관함수이다. 식 (1)에서 계산된 상호상관함수는 식 (2)와 같이 마이크로폰이 존재하는 공간에 따른 적절한 사상함수, $\Theta: \tau \rightarrow \theta$ 를 이용하여 음원 위치에 대한 공간 좌표로 사상되고, 사상된 함수를 공간좌표로 사상된 GCC 함수(spatially mapped GCC function, $mGCC$)라고 정의한다.

$$mGCC_i(\theta) = \Theta(R_i(\tau)) \quad (2)$$

각 마이크로폰 쌍에서 얻어진 $mGCC$ 함수는 정의된 전체 좌표(global coordinate)로 변환 후, 식 (3)과 같이 동일한 공간 좌표에 대해 합하여 하나의 함수로 표현된다. 이를 합 GCC 함수(summed GCC function, $sGCC$)라고 정의한다. 이때 각 $mGCC$ 함수를 동일한 공간좌표로 표현하는 방법은 여러 가지가 있을 수 있으며, 본 논문에서는 전체 좌표에 대해 각 $mGCC$ 함수 값을 더하여 표현한다.

$$sGCC(\theta) = \sum_{i=1}^M mGCC_i(\theta) \quad (3)$$

M 은 적용된 마이크로폰 쌍의 개수이다. 여기에서 단 음원인 경우에는 $sGCC$ 함수에서 최대값을 갖게 하는 θ 가 음원의 위치가 되었다. 하지만 다 음원인 경우에는 위치 추정 방법이 식 (4)와 같이 수정된다.

$$[\theta] = \{\theta \mid sGCC(\theta) \geq TH\}, \quad (4)$$

$$TH \equiv c \cdot \max(sGCC(\theta))$$

식 (4)는 $sGCC$ 함수가 특정 값, TH 이상의 값을 갖는 위치를 음원의 위치로 결정하며, 본 논문에서는 TH 값을 $sGCC$ 함수의 최대값에 보정계수, c 를 곱한 값으로 정하였다. c 값은 마이크로폰의 위치와 음원의 개수 및 특징, 적용 공간에 대해 달라질 수 있는 값으로 사전 실험을 통해서 결정될 수 있다.

이상에서 설명한 방법이 공간좌표로 사상된 GCC 함수를 이용한 다 음원 위치 추정 방법이다.

III. 다 음원에 대한 공간좌표로 사상된 GCC 함수의 특성

제안된 음원 위치 추정 방법의 다 음원에 대한 성능 평가에 앞서, 다 음원의 경우 공간좌표로 사상된 GCC 함수가 어떠한 특성을 갖는 지에 대해서 살펴보고자 한다. 이를 위해서 두 개의 음원이 각각 $\pm 30^\circ$ 에 위치하고 있는 경우에 두 개의 마이크로폰을 이용하여 측정된 신호를 이용하여 공간좌표로 사상된 GCC 함수를 계산해 보았다. 이때 두 음원의 신호를 correlated white noise와 uncorrelated white noise로 하였으며, 두 음원 사이의 파워 비(source power ratio, SPR)를 0dB에서 6dB로 변화시켜 가면서 측정하였다. 그 결과는 그림 1, 2와 같다. 그림 1의 correlated white noise의 경우에는 실제 음원의 위치에 해당하는 $\pm 30^\circ$ 위치뿐만 아니라 0° 에서도 공간좌표로 사상된 GCC 함수의 peak값이 존재하는 것을 알 수 있다.

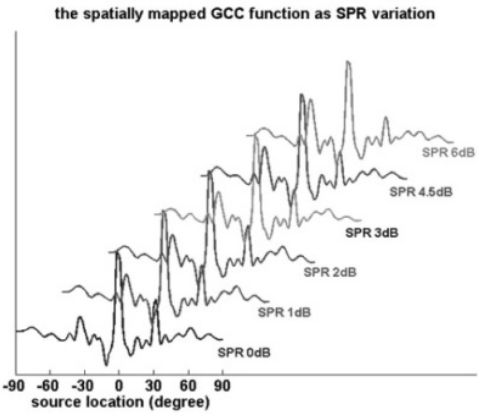


그림 1. Correlated white noise인 경우의 두 음원 사이의 파워 비에 따른 공간좌표로 사상된 GCC 함수.
Fig. 1. The spatially mapped GCC function as SPR variation when two source signals are correlated.

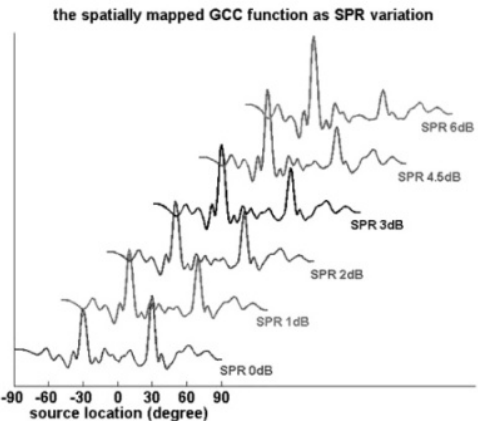


그림 2. Uncorrelated white noise인 경우의 두 음원 사이의 파워 비에 따른 공간좌표로 사상된 GCC 함수.
Fig. 2. The spatially mapped GCC function as SPR variation when two source signals are uncorrelated.

이는 두 음원에 의해서 생성된 가상 음원의 위치에 peak가 발생하게 됨을 의미하고, 본 논문에서 제안된 방법으로는 실제 음원의 위치를 추정하기가 어렵게 된다. 그림 2의 uncorrelated white noise의 경우에는 실제 음원의 위치 $\pm 30^\circ$ 에 정확하게 peak가 발생하는 것을 알 수 있다. 뿐만 아니라, 두 음원의 파워 비에 의해서 peak의 크기가 달라지는 것을 알 수 있는데, 이는 제안된 다 음원 위치 추정 방법에서 음원 사이의 파워 비가 실제 음원의 위치 추정에 영향을 미칠 수 있음을 의미한다. 여기에서 음원 사이의 파워 비와 공간상으로 사상된 GCC 함수의 peak의 크기 비는 정량적으로 계산될 수 있지만, 측정되는 공간에 따라 달라지기 때문에 본 논문에서는 자세하게 언급하지 않는다.

이상에서 살펴본 비에 의하면, 제안된 음원 위치 추정 방법을 적용하여 다 음원의 위치를 추정하기 위해서는 음원들 사이에 상관성이 없거나 적어야 하며, 각 음원의 크기가 비슷해야만 한다. 또한 두 음원의 파워 비에 의해 peak의 크기 비가 결정되므로 이를 고려하여, 식 (4)의 보정계수, c 를 결정해야 한다.

IV. 다 음원 위치 추정 실험

앞 절에서 설명한 다 음원 위치 추정 방법에 대한 성능 평가를 위해서 마이크로폰이 설치되어 있는 로봇 플랫폼을 이용하여 두 개의 음원의 위치를 추정하는 실험을 수행하였다. 실험 환경은 잔향시간이 약 0.3sec인 연구실로 그림 3과 같다. 두 개의 스피커를 로봇 플랫폼에서 2m 떨어진 곳에 각각 +10°, -60°에 위치시켰다. 각각의 스피커에는 남성과 여성의 음성을 5초간 재생시켜 주었으며 그 신호는 그림 4와 같다. 이 때 배경 잡음에 대해서 신호의 크기가 10dB 이상이 되도록 하였다. 음원의 수평각을 추정하기 위해서 로봇 플랫폼에 위치한 세 개의 마이크로폰에서 샘플링 주파수를 16kHz로 하여 신호를 측정하였으며, 측정된 신호에서 매 0.25sec 마다 음원의 위치를 추정하게 하였다. 또한 제한된 방법의 성능 반복성(repeatability)을 검증하기 위해서 5번 실험을 통해 성능을 평가하였다.

제안된 방법으로 다 음원의 위치를 추정하기에 앞서 보정 계수 c , 값을 정해야 한다. 본 실험에서 c 값은 그림 3의 실험 조건과 동일한 조건에서 각 스피커에 white noise를 재생시켰을 때, 측정된 신호를 이용하여 계산된 합 GCC 함수에서

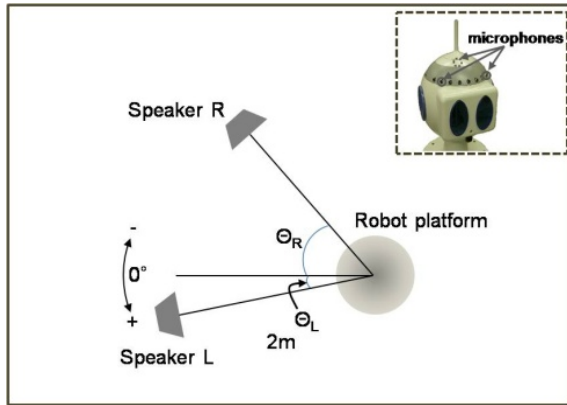


그림 3. 실험 환경; Speaker L: +10°, Speaker R: -60°, 로봇 플랫폼 (우상).

Fig. 3. Experimental condition; Speaker L: +10°, Speaker R: -60°, Robot platform (in the dotted line box).

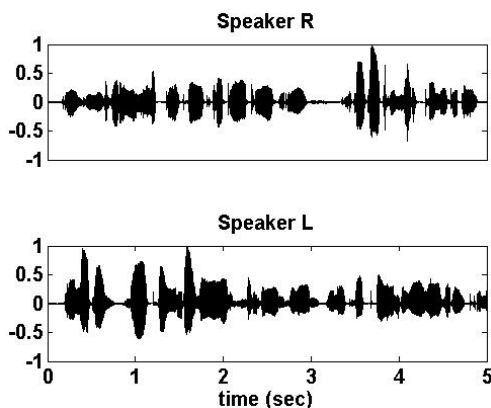


그림 4. 각 스피커에서 재생된 음원 신호; (Speaker R: 여성, Speaker L: 남성).

Fig. 4. Source signals for each speaker; (Speaker R: woman speech, Speaker L: man speech).

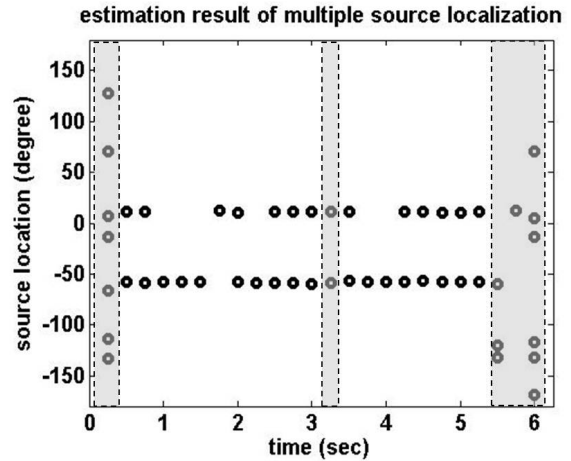


그림 5. 다 음원 위치 추정 결과 (회색 박스 부분 - 비음성 구간).

Fig. 5. Experimental result for multiple source localization (gray box - non-speech period).

각 음원의 위치에 해당하는 peak값의 비로 결정하였다. 실험 결과 5번 실험의 평균값인 0.7을 본 실험에서의 보정계수로 정하였다[14].

예비 실험을 통해 얻어진 보정계수 값을 이용하여 음원의 위치를 추정한 결과는 그림 5와 같다. 측정된 신호에서 음성 구간에서만 음원 위치 추정을 하기 위해서 음성/비음성 구분 알고리즘을 적용하였다[15]. 모든 구간에서 음원 위치 추정 알고리즘을 적용하되, 음성 구간에서의 결과만을 신뢰하도록 하였다. 그림 5는 5번의 실험 중에서 1번의 실험 결과이며, 측정된 신호에서 각 구간의 음원 위치 추정 결과이다. 회색 박스 부분은 음성/비음성 구분 알고리즘에 의해서 비음성 구간으로 검지되어 추정 성능 결과에서는 배제하였다. 5번의 실험 결과를 바탕으로 음원 위치 추정 성능을 정량적으로 나타내기 위해서 본 논문에서는 식 (5), (6)과 같이 정의된 추정 오차 평균(MEE: Mean value of Estimation Error)과 다 음원 추정률 (Percentage of Multiple Source Localization, MSL)을 계산하였다. 추정 오차 평균은 실제 음원의 위치와 추정된 음원의 위치 사이의 절대적 오차 평균으로 음성 구간 내에서 모든 음원의 위치에 대해서 평균값으로 계산되었다. 또한 다 음원 추정률은 실제 다 음원이 존재하는 경우에 정의된 오차 범위 내로 음원의 위치가 동시에 추정된 경우의 백분율로 다 음원 추정 성능을 나타낸다.

$$MEE = \frac{1}{\sum m_j} \sum_{j=1}^n \sum_{i=1}^{m_j} |\theta_{jiTRUE} - \theta_{jiESTIMATION}| \quad (5)$$

$$MSL(\%) = \frac{\text{No. of multiple source estimation}}{\text{No. of all estimation in speech period}} \times 100 \quad (6)$$

식 (5)에서 n 은 음원의 개수이며, m_j 는 각 음원에 대해 음성 구간에서 음원의 위치를 추정한 총 횟수이다. 5번의 실험 결과에 대해서 추정 오차 평균과 다 음원 추정률을 표 1에 정리하였다. 다 음원 추정률에서 추정오차범위는 $\pm 5^\circ$ 로 하였다. 5번의 실험결과에서 추정 오차 평균은 약 1.4°로 추정된 다 음원의 위치는 상당히 정확함을 알 수 있다. 뿐만 아니라, 다

표 1. 실험 결과.

Table 1. Experimental results.

trials	1	2	3	4	5	avg.
MEE(°)	1.4	1.3	1.2	1.5	1.5	1.4°
MSL(%)	71	75	71	38	53	62%

음원 추정률은 평균 62%의 정확성을 보여주고 있다. 이상의 결과는 기존의 공간좌표로 사상된 GCC 함수를 이용한 단일 음원의 위치 추정 방법에서의 실험 결과와 유사한 추정 오차 평균 값을 보이며, 추정 방법의 수정으로 다 음원의 위치를 정확하게 추정할 수 있게 되었음을 보여준다. 또한 다 음원 추정률은 실험 조건에 따라 두 음원 사이의 파워비가 일정하지 않기 때문에 편차가 발생함을 볼 수 있다. 여기에서 다 음원에 대한 추정 오차 평균값과 다 음원 추정률 결과는 식 (4)의 다 음원 추정 방법에서 보정계수 c , 값에 의존적이다. 보정계수 값이 커지면 추정 오차 평균값은 작아지며, 다 음원 추정률은 낮아진다. 반대로 보정계수 값이 작아지면 추정 오차 평균값은 커지는 반면, 다 음원 추정률은 높아지게 된다. 그 예로 4번째 실험에서 보정계수를 0.6, 0.8로 적용한 경우 다 음원 추정률은 각각 81%, 18%이고, 추정 오차 평균은 4.3°, 1.6° 였다. 기존의 연구에서 제안된 공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법이 추정 오차 평균 관점에서는 상당히 강인하므로 보정계수에 따른 추정 오차 평균은 덜 민감하지만, 다 음원 추정률에 있어서는 상당히 민감함을 알 수 있다.

V. 결론

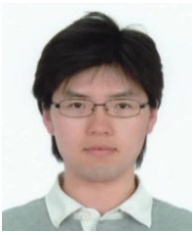
본 논문에서는 기존의 연구에서 제안된 공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법을 다 음원 위치 추정에 적용할 수 있도록 수정하였다. 두 음원이 서로 상관되어 있는 경우에는 음원의 위치를 추정하기 위해 계산된 공간좌표로 사상된 GCC 함수에서 실제 음원의 위치뿐만 아니라 두 음원에 의해 발생하는 가상 음원 위치에도 peak가 발생한다. 하지만 두 음원이 서로 상관되어있지 않은 경우에는 실제 음원의 위치에서만 peak가 발생됨을 알 수 있었다. 또한 공간좌표로 사상된 GCC 함수에서의 peak은 두 음원의 파워 비에 따라 그 크기가 달라짐을 알 수 있었다. 따라서 공간좌표로 사상된 GCC 함수를 이용하여 다 음원의 위치를 추정하기 위해서는 음원들 사이에 상관성이 없어야 하며, 서로 비슷한 크기를 갖는 음원이어야 한다. 다 음원 위치 추정을 위해 수정된 방법의 성능을 평가하기 위해서 실제환경에 남성과 여성의 음성신호를 이용하여 실험을 수행하였다. 그 결과 실제 음원의 위치와 추정된 음원의 위치 사이의 절대적 오차 평균으로 정의된 추정 오차 평균은 1.4°이며, 다 음원 추정률은 평균 62%의 정확성을 보여주었다. 추정 성능은 다 음원 위치 추정 방법에서 보정계수의 의존적인 성향을 보이며, 추정 오차 평균보다는 다 음원 추정률에 대해 더욱 민감함을 알 수 있었다.

참고문헌

[1] M. S. Brandstein and H. F. Silverman, "A practical methodology for speech source localization with microphone arrays,"

Computer Speech and Language, vol. 11, no. 2, pp. 91-126, 1997.

- [2] J.-H. Jeon and Y.-H. Kim, "Localization of moving periodic impulsive source in a noise environment," *Mechanical Systems and Signal Processing*, vol. 22, Issue 3, pp. 753-759, 2008.
- [3] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," *Acoustics, Speech, and Signal Processing, ICASSP-97*, vol. 1, pp. 187-190, 1997.
- [4] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano, "Active audition for humanoid," *Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000)*, pp. 832-839, 2000.
- [5] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Kitano, "Applying scattering theory to robot audition system: robust sound source localization and extraction," *Proceedings of the International Conference on Intelligent Robots and Systems (IROS2003)*, vol. 2, pp. 1147-1152, 2003.
- [6] S. Hwang, Y. Park, and Y. Park, "Sound source localization using HRTF database," *Proc. of International Conference on Control, Automation, and Systems*, 2005.
- [7] Y. Sasaki, S. Kagami, and H. Mizoguchi, "Multiple sound source mapping for a mobile robot by self-motion triangulation," *Proceeding of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 380-385, 2006.
- [8] S. Hwang, Y. Park, and Y. Park, "Sound direction estimation using artificial ear," *International Conference on Control, Automation and Systems*, pp. 405-409, 2007.
- [9] M. Kumon, T. Shimoda, R. Kohzawa, I. Mizumoto, and Z. Iwai, "Audio servo for robotic system with pinnae," *Proceedings of the International Conference on Intelligent Robots and Systems (IROS2005)*, pp. 1881-1886, 2005.
- [10] H. Li, T. Yosiara, Q. Zhao, T. Watanabe, and J. Huang, "A spatial sound localization system for mobile robots," *IEEE Instrumentation and Measurement Technology Conference*, pp. 1-6, 2007.
- [11] C. Liu, B. C. Wheeler, W. D. O'Brien, Jr. and R. C. Bilger, "Localization of multiple sound sources with two microphones," *Journal of Acoustical Society of America*, vol. 108, issue 4, pp. 1888-1905, 2000.
- [12] 권병호, 박영진, 박윤식, "공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법," 한국소음진동공학회 논문집 제19권 제4호, pp. 355-362, 2009.
- [13] 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. AS-24, no. 4, pp. 320-327, 1976.
- [14] B. Kwon, Y. Park, and Y. Park, 2009, "Multiple sound sources localization using the spatially mapped GCC functions," *Proc. of International Conference on Control, Automation and Systems*, pp. 1773-1766, 2009.
- [15] L. R. Rabiner and M. R. Sambur, "An algorithm for determining the endpoints of isolated utterances," *The Bell System Technical Journal*, vol. 54, no. 2, pp. 297-315, 1975.
- [16] 권병호, 박영진, 박윤식, "공간좌표로 사상된 GCC 함수를 이용한 다 음원 위치 추정," 한국자동제어학술회의 논문집(KACC 2009), 부산, 한국, 2009.



권병호

2006년 한양대학교 기계공학과 졸업.
2008년 한국과학기술원 석사. 2008년~
현재 동 대학원 박사 과정. 관심분야는
자동제어, 능동 소음 제어, 3차원 입체
음향, 신호처리.



박영진

1980년 서울대학교 기계공학과 졸업.
1982년 동 대학원 석사. 1987년 the
university of Michigan 박사. 1987년~1988
년 the university of Michigan research
fellow. 1988년~1990년 New Jersey
Institute of Technology 조교수. 1990년~현
재 한국과학기술원 기계공학과 교수. 관심분야는 자동제어, 3
차원 입체음향, 능동 소음 및 진동제어, 차량 사시 제어.



박윤식

1972년 서울대학교 기계공학과 졸업.
1979년 Iowa state university 석사. 1981년
동 대학원 박사. 1981년~1984년 한국과
학기술원 선임연구원. 1984년~현재 한국
과학기술원 기계공학과 교수. 관심분야
는 동역학, 진동 및 소음 제어, 구조물

동특성 변경법, 기계 및 매니퓰레이터의 동역학, 모달해석,
시스템 규명.