

논문 2009-46SP-2-3

음원 위치 추정 시스템의 정확도 향상 방법

(The Method of Elevation Accuracy In Sound Source Localization System)

김 용 은*, 정 진 균**

(Yong-Eun Kim and Jin-Gyun Chung)

요 약

음원 추정 시스템은 로봇, 화상회의, CCTV(Closed-circuit television) 시스템에 쓰인다. 이러한 음원 추정 시스템은 사람을 대상으로 하며 사람이 말하는 동안 여러 개의 음성 데이터 프레임을 입력받을 수 있다. 본 논문에서는 입력된 음성 데이터 프레임으로부터 정확한 각도를 계산 할 수 있는 음성 데이터 프레임을 선별하여 각도 추정 오차를 줄이는 방법에 대해서 제안한다. 또한 선별된 데이터를 각도로 변환한 후 메디언 필터를 적용하여 음원 추정 시스템의 오차를 줄일 수 있다. 제안된 시스템을 이용하여 실험한 후 각도 추정 오차 평균이 31%감소함을 보인다.

Abstract

Sound source localization system is used in a robot, a video conference and CCTV(Closed-circuit television) systems. In this Sound source localization systems are applied to human and they can receive a number of sound data frames during speaking. In this paper, we propose methods which is reducing angle estimation error by selecting sound data frame which can more precisely compute the angles from inputted sound data frame. After selected data converted to angle, the error of sound source localization recognition system can be reduced by applying to medium filter. By the experiment using proposed system, it is shown that the average error of angle estimation in sound source recognition system can be reduced up to 31 %.

Keywords : 음원 위치 추정, 오차 감소, 데이터 선별, 메디언 필터

I. 서 론

사람과 기계 사이의 원활한 상호작용을 위하여 음원 추정을 활용하는 연구가 오래전부터 진행되고 있고 응용 제품들이 제작되고 있다^[1~3]. 이러한 음원 추정은 화자의 위치를 파악하는 애완용로봇, 화상회의 도중 발표하는 사람의 얼굴을 향해 카메라를 이동시키는 화상회의 시스템, 보안카메라가 음원을 향해 카메라를 이동시

키는 보안시스템에 특히 널리 적용되고 있다^[4~5]. 그러나 이러한 제품들이 사용되는 환경은 대부분 잡음이 존재하고, 장애물로 인하여 소리의 지연이 발생하여 음원 추정 성능을 떨어뜨리게 된다.

사람과의 인터페이스를 위하여 음원을 추정하는 경우 화자가 음성을 기계에 전달할 때 보통 0.5초 이상의 음성을 발생한다. 이 때 이 시간 동안 많은 양의 음성 데이터가 입력된다. 이러한 음성 데이터를 일정 샘플 길이로 나누어 프레임으로 만들고 각각의 프레임들을 각도로 환산하므로 한번 음원이 입력될 때 여러 개의 각도가 출력된다. 만약 여러 개의 각도 출력 중 음원 추정이 잘 되었다고 판단된 각도 계산 값을 찾을 수 있다면 그 각도를 선택하여 보다 더 정확한 음원 추정 시스템을 구현할 수 있다.

* 학생회원, ** 정회원, 전북대학교 전자정보공학부 (Div. of Electronic & Information Engineering Chonbuk University)

※ 이 연구에 참여한 연구자는 2단계 BK21사업의 지원비를 받았음, This work was supported by the second stage of Brain Korea 21 Project.

접수일자: 2008년10월14일, 수정완료일: 2009년2월27일

본 논문에서는 3개의 마이크로폰을 이용한 시간지연 방식의 음원추정시스템에서 각 마이크로폰 쌍으로부터 계산된 상관계수 관계를 검사하여 음원 추정이 정확하게 이루어졌는지를 판단하고 이를 메디언 필터를 이용하여 확률적으로 정확한 음원 추정 값을 출력하는 방법에 대해서 제안한다.

II장에서는 음성 도달 지연 시간을 이용한 기존의 음원 추정 방법, III장에서는 제안된 음원 추정 오차 감소 방법에 대해 설명한다. IV장에서는 시뮬레이션 결과를 제시하고 V장에서는 결론을 맺는다.

II. 음성 지연 시간을 이용한 음원 추정 방법

음원의 위치를 추정하기 위해서는 도달시간지연을 이용한 방법이 널리 사용된다. 음원에서 나온 음파는 두 마이크로폰에서 측정된 신호 $x_1(t)$ 와 $x_2(t)$, 음원에서 첫 번째 마이크로폰에 도달한 신호 $s_1(t)$, 잡음 $n_i(t)$ 를 사용하여 식 (1)과 같이 나타낼 수 있다^[6].

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

식 (1)에서 α 는 감쇠계수이고, D 는 추정하고자 하는 두 신호사이의 시간지연이다^[6].

두 신호 $x_1(t), x_2(t)$ 의 상호 상관계수는 식 (2)와 같다.

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

그림 1에서와 같이 두 마이크로폰사이의 거리를 l_{mic} 음원 위치 각도를 ϕ 라 하면 지연거리 d 는 다음과 같이 표현된다^[6].

$$d = l_{mic} \cos \phi \quad (3)$$

음속을 ($c = 343m/s$)라고 하면 지연거리 d 에 해당하는 지연시간 τ 는 다음과 같다^[6].

$$\tau = \frac{d}{c} = \frac{l_{mic} \cos \phi}{c} \quad (4)$$

따라서 구하고자 하는 음원위치각도 ϕ 는 다음과 같이 표현된다^[6].

$$\phi = \cos^{-1} \frac{\tau c}{l_{mic}} \quad (5)$$

그림 1과 같이 마이크 2개만을 이용하여 음원의 위치를 추정하면 전방과 후방을 구분하지 못하는 앞뒤 혼동 현상이 발생한다. 따라서 그림 2와 같이 마이크로폰 3개 이상을 이용하여 음원 추정 시스템을 제작해야 한다. 3개의 마이크로폰이 존재하는 음원 추정 시스템에서는 각각 3개의 마이크로폰 쌍에서 각도 값들이 출력되게 되고 이 세 개의 각도 값은 일정한 관계가 있다. 예를 들어 그림 3에서 90도에서 소리가 입력되었을 때 각각의 마이크로폰 쌍에서 절대 축으로 계산된 각도는 LR에서 90도, 270도(앞뒤 혼동각도)를 BL에서 330도, 30도(앞뒤 혼동각도) BL에서 210도, 150도(앞뒤 혼동각도)를 출력하게 된다. 이렇게 출력된 각도를 상대 축으로 변환하기 위해서는 표 1과 같이 LR마이크로폰 쌍을 기준으로 검출된 각도에 LB는 120도를 더한 후 360도를 모듈레이션 취하고 BL에서는 240도를 더한 후 360도를 모듈레이션 하여야한다. 이렇게 상대 축으로 각도를 계산하고 LB, BL, LR에서 계산된 상대 축 각도들을 더하여 가장 큰 값을 찾으면 최종각도가 출력된다. 그림 3-(a), (b), (c)는 90도에서 음원이 발생했을 때 22개의 프레임의 평균으로 각 마이크로폰의 각도가 계산된 값이며 최종각도는 3-(d)처럼 3-(a), 3-(b), 3-(c)가 더해진 값이다.

데이터 1개의 샘플을 이용하여 각도 추정 연산을 수행하기 보다는 데이터 여러 개를 이용하여 각도 추정

표 1. 각 마이크로폰 쌍에서 출력된 각도 관계(θ : LR 쌍에서 계산된 각도, mod: 나머지 값 연산)
Table 1. The degree relation.(θ : computed angle for LR pair, mod : remainder computation).

마이크로폰쌍	각도
LR	θ
BL	$(\theta + 120\text{도}) \bmod 360$
BL	$(\theta + 240\text{도}) \bmod 360$

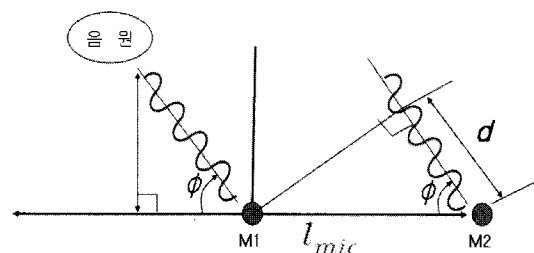


그림 1. 마이크로폰 2개를 이용한 음원 위치 추정
Fig. 1. Sound source localization system using two microphone.

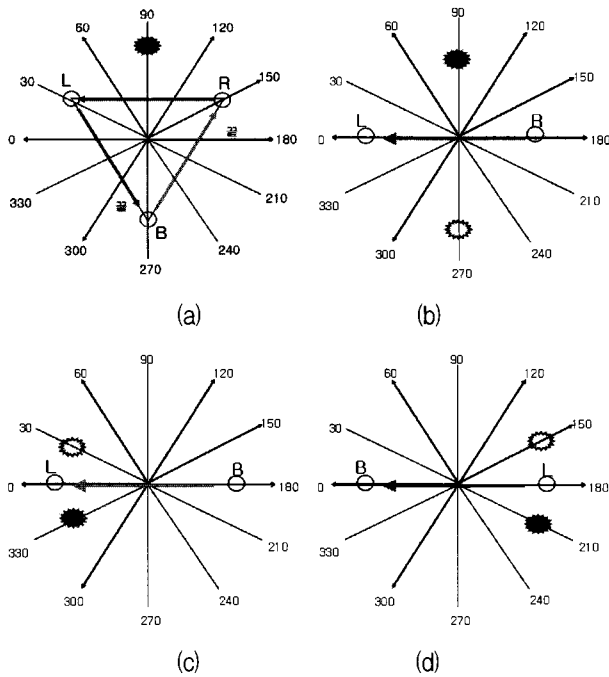


그림 2. 마이크론 3개를 이용한 음원 추정
Fig. 2. Sound source localization system using tree microphone.

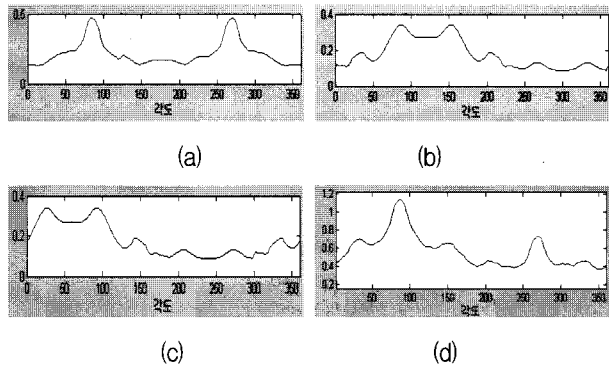


그림 3. 음원이 90° 일 때 각 마이크 쌍에서 구해진 각도: (a)LR, (b)BL (c)RB, (d)LR+BL+RB
Fig. 3. Angle estimation results obtained When sound source is 90°: (a)LR, (b)BL (c)RB, (d)LR+BL+RB.

연산을 한 후 이 값들의 평균을 음원 각도로 출력하면 오차를 줄일 수 있다. 이 시스템에서 프레임을 증가 시키면 오 추정 확률이 작아지지만 연산될 데이터의 증가 때문에 연산속도 및 파워소모에서 오버헤드가 커진다. 따라서 데이터와 각도 추정오차와의 trade-off가 필요하다.

III. 음원 추정 오차를 줄이기 위해 제안한 방법

1. 프레임의 파워 값 계산

사람이 음성으로 명령을 할 때는 대부분 음절이 존재한다. 기존의 시스템에서는 프레임 파워 초기 값의 경계 값을 계산하고 그 값이 경계 값 이상 크기로 판단되면 그 때부터 일정 프레임의 각도를 연산한다. 만약 '전주'라는 음성을 샘플링하면 그림 4에서와 같이 파형이 출력된다. 이 때 처음 프레임의 파워 값이 경계 값 이상의 파워로 측정되면 그 때부터 몇 개의 프레임 각도를 계산하여 각도를 계산한다. 예를 들어 그림 4에서 각도 연산 프레임이 10개라고 하면 1프레임(1F)에서 경계 값 이상의 파워가 출력 되고 그 다음 10프레임(10F)까지 각도 값이 계산된다. 하지만 사람의 음성은 일정하게 출력되는 것이 아니고 중간에 끊어지는 부분이 존재한다. 따라서 그림 4에서 6F~7F는 잡음 각도로 계산하게 된다. 이러한 문제를 해결하기 위해 각 프레임 단위의 파워를 계산하여 프레임 개별적인 파워가 경계 값 이상인지를 판단하고 경계 값 이상이면 그 프레임은 소리라고 판단하고 그 이하이면 무시하도록 한다. 따라서 그림 4의 6F~7F 음성 프레임의 각도 값을 버리므로 더욱더 정확한 음원 추정이 가능하다.

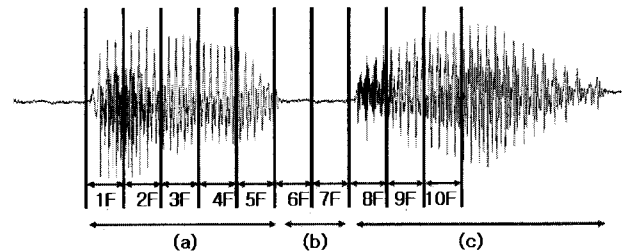


그림 4. 샘플링된 음성데이터 예제: (a)'전'발음 구간, (b) 다음발음을 위한 시간지연구간, (c)'주'발음 구간
Fig. 4. Example of sampled sound data: (a)'JEON' signal section, (b) time delay section for next signal, and (c) 'JU' signal section.

2. 정확한 음성 프레임 추출하기

주위 잡음과 소리의 반사 또는 주위 환경 방해 요소나 마이크론이 장착되는 모형으로 인해 지연거리 d 의 계산과정에서 오차가 발생할 수 있다^[8]. 사람은 보통 0.5초 이상 음성 명령을 내리고 16KHz로 음성을 샘플링하며 256샘플 데이터를 1프레임으로 가정하면 1초에 30번 이상 각도 연산이 가능하다. 30번 이상 각도 연산 값 중 주위 환경으로 인해 오차가 큰 값과 정확한 각도 값이 존재 한다. 이때 각 프레임의 각도 연산 값 중 정확한 각도 결과를 찾는다면 오차가 감소될 것이다.

이상적이 환경에서는 그림 3에서와 같이 각각 마이

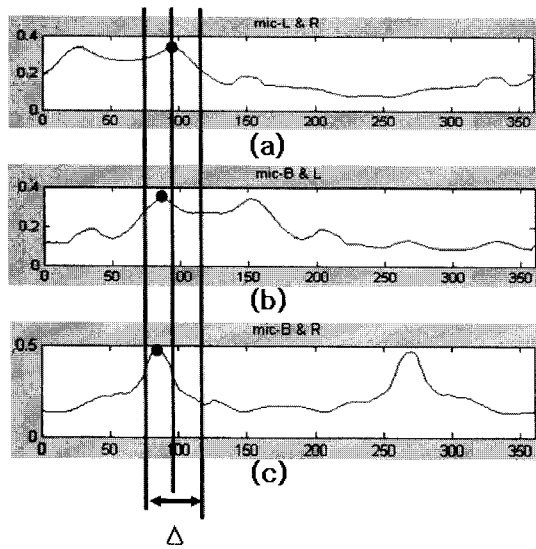


그림 5. 각 마이크론 쌍에서 연산된 정확한 각도 값 검사 방법: (a) LR, (b) BL (c) RB
 Fig. 5. Detection method of accurate angle value calculated from each microphone pair.

크로폰 쌍의 가장 큰 값 2 포인트 위치 중 1개의 포인트 위치가 동일해야 한다. 하지만 잡음 또는 방해물로 인하여 위치가 조금씩 틀어지게 된다. 따라서 경계 값 Δ 를 설정하고 최대값의 위치가 경계 값 Δ 이상 차이가 나면 주위환경에 의해 크게 손상된 음성 프레임으로 간주하여 그 프레임의 연산 결과 각도 값은 버리고 경계 값 Δ 이하 차이값을 가지는 프레임은 정확한 프레임으로 판단하여 그 데이터 프레임에서 계산된 각도 값을 사용하도록 한다. 그림 5는 이러한 과정을 예로 보여준다. 그림 5에서 LR, BL, RB의 최대 값의 위치가 Δ 범위 안에 존재하므로 그림 5의 음성 데이터 프레임을 이용한 각도 변환은 정확한 것으로 간주한다.

3. 메디언 필터 사용하기

Ⅲ-1, Ⅲ-2 과정을 거처도 혹시 잘못된 음성 프레임을 이용한 각도 연산이 이루어 질 수 있다. 보다 정확한 각도 연산을 위해서는 Ⅲ-1, Ⅲ-2 과정을 거처서 계산된 각도에 메디언 필터를 이용하면 된다. 기존의 방법은 프레임들에서 연산된 각도의 평균을 최종각도로 출력한다. 이 때 1개라도 프레임 각도가 엉뚱하게 출력되는 경우 프레임 각도의 평균 값 전체에 영향을 미칠 수 있다. 이러한 문제를 해결하기 위하여 메디언 필터^[7]를 이용할 수 있다. 예를 들어 30도에 음원이 존재하고 30개의 프레임이 입력되고 Ⅲ-1, Ⅲ-2 과정에서 필터링된 프레임 각도가 {30, 30, 120, 31, 32, 29, 30}(i는 각도 개

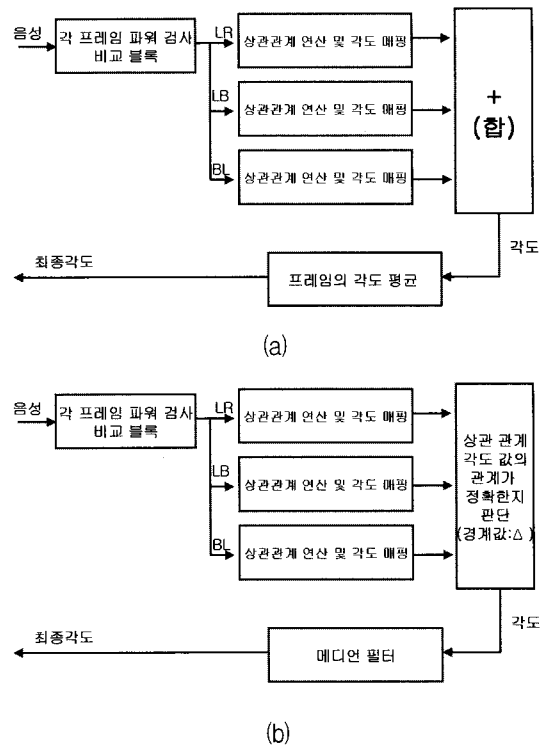


그림 6. 기존의 음원인식 시스템 블록도와 제안한 음원인식 시스템 블록도: (a)기존의 음원인식 시스템 블록도, (b)제안한 음원인식 시스템 블록도
 Fig. 6. Block diagram of sound source localization: (a) conventional method and (b) proposed method.

수) 일 때 오추정 각도가 120도임을 알 수 있다. 기존의 방법을 이용하여 평균 각도를 구하면 45도로 오차가 크다. 하지만 메디언 필터를 이용하면 주어진 각도를 소트 시키고 정수로 픽스된 가운데 값($i/2$)를 선택하므로 30도가 출력 된다. 따라서 메디언 필터를 사용하면 음원 추정 시스템에서 음성 데이터 프레임이 몇 개가 틀릴지라도 안정적으로 각도를 연산할 수 있게 된다. 이와 같이 Ⅲ-1, Ⅲ-2, Ⅲ-3 세 가지 방법을 적용하여 보다 정확한 음성 샘플 데이터를 추출하고 안정성을 고려하여 음원을 추정하여야 올바르게 각도 추정할 확률이 크다. 그림 6은 기존의 블록도와 제안한 블록도를 보여준다.

그림 7은 360도에서 음성이 입력되었을 때를 예제로 들어 제안한 방법을 적용하는 과정을 보여준다. 기존의 시스템에서 각도의 평균값을 구하였을 때 데이터 프레임 30개 중 9개 프레임의 오차로 인하여 256도로 오추정하지만 제안한 방법은 데이터 프레임을 선별하기 때문에 360도로 정확한 각도가 연산됨을 알 수 있다.

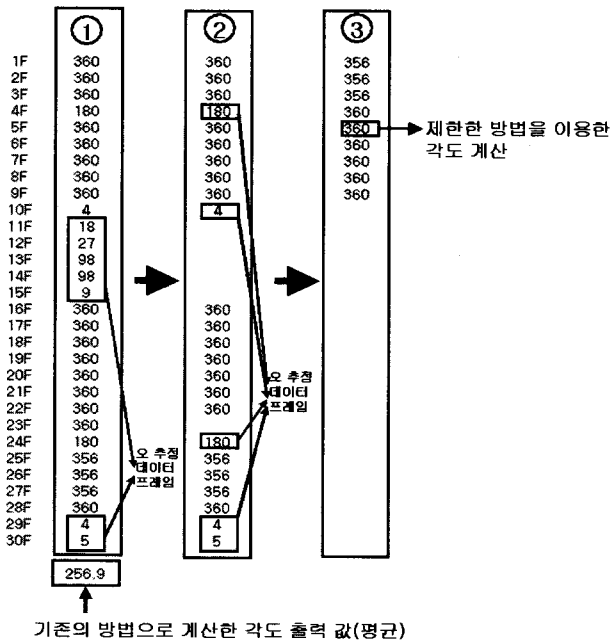


그림 7. 360도에서 음성이 들어왔을 때 제안한 방법이 적용되어 계산되는 과정
 Fig. 7. Process of proposed angle computation when inputting sound at 360°.

IV. 실험결과

그림 8은 음원추정을 위한 테스트 환경이다. 정삼각형 모양으로 배치된 3개의 마이크에 소리를 내면 이 소리가 증폭기를 통해 증폭된 후, DAQ(ADC)로 보내진다. DAQ에서는 음성신호를 16 KHz로 샘플링 한 후 PC로 보낸다. PC에서는 음성 신호를 텍스트 형태로 받아 한 개의 데이터 프레임을 256 샘플수로 구성해서 매트랩으로 작성된 음원 추정 프로그램에서 각도를 계산하도록 하였다. 360도에서 1m~5m까지 거리를 두어서 측정한 결과 표 2와 같은 결과를 얻을 수 있었다. 비록 실험 횟수가 한정적으로 이루어지긴 했으나, 명확한 사실은 제안한 방법을 이용하면 오추정 가능한 데이터 프레임을 제거시켜 더 정확하게 음원 추정을 수행할 수 있다는 것이다. 논문에서 주어진 예제를 제안한 알고리

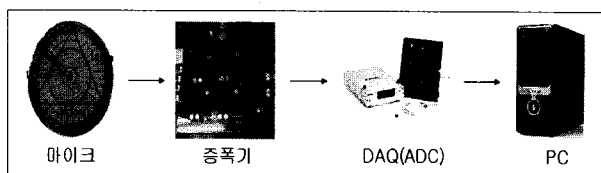


그림 8. 실험 환경
 Fig. 8. Test environment.

표 2. 실험결과

Table 2. The Simulation result.

		1m	2m	3m	4m	5m
기존 방법	그림 7의 ①상태의 정확도 (1 - 오추정각도/정확한각도)	0.72	0.65	0.68	0.68	0.72
	최종 평균 출력각도	275	256	277	269	269
제안한 방법	그림 7의 ③상태의 정확도 (1 - 오추정각도/정확한각도)	1	1	1	1	1
	최종 메디언필터를 거친 각도	360	360	360	360	360

즘을 이용하여 실험 하였을 때 31% 오차 감소 효과가 있다.

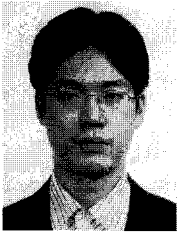
V. 결론

음원 추정 시스템에서 음성 데이터 프레임 중 음원을 정확하게 추정할 수 있는 음성 데이터 프레임을 추출하여 각도로 변환하고 각도로 변환된 결과에 메디언 필터를 적용하면 기존의 음원 추정 시스템보다 더 정확한 음원 위치를 추정할 수 있음을 보였다. 하지만 이러한 음원 추정 시스템은 고속으로 동작 되어야 하므로 IP개발이 필수적이다. 따라서 추후 본 시스템을 IP로 개발할 것이다.

참고 문헌

- [1] K. Nakadai, et al, "Active audition for humanoid," *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence*, pp. 832-839, 2000.
- [2] 황성목, "머리전달함수를 이용한 로봇의 화자 위치 추정", *한국소음진동공학회 추계학술대회논문집*, pp. 637-640, 2005.
- [3] 권병호, 김경호, 박영진, "로봇 플랫폼에서 마이크로 폰 위치를 고려한 음원의 방향 감지 방법" *한국소음진동공학회 2007년 추계학술대회 논문집*, pp. 1080-1084, 2007.
- [4] <http://211.189.19.10/exbi2007/view/1471/2>
- [5] http://www.ytn.co.kr/_ln/0105_200808231338267820
- [6] C. H. Knnapp and G. C. Cater, "The generalized correlation method for estimation of time delay," *IEEE Trnas. on acoustics, speech and signal processing*, Vol. Assp-24, No. 4, 1976.
- [7] S. J Ko. Y.H Lee "Center Weighted Median Filters and Their Application to Image Enhancement" *IEEE Trans. Circuits and System*, VOL. 38, No. 9, September 1991.

저 자 소 개



김 용 은(학생회원)
2005년 전북대학교 전자정보
공학부 학사 졸업
2007년 전북대학교 정보통신
공학과 석사 졸업
2007년~현재 전북대학교
전자정보공학부 박사과정

<주관심분야 : 통신, 신호처리, 반도체>



정 진 균(정회원)
1985년 전북대학교 전자공학
학사 졸업
1989년 미국 미네소타 주립대학
전기공학 석사 졸업
1991년 미국 미네소타 주립대학
전기공학 박사 졸업

<주관심분야 : 통신, 컴퓨터, 신호처리, 반도체>