휴모노이드 로봇을 위한 시청각 정보 기반 음원 정위 시스템 구현

Implementation of Sound Source Localization Based on Audio-visual Information for Humanoid Robots

박 정 옥* · 나 승 유* · 김 진 영*
Jeong Ok Park · seung You Na · Jin Young Kim

ABSTRACT

This paper presents an implementation of real-time speaker localization using audio-visual information. Four channels of microphone signals are processed to detect vertical as well as horizontal speaker positions. At first short-time average magnitude difference function(AMDF) signals are used to determine whether the microphone signals are human voices or not. And then the orientation and distance information of the sound sources can be obtained through interaural time difference. Finally visual information by a camera helps get finer tuning of the angles to speaker. Experimental results of the real-time localization system show that the performance improves to 99.6% compared to the rate of 88.8% when only the audio information is used.

Keywords : 음상정위, 시청각 정보, Sound Source Localization

I. 서 론

최근 인간의 삶의 질과 관련된 시스템의 연구가 다양하게 진행되고 있다. 그 중에서 단연 활발한 분야는 인간 친화형 로봇의 개발이다. 대표적인 예로 일본의 AIBO, 아이로보 등을 들 수 있다. 이들 로봇은 화자의 위치를 추정하여 마주보고 음성을 인식하여 인간과 더 친화적인 느낌을 유발시킨다. 여기서 사용하는 방법이 음상정위인데, 음상정위라 함은 간단히 말해 음파를 방사하는 음원의 방향이 어디인지를 판별하는 것을 말한다[1]. 그런데 인간은 두 귀를 가지고 음이 발생한 방향(상/하, 좌/우)과 거리를 알 수 있다. 동일하게 시스템으로 구현할 경우 인간의 두 귀를 대신한 두 개의 마이크를 사용하면 상/하, 혹은 좌/우 추적 중 하나만을 사용할 수밖에 없다. 2 차원이 아닌 3 차원적으로 음원을 추적하기 위해서는 최소 3 채널 이상의 마이크 입력을 이용하여야 한다. 한편, 실내에서 발생하는 잡음에 강인하게 반응하도록 구현하기 위해서는 시스템이 사람의 음성에만 반응하도록 하여야 한다. 그렇게 하기 위한 해결 수단으로 입력된 신호의 피치를 구한 후 일정 영역에 들어오는지 확인하여 음성유무를 판단할 수 있는 방법이 있다. 음원의 정위를 구하기 위한 주된 방법은 두 마이크 각각에 도달하는 음으로 시간차(Interaural Time Difference: ITD)와 레벨차

* 전남대학교 공과대학 전자정보통신학과

(Interaural Level Difference : ILD)를 구해 방향과 거리를 알아낸다[2].

음성 신호만을 이용하여 음원의 정위를 추정할 경우 음의 고유 성질인 회절, 반사, 굴절 등에 의하여 오차가 발생할 수 있다. 이런 오차를 최소화하기 위한 보정수단으로 카메라를 이용하여 얼굴을 검출 화자의 정확한 위치를 추적하여 오차를 보정할 수 있다. 이러한 얼굴 검출 기술은 시점 변화, 자세 및 조명 변화 등에 존재하는 다양성을 수용하는 알고리즘을 사용하며, 이는 시스템의 처리 속도에 큰 영향을 미친다. 따라서 검출 알고리즘의 처리 속도를 향상시키는 것은 실질적인 응용 시스템을 개발할 때 고려해야 할 중요한 문제이다[3].

본 논문에서는 좌/우, 상/하 4 채널, 피치를 검출하여 음성의 유무를 판별하고 음성일 경우 음상 정위를 판별하는 시스템을 구현하였다. 그리고 추적 성능 향상을 위하여 카메라를 통해 입력된 영상 정보를 바탕으로 화자의 위치를 정확히 추정할 수 있도록 모터를 제어 보정하였으며, 이를 실시간으로 구현하였다.

II. 음원 정위 시스템

본 논문에서는 모든 소리에 반응하는 민감한 시스템이 아닌 사람의 음성에만 반응하도록 음성 판단 여부를 알 수 있는 모듈을 사용하였으며, 4 채널(상/하, 좌/우) 마이크로 입력되는 신호의 음성 여부를 판단하여 음원 추적을 구현하였다[4]. 그 다음, 음원 추적의 보정을 위하여 카메라로 입력되는 영상 신호를 이용 얼굴 영역을 검출하여 보다 정확한 음원 정위를 추정하였다.

1. 음성의 주기성을 이용한 음성여부 결정

본 논문에서 구현한 피치검출 프로그램은 음성 데이터가 들어오면 각 프레임 별로 에너지를 추출하는데 그 에너지가 최대가 되는 프레임(즉, 유성음)을 찾아 그 정보를 가지고 short-time average magnitude difference function(AMDF) 신호를 이용하여 아래 그림 1과 같이 null points(AMDF 신호에서 꼭지점들)를 찾아서 범위 결정을 하였으며, 다음 식 1은 $AMDF(\gamma_n(k))$의 정의 식이다.

$$\gamma_n(k) = \sum_{m=-\infty}^{\infty} |x(n+m)w_1(m) - x(n+m-k)w_2(m-k)| \quad (1)$$

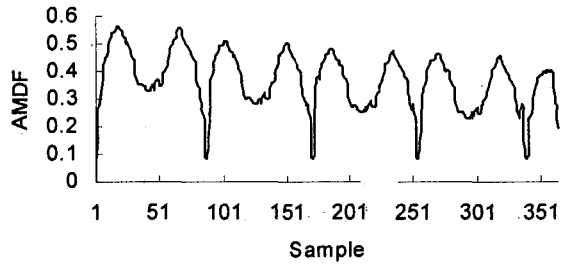


그림 1. AMDF 처리 결과의 예시

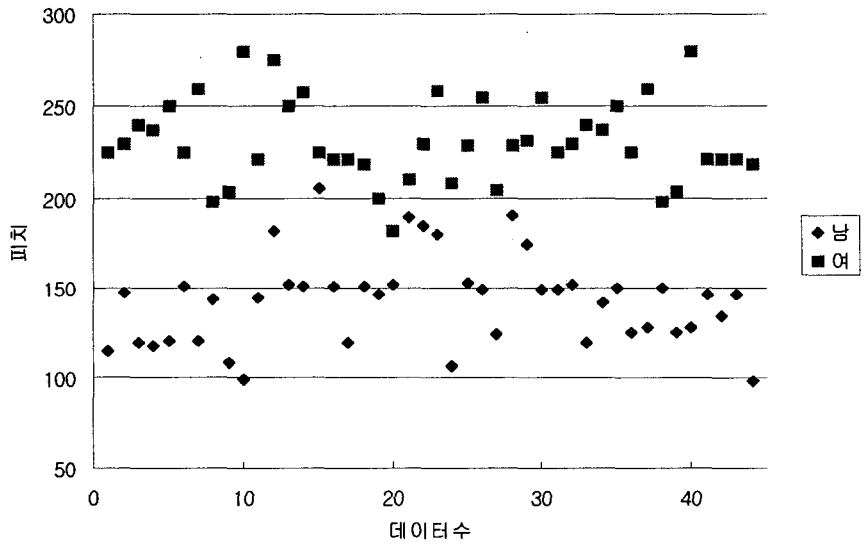


그림 2. 피치 검출 결과

여기서 $x(n)$은 입력신호이고 $w_1(m)$, $w_2(m)$은 윈도우들이고 k는 delay이다. 그림 1은 AMDF 처리 결과의 예시이다. 그리고 음성여부를 판단한 피치의 범위(70 Hz~350 Hz)는 학습데이터의 실험으로 결정되었으며 실험결과는 그림 2와 같다. 따라서 본 실험에서는 그림 3과 같은 비주기 잡음, 고주파, 저주파 잡음 신호를 제거된 사람 음성 데이터만을 가지고 음원 정위를 구하였다.

2. 음원 정위에 사용된 알고리즘

앞에서 제시한 음성의 유무를 판별한 다음, 음성일 경우에 음원이 발생한 지점을 추정하기 위해서 음원을 특정 각도에 정위시켜야 한다. 이 과정을 수행하기 위한 가장 중요한 단서는 수평면에 위치한 두 귀에서 파면의 상대적인 차이, 즉 두 귀에 입사하는 신호의 차이를 이용한다[5]. 그림 4는 음원이 발생한 각을 추정하기 위한 것이다. 여기서 입사음은 평면파라고 가정을 한다. 하지만 음

성은 구면파이기 때문에 오차가 발생하게 된다.

또한 마이크로 입력받은 데이터에는 마이크의 전기적인 특성으로 인한 60 Hz의 잡음과

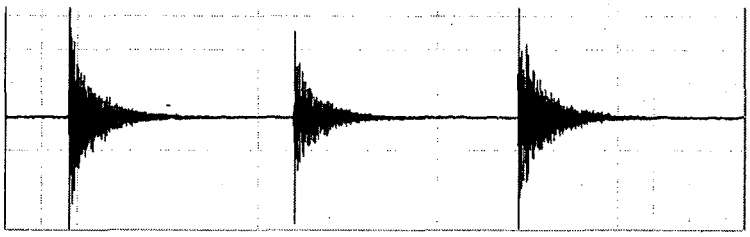


그림 3. 비음성으로 분류되는 박수소리의 파형

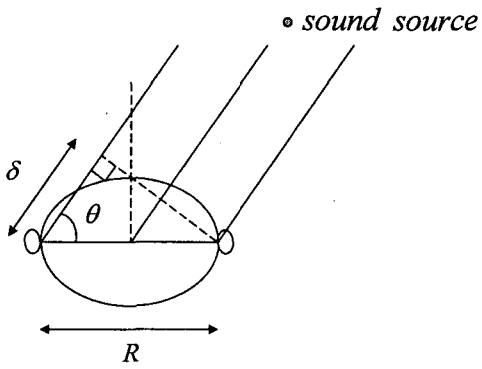


그림 4. 음원 정위의 원리

외부의 잡음도 함께 들어와 각을 추정하기에 상당한 애로사항이 발생한다. 이 점을 보완하기 위하여 2.2 kHz 저역 통과 필터링을 하였다. 저역 통과 필터링을 하게 된 이유는 위와 같은 잡음 원인 외에 ITD를 이용하여 음원 정위를 구하기 위한 최적의 주파수 대역이 1.5 kHz 이하의 저주파 대역이기 때문에 2.2 kHz의 저역 통과 필터를 사용하였다.

입력신호의 총 데이터 사이즈와 시간을 이용하면, 음파가 전달되는데 걸리는 시간을 구할 수 있다. 음속(C)은 340 m/s이므로 식 (3)에 대입하면 음원의 발생 방향(θ)을 알아낼 수 있다[6].

$$\cos\theta \approx \frac{\delta}{R} \tag{3}$$

한 음원에서 발생하는 신호를 분리된 두 마이크로 비상관 잡음과 함께 수신하는 과정을 다음과 같이 수학적 모델로 표현할 수 있다.

$$r_1(t) = \alpha_1 s(t-\tau_1) + n_1(t) \tag{4}$$

$$r_2(t) = \alpha_2 s(t-\tau_2) + n_2(t) \tag{5}$$

여기서 신호 $s(t)$는 잡음 $n_1(t)$와 $n_2(t)$에 비상관관계인 음원신호라고 가정하고, 두 마이크에서 받은 신호의 상호상관함수는

$$R_{r_1r_2}(\tau) = E[r_1(t)r_2(t-\tau)] = \int_{-\infty}^{\infty} r_1(t)r_2(t-\tau)dt \tag{6}$$

이고 여기서 $E[\ ]$는 기대값 (시간평균)을 나타낸다. 마이크로폰 신호 r_1에 대한 마이크 신호 r_2의 상대적 지연시간은 $\delta_{12} = \tau_2 - \tau_1$이므로, $D \equiv \delta_{12}$라 하면 D는 식(6)을 최대로 하는 변수 τ임을 알 수 있다. 그러나 실제로는 무한구간에서 측정할 수 없으며 유한관찰

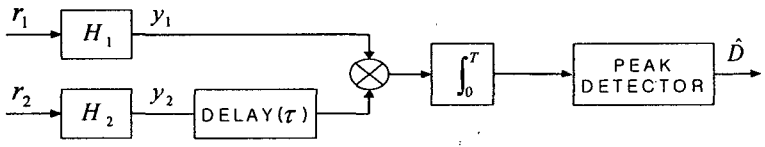


그림 5. 일반적 상관함수 방법을 이용한 피크 검출

시간에 대해서 식(6)을 근사화 해야 한다. 즉,

$$\hat{R}_{r_1r_2}(\tau) = \frac{1}{T-\tau}\int_{|\tau|}^{T} r_1(t)r_2(t-\tau)dt, \qquad -T<\tau\le 0 \tag{7}$$

$$\hat{R}_{r_1r_2}(\tau) = \frac{1}{T-\tau}\int_{0}^{T-\tau} r_1(t)r_2(t-\tau)dt, \qquad 0<\tau\le T \tag{8}$$

여기서 T는 관찰시간이고 τ의 부호에 따라 식(7) 또는 식(8)를 사용하여 $\hat{R}_{r_1r_2}$의 피크값의 위치를 검출함으로써 근사화된 시간지연 $\hat{D}$를 구할 수 있다. 측정된 시간지연 $\hat{D}$의 정확성을 향상시키기 위해서 식(7) 또는 식(8)의 적분을 수행하기 전에 $r_1(t)$와 $r_2(t)$는 그림 5와 같이 필터 H_1, H_2를 각각 통과시키면, 신호 r_i는 필터 H_i를 통과하여 출력 y_i ($i=1, 2$)이 된다. 그리고 한 신호(y_2)를 τ 만큼 지연시켜 각각의 y_i는 서로 곱해진 다음 적분되며 ($\hat{R}_{r_1r_2}$), 이 과정을 τ를 변화시켜가면서 피크 값을 얻을 때까지 반복한다. 검출된 피크값의 τ축에서의 위치가 예상 시간지연 $\hat{D}$가 된다. 이 방법을 "일반화된 상호상관법"이라고 한다. 그림 5는 일반화된 상호상관법을 이용한 피크 검출법이며 본 논문에서 실시간으로 음원을 추적하기 위하여 사용된 알고리즘이다.

3. 음원 정위 보정을 위한 얼굴 영역 검출 및 추적 알고리즘

음원 추적 과정에서 마이크에 입력된 음의 굴절, 반사, 회절로 인한 계산 오차를 보완하기 위하여 음원 추적 후 얼굴 영역 검출 알고리즘을 사용하여 추적 성능을 높일 수 있다. 얼굴 영역 검출 알고리즘은 학습된 계층적 분류기를 통해, 빠르고 효율적으로 얼굴 영역을 찾아낸다. 검출 알고리즘은 전처리, 실시간 얼굴 영역 검출의 두 단계로 구성된다[7].

전처리에서는 고유한 얼굴의 특징을 추출할 수 있는 사각형 특징 마스크를 이용한 분류기를 사용하였다. 이것은 픽셀 기반 분류보다 계산적인 면에서 효율적인 처리가 가능하고, 특정 영역에 대한 에지나 라인 등의 구조적인 정보를 제공한다. 사각형 특징 마스크는 그림 6과 같이 5 개의 사각형 형태로 구성되며, Haar wavelet 함수를 다양하게 변형하여 구성된다[8]. 제안된 사각형 특징 마스크를 이용하여, 입력 패턴에 대한 방대한 특징 집합을 생성하고, 각 특징은 20×20 크기의 입력 패턴 내에서 다양한 위치와 크기로 존재하게 된다. 따라서 제안된 사각형 특징마스크는 얼굴 영상에서 발생하는 다양성을 수용할 수 있으며, 얼굴패턴의 중요한 특징 정보를 추출할 수 있다.

각 사각형 특징 마스크는 전체 사각형 영역을 포함하는 흰색 사각형 영역과, 이 위에 오버랩된 형태로 존재하는 검정색 사각형 영역으로 구성된다. 각 레벨의 영상은 5×5 가우

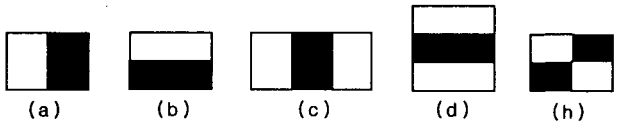


그림 6. 사각형 특징 (rectangular feature) 마스크

시안 필터를 적용하여 잡음을 제거한 후, 20×20 크기로 일반화하여 입력 패턴을 생성한다. 입력 패턴은 대비 확장(contrast stretching) 알고리즘을 통해 색상을 보정한 후, 이에 대한 SAT (summed-area table)를 생성하여 특징값을 효율적으로 계산한다[9].

전처리 과정에서 생성된 입력 패턴을 분류하는 계층적 분류기는 앙상블 분류기 $H(x)$가 계층적으로 구성된 형태이며, 앙상블 분류기 $H(x)$는 약한 분류기 h_t의 선형적인 결합으로 구성된다. 즉, 앙상블 분류기는 사각형 마스크의 결합으로 구성되며, 얼굴 패턴의 특징을 추출하는데 결정적인 역할을 한다. 앙상블 분류기는 식 (9)와 같이 정의된다[10].

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{9}$$

추적 알고리즘은 서보모터를 통해 동적으로 검출 영역을 확장시키며, 실시간으로 추출된 얼굴 영역의 위치 정보만을 이용함으로써 시스템의 처리 속도를 증가시킨다. 구현된 추적 알고리즘은 분류기 학습, 실시간 얼굴 검출, 얼굴 추적의 세 단계로 구성된다.

여기서 얼굴 검출 단계는 생성된 계층적 분류기를 이용하여, 실시간으로 얼굴 영역을 찾아낸다. 계층적 분류기는 특징들을 이용하여, 입력 패턴에 대한 특징 값을 계산한다.

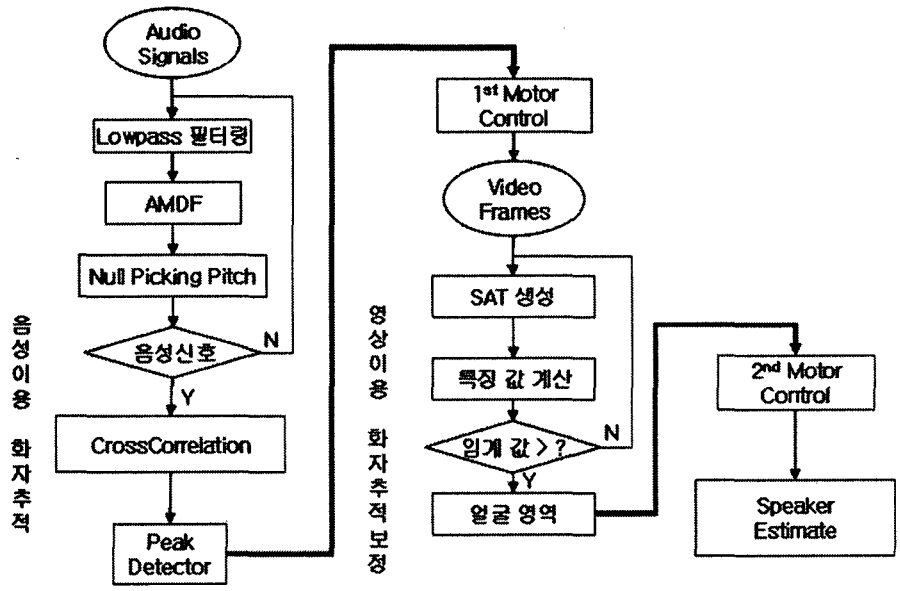


그림 7. 구현 시스템 블록 다이어그램

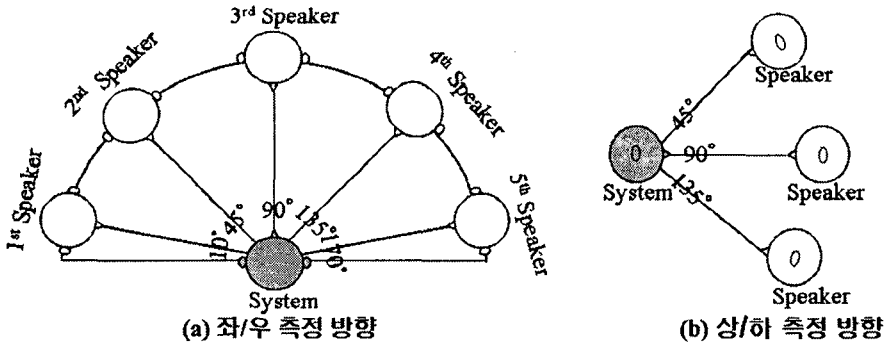


그림 8. 실험 측정한 방위

따라서, 입력 패턴을 계산된 특징 값이 임계값을 만족하면 얼굴 영역으로 분류한다. 마지막 단계인 얼굴 추적 단계에서는 얼굴 영역의 위치 정보를 이용하여 서보 모터를 제어한다. 최종적으로 구현된 음원 정위 시스템의 전체적인 블록 다이어그램은 그림 7과 같다.

Ⅲ. 실험 방법

실험은 20 대의 남성화자와 여성화자 10 명으로 화자의 위치는 그림 8과 같이 좌/우 10, 45, 90, 135, 170 도에 자리하게 하였으며 상/하 45, 90, 135 도의 위치에서 발음하게 하였다. 그리고 동일한

실험 환경을 위하여 마이크와 화자의 간격을 1 m로 고정하여 위치하였으며, 화자는 일정한 순서에 의해 그림 9와 같은 "무궁화 꽃이 피었습니다."의 문장을 1 회 발음하게 하였다.

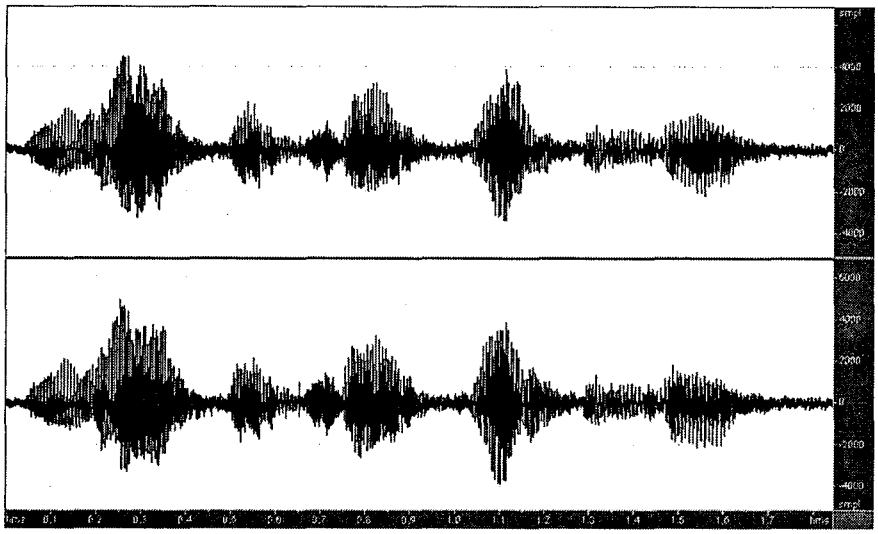


그림 9. 실험에서 사용된 2채널 음성 파형 ("무궁화 꽃이 피었습니다")

표 1. 구현에 사용된 실험 환경

	음 성 신 호	영 상 신 호
샘플링 rate	16 bit, 44.1 kHz	초당 20 프레임
입 력 채 널	4 채널 (상/하/좌/우)	1 대의 카메라 (시야각 : 45°)
분석데이터크기	입력된 신호의 음성구간	172×144 (width×height)
실 험 대 상	20 대의 남녀화자 10 명	20 대의 남녀 10 명

사용한 시스템은 표 1과 같이 4채널 오디오 입/출력 카드와 일반적으로 사용하는 카메라로 구현하였으며, 이 시스템의 구성은 그림 10, 11과 같다.

음성 신호 입력은 4채널에서 동시에 16 bit, 44.1 kHz로 샘플링하며, 입력된 신호의 피치를 구하여 음성 여부를 판단한 후, 비음성일 경우에는 동작을 하지 않았으며 사람의 음성일 경우의 구간만을 사용하여 음원의 정위를 추정하였다.

영상신호 입력은 쉽게 접할 수 있는 시야각이 45 도 이내인 USB 웹 카메라를 사용하였으며, 가로/세로 172×144의 이미지를 초당 20 프레임으로 입력 받은 영상을 이용하여 얼굴영역을 검출하였다. 실제적인 모터를 제어하기 위하여 사용된 MCU(micro control unit)는 ATMEL사의 ATMEGA 103을 이용하여 RS-232C를 통하여 PC에서 전송받은 위치 신호를 입력받아 2 채널의 PWM(pulse width modulation) 신호를 출력하여 상/하, 좌/우 서보모터를 제어하였다. 실험환경은 잡음(실험장비의 기계적 노이즈)이 존재하는 사무실 환경에서 실험을 실시하였다.

실험은 크게 2 단계로 음성신호만을 이용한 상/하, 좌/우 화자 추적과 음성신호와 영상신호를

이용하여 음원정위를 보정한 화자 추적으로 구분하여 실험하였다. 본 실험에서 오차범위는 ±5° 이내로 하였으며 범위를 벗어났으나 화면에 화자의 얼굴이 표시되는 경우는 위치에 따라 ±5의 페널티를 부여하였다. 표 2는 실험 출력 영상의 예시이다.

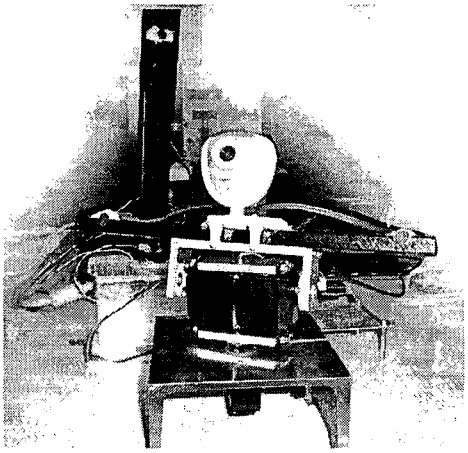


그림 10. 실험용 구현 시스템

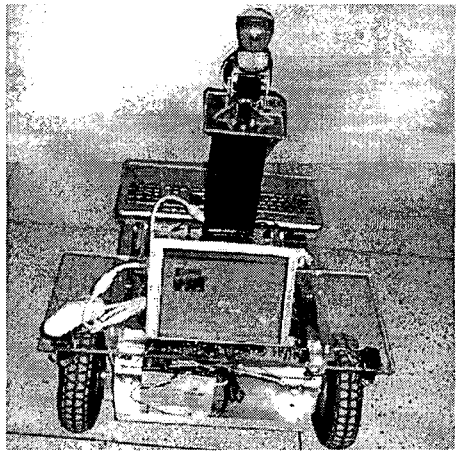
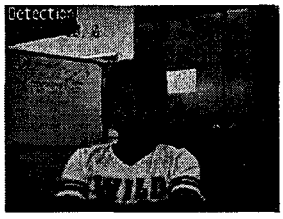


그림 11. 로봇베이스 구현 시스템

표 2. 음성/영상신호를 이용한 실험 방법

출력영상	영상에 관한 설명
	화자 1 차 발음 후 시스템이 정확한 추정을 하지 못하여 오차가 발생한 경우
	화자 1 차 발음 후 시스템이 정확히 추정 화자가 화면 정중앙에 위치한 경우
	1 차 화자 발음 후 각을 추정하여 위치를 찾은 후 영상을 이용 보정한 경우

IV. 실험 결과

본 실험을 3 단계로 구성하였다. 1 차 실험은 마이크에 입력된 소리가 음성인지의 유무를 판단하는 전처리 과정이며, 2 차 실험은 전처리를 통하여 입력된 사람의 음성을 토대로 각을 추정하는 실험이며, 계산되어진 음원 추정 각을 보정하기 위해 추정된 결과에 영상신호를 이용하여 얼굴을 검출하여 화자의 위치를 정확히 찾는 실험을 하였다. 그리고 부가적으로 동일한 조건에서 사람을 대상으로 위치추정을 함으로써 구현된 시스템과 비교하였다.

1. 음성유무의 판단 실험 결과

실험 환경과 동일한 위치에서 음성 외에 박수소리와 같은 고주파 신호를 50 회씩 10 회 발생시킴으로서 시스템의 동작여부를 관찰하였다. 그 결과 박수소리와 같은 고주파 신호에는 100% 반응하지 않음을 확인할 수 있었다.

2. 상/하, 좌/우 음성 신호 및 영상 신호를 이용한 추적 실험 결과

그림 12는 총 12 명의 화자에서 1 명씩 0, 30, 90, 150, 180 도의 위치에서 시스템을 향해 "무궁

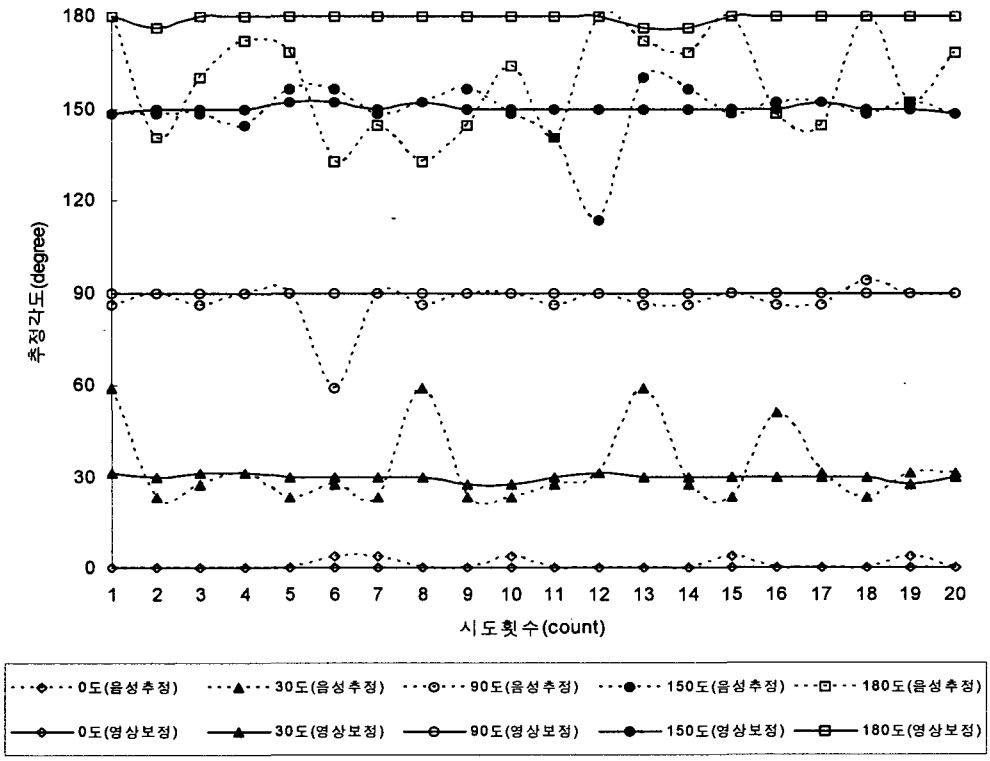


그림 12. 화자 추적 실험 결과

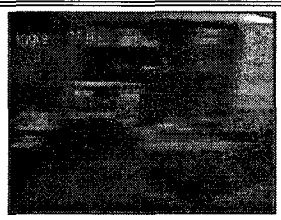
화꽃이 피었습니다."를 각 5회씩 발음하게 한 후 시스템이 화자의 위치를 추정한 계산 결과를 나타낸 그림이다.

그림에서 점선으로 표시된 부분은 음성신호만을 이용하여 화자의 위치를 계산한 결과이며 30도와 150 도의 위치에서 30 도에 가까운 오차가 발생함을 볼 수 있다. 그리고 실선으로 표시된 부분은 음성신호에서 발생한 오차를 보정하기 위하여 영상신호를 사용한 것을 보여주고 있는데, 앞서 실험한 음성만을 이용한 결과의 오차가 현격히 줄어들었음을 확인할 수 있다.

이러한 음성에서 발생한 오차를 영상을 이용하여 보정하는 과정을 표 3에 나타내었다.

표에서 보는 바와 같이 2 프레임에서는 180°의 위치에서 발음된 화자의 음성을 172°로 계산하여 -8°의 오차가 발생하였다. -8°의 오차를 보정하기 위하여 카메라에 입력된 영상을 이용하여 화자의 얼굴 영역을 검출한 후, 3 프레임에서 보이는 바와 같이 검출된 얼굴 영역이 화면 중앙에 위치할 수 있도록 모터를 +8° 만큼 움직여 화자의 얼굴이 카메라의 정 중앙에 위치할 수 있도록 하였다. 4 프레임은 마이크와 카메라가 화자의 위치에 수렴한 것을 보여주고 있다.

표 4. 음성/영상신호를 이용 화자 추적 과정

단계(Frame)	출력영상	시스템동작
1 프레임		초기단계 추적각도 : 90.0°
2 프레임		음성신호를 받음 172°로 -8° 오차 발생 추적각도 : 172.2°
3 프레임		화자의 얼굴영역을 찾은 후 +8°만큼의 이동을 통하여 오차 보정 추적각도 : 180.0°
4 프레임		오차 보정한 결과

3. 휴먼 추적 실험 결과

구현된 시스템의 성능을 사람과 비교하기 위하여 7 명을 대상으로 0, 10, 30, 50, 70, 90, 110, 130, 150, 170, 180 도의 위치에서 녹음된 음성을 실험에 앞서 순차적으로 듣게 하였다. 그런 다음 임의의 순서대로 녹음된 음성을 들려준 후 음성이 녹음된 방향을 가리키도록 하였다. 그 결과 청각이 예민한 사람의 경우 음성이 녹음된 정확한 방향을 가리켰지만 그렇지 못한 사람의 경우는 상당히 틀린 각을 가리켰다.

그림 13은 지금까지 실험에서 발생한 오차를 나타낸 것이다. 휴먼 추적은 사람의 개별 특성으로 인한 편차가 심하여 오차가 최고 28 도까지 발생하였음을 알 수 있다. 그리고 좌/우 추적에 비해 상/하 추적의 오차가 큰 것을 볼 수 있는데 그 원인은 좌/우와 동일하게 상/하추적 실험에서도 0~180 도의 위치를 추적하게 하여 상/하 마이크가 놓여 있는 탁자에서 음이 반사되어 생긴 오차일 것이라 추정된다. 그리고 마지막으로 영상을 이용하여 보정된 결과를 확인할 수 있다.

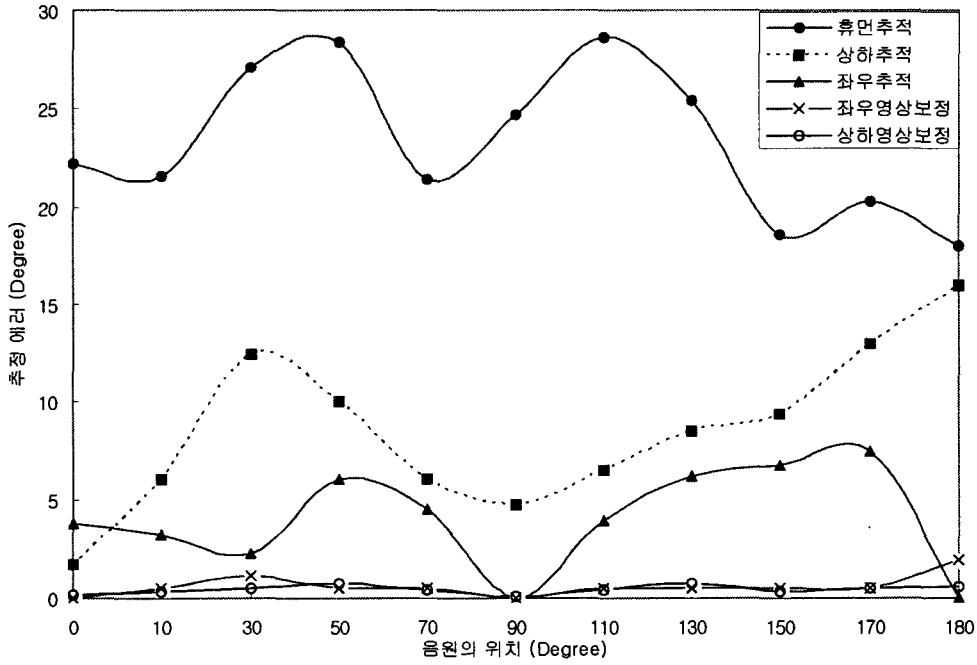


그림 13. 화자 추적에서 발생하는 에러

V. 결 론

본 논문에서는 기 구현 시스템[11]에서 새롭게 피치검출 부분을 첨가하여 사람의 음성에만 반응하며, 상/하 추적이 가능하도록 기능을 확장하였다. 음성 여부를 판단하는 모듈을 사용하지 않은 경

우 일상생활에서 발생하는 잡음(예: 문 여닫는 소리, 의자가 움직이며 발생하는 소리 등)에 반응하지 않음을 확인할 수 있었다. 하지만 음원 추적과정에서 발생하는 잡음, 방향차로 인한 추적 오차를 줄이지는 못하였으며, 이러한 오차는 영상을 통한 얼굴 추적으로 오차를 보정하여 성능을 향상 시켰다. 향후 다수의 화자가 동시에 발음을 하였을 경우 화자들의 발음을 구분하여 특정 화자에 반응할 수 있는 화자 인증 및 화자가 특정 단어를 발음하였을 때 해당 동작을 할 수 있는 단어 인식, 끝으로 영상 신호를 이용하여 화자의 얼굴 영역을 검출 한 후 실시간으로 화자를 인식할 수 있게 하는 연구가 필요하다.

감사의 글

본 연구는 한국과학재단 지정 전남대학교 고품질 전기전자부품 및 시스템연구센터의 연구비 지원에 의해 연구되었음.

참 고 문 헌

[1] Schauer, C., Gross, H. M. 2001. "Model and application of a binaural 360° sound localization system." in *Proceedings of the International Joint Conference on Neural Networks*, Vol. 2, pp. 1132-1137.

[2] Knapp, C. H., Carter, G. C. 1976. "The Generalized Correlation Method for Estimation of Time Delay." in *Peoceedings of the IEEE Trans On Acoustic, Speech, and Signal Processing*, Vol. 24, pp. 320-327.

[3] Viola, P., Jones, M. 2001. "Robust real-time face detection." in *Proceedings of International Conference on Computer Vision*, Vol. 2, pp. 747-747.

[4] Chang, P. S. 1997. "Performance of 3D Speaker Localization Using a Small Array of Microphones." in *Proceedings of IEEE International Conference on Thirty-First Asilomar*, Vol. 1, pp. 2-5.

[5] Etter, D. M., Stearns, S. D. 1981. "Adaptive Estimation of Time Delays in Sampled data Systems," in *Proceedings of IEEE Trans On Acoustic, Speech, and Signal Processing*, Vol. 29, pp. 582-587.

[6] Chan, Y. T. & Riley, J. M. F. & Plant, J. B. 1981. "Modeling of Time Delay and Its Application Estimation of Nonstationary Delays." in *Proceedings IEEE Trans OnAcoustic, Speech, and Signal Processing*, Vol. 29, pp. 577-581.

[7] 김수희, 이배호. 2003. "계층적 분류기를 이용한 실시간 얼굴 검출 및 추적." 한국정보처리학회 추계학술발표대회, 제10권, 제2호, pp. 137-140.

[8] Oren, M., Papageorgiou, C., Sinha, P., Osuna, E. & Poggio, T. 1997. "Pedestrian detection using wavelet templates." in *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 193-199.

[9] Crow, F. 1984. "Summed-area tables for texture mapping." in *Proceedings of SIGGRAPH*, Vol. 18(3), pp. 207-212.

[10] Freund, Y., Schapire, R. E. 1999. "A Short Introduction to Boosting." in *Journal of Japanese Society for Artificial Intelligence*, Vol. 14(5), pp. 771-780.
[11] 박정옥, 나승유, 김진영. 2004. "휴머노이드 로봇을 위한 시청각 정보 기반 음원 정위 시스템 구현." 제17회 신호처리합동 학술대회, pp. 84.

접수일자: 2004. 11. 01
게재결정: 2004. 11. 29

▲ 박정옥
광주광역시 북구 용봉동 전남대학교 공대 6호관 603호 (우: 550-757)
전남대학교 공과대학 전자정보통신공학과
Tel: +82-62-530-0472 (O), +82-61-743-3825 (H), 011-9119-0709 (H/P)
Fax: +82-62-530-1759
E-mail: aneohero@hanmail.net

▲ 나승유
광주광역시 북구 용봉동 전남대학교 공대 6호관 616호 (우: 550-757)
전남대학교 공과대학 전자정보통신공학과
Tel: +82-62-530-1753 (O), 010-9821-1110 (H/P), Fax: +82-62-530-1759
E-mail: syna@chonnam.ac.kr

▲ 김진영
광주광역시 북구 용봉동 전남대학교 공대 6호관 614호 (우: 550-757)
전남대학교 공과대학 전자정보통신공학과
Tel: +82-530-1757 (O), 018-615-3214 (H/P), Fax: +82-62-530-1759
E-mail: beyondi@chonnam.ac.kr