
음원 위치 검출기의 구현

이 종 혁*, 김 진 천**

Implementation of Sound Source Location Detector

Jong-Hyeok Lee, Jin-Chun Kim

본 연구는 1998학년도 경성대학교 연구년 지원사업에 의하여 연구되었음.

요 약

인간의 청각시스템은 두 가지 요소 즉, ITD(Interaural Time Difference)와 IID(Interaural Intensity Difference)를 처리하여 음원의 위치와 추적을 하고 있다.

본 연구에서는 음원의 위치 검출을 위하여 ITD와 IID뿐만 아니라 이전의 위치 정보를 이용하여 정확한 음원의 방향을 결정할 수 있는 TEPILD(Time Energy Previous Integration Location Detector) 모델을 제안하였다. TEPILD 모델에서 time function generator는 ITD, energy function generator는 IID를 처리할 수 있도록 하였다. 음원은 정현파(500Hz, 1KHz, 2KHz, 3KHz), White noise, Pink noise, News, Music으로 하고 음원의 방향은 right, front right, front, front left, left로 하였다. 실험 결과 전체 평균 정확도가 99.2로 좋은 결과를 얻을 수 있었으며, TEPILD가 음원 위치 검출기에 이용될 수 있음을 확인하였다.¹⁾

Abstract

The human auditory system has been shown to possess remarkable abilities in the localization and tracking of sound sources. The localization is the result of processing two primary acoustics cues. These are the interaural time difference(ITD) cues and interaural intensity difference(IID) cues at the two ears.

In this paper, we propose TEPILD(Time Energy Previous Integration Location Detector) model. TEPILD model is constructed with time function generator, energy function generator, previous location generator and azimuth detector. Time function generator is to process ITD and energy function generator is to process IID.

* 경성대학교 전기전자·컴퓨터공학부 교수

** 경성대학교 전기전자·컴퓨터공학부 조교수

접수일자 : 2000. 9. 7.

Total average accuracy rate is 99.2%. These result are encouraging and show that proposed model can be applied to the sound source location detector.

I. 서론

인간이 가지고 있는 오감(五感)은 여러 학문 분야 특히 인공지능(Artificial Intelligence) 분야에서 관심의 대상이 되고 있으며, 청각과 밀접한 관련이 있는 분야는 음원 위치 추정(Sound Source Location Estimation), 음성분석, 음성인식 및 음성합성 등이다.

최근 각광을 받고 있는 영상회의 시스템에서 음원 위치에 따라 움직이는 카메라를 사용하므로 더욱 자연스러운 영상회의를 진행할 수 있으며[1], 더욱 효율적인 로봇제어를 위하여 카메라에서 얻은 영상 정보뿐만 아니라 마이크에서 얻은 음원 위치 정보를 이용하고 있다.[2] 다른 산업 현장 즉, 수중 전파탐지기(sonar), 레이더, 음향학(acoustics), 지구물리학(geophysics) 및 생의학 공학(biomedical engineering) 등에서도 음원의 위치 추정에 관련된 연구가 많이 발표되고 있다.[2][3][4]

일반화된 상호상관에 관계된 연구는 두 귀의 입력 신호사이의 유사성을 측정하여 처리하는 모델로 Basic cross-correlation methods, Roth processor, Smoothed coherence transform(SCOT), Phase transform, Eckart processor, Maximum likelihood estimator, 시간 영역에서의 LMSTDE(least mean square time delay estimation) algorithm, SCOT algorithm과 비슷한 적응방법인 adaptive squared coherence function for TDE algorithm[5], 신호들의 위상차에 관한 정보를 얻을 수 있는 Crosspower Spectrum Phase(CSP)에 기초한 방법[6], 음원의 방향뿐만 아니라 파형도 예측할 수 있는 새로운 composite complex sinusoidal modeling방법[7] 등이 있다.

일반화된 상호상관에 관계된 연구는 높은 주파수에서는 좋은 결과를 얻을 수 없다. 이를 해결하기 위하여 인간의 청각 메카니즘을 도입하였다[8][9]. IID(Interaural Intensity Differences)는 높은 주파수에서 ITD (Interaural Time Differences)는 낮은 주파수에서 정보를 추출하는 역할을 하므로 높은 주파수에서도 좋은 결과를 얻을 수 있는 HRTF(head-related

transfer function) 방법에 대한 연구도 발표되고 있다 [10][11].

종래의 많은 연구는 여러 개(3개 이상)의 마이크로폰에서 수신되는 신호들의 콘볼루션을 구하여 ITD를 구하고, 여기에 음파의 속도를 이용하여 음원의 방향을 결정하였다.

본 연구에서는 두 개의 마이크로폰만을 사용하고, 이것으로 발생하는 단점을 극복하기 위하여 인간이 음원의 방향을 결정하는 두 가지 요소 즉 ITD, IID를 동시에 이용할 뿐만 아니라 바로 전 검출한 음원의 위치 정보도 이용하므로 정확한 음원의 위치를 구하고자 한다.

II. 인간의 청각 시스템

소리가 귀에 전달되는 과정은 음파의 떨림이 자유공간을 통하여 전달되고 전달된 신호가 고막(eardrum)을 진동시키고 이 고막의 진동이 내이에 전달되고 내이에서 이를 신경섬유에 펄스(pulse)로 전달하여 신경경로를 통하여 뇌에 전달하게 된다.

2.1 소리의 전달 통로

외부의 소리는 일정한 통로를 거쳐 인간의 귀에 전달된다. 외부의 소리가 귀에 전달되기 위해서는 얼굴부분을 반드시 통과하게되므로 귀와 얼굴 부분에 대한 정보를 알아볼 필요가 있다. 양쪽 귀를 포함한 얼굴의 단면도는 사다리꼴 모양을 하고 있다. 밑변의 길이 즉, 두 귀 사이의 거리는 대략 160mm정도이며, 윗변의 길이 즉, 두 뺨 사이의 거리는 130mm이며, 높이 즉, 얼굴의 옆면에서 볼 때 옆면의 폭은 75mm정도이다. 그리고 앞에서 본 귓바퀴의 폭은 10mm정도이며, 옆면에서 볼 때 귓바퀴의 폭은 25mm정도이다.

위와 같은 얼굴 구조에서 두 귀에 소리가 직접적으로 전해지려고 하면 소리의 위치는 얼굴 정면 중앙 기준 20도 이내, 270mm정도 이상 떨어져 있어야 한다. 그리고 소리의 위치는 거리와 각도로써

표시되지만 각도가 일정값 이상이 되면 가까운 귀에는 소리가 직접적으로 전달되지만 다른 쪽 귀에는 직접적으로 전달되지 못하고 제 3의 물체에 반사되어 오는 반사파가 전달되므로 에너지는 많이 감소되고 그리고 도달하는 시간도 많이 걸린다. 이런 특성을 이용하여 소리의 위치를 파악할 수 있게 된다.

2.2 귀의 구조

귀는 조직의 전달 성분별로 외이, 중이, 내이의 세 부분으로 나눌 수 있다. 외이는 음원을 수집하는 역할을 담당하는 귓바퀴와 고막에 이르기까지 음을 전달하는 전달관으로 구성된다. 그리고 고막은 음파 신호를 중이에 있는 작은 뼈들의 기계적 떨림으로 전달한다. 고막을 통해 전달된 음파를 이 뼈들의 움직임에 의해 다시 내이의 전정관으로 전달한다. 내이는 전정기관과 와우각으로 구성된다.

2.3 와우각의 구조

와우각은 전체 길이가 3.5센티미터 정도이고 두 바퀴반 감긴 튜브모양을 하고 있다. 와우각관은 기저막, 라이스너막 및 와우각 벽으로 둘러싸여 있는데 기저막과 텍토리알막 및 섬모 세포들로 구성된 코티기관(organ of Corti)을 포함하고 있다. 여기서 유체의 진동에 대한 반응으로 기저막은 진행파를 생성하게 된다. 그런데 기저막의 특성은 타원창 부근에서 좁고 딱딱하며 정점 부근에서는 넓고 덜 딱딱한 판의 형태를 하고 있다. 이러한 특성으로 인해 타원창 부근에서부터 정점 부근까지 고주파에서 저주파의 신호를 추출하는 일종의 필터뱅크를 구성한다고 볼 수 있다. 각 섬모 세포는 대략 1000개의 섬모들을 가지고 있으며 기저막이 진동하면 이 진동에 맞춰 섬모세포들이 펄스를 발한다. 일반적으로 내 섬모세포의 역할은 기저막의 진동에 맞추어 펄스를 발생시켜 청각신경정보를 전달한다고 알려져 있다.

2.4 청각 신경경로

와우각의 신경섬유에서 전달된 신호는 신경세포

가 다발로 연결된 여러 단계, 즉 cochlear nucleus, superior olive complex, inferior colliculus 그리고 medial geniculate와 같은 여러 단계의 신경 다발을 통해 청각피질(auditory cortex)에 전달된다. 여기에서 나타나는 일반적인 특성은 많은 부분의 세포들이 주파수의 반응에 따라 위치가 순차적으로 정해져 있고(tonotopic mapping), 청각 신경경로를 따라 올라감에 따라 복잡한 음성특징들이 나타나는 것이다.

Ⅲ. 음성 위치 추정

우리는 양쪽 귀에서 들려오는 소리의 미묘한 차이 즉, 도달 시간과 세기의 차이를 가지고 소리의 위치를 찾아내고 있다. 이를 구현할 수 있는 모델을 제시한다.

3.1 위치변수 모델

양쪽 귀에서 생기는 소리의 시간지연차이 (ITD)와 세기차이(IID)의 두 가지 정보를 찾아내고 이로부터 소리의 위치를 결정하는 메커니즘 중에 측음화 현상을 정량적으로 나타내면 다음 식과 같다.

$$E[L(\tau, f)] = E[L(\tau, t, f)]$$

$$= \int_{-\infty}^t r_L(a) r_R(a - \tau) w_c(t - a) p(\tau) da \dots\dots\dots (1)$$

여기서 $r_L(a)$ 와 $r_R(a)$ 는 양쪽 귀에서 청신경 세포의 시간에 따른 발화함수를 나타내며, $w_c(t)$ 는 시간에 대한 교차상관함수의 적분가중치를 뜻하는데, 5Hz의 차단 주파수를 가지는 일차의 저역 통과 필터로 $w_c(t)$ 를 근사하였다. 그리고 $p(\tau)$ 는 중앙함수를 의미하며, 작은 크기의 내부 지연을 강조하기 위해 $e^{(-k\tau)}$ 형태의 값을 가지도록 했다. 이러한 상호상관에 의한 ITD와 IID를 고려한 위치변수(position-variable) 모델을 제시되었다[12]. 이 모델에서는 ITD 정보와 IID 정보의 곱으로 위치함수를 정의하였고, 이 함수의 중앙값을 측음화가 일어나는 위치로 보았다.

3.2 TEPILD 모델

자유공간에 있는 음원(sound source)과 두 개의 마이크로폰을 가진 시스템에서 두 마이크로폰 사이의 거리, 음원과 마이크로폰과의 거리 및 방위각의 관계를 그림 1에 나타내었다.

그림 1에서 두 마이크로폰의 중점에서 본 음원과 수평축이 이루는 각도 즉, 방위각은 삼각함수로써 구할 수 있으며 식 (2)와 같다.

$$azimuth = \arccos(e_{dist}/L) * 180.0/\pi \dots\dots\dots (2)$$

그러나 삼각함수로 구한 음원의 위치 정보는 음원의 위치에 따라 에러를 발생할 수 있다. 이 때는 a_{dist} 가 e_{dist} 보다 크므로 이론치보다 실측치가 클 수 있다. 그러나 음원이 장치의 중앙축에 존재하던지 아니면 두 마이크의 거리보다 매우 떨어진 거리에 있으면 즉, $r \gg L$ 이면 l 은 작아지므로 e_{dist} 와 a_{dist} 는 거의 같아지며 식 (3)과 같이 나타낼 수 있다.

$$e_{dist} \cong a_{dist} = \text{abs}(\text{offset}) * V_{\text{sound}}/fs \dots\dots\dots (3)$$

여기서 offset은 두 입력신호의 상호상관에 의해서 구해지는 값이며 V_{sound} 는 음속, fs 는 샘플링 주파수를 나타낸다. 식 (3)을 식(2)에 대입하면 음원의 방위각을 구할 수 있게 된다.

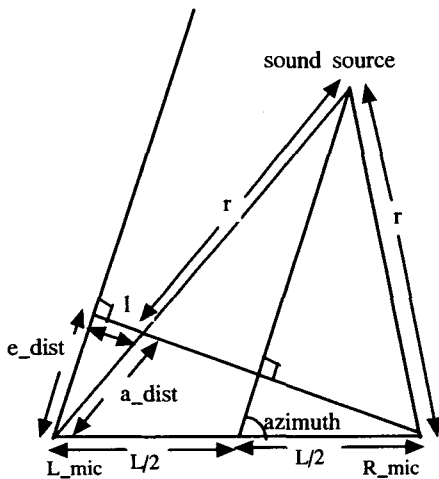


그림 1. 두 개의 마이크로폰시스템에서 방위각
Fig. 1. Azimuth at two microphone system

일반적으로 신호들은 에너지가 연속적으로 변하며, 위치는 급격하게 변하지 않으므로 본 연구에서는 두 마이크로폰에 도달하는 음파의 시간차를 이용하여 구한 현재 음원의 위치, 에너지 비를 이용한 음원의 위치와 이전의 음원의 위치 정보를 종합적으로 이용하여 음원의 방위각을 검출할 수 있는 TEPILD(Time Energy Previous Integrated Location Detector) 모델을 제안하였으며 이를 그림 2에 나타내었다.

Time Function Generator에서는 두 마이크로폰에 도달하는 음파의 시간차를 이용하여 현재 음원의 위치를 구할 수 있도록 하고, Energy Function Generator에서는 두 마이크로폰의 에너지 비를 이용한 음원의 위치를 구할 수 있도록 하였으며 Previous Location Generator에서는 음원의 이전 위치와 현재 위치의 관계성을 구할 수 있도록 하였으며 Azimuth Detector에서는 이들 정보를 종합하여 최종적으로 음원의 위치를 출력할 수 있도록 하였다.

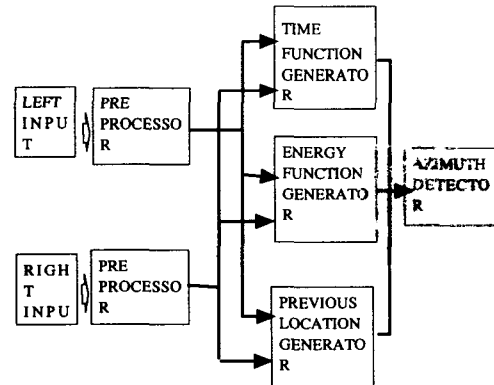


그림 2. TEPILD 모델
Fig. 2. TEPILD Model

음원이 이동할 때 이동속도가 빠르지 않다고 할 때 현재의 위치와 이전의 위치와는 별 차이가 없어야 한다. 그러므로 이 두 정보의 차로써 시간차를 이용하여 구한 현재의 음원 위치에 대한 신뢰성을 확인하고자 하며 이를 가우스 함수로 나타내면 다음과 같다.

$$P_f = \frac{1}{\epsilon ((C_{lc} - P_{lc}) * Pass_{factor} / T_{avg})^2} \dots\dots\dots (4)$$

여기서 C_{loc} 는 현재의 위치를 나타내며, P_{loc} 는 이전의 위치를 나타낸다. 출력 분해능을 30도로 했으므로 허용 오차 각 T_{ang} 은 출력 분해능의 1/2 이면 되지만 본 연구에서는 음원의 이동속도를 빠르지 않는 것으로 한정하였으므로 출력 분해능의 $1/\sqrt{2}$ 즉, T_{ang} 을 $30/\sqrt{2}$ 로 하였다. 그리고 ($C_{loc} - P_{loc}$)값이 T_{ang} 과 같아지면 $P_f = 1/\sqrt{2}$ 즉, 통과와 차단 관계값이 되도록 하였다. 그러므로 $\epsilon^{Pass_{factor}^2} = \sqrt{2}$ 관계가 성립하고 따라서 통과 임계값 $Pass_{factor} = \sqrt{\ln \sqrt{2}}$ 이 된다. 식 (4)에서 제공하는 함수를 이용하여 Previous Location Generator를 구현하도록 한다.

에너지 비와 음원 위치의 관계에서 먼저 에너지 비를 이용한 음원의 위치를 시그모이드 함수를 이용하여 예측해보면 다음과 같다.

$$E_{ang} = \frac{180}{(1 + \epsilon^{-dir(((L_E + TH)/(R_E + TH))^{dir} - 1) * R_{factor}})} \dots\dots (5)$$

여기서 E_{ang} 는 두 마이크로폰의 에너지 비에 의해서 예측되는 음원의 위치를 나타내며, L_E 는 왼쪽 마이크로폰의 에너지를 R_E 는 오른쪽 마이크로폰의 에너지를 나타낸다. 그리고 TH 는 분모가 "0"으로 되면 연산이 되지 않으므로 이를 처리하기 위한 임계값이며, dir 은 에너지비에 따른 값으로 L_E 가 R_E 보다 크면 +1 아니면 -1이다. 음원의 위치가 전방 45도에 있다고 할 때 좌우측 에너지의 비가 1/2이 된다면 다음의 수식이 성립한다.

$$\frac{180}{45} = 1 + \epsilon^{-dir(((L_E + TH)/(R_E + TH))^{dir} - 1) * R_{factor}} \dots\dots (6)$$

여기서 R_{factor} 를 구하면 $\ln 3$ 이 된다. 에너지 비에 의해서 구해진 음원의 위치와 이전의 음원 위치의 동일의 정도를 가우시안 함수를 이용하여 나타내면 다음과 같다.

$$E_f = \frac{1}{\epsilon^{((E_{ang} - P_{loc}) * Pass_{factor} / T_{ang})^2}} \dots\dots (7)$$

여기서 E_{ang} 는 두 마이크로폰의 에너지 비에 의해서 예측되는 음원의 위치를 나타내며, P_{loc} 는 이전의 위치를 나타낸다. T_{ang} 은 허용 오차 각을

나타내며 $Pass_{factor}$ 는 통과 임계값을 나타낸다. 식 (5), (6) 및 (7)에서 제공하는 함수를 이용하여 Energy Function Generator를 구현하도록 한다.

현재의 위치, 이전의 위치 및 두 마이크에서 수신된 신호의 에너지 비의 함수로써 음원 위치 정보의 신뢰성을 확정 할 수 있으며 이를 시그모이드 함수로써 표현하면 다음과 같다.

$$A_{dec} = 1 / (1 + \epsilon^{-(X - Pass_{value})}) \dots\dots (8)$$

여기서 $X = P_f + E_f + Distinguish_{part}$ 이며 $Pass_{value} = 2/\sqrt{2}$ 이다. P_f 는 현재 위치와 이전 위치의 관계에 의해서 결정 지는 값이며 E_f 는 에너지와 시간차에 의해서 구해지는 위치 정보로부터 결정지는 값이며, $Distinguish_{part}$ 는 시간차에 의해서 구한 음원 위치 정보의 정확성에 따라 "0" 혹은 " $1/\sqrt{2}$ "이다. X 값은 세 개의 항으로 구성되므로 두 개이상의 항에서 통과를 할 때 이 함수의 값이 의미가 있는 것으로 하기 위해서 $Pass_{value}$ 를 $2/\sqrt{2}$ 로 하였다. 식 (9)에서 제공하는 함수를 이용하여 Azimuth Detector를 구현하도록 한다.

TEPILD 모델에서 Pre Processor는 고주파 잡음을 감소시키기 위하여 사용하며 가우시안 필터로 이를 구현하였다.

일반적인 Gaussian 함수를 차단 각주파수를 포함하는 시간 함수로 나타내면 다음과 같다.

$$h(t) = \frac{\omega_c}{\sqrt{2\pi * \ln 2}} e^{-t^2 * \omega_c^2 / 2.0 * \ln 2} \dots\dots (9)$$

식 (10)에서 차단 주파수는 다음과 같다.

$$f_c = \frac{\sqrt{\ln 2} f_s}{2\pi\sigma} \dots\dots (10)$$

여기서 샘플링 주파수를 44.1kHz로 했을 때 차단 주파수를 1kHz 이내로 하기 위해서는 시그마는 5.8보다 커야 한다. 따라서 본 연구에서는 6으로 하였다.

IV. 실험 및 고찰

4.1 실험 장치 및 방법

본 실험에 사용된 장치는 감각기관의 상호 작용

에 관한 연구의 일부분으로 Octane 컴퓨터(2*64 bit MIPS R 10000 processor)와 무선으로 연결되어 작동되는 Trilobot사에서 제공하는 이동 로봇이며 이를 그림 3에 나타내었다.

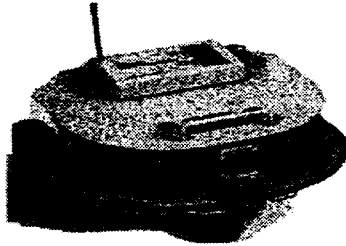


그림 3. Trilobot 이동 로봇
Fig. 3. Trilobot Mobile Robot

이동 로봇에는 부착되어 있는 센서로는 비디오 카메라(640×486), 디지털 나침반(North, NE, East, SE, South, SW, West, NW), omni directional electric condenser (powered by system) 마이크로폰 2개, 초음파 거리 측정기(6" to 255"), 온도 센서(32 °F-120 °F), 밝기 센서(10Lux-100Lux), 범퍼 스위치, 경사 센서, 바퀴 경로 엔코더, 초기 상태 복귀 스위치, 배터리 상태 감지기 등이 있다.

이동 로봇에서 초음파 거리 측정기 및 주변 지물로 인한 수신 장애물과 두 마이크로폰과의 치수 관계를 그림 4에 나타내었다.

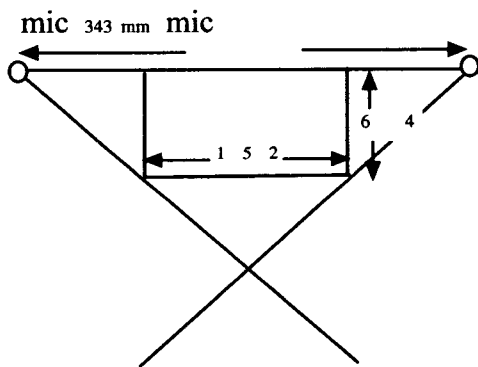


그림 4. 이동 로봇에서 치수
Fig. 4. Dimension at mobile robot

두 마이크로 폰 사이의 거리를 약 340mm로 하

고 음속을 340m로 했을 때 두 마이크로폰에 음파가 도달하는 시간의 차이는 1ms이며, 이 때의 방위각은 90도이다. 그러므로 1도의 해상도를 얻기 위한 최저 샘플링 속도 fs는 90KHz가 된다. 그러나 일반적으로 사운드 카드에서는 위와 같은 높은 주파수는 지원되지 않는다.

표 1에는 일반적으로 지원되는 최고 샘플링 주파수 및 많이 사용되는 주파수에서의 각 해상도를 나타내었으며, 본 실험에서는 샘플링 주파수를 44.1KHz로 하였다.

표 1. 샘플링 주파수와 각 해상도관계
Table 1. Sampling frequency vs degree resolution

샘플링 주파수[Hz]	각 해상도[deg]
44,100	90,000/44,100= 2.04
8,000	90,000/8,000=11.25

실험을 위한 조건으로는 한 개의 음원이 있는 것으로 하고 음원은 매우 천천히 움직일 수 있으며 조용한 실험실 환경(SNR 20dB 이상)으로 가정하고 단지 두 개의 마이크로폰만 사용한다.

표 2에 방위각의 범위와 출력의 관계를 나타내었다. 여기서 방위각은 Azimuth Detector에서 구하며 출력은 본 실험에서 사용된 이동 로봇이 움직여야할 방향을 나타낸다.

표 2. 방위각 범위와 출력 관계
Table 2. Azimuth range vs output

Azimuth Range	Output
0 - 30	Right
30 - 60	Right or Front
60 - 120	Front
120 - 150	Front or Left
150 - 180	Left

실험에 사용된 신호는 정현파(500Hz, 1KHz, 2KHz,

3KHz), White noise, Pink noise, News, Music이며 이들 신호를 녹음하여 이동 로봇과 1미터에서 3미터 거리의 임의의 점에 두고 음원 위치를 구하도록 하였다.

4.2 실험 결과 및 고찰

표 3에 신호원을 두 마이크로폰의 중앙을 기준으로 거리는 2미터, 방위각은 각각 15도, 45도, 90도, 135도 및 165도에 두었을 때 상호상관에 의해서 구해진 방위각 정보를 소수 첫째자리에서 반올림하여 나타내었다.

표 3. 상호상관에 의해서 구해진 방위각

Table 3. Azimuth obtained by cross correlation

	15°	45°	90°	135°	165°
500Hz	12	43	90	137	168
1KHz	15	45	90	135	165
2KHz	12	45	92	135	168
3KHz	10	43	88	137	170
White noise	8	41	94	137	172
Pink noise	10	43	84	139	174
News	10	43	90	137	170
Music	8	43	86	139	170

음원이 전방측에 위치했을 때 좋은 결과를 보였지만 주파수가 높은 신호에서는 오차가 상대적으로 크다. 이는 상호상관이 가지고 있는 문제로 생각된다. 그리고 신호원이 측면에 있을 경우는 먼쪽의 마이크로폰에 수신된 신호가 반사파이므로 시간이 더 지연 되었으며 이로 인해 오차가 크다고 생각된다.

표 4에 신호원을 두 마이크로폰의 중앙을 기준으로 거리는 2미터, 방위각은 각각 15도, 45도, 90도, 135도 및 165도에 두었을 때 두 마이크로폰에서 수신된 에너지의 비를 오른쪽 마이크로폰 기준으로 나타내었다. 각 측정값은 소수 셋째자리에서 반올림하였다.

표 4. 두 채널간의 에너지 비

Table 4. Energy ratio between two channels

	15°	45°	90°	135°	165°
500Hz	0.30	0.52	1.02	2.05	3.19
1KHz	0.32	0.51	1.01	2.02	3.21
2KHz	0.32	0.52	1.02	2.01	3.21
3KHz	0.30	0.48	1.03	2.11	3.20
White noise	0.31	0.54	1.03	1.98	3.19
Pink noise	0.30	0.46	1.04	2.02	3.20
News	0.32	0.52	1.02	2.02	3.21
Music	0.30	0.47	1.03	1.94	3.19

표 4의 결과에서 높은 주파수 및 측면에 있는 신호에서도 충분한 변별력이 있다. 그러므로 상호상관에서 해결할 수 없었던 정보를 제공해 줄 수 있다고 생각된다.

표 5. 신호위치에 따른 최종결과의 정확도

Table 5. Accuracy rate of final result according to sound location

	right	front right	front	front left	left	평균
1.0m	96	98	100	98	98	98.00
1.5m	100	100	100	100	98	99.60
2.0m	100	100	100	100	100	100.0
2.5m	98	100	100	100	98	99.20
3.0m	98	100	100	100	98	99.20
평균	98.40	99.60	100.0	99.60	98.40	99.20

표 5에 News 신호원을 right(15도, 25도), front right(35도, 55도), front(65도, 115도), front left(125도, 145도) 및 left(155도, 165도)방향의 범위에 두고 거리를 변동했을 때 최종 출력 정보를 구하고 이들을 각각 25회 실시한 후(각 거리에서 방향별 실시회수 : 50) 정확도를 평균하여 백분율로 나타내었다.

전체 평균이 99.2%로 아주 좋은 결과를 얻을

수 있었다. 그러나 가까운 거리와 측면쪽으로 갈수록 상대적으로 나뉘었다. 그러나 이 부분은 이동로봇에서는 여러 센서가 상호 작용을 하고 특히 카메라 센서가 가까운 부분에서는 정보를 처리할 수 있으므로 문제가 없었다.

V. 결론

본 연구에서는 두 개의 마이크로폰만을 사용하고, 이것으로 발생하는 단점을 극복하기 위하여 인간이 음원의 방향을 결정하는 두 가지 요소 즉, ITD와 IID를 동시에 이용할 뿐만 아니라 바로 전 검출한 음원의 위치 정보를 이용하는 TEPLD 모델을 제안하여 정확한 음원의 위치를 구하고자 하였다.

현재의 위치는 ITD 기능을 시뮬레이션 하여 구하고 IID 기능에 기반을 둔 두 마이크로폰의 에너지 비에 매칭되는 예측 방향 정보는 시그모이드 함수를 이용하여 구하였다. 일반적인 소리는 연속적이며 급격하게 위치가 변화되지 않으므로 현재의 위치와 과거의 위치와는 상관 관계가 있으므로 지수 함수를 이용하여 과거의 위치를 고려한 현재 위치의 타당성을 구하도록 하였다. 즉, 현재의 위치, 이전의 위치 및 두 마이크에서 수신된 신호의 에너지 비의 함수로써 음원 위치 정보 구할 수 있도록 하였다. 그리고 잡음의 영향을 줄이기 위하여 가우시안 함수를 이용하였다.

음원은 정현파(500Hz, 1KHz, 2KHz, 3KHz), White noise, Pink noise, News, Music으로 하고 음원의 방향은 right, front right, front, front left, left 범위로 하였으며 거리는 1미터에서 3미터 범위로 하였다. 최종 결과의 정확도 전체 평균이 99.2%로 아주 좋은 결과를 얻을 수 있었다.

향후 영상회의 시스템용 화자 추적에 관한 연구도 계속되어져야 되리라 생각한다.

참고문헌

[1] K. Okamura, S Shimada, and H Hokari, "Estimation of a Sound Location using of randomly positioned microphones," 電子情報通信學會論文集 A, Vol. J80-A No. 11,

pp.2054-2058, 1997.

- [2] Nicholas Ayache, "Artificial Vision of Mobil Robotics: stereo vision and multisensory perception," MIT press, 1991.
- [3] L. Jiaxiang, "A new cross spectral method to locate sound sources," Proceeding of ICSP' 96, pp.233-236, 1996..
- [4] M. Abe, Y. Nagata, and K. Kido, "A New Method to Locate Sound Sources by Searching the Minimum Value of Error Function," ICASSP 86, TOKYO, pp.933- 936, 1986..
- [5] D. H. Youn and N. Ahmed, "Time delay estimation via coherence : An adaptive approach," J. Acoust. Soc. Am. 75 (2), Feb. 1984.
- [6] H. Wang and P. Chu, "Voice Source Localization for Automatic Camera Pointing System in Vide Conferencing, Proceeding of ICSP' 97, pp.187-190, 1997.
- [7] M. Abe, J. Liu, and K. Kido, "An Application of Composite Complex Sinusoidal Modeling to the Estimation of Directions and Waveforms of Incident Plane Waves," J. Acoust. Soc. Jpn. (E) 11, 2, pp. 107-110, 1990.
- [8] J. Huang, N. Ohnishi, and N. Sugie, "A Biomimetic System for Localization and Separation of Multiple Sound Sources," IMTC '94, pp.967-970, May, 1994.
- [9] W. Chau and R. O. Duda, "Combined Monaural and Binaural Localization of Sound Sources," Proceedings of ASILOMAR_29, pp.1281-1285, 1996.
- [10] D. Nandy and J. Ben-Arie, "Estimating the Azimuth of a Sound Source from the Binaural Spectral Amplitude," IEEE TRANS. on speech and audio processing, Vol. 4, No. 1, 1996.
- [11] P. Zakarauskas and M. S. Cynader, "A computational theory of spectral cue localization," J. Acoust. Soc. Am. 94, (3), pp.1323-1331, Sept. 1993.
- [12] R. M. Stern, Jr. and H. S. Colburn, "Theory

of binaural interaction based on auditory nerve data," J. Acoust. Soc. Am. 64(1), 1978.



이 종 혁(Jong-Hyeok Lee)

1975년 2월 부산대학교 전자공학과(공학사)

1980년 2월 부산대학교 대학원 전자공학과(공학석사)

1991년 2월 부산대학교 대학원

전자공학과(공학박사)

1980년 3월 ~ 1990년 2월 동의공업대학 전자과 부교수

1990년 3월 ~ 현재 경성대학교 전기전자·컴퓨터공학부 교수

1998년 7월 ~ 1999년 6월 미국 Beckman Institute, University of Illinois 객원연구원

※ 주관심분야 : 인공지능, 음성인식, 신호처리



김 진 천(Jin-Chun Kim)

1983년 2월 한양대학교 전기공학과(공학사)

1985년 8월 미국 미시간주립대학교 전자 및 시스템공학과(공학석사)

1996년 2월 한국과학기술원 전산학과(공학박사)

1988년 4월 ~ 1993년 5월 삼성종합기술원 정보시스템연구소 선임연구원

1996년 3월 ~ 현재 경성대학교 전기전자·컴퓨터공학부 조교수

※ 연구분야 : 멀티미디어 통신, ATM 스위치 구조, 컴퓨터 구조, 초고속 네트워크