

## 감성로봇을 위한 음원의 위치측정 및 분리

김경환\*, 김연훈\*, 곽윤근#

### Sound Source Localization and Separation for Emotional Robot

Kyoung-Hwan Kim\*, Yeon Hoon Kim\*, Yoon Keun Kwak#

#### ABSTRACT

These days, the researches related with the emotional robots are actively investigated and in progress. And human language, expression, action etc. are merged in the emotional robot to understand the human emotion. However, there are so many sound sources and background noise around the robot, that the robots should be able to separate the mixture of these sound sources into the original sound sources, moreover to understand the meaning of voice of a specific person. Also they should be able to turn or move to the direction of a specific person to observe his expression or action effectively. Until now, the researches on the localization and separation of sound sources have been so theoretical and computative that real-time processing is hardly possible. In this reason for the practical emotional robot, fast computation should be realized by using simple principle. In this paper the methods for detecting the direction of sound sources by using the phase difference between peaks on spectrums, and the separating the sound sources by using fundamental frequency and its overtones of human voice, are proposed. Also by using these methods, it is shown that the effective and real-time localization and separation of sound sources in living room are possible.

**Key Words** : Emotional robot (감성로봇), Sound source localization (음원의 방향측정), Sound source separation (음원의 분리)

#### 기호설명

$\theta$  = direction of sound a source

$l$  = length from a sound source to a microphone

$d$  = distance between 2 microphones

$\Delta_p$  = path-difference from a sound source to 2 microphones with plane-wave assumption

$\Delta_r$  = path-difference from a sound source to 2 microphones with round-wave assumption

$\Delta_\phi$  = phase-difference from a sound source to 2 microphones with plane-wave assumption

$f_p$  = frequency of a peak on spectrum

$\lambda$  = wave length of sound

$v$  = wave speed of sound

#### 1. 서론

2002년 3월 29일 접수  
\* 한국과학기술원 기계공학과 대학원  
교신저자, 한국과학기술원 기계공학과  
Email kyk@kaist.ac.kr Tel. (042) 869-3212

감성공학은 출현한지 10 여년이 지났으나 아직 정확한 의미가 정의되어 있지 않은 상태로서 단순히 인간의 감성에 맞는 기술이나 제품을 개발하는 분야로 이해되고 있다.<sup>1</sup> 감성공학에 대한 연구는 아직 데이터베이스 구축 수준에 머물고 있는 정도이며, 감성공학의 실질적인 응용분야로서의 감성 로봇은 이제 시작 단계라고 할 수 있다.

감성로봇과 인간과의 관계에서 가장 중요한 것은 상호간의 감성교류이다. 특히 감성로봇이 인간의 감성 상태를 판단하고 상황에 맞게 대처하는 능력은 기본적인 기능의 하나로서 중요하다. 인간의 감성 상태를 판단하기 위하여 이용할 수 있는 것에는 가장 기본적인 것으로서 언어를 통해 이해하는 방법과 얼굴 표정이나 몸짓, 행동을 분석하는 방법 등이 있다.

로봇이 인간의 감성을 원활하게 파악하기 위해서는, 서로의 표정과 함께 소리를 이해할 필요가 있다. 이를 위하여 음원에서 발생하는 음파를 마이크로폰으로 측정하고 측정된 신호를 이용하여 음원의 방향을 알 수 있다. 이렇게 로봇에서 인간으로의 방향을 알아내는 것을 음원의 방향측정(sound source localization)이라고 한다.

나아가, 로봇이 언어를 이용하여 감성교류를 하기 위해서는 인간의 언어를 이해해야 한다. 일상적인 생활공간에서는 로봇이 상대하고 있는 인간 이외에도 다른 대상이 발하는 음성과 배경소음이 함께 존재하므로 로봇의 마이크로폰에서 측정되는 신호에는 이러한 것들이 모두 혼합되어 있다. 따라서 특정인의 음성을 인식하기 위해서는 혼합된 신호를 음원별로 구별하여 혼합전의 신호 상태로 복원하고, 이렇게 구별된 신호의 방향을 이용하여 그 사람과 일치하는 방향에 있는 음원을 선택한다. 여기서 혼합된 신호로부터 각 음원별로 혼합되기 전의 신호를 복원하는 것을 음원의 분리(sound source separation)라고 한다.

인간과 로봇의 감성교류가 원활하게 이루어지려면 인간이 음성을 발생시키는 즉시 로봇이 이것을 감지하여 즉각적으로 반응을 해야 한다. 즉 인간이 느끼기에 로봇이 실시간적으로 동작을 하고 있어야 한다. 또한 음성이 연속되어 발생하므로 이것을 측정할 때도 연속적으로 해야 한다.

감성로봇의 동작환경은 일반적인 실내공간이다. 여기에는 대상이 되는 인간의 음성과 다른 여러 음원이 함께 존재할 뿐만 아니라 음파가 진행

하는 도중에 있는 장애물이나 벽에 반사되어 매우 복잡한 음향환경을 만들어 낸다.

Bregman은 인간의 청각 시스템이 여러 음원과 잡음이 섞인 환경에서 원하는 음원만을 분리해 낼 수 있는 능력을 음원장 분석(ASA; Auditory Scene Analysis)이라고 하였다.<sup>2</sup> Bregman은 ASA 문제를 해결하기 위하여 시작점(onset)과 끝점(offset), 변화의 완만함(gradualness of change), 하모닉 성질(harmoniccity), 음향학적 이벤트의 변화(changes occurring in the acoustic event)와 같은 4 가지 규칙성을 이용하였다.<sup>3</sup> ASA에 기초한 몇 편의 연구에서 제한 조건들을 이용하여 음원 분리 문제를 푸는 것이 가능함을 보였으며 이러한 방법을 CASA (Computational Auditory Scene Analysis)라고 한다.

Nakatani<sup>4</sup>는 음성이 기본주파수(fundamental frequency)와 배음(overtone)으로 이루어져 있음을 이용하여 bi-HBSS(binaural-Harmonic Based Stream Segregation) 방법을 제안하였다. Huang<sup>5</sup>은 신호 대 반사파의 비를 이용하여 선행 효과(precedence effect)의 계산 모델을 제시하였다. Unoki<sup>6</sup>는 혼합된 두 음원을 분리하는 것을 부적절하게 정의된 역전형 문제(ill-posed inverse problem)로 정의하고 Bregman의 4 가지 제한 조건을 이용하여 풀 수 있음을 보였다.

Nakadai<sup>7</sup>는 휴머노이드 SIG의 음원의 위치 측정 및 분리를 위하여 두 마이크로폰에서 측정된 신호의 스펙트럼에서 같은 주파수를 갖는 피크(peak)들 사이의 위상차를 이용하여 음원의 방향을 계산하였다. 또한 같은 방향을 갖는 피크들을 하나의 음원에서 나온 것으로 보고 이것을 역 푸리에 변환(inverse Fourier transform)하여 음원을 분리하는 방법을 제안하였다.

인간에게 친숙하게 다가오는 감성로봇은 다양한 기능과 함께 독립적인 동작이 요구되고 친화적인 형태를 갖기 위해서 그 크기와 디자인에 제약을 많이 받으며 경제적인 면 또한 중요한 요소이다. 따라서 음원의 방향측정과 분리에 이용되는 장비는 가능한 한 간단하게 하여 공간과 무게를 적게 차지하면서도 적은 비용으로 가능해야 한다.

본 연구에서는 일반적인 실내환경 내에서 비교적 간단한 장비를 사용하여 실시간, 연속적으로 음성 음원의 방향을 알아내고 여러 음성과 배경소음이 혼합된 신호를 혼합되기 전의 신호로 분리하는 방법에 대한 연구를 수행하고자 한다.

## 2. 음원의 방향측정 및 분리

### 2.1 음원의 방향측정

음원에서 발생한 음파가 마이크론까지의 전파시간은 음원에서 마이크론까지의 거리에 비례한다. 서로 다른 곳에 위치한 마이크론이 2 개 있으면 음원으로부터 각각의 마이크론까지의 거리가 다르므로 음파가 마이크론까지 도달하는데 걸리는 시간도 다르다. 두 마이크론에서 측정되는 신호를 주파수 영역에서 비교하면 음파의 주파수에 해당하는 곳에 피크가 있고 두 신호 사이에 일정 크기의 위상차가 있음을 알 수 있다. 이 위상차는 음원에서 각 마이크론까지의 경로차에 비례하고 이를 이용하면 마이크론에서 음원으로의 방향각을 알 수 있다.

음파는 공간상에서 구면파의 형태로 전파되며 음원의 방향과 음원으로부터 마이크론까지의 경로차는 다음과 같은 관계가 있다.

$$\Delta_r = \sqrt{\frac{d^2}{4} + l^2 + dl \sin \theta} - \sqrt{\frac{d^2}{4} + l^2 - dl \sin \theta} \quad (1)$$

계산을 간단하게 하기 위하여 평면파로 가정할 때의 경로차는 다음과 같다.

$$\Delta_p = d \sin \theta \quad (2)$$

평면파 가정의 오차를 알아보기 위하여 두 경로차의 비를 구하면 다음과 같다.

$$\frac{\Delta_p}{\Delta_r} = \frac{\sin \theta}{\sqrt{\frac{1}{4} + (\frac{l}{d})^2 + \frac{l}{d} \sin \theta} - \sqrt{\frac{1}{4} + (\frac{l}{d})^2 - \frac{l}{d} \sin \theta}} \quad (3)$$

여기서  $l/d$  가 4 이상이면 식의 최대값이 1.01 이하이므로 평면파 가정의 오차는 매우 작다고 할 수 있다.

또한 경로차  $\Delta_p$  와 두 마이크론에서 측정된 신호의 피크에서의 위상차  $\Delta_\phi$  사이에는 다음과 같은 관계가 있다.

$$\Delta_\phi = \frac{\lambda}{2\pi} \Delta_p = \frac{v}{2\pi f_p} \Delta_p \quad (4)$$

두 식 (2)와 (4)를 정리하면  $\theta$ 와  $\Delta_\phi$  사이의 관계를 얻을 수 있다.

$$\theta = \sin^{-1} \frac{\Delta_p}{d} = \sin^{-1} \left( \frac{v}{2\pi d f_p} \Delta_\phi \right) \quad (5)$$

두 마이크론에서 측정된 신호의 피크에서의 위상차  $\Delta_\phi$ 는 다음의 식에 의해 구할 수 있다.

$$\Delta_\phi = \tan^{-1} \frac{\text{Im}[F_1(f_p)]}{\text{Re}[F_1(f_p)]} - \tan^{-1} \frac{\text{Im}[F_2(f_p)]}{\text{Re}[F_2(f_p)]} \quad (6)$$

여기서  $F_i()$ 는  $i$  번째 마이크론의 신호의 푸리에 변환식,  $f_p$ 는 피크의 주파수,  $\text{Im}[]$ ,  $\text{Re}[]$ 는 각각 복소수의 허수부, 실수부이다.

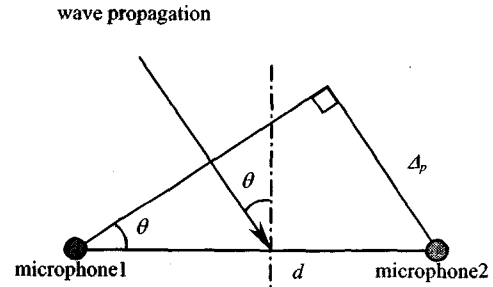


Fig. 1 Phase difference in frequency domain and the direction of sound source

### 2.2 음원의 분리

모음으로 이루어진 음원의 스펙트럼을 보면 특정 주파수에서만 피크가 생김을 알 수 있다. 특정 모음의 스펙트럼을 볼 때 두 번째 피크의 주파수는 첫 번째 피크의 주파수의 2 배이고, 세 번째는 3 배가 된다. 즉, 두 번째 이후의 피크의 주파수는 첫 번째 피크의 주파수의 배수가 된다. 여기서 첫 번째 피크를 기본주파수(fundamental frequency), 두 번째 이후의 피크를 배음(overtone)이라고 한다. 다섯 개의 스펙트럼들을 비교해보면 모두 같은 기본주파수를 가지며, 배음의 주파수도 모두 일치한다. 따라서 기본주파수와 배음의 관계를 이용하여 음성으로 이루어진 음원의 분리를 할 수 있다.

만약 기본주파수가 일치하는 음원이 있다고 가정하면 제시된 방법으로는 분리가 불가능하다. 그러나 일반적으로 여러 사람들의 음성이 섞여 있을 때, 각 사람들이 발화하는 음성의 높이가 다르므로 기본주파수가 다르게 되고 그 차이가 어느 정도 크기인가에 따라 구분 가능성이 결정된다. 즉 푸리에 변환의 주파수 분해능이 기본주파수 차이보다 작을 때 그 음원들을 분리할 수 있다.

음원을 분리하기 위하여 마이크로폰에서 측정된 신호의 스펙트럼에서 나타나는 피크의 주파수 정보를 이용하므로 음원의 주파수는 매우 중요한 요소이다. 그런데 마이크로폰에서 측정된 신호를 주파수 분석하기 위하여 FFT 를 이용하므로 스펙트럼에서 주파수가 일정한 분해능을 가지고 불연속적으로 나타나게 된다. 따라서 기본주파수만으로 이루어진 두 음원이 있을 때, 두 기본주파수 차이가 충분히 크면 스펙트럼 상에서 두 피크가 완전히 구분되어 나타나지만 그 차이가 점점 감소하여 주파수 분해능의 1.5 배 미만이면 두 피크가 겹쳐져 하나로 보이게 되고 따라서 음원을 분리할 수 없다.

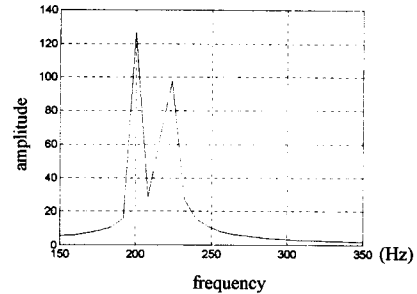
Fig. 2 는 이러한 현상을 그래프로 나타낸 것이다. 감성로봇이 분리하고자 하는 대상은, 신호 크기의 차가 크지 않은 사람의 음성이므로, 신호 크기가 비슷한 두 정현파의 피크에 대한 구분만을 다루었다. 동일한 창문함수를 사용하고, 비슷한 주파수를 갖는 두 음원이 혼합된 신호를 2048Hz 로 256 개 샘플링하여 스펙트럼을 구하였다. 두 음원의 주파수 차이가 클 때는 두 피크가 완전히 구분되어 나타난다. 그러나 주파수 차이가 감소하여 Fig. 6 (a)와 같이 각각 200Hz, 220Hz 가 되면 두 피크가 겹쳐지기 시작하고 Fig. 6 (b)와 같이 각각 200Hz, 211Hz 가 되면 피크가 하나로 되어 두 음원의 분리가 불가능해진다.

### 3. 실험

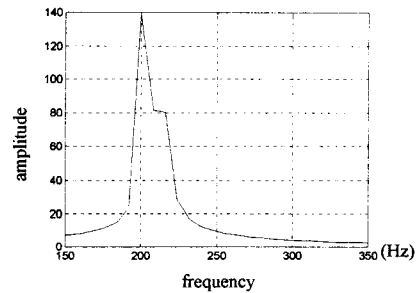
#### 3.1 실험장치

감성로봇이 있는 공간은 3 차원이기 때문에 정확한 방향측정은 3 차원적인 것이어야 한다. 그러나 대개의 경우 로봇과 인간의 행동은 2 차원 평면상에서 이루어지는 것으로 생각할 수 있다. 따라서 로봇에서 음원으로서의 방향측정은 2 차원에서의 방향각을 측정하는 것으로 한정하도록 한다.

사람과 로봇의 높이 차는 1m 이내로 가정한다.



(a) spectrum of 2 sound sources of 200Hz and 220Hz



(b) spectrum of 2 sound sources of 200Hz and 211Hz

Fig. 2 Spectrum of 2 sound sources having similar frequencies

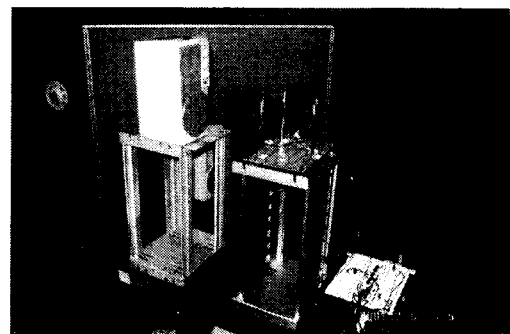


Fig. 3 Experimental setup

Fig. 3 과 Fig. 4 는 각각 실험장치의 사진과 개념도이다. 정삼각형의 꼭지점 위치에 3 개의 마이크로폰이 있고 같은 평면상에 음원이 존재한다. 사용된 장비는 BSE 무지향성 마이크로폰, National Instrument 6035E DAQ 보드, IBM PC Duron-650 등이다.

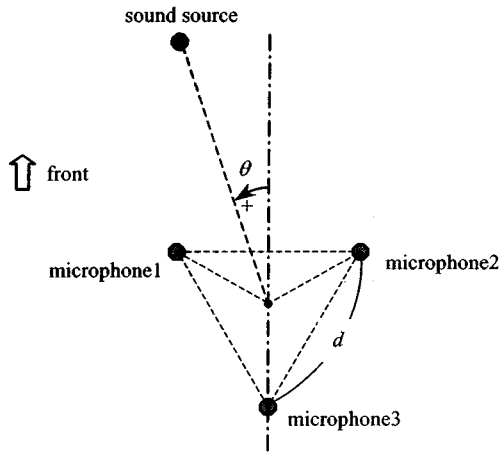


Fig. 4 Schematic diagram of experimental setup

마이크로폰 사이의 거리  $d$  는 나이퀴스트 정리 (Nyquist theorem)에 의하여 측정할 음파 파장  $\lambda$ 의 1/2 이하가 되어야 하므로 다음의 식이 성립한다.

$$d \leq \frac{\lambda}{2} = \frac{v}{2f} \quad (7)$$

음파의 속도를 340m/s, 음파의 최대 주파수를 1kHz 라 하면 마이크로폰 사이의 거리는 17cm 이하가 되어야 하며 실제 실험에서는 10cm 로 설정하였다. 또한 마이크로폰 사이의 거리에 대한 마이크로폰에서 음원까지의 거리의 비  $l/d$  는 5 가 되도록 하였다.

### 3.2 데이터 획득

일반적인 성인 남성의 경우 음성의 주파수 대역은 대략 1kHz 정도이다. 따라서 나이퀴스트 정리에 의하면 샘플링 주파수  $f_s$  는 2kHz 이상이 되어야 한다.

주파수 영역에서 피크를 찾을 때 주파수 분해능이 높을수록 정확하게 피크를 찾을 수 있다. 이산 푸리에 변환(discrete Fourier transform)에 들어가는 데이터 개수를  $N$ , 주파수 분해능을  $\Delta_f$  라고 하면 다음과 같은 관계가 성립한다.

$$\Delta_f = \frac{f_s}{N} \quad (8)$$

$N$  개의 데이터를 샘플링 하는데 걸리는 시간을  $T_b$  라고 하면  $T_b = N/f_s$  이므로 시간  $T_b$  내에 모든 데이터 처리가 이루어져야 한다. 따라서 계산 결과의 정확도에 큰 영향을 주지 않는 범위 내에서 계산량을 최소화 해야 한다.

이산 푸리에 변환에 걸리는 시간은  $N$  이 증가함에 따라 기하급수적으로 증가하게 된다. 데이터를 처리하는 데 걸리는 시간에서 푸리에 변환이 차지하는 비중이 크기 때문에  $N$  을 작게 하여 계산 시간을 줄일 수 있다.

위의 고찰에 따라서 다음과 같이 데이터 획득에 필요한 변수를 정하였다.

$$\begin{aligned} f_s &= 2048\text{Hz} \\ N &= 256\text{개} \\ T_b &= 1/8\text{초} \\ \Delta_f &= 8\text{Hz} \end{aligned} \quad (9)$$

방향측정 및 분리가 실시간적으로 이루어지기 위해서는 데이터 처리가 연속적으로 이루어져야 한다. 이를 위하여 데이터 획득 보드의 버퍼를 이용하였다. 버퍼를 반으로 나누어 한 쪽(halfbuffer1)에는 센서에서 측정된 데이터를 저장하고 다른 쪽(halfbuffer2)에서는 버퍼에 저장된 데이터를 PC 로 전송한다. halfbuffer1 에 데이터가 다 채워지면 PC 로 전송하고 halfbuffer2 에 데이터를 저장한다. 이러한 과정을 반복하면 시간을 버퍼 크기에 해당하는 일정 단위로 나누어 연속적으로 데이터를 획득할 수 있다.

### 3.3 데이터 처리

앞에서 DAQ 보드를 이용하여 PC 에서 받아들이는 데이터를 이용하여 음원의 방향측정과 분리가 이루어지는 과정은 다음과 같다. 각각의 과정을 LabWindows/CVI™ 를 이용하여 C 로 코딩하였다.

- 1) 데이터 획득
- 2) 푸리에 변환

내장된 FFT(Fast Fourier Transform) 함수를 이용하여 변환한다.

3) 피크 탐지

기울기가 양수에서 음수로 변하는 곳에서 극대값을 가지므로 이를 이용하여 피크를 찾는다. 그런데 모든 극대값이 음원에서 기인하는 것이 아니고 배경소음, 반사파, 노이즈에 의한 것도 포함되어 있으므로 일정 수준의 한계치를 두어 이보다 큰 값을 갖는 것만 피크로 인정한다. 또한 60Hz 성분은 교류전원에 의한 배경소음이므로 이 이하의 주파수를 갖는 피크는 제거한다.

4) 피크의 위상차 계산

2 개의 마이크로폰에서 측정된 신호를 주파수 영역에서 분석하면 탐지되는 피크의 주파수는 같지만 위상은 다른 값을 갖는다. 각각의 피크의 위상차를 식 (6)에 따라 계산한다.

5) 피크의 방향 계산

식 (5)에서 보듯이 음원의 방향  $\theta$ 는 피크의 주파수  $f_p$ 의 함수이다. 그런데 앞에서 주파수 분해능이 8Hz로 정해졌기 때문에  $\theta$ 의 분해능을 그 이상으로 올릴 수 없다. 따라서 보간법(interpolation)을 이용하여  $f_p$ 와  $\theta$ 의 정확도를 높이도록 한다.

6) 전후 구별

마이크로폰 2 개로 된 조합은 좌우 대칭이므로 전후 대칭 위치에 있는 음원을 같은 위치에 있는 것으로 인식하게 된다. Fig. 5 와 같이 마이크로폰 3 개를 이용하면 이 중에서 2 개를 선택하는 조합은 3 가지가 있고, 각각의 조합에서 2 개씩 음원의 방향각이 계산되므로 모두 6 개의 방향각을 얻는다. 이 방향각 중에서 일치하는 3 개의 방향각이 실제 음원의 방향각과 같다.

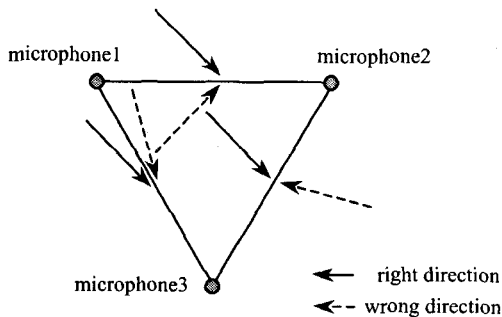


Fig. 5 6 detected directions using 3 microphones

7) 피크를 음원별로 구분

하나의 음원에서 나온 음파는 기본주파수와 배음의 관계를 갖게 된다. 앞에서 탐지된 피크들의 주파수를 비교했을 때, 배수 관계에 있는 피크들이 존재하면 이것들은 하나의 음원에서 나온 것이라고 할 수 있다.

8) 음원의 방향 계산

같은 음원에서 나온 피크라고 판단된 것들의 방향의 평균을 내어 그 음원의 방향을 계산한다.

9) 역 푸리에 변환

앞에서 계산된 푸리에 변환 결과에서 같은 음원에서 나온 피크라고 판단된 것들만 남기고 나머지는 모두 0 으로 한다. 이것을 역 푸리에 변환하면 분리된 결과를 얻을 수 있다.

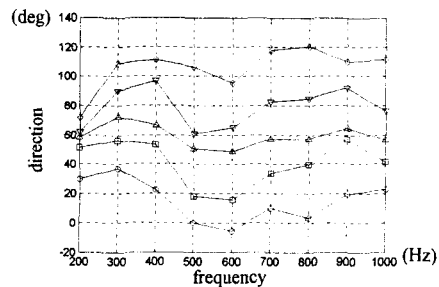
10) 결과 저장

4. 실험결과 및 토의

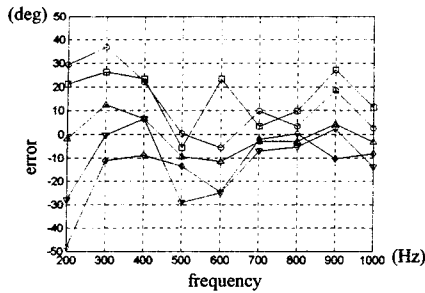
4.1 음원의 방향측정

하나의 음원에 대하여 주파수와 위치를 변화시키면서 음원의 방향을 측정하였다. 음원이 각각 0°, 30°, 60°, 90°, 120° 방향에 위치할 때 음원의 주파수를 200 에서 1000Hz 까지 100Hz 간격으로 증가시키면서 실험을 수행하였다.

Fig. 6 은 실험결과를 그래프로 나타낸 것이다. 각각의 경우에 50 번 방향값을 측정하고 이것의 평균값과 실제의 음원 방향과의 오차를 계산하였다. 전체 평균은 약 ±13°, 분산은 11°이며, 일반 환경의 반사된 음파의 영향과 수식과 관련된 오차 등의 영향으로 방향 측정에 많은 오차 값을 보여주고 있다. 오차의 크기가 ±10°보다 작은 경우는 약 56%, ±20°보다 작은 경우는 약 78%로 나타났다.



(a) average of detected direction



(b) error of detected direction compared to real direction

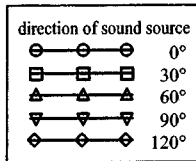


Fig. 6 Detected direction of sound sources

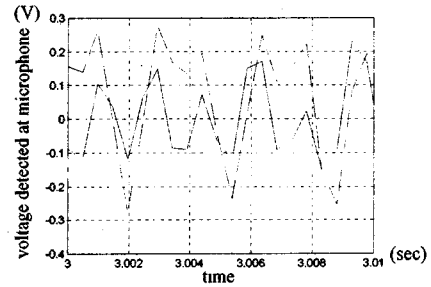
## 4.2 음원의 분리

### 4.2.1 두 음원의 분리

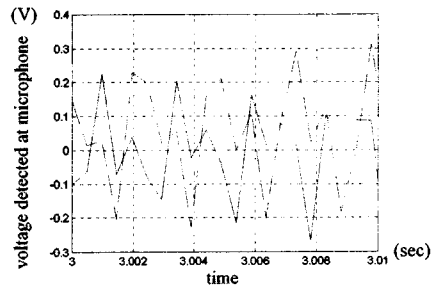
Fig. 7 은 2.2 장에서 설명한 방법으로 두 음원이 혼합된 신호로부터 두 음원을 분리한 결과이다. 음원 1 은 10°의 방향에 있으며 300Hz 와 600Hz 의 사인파를 동시에 발생시키고, 음원 2 는 -10°의 방향에 있으며 400Hz 와 800Hz 의 사인파를 동시에 발생시킨다. 마이크로폰에서 이 음파를 측정하면 두 음원이 혼합된 상태로 나타난다. 이 신호를 푸리에 변환하여 주파수 영역에서 보면 300, 400, 600, 800Hz 에서 피크가 존재한다. 이 피크의 주파수 중에서 가장 낮은 300Hz 는 음원 1 의 기본주파수가 되고 그 다음으로 낮은 400Hz 는 30 의 배음이 아니므로 음원 2 의 기본주파수이다. 나머지 600Hz 는 300Hz 의 배수이므로 음원 1 의 배음이며 800Hz 는 400Hz 의 배수이므로 음원 2 의 배음이다. 따라서 음원 1 에 속하는 피크는 300, 600Hz 이며, 음원 2 에 속하는 피크는 400, 800Hz 이다. 이렇게 분리된 피크의 스펙트럼을 역 푸리에 변환하면 시간 영역에서 분리된 음원 1, 2 의 신호를 얻을 수 있다.

분리된 결과와 원래의 신호를 비교하면 샘플링 주파수가 낮기 때문에 비교에 어려움이 있지만 전체적으로 파형은 비슷하고 위상에 차이가 있음을 알 수 있다. 실제로 음성인식에서 이용하는 것은 음성의 스펙트럼에서 피크가 나타나는 주파수와

피크의 크기에 관련된 정보이므로 음원의 분리를 수행하면서 파형이 유지되는 것은 중요하다고 할 수 있다.



(a) sound source 1



(b) sound source 2

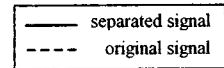


Fig. 7 Original and separated signal of sound sources

### 4.2.2 주파수가 비슷한 두 음원의 분리

Table 1 은 두 음원의 주파수가 어느 정도로 근접했을 때 분리가 불가능한가를 나타낸 것이다. 먼저 15°의 방향에 위치한 음원 1 의 주파수를 200Hz 로 고정시키고 -15°의 방향에 위치한 음원 2 의 주파수를 180~220Hz 범위에서 1Hz 씩 증가시키면서 측정되는 피크의 개수를 조사하였다. 그 결과 음원 2 의 주파수가 188~212Hz 일 때 1 개의 피크가 측정되었다. 음원 1 의 주파수가 400, 600, 800Hz 일 때에 대하여 같은 실험을 수행하였을 때 음원 2 의 주파수가 대략 음원 1 의 주파수의  $\pm 12\text{Hz}$  이내의 범위에 있으면 1 개의 피크가 나타났다. 즉 두 음원의 주파수 차이가 12Hz 이하이면 두 개의 음원이 아닌 하나의 음원으로 인식하게 되어 음원을 분리할 수 없게 된다.

2048Hz 로 256 개의 데이터를 샘플링하여 푸리에 변환하면 주파수 분해능이 8Hz 이다. 두 피크의 주파수 차이가 16Hz 미만이면 주파수 분해능의 2 배보다 작으므로 스펙트럼 상에서 일부가 겹쳐져 나타나고, 만약 주파수 차이가 12Hz 미만이면 주파수 분해능의 1.5 배보다 작으므로 두 피크가 완전히 겹쳐져서 하나의 피크로 나타난다. 즉 두 음원의 주파수 차이가 푸리에 변환의 주파수 분해능의 1.5 배보다 큰 경우에만 음원의 분리가 가능하다.

Table 1 Range of folding frequency of 2 sound sources

Frequency of sound source 1 (Hz)	Range of folding frequency of sound source 2 (Hz)
200	188~212
400	388~412
600	588~612
800	785~813

### 5. 결론

감성로봇은 감성공학의 한 응용분야로서 요즘 그 중요성이 부각되고 있다. 감성로봇이 인간의 감성 상태를 판단하기 위하여 언어와 얼굴표정, 몸짓 등을 이용한다. 일상적인 실내 공간에서 이러한 기능을 하기 위하여 음원의 방향측정과 분리를 실시간적, 연속적으로 수행할 필요가 있다.

음원의 방향측정을 위하여 두 마이크로폰에서 측정된 신호의 스펙트럼에 나타나는 피크의 위상차를 이용하며, 음원의 분리를 위하여 음성이 기본주파수와 그 배음으로 이루어짐을 이용한다.

실험장치를 구성함에 있어서 마이크로폰과 증폭회로, 데이터 획득 보드, PC 만을 이용하여 간단하면서도 저비용으로 가능하게 하였다. 또한 마이크로폰을 3 개 사용하여 360° 전 방향에 걸쳐서 음원의 방향측정이 가능하도록 하였고, 효과적이고 연속적인 데이터 획득이 가능하도록 샘플링 변수를 정하였다. 또한 획득된 데이터를 처리하여 원하는 기능을 구현하도록 C 언어로 프로그래밍하였다.

마지막으로 실험을 통하여 실시간적이고 연속적으로 다양한 음원의 방향과 주파수에 대하여 방향을 측정하였으며, 두 음원이 혼합된 신호로부터

음원을 분리하고 두 음원의 주파수가 어느 정도 근접할 때까지 음원 분리가 가능한가에 대하여 알아보았다.

### 후 기

본 연구는 BK21 사업의 지원으로 수행되었습니다.

### 참고문헌

1. Lee, K. H., "Human stability and sensible technology," J. of KSPE, Vol. 18, No. 2, pp. 9-13, 2001.
2. Bregman, A. S., "Auditory scene analysis: the perceptual organization of sound," MIT press, Cambridge, MA, 1990.
3. Bregman, A. S., "Auditory scene analysis: hearing in complex environment" in McAdams and Bigand (eds.) "Thinking in sounds," Oxford University press, New York, 1993.
4. Nakatani, T. and Okuno, H. G., "Harmonic sound stream segregation using localization and its application to speech stream segregation," Speech Communication 27, pp. 209-222, 1999.
5. Huang, J., et al., "Echo avoidance in a computational model of the precedence effect," Speech Communication 27, pp. 223-233, 1999.
6. Unoki, M. and Akagi, M., "A method of signal extraction from noisy signal based on auditory scene analysis," Speech Communication 27, pp. 261-279, 1999.
7. Nakadai, K. et al., "Humanoid active audition system," IEEE Humanoids, 2000.