

# 능동 시청각 시스템에서 하모닉 정보를 이용한 음원의 위치추정

\*황민, 임성길, 이현수

경희대학교 전자계산공학과

e-mail : hm8106@hotmail.com skan72@khu.ac.kr, leehs@khu.ac.kr

## Sound Localization using Harmonic Structure in Active Perception System

\*Min Hwang, Sung-Kil Lim, Hyon-Soo Lee

Dept. of Computer Engineering, Kyung Hee University

### Abstract

In this paper, we propose a new sound localization algorithm for an active perception system. In an active perception system, an acquired sound is mixed with the sound of motors. So a sound localization algorithm for an active perception system requires a robustness for the noise and a computational efficiency. The proposed localization algorithm can achieve robustness and efficiency to use only sub-band channels that are contained harmonic structure of the target speech.

### I. 서론

능동 시청각 시스템이란 시각 및 청각적인 인식과정과 정확한 시/청각적 데이터를 관측하기 위한 움직임을 가진 시스템을 의미한다[1][2]. 능동 시청각 시스템은 주변의 환경을 인식/판단하고 그에 따른 적절한 동작을 해야 하는 로봇에서 필수적인 기술이다.

능동 인식 시스템에서 음원의 위치를 판단하기 위해서 해결해야 하는 문제는 움직임에 따른 잡음이 발생한다는 것과, 움직임에 따라 음원의 상대적 위치가 변경되기 때문에 실시간에 가까운 음원의 위치 판단이 필요하다는 것[3]이다.

본 연구에서는 능동 시/청각 시스템에서 잡음에 강하고 빠르게 음원의 위치를 판단하기 위하여, 대상 음성의 피치를 추정한 후에, 추정된 피치의 배음 구조에 해당하는 채널들에 대해서만 시간지연을 계산하여 음원의 위치를 추적하는 방법을 제안한다.

### II. 피치 추정 알고리즘

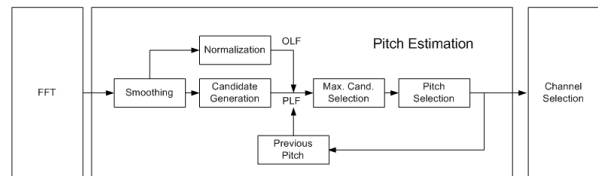


그림 1. 제안하는 피치 추정 방법의 구조도

잡음 환경에서 피치를 추정하기 위해 제안하는 방법의 구조도는 그림 1과 같다. 먼저 주파수 영역으로 분석된 스펙트럼을 평탄화(smoothing)하여 피치가 될 후보 채널을 선택하고, 평탄화된 스펙트럼을 지역적인 최대점에 대하여 평균화하여 OLF(Overtone Likelihood Function)를 정의한다. 치가 될 후보 채널은 피치가 존재 가능한 주파수 영역의 지역적인 피크들로 결정하며, OLF의 정의는 다음과 같다.

$$OLF(x) = \begin{cases} \frac{(y - LV_y)}{(LP_y - LV_y)} & \text{if } x \in [LV_x, LP_x] \\ \frac{(y - RV_y)}{(LP_y - RV_y)} & \text{if } x \in [LP_x, RV_x] \end{cases}$$

이때,  $LP$ 는 지역적인 피크를 의미하며,  $LV$ 와  $RV$ 는 각각 지역 피크의 좌/우측에 위치하는 골을 나타낸다. 아래첨자  $x$ 는 주파수 채널이며  $y$ 는 그에 해당하는 스펙트럼의 크기이다.

각 피치 후보가 피치인가를 결정하기 위하여 PLF(Pitch Likelihood Function)을 정의한다. PLF는 후보 피치의 정수배에 해당하는 OLF의 평균으로 정의하며, 피치의 연속성과 더불어 피치 문제를 해결하기 위하여 가중치를 적용하여 사용한다.

### III. 하모닉 정보를 이용한 음원 위치추적

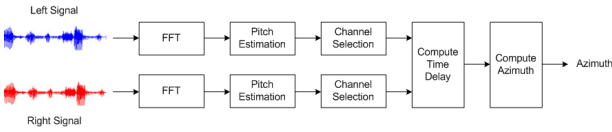
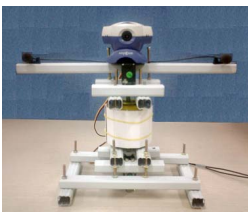


그림 2. 제안하는 음원 위치 추정 방법의 구조도

본 논문에서는 잡음에 강인한 위치 추적을 위하여 음향 신호를 채널별로 나누어 처리하였으며, 음성 신호에 해당하는 채널을 선택하기 위해 음성의 피치를 사용한다. 제안하는 음원 위치 추정 방법의 구조는 그림 2와 같다.

두개의 마이크로부터 입력된 음향 신호는 주파수 분석을 통하여 피치 추정 부분으로 입력된다. 피치가 추정되면 피치의 정수배에 해당하는 채널을 음성에 해당하는 채널로 선택한 후에, 양쪽 신호에서 선택된 채널들에 대하여 시간지연을 계산한다. 시간지연을 계산하는 방법은 시간영역에서 Cross Correlation을 계산하는 방법과 주파수 영역에서 위상 차이를 이용한 방법이 있는데, 본 논문에서는 주파수 영역에서 피치를 추정하였으므로, 위상차를 이용한 시간 지연을 계산한다. 각 채널 별로 시간 지연이 계산되면, 잡음의 혼재 여부에 따라 시간 지연의 차이가 발생하게 된다. 전체적인 시간 지연을 계산하기 위하여 각 채널의 시간 지연에 가우시안 분포를 적용하고, 그 합의 최대값이 되는 시간을 전체 시간지연으로 결정한다. 시간 지연이 계산되면, 음속과 시간지연의 관계를 고려하여 방위각을 계산한다.

### IV. 능동 청각 시스템 적용 및 실험



장비	
Microphone	AKG C417 x2
Camera	MPC-C10 x1
Motor	RC Servo Moter x2
PC	P4 1.9GHz 512MB RAM

그림 3. 적용된 능동 청각 시스템 및 하드웨어 사양

제안하는 위치 추정 알고리즘은 본 연구실에서 시험 제작한 그림3의 능동 시청각 시스템에 적용하여 실험하였다. 그림 3의 능동 시청각 시스템은 두개의 마이크로부터 음성을 입력받으며, 1개의 CCD 카메라를 이용하여 영상을 입력 받는다. 움직임은 두개의 서보 모터를 이용하여 360도 회전운동을 하도록 제작하였다.

제안하는 알고리즘이 잡음에 강인하게 동작한다는

것을 보이기 위하여 잡음이 없는 상태에서의 위치추정(그림4(a)), 모터가 동작하는 환경에서의 cross-correlation 계산에 의한 위치추정(그림4(b)), 제안한 알고리즘에 의한 위치 추정(그림4(c))에 대하여 실험하였다. 실험 결과 제안한 방법을 적용한 경우에 안정적인 위치 추정 결과를 확인 할 수 있었다.

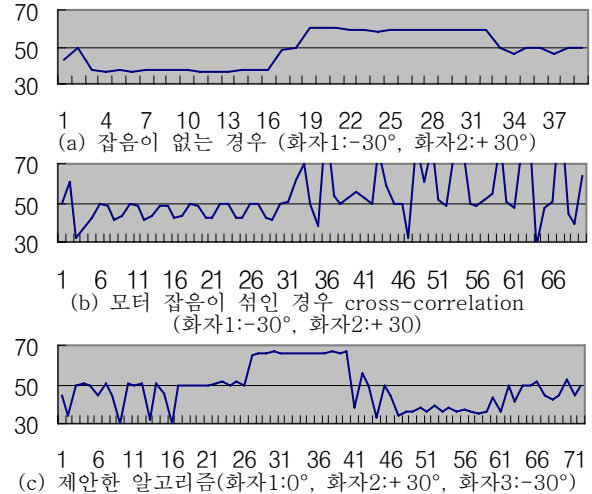


그림4. 위치추정 실험 결과

### V. 결론 및 향후 연구 방향

본 논문에서는 능동 인식 시스템에 적용하기 위한 음원 위치 추정 알고리즘을 제안하였다. 먼저 대상 음성의 피치를 추정하고, 추정된 피치의 배수에 해당하는 채널만을 선택하여 시간 지연을 계산함으로써 잡음에 강하고, 모든 채널에 대하여 시간 지연을 계산하는 방법보다 계산의 속도를 줄일 수 있었다.

향후에는 한 명이상의 화자가 동시 발화하는 상황에서 각 화자의 위치를 판단하고, 지능적으로 동작하는 능동 시청각 시스템을 구현하는 연구가 필요하다.

#### 참고문헌

[1] Y. Aloimonos, I. Weiss, A. Bandyopadhyay, "Active vision", International Journal of Computer Vision, 1987

[2] K. Nakadai, T. Matsui, H.G. Okuno, H. Kitano, "Active audition system and humanoid exterior design", Proc. of IEEE International Conference on Intelligent Robots and Systems, pp. 1453-1461, 2000

[3] K. Nakadai, H.G. Okuno, H. Kitano, "Auditory Fovea Based Speech Separation and Its Application to Human-Robot Dialog System", Proc. of 7th International Conference on Spoken Language Processing, pp. 1817-1820, 2002