한국소음진동공학회 2009년 추계학술대회논문집, pp.706~707

음성 인식 방법을 활용한 화자의 위치를 추정하는 로봇 청각 시스템 구현

Implementation of Robot Auditory System for Speaker Localization using Speech Recognition Technique

권병호† · 조현* · 이상문* · 박영진* · 박윤식*

Byoungho Kwon, Hyun Jo, Sangmoon Lee, Youngjin Park and Youn-sik Park

1. 서 론

로봇에 대한 연구가 진행되면서, 사람과 로봇의 원활한 상호작용을 위해서 시각 정보뿐만 아니라 청각 정보를 이용하려는 연구들이 시도되고 있다. 청각 정보에는 화자가 어떤 말을 하고 있는지를 인지하는 음성 인식(speech recognition) 부분과 화자의 위치를 인지하는 음원 위치 추정(sound source localization) 부분을 포함한다. 이들 각 기능은 그 자체만으로도 로봇에게 상황 인지 능력을 부여하지만, 이 둘이 융합될 경우에는 시너지 효과로 인해 상황 인지 능력이 배가 되는 효과를 얻을 수 있다. 이와 같은 로봇 청각시스템의 개발 중에 가장 대표적인 사례는 사람과 로봇의 커뮤니케이션을 위해 개발중인 일본 교토 대학의 SIG 로봇이다. 이 로봇은 다수의 사람의 위치를 추정하면서 사람의 음성도 인식할 수 있어 사람과 원활한 상호작용이 가능하게 하였다.[1]

본 논문에서는 개별적으로 연구된 공간좌표로 사상된 GCC 함수를 이용한 음원위치 추정 방법과, 인공귀를 이용한 음원위치 추정 방법, 그리고 고립어에 대한 음성/화자 인식 방법들을 통합한 통합형 로봇 청각 시스템을 제안하고, 이를 실제 로봇 플랫폼에 적용하여 성능 평가를 하고자 한다.

2. 로봇 청각 시스템

2.1절 음성/화자 인식 모듈

통합형 로봇 청각 시스템에 적용될 음성/화자 인식 모듈은 사용자의 이름이나 단발적인 명령과 같은 고립단어(Isolated word)에 대한 음성/화자 인식이 가능한 방법이다. 음성인식(Speech recognition)을 위해서는 MFCC (Mel-Frequency Cepstral Coefficient)가 이용되었으며, 화자인식(Speaker Identification)을 위해서는 성대의 첫 번째 고유 진동수를 나타내는 Pitch 값을 이용하였다. 음성/화자 인식을 위해서 특정 화자의 참조 모델(Reference model)을 이 두 파라미터를 이용하여 수립하였고, 측정된 음성과 참조 모델과의 비교를 위해서는 동적 시간 정합 알고리즘(Dynamic Time Warping, DTW)을 이용하였다. 인식 성능을 향상시키기 위해서 각 파라미터들의 확률 모델을 이용하는 방법을 적용하였다[2].

2.2절 음원 위치 추정 모듈

음원 위치 추정 모듈에는 공간좌표로 사상된 GCC 함수를 이용한 음원위치 추정 방법과 인공귀를 이용한 방법이 적용되었다.

(1) 수평각 추정

화자의 위치 중 수평각을 추정하기 위해서는 로봇 플랫폼에 설치되어 있는 세 개의 마이크로폰을 이용하여 공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법을 적용하였다[3]. 이 방법은 저가의 지능형 로봇의 청각 신호 처리용 SoC 에 적용되기 위해 개발된 방법으로 적은 계산량과, 다양한 로봇 플랫폼에 쉽게 적용할 수 있는 장점이 있다. 또한 배경잡음이 존재하는 환경에서도 강인하게 음원의 위치를 추정할 수 있음이 검증되었다.

(2) 고도각 추정

화자의 위치 중 고도각을 추정하기 위해서는 두 개의 마이크로폰이 설치된 한 쌍의 인공귀가 적용되었다[4]. 이는 다양한 로봇에 쉽게 설치할 수 있도록 모듈화 되어있으며, 로봇 플랫폼에 의해 발생하는 반사파의 영향을 최소화할 수 있는 신호처리 기법이 적용되었다.

이상의 두 가지 모듈을 하나의 청각 시스템으로 통합하였으며, 통합된 청각 시스템의 순서도는 Fig. 1 과 같다. 먼저 특정 크기 이상의 신호가 측정되면, 측정된 신호의 음성/비음성 구분을 한다. 이는 로봇이 사람의 음성에만 반응할 수 있도록 해 불필요한 계산을 줄이기 위함이며, 사람의 Pitch 특성을 이용하여 쉽게 구현하였다. 음성 신호가 인지되면 음성/화자 인식 과정을 수행하고 동시에 화자의 위치를 추정하도록 하였다. 음성/화자 인식과 음원 위치 추

† 교신저자; KAIST
E-mail : bhkwon@kaist.ac.kr
Tel : (042)350-3076, Fax : (042) 350-8220
* KAIST

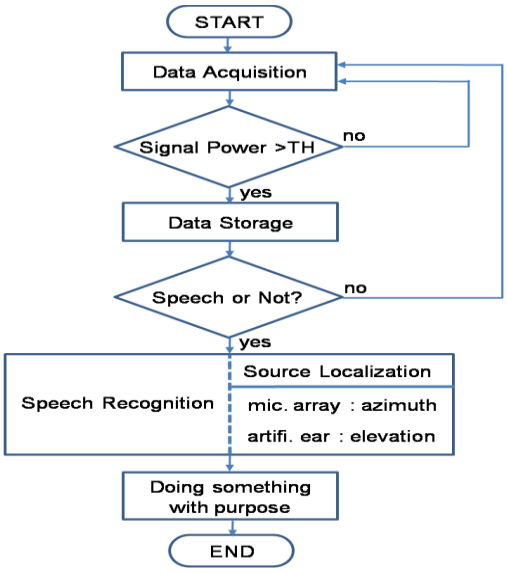

Fig. 1 Flow chart of the unified robot auditory system

정이 끝나면, 이들 정보를 이용하여 적절한 행동을 할 수 있도록 하였다.

3. 성능 평가

앞서 설명한 통합 청각 시스템은 Fig. 2 와 같이 로봇 플랫폼에 장착되었으며, NI-DAQ 보드와 PC 상의 MATLAB™ 프로그램 상에서 구현되었다. 이는 실제 환경에서 세 명의 화자가 “ 꾀돌아” , “ 이리와” , “ 청소해” , “ 안녕” , “ 반가워” 이상의 다섯 단어에 대해서 수평각은 0°~180° 사이에서 30° 간격으로, 고도각은 사람이 서 있는 경우와 앉아 있는 경우에 대해서 실험하였으며, 실환 환경은 Fig. 3 과 같다. 세 명중 한 명의 음성만을 인식하게 한 경우 음성/화자 인식률은 특정인의 특정 음성을 정확하게 추정한 경우가 91%, 특정인의 다른 말들을 인식한 경우가 8%였다. 또한 음원 위치 추정 성능은 수평각의 경우 ±10° 오차 범위로 100% 인식했으며, 인공귀를 이용한 위-아래 추정도 100% 정확성을 보여주었다.

(a) **(b)**

Fig. 2 The unified robot auditory system; (a) microphone arrays for azimuth angle estimation, (b) artificial ear for elevation angle estimation of speaker.

Fig. 3 The unified robot auditory system; (a) microphone arrays for azimuth angle estimation, (b) artificial ear for elevation angle estimation of speaker.

4. 결 론

본 논문에서는 기존에 개발된 음성/화자 인식 방법과 마이크로폰 어레이와 인공귀를 이용한 음원 위치 추정방법을 통합한 로봇 청각 시스템을 구현하고 그 성능을 검증하였다. 실제 환경에서의 실험에서 음성/화자 인식률은 90% 이상이었으며, 두 모듈을 이용한 음원 위치 추정 모듈은 특정 오차범위로 100% 인식률을 보여주었다. 본 논문에서 구현된 로봇 청각 시스템을 이용해 사람과 로봇 사이의 더욱 원활한 상호작용이 가능해 질 것으로 기대된다.

후 기

이 논문은 지식경제부 및 정보통신연구진흥원의 IT 핵심기술개발사업[2008-F-044-01, 전자파, 음향 및 건물 환경을 개선하는 지능형 건설 IT 융합 신기술 개발], 한국과학재단을 통해 교육과학기술부의 국가지정연구실 사업(R0A-2005-000-10112-0) 으로 수행하였음.

참고 문헌

[1] Kazuhiro Nakadai, Tino Lourens, Hiroshi G. Okuno, and Hiroaki Kitano, “ Active Audition for Humanoid," Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000), 832-839, 2000

[2] 조현, 김경호, 박영진, “ 로봇 시스템에의 적용을 위한 음성 및 화자인식 알고리즘,” 한국소음진동공학회 2007 년도 춘계학술대회 논문집, 2007

[3] 권병호, 박영진, 박윤식, “ 공간좌표로 사상된 GCC 함수를 이용한 음원 위치 추정 방법,” 한국소음진동공학회논문집, 제 19 권, 제 4 호, pp. 355~362, 2009

[4] Sangmoon Lee, Sungmok Hwang, Youngjin Park, and Youn-sik Park, “ Sound source localization in Median Plane using Artificial Ear,” International Conference on Control, Automation and Systems, 246-250, 2008