

# 오디오 기반 인간로봇 상호작용 기술

Audio-Based Human-Robot Interaction Technology

## 지능형 로봇 특집

|                 |                   |
|-----------------|-------------------|
| 곽근창 (K.C. Kwak) | 인간로봇상호작용연구팀 선임연구원 |
| 김혜진 (H.J. Kim)  | 인간로봇상호작용연구팀 연구원   |
| 배경숙 (K.S. Bae)  | 인간로봇상호작용연구팀 연구원   |
| 윤호섭 (H.S. Yoon) | 인간로봇상호작용연구팀 책임연구원 |

## 목 차

- .....
- I. 서론
  - II. 기술별 현황
  - III. 오디오 기반 상호작용 기술발전
  - IV. 결론

인간로봇 상호작용 기술(human-robot interaction)은 다양한 의사소통 채널인 로봇 카메라, 마이크로폰, 기타 센서를 통해 인지 및 정서적으로 상호작용할 수 있도록 로봇 시스템 및 상호작용 환경을 디자인하고 구현 및 평가하는 지능형 서비스 로봇의 핵심 기술이다. 본 고에서는 오디오 기반 인간로봇 상호작용 기술 중에서 음원 추적(sound localization)과 화자인식(speaker recognition) 기술의 국내외 기술동향을 살펴보고 최근 ETRI 지능형로봇연구단에서 상용화를 추진중인 시청각 기반 음원 추적(audio-visual sound localization)과 문장독립 화자인식(text-independent speaker recognition) 기술들을 다룬다. 또한 이들 기술들을 가정환경에서 효과적으로 사용하기 위해 음성인식, 얼굴검출, 얼굴인식 등을 결합한 시나리오에 대해서 살펴본다.

## I. 서론

인간로봇 상호작용(HRI) 기술은 로봇이 가지는 자율성, 상호작용의 양방향성, 상호작용 또는 제어 수준의 다양성 측면 등에서 인간컴퓨터 상호작용(HCI)과는 근본적인 차이점을 가지고 있다[1]. 이러한 인간로봇 상호작용 기술은 다양한 의사소통 채널을 통해 인간과 로봇이 자연스럽게 상호작용할 수 있는 지능형 서비스 로봇의 핵심적인 기술이다. 국내 연구동향을 살펴보면, 21세기 프론티어 과제인 인간기능 생활지원 지능로봇 개발사업 내에 인간로봇 상호작용 과제를 수행중이며, 우수연구센터 사업으로 인간로봇 복지로봇시스템 연구센터에서 관련 분야를 연구중에 있다. 또한, 삼성전자, LG전자, 한울로보틱스, 유진로보틱스 등에서 청소로봇과 같은 가정용 서비스 로봇의 개발과 함께 일부 업체에서는 인간로봇 상호작용 기술의 상용화에 관심을 보이고 있다. 한편, 국외 연구동향을 살펴보면, 미국에서는 DARPA와 NSF 주도 하에 대형 국책 프로젝트로서 인간로봇 상호작용에 대한 연구를 진행하고 있으며, 일본에서는 Sony, 혼다, NEC, 마쓰시타, 미쯔비시 등에서 인간로봇 상호작용 프로젝트와 관련된 연구를 진행하고 있다.

일반적으로 인간로봇 상호작용 기술은 비디오 기반 상호작용 기술, 오디오 기반 상호작용 기술, 기타 상호작용 통합 및 응용 기술로 분류할 수 있다. 먼저 비디오 기반 상호작용 기술은 로봇카메라로부터 얻어진 영상정보를 근거로 하여 얼굴검출(face detection), 얼굴인식(face recognition), 얼굴검증(face verification), 얼굴 표정인식(facial expression recognition), 인간 추출(human extraction), 사용자 인식(user identification), 특정인 추적(human tracking), 호출자 식별(caller identification), 제스처인식(gesture recognition), 포스처인식(posture recognition), 행동인식(behavior recognition) 등을 수행한다. 한편, 오디오 기반 상호작용 기술은 로봇에 부착된 마이크로폰으로부터 얻어진 음성정보에 근거해서 음성인식(speech recognition), 화

자인식(speaker recognition), 음원 추적(sound localization), 음원 분리(sound separation) 등을 포함하고 있다. 마지막으로 기타 상호작용 통합 및 응용 기술로는 멀티모달 사용자인식(multimodal user identification), 인지적 상호작용 통합(cognitive interaction), PDA 혹은 휴대폰을 통한 HRI 기술 통합 및 응용과 관련되어진다.

본 고에서는 인간로봇 상호작용 기술 가운데 오디오 기반 상호작용 기술의 연구동향 및 기술발전에 초점을 두고 있다. 음성신호의 인식 기술은 크게 음성인식과 화자인식으로 분류되며 음성인식 시스템은 특정한 화자에 대해서만 인식하는 화자종속(speaker-dependent) 시스템과 화자에 상관없이 인식하는 화자독립(speaker-independent) 시스템으로 나뉘어진다. 또한, 시스템이 말하는 화자가 누구인지를 알게 하는 화자인식 시스템은 크게 화자식별 시스템(speaker identification)과 화자검증(speaker verification)으로 분류되는데 화자식별 시스템은 등록된 화자들 중에서 가장 유사한 화자를 찾아가 그 화자가 누구인지를 알아내는 것이고, 화자검증은 입력된 음성이 등록된 화자의 음성인지를 판별하는 시스템이다. 이러한 화자인식 기술은 무인경비시스템의 보안장치, 전화망 서비스에서의 암호 키, 인터넷 보안장치를 비롯한 모든 종류의 보안시스템에서 사용되고 있다[2]. 한편, 음원 추적 기술은 주의집중을 위해 사용자와의 상호작용을 시작하는 가장 첫 번째 단계로 호출에 응답하도록 하는 로봇의 가장 기본적인 기능을 구현하는 필수 기술이다. 음원 추적에서 비디오 및 오디오 정보는 중요하며, 먼저 오디오 정보에 의해 음원 추적이 수행되고 비디오정보에 의해 추적오차를 줄여 나감으로써 좀 더 빠른 수행시간과 정확성을 기대할 수 있다. 음원 분리 기술은 로봇주위에서 여러 명이 동시에 대화를 나눌 때 각 마이크로폰으로부터 얻어진 혼합된 음성신호를 근거로 해서 각 음원을 분리하고, 그 분리된 신호로부터 음성인식 및 화자인식을 수행하여 누가 어떤 위치에서 어떠한 말을 하고 있는지를 파악할 수 있는 중요한 기술이다.

본 고의 구성은 다음과 같다. II장에서는 음원 추적과 화자인식의 기술동향을 살펴보고, III장에서는 ETRI 지능형로봇연구단에서 개발된 시청각 기반 음원 추적과 문장독립 화자인식 기술과 함께 시나리오를 다룬다. 마지막으로 IV장에서 결론으로 맺는다.

## II. 기술별 현황

### 1. 음원 추적 기술

지능형 로봇은 음원 추적 시스템을 이용하여 공공장소나 가정에서 주위상황을 인지하고 판단하여 도움을 필요로 하는 근처로 이동하여 적절한 대응조치를 취할 수 있도록 해준다. 이와 관련해서 먼저 국내 연구동향을 살펴보면, Choi[3]는 오디오와 비디오 정보를 융합한 시청각 기반의 음원 추적 기술을 수행하였다. 먼저 세 개의 마이크로폰으로부터 음원 추적을 수행한 후에 추적오차를 보정하기 위해 얼굴검출을 이용하여 호출자에게 다가간다. 또한, Park[4]은 휴머노이드 로봇에서 2차원이 아닌 3차원적으로 음원을 추적하기 위해 세 개의 마이크로폰을 이용하였으며, 상하 혹은 좌우를 추적하고 마찬가지로 얼굴검출을 통해 추적오차를 보정하였다. 그러나, 이들 방법들은 음원 추적 후 카메라의 영상 안에 여러 명이 존재하는 경우에는 얼굴검출만으로는 정확히 호출한 사용자에게 다가가기 어렵다. 또한, 근거리(2m 이내)에서 음원 추적을 수행하였지만 실제환경에서는 원거리(5m 이내)까지 고려해야 한다. 게다가 호출하는 음성을 듣고 바로 음원 추적을 수행하는 것보다는 그 음성이 호출음성인지 아닌지를 인식하고 호출자가 또한 누구인지를 알고 음원 추적을 수행하는 것이 좀 더 인간친화적인 방법이 된다.

한편 국외의 연구동향을 살펴보면, 시청각 기반 음원 추적[5]뿐만 아니라 에코와 반향에 강인한 알고리즘에 관한 연구가 많이 진행되고 있으며, 음원 분리 기술과 함께 활발한 연구가 진행되고 있다. Huang[6]은 에코와 반향에 대처하기 위해 인간 청각시스템의 선행 효과를 기반으로 한 모델기반 음원

추적(model-based sound localization) 시스템을 연구하였다. 세 개의 마이크를 이용하고 시간차 방법(ITD)을 사용하여 기존의 빔 형성 방법, MUSIC, 최대 우도 방법과 비교하여 적은 계산시간과 정확한 추적을 하였다. Valin[7],[8]은 steered beamformer를 이용하여 8개의 마이크와 3차원 공간에서 강인한 음원 추적을 수행하였으며 5m에서 1.4도의 추적오차를 보였다. Nakadai[9]는 두 쌍의 마이크를 SIG 휴머노이드 로봇에 음원 추적을 적용하였다. 한 쌍은 머리의 양 측면에서 설치되어 있고, 다른 쌍은 잡음제거를 위해 모터 잡음과 같은 내부 음원을 기록하기 위해 머리 안쪽에 위치되어져 있다. 이 로봇은 인간과 같이 3차원공간에서 음원 소스를 위치하기 위해 스테레오 방식의 음원 추적방법을 사용하였다. 한편, Yamamoto[10]는 ASIMO에서 동시에 발생된 혼합된 음성을 가지고 음원 추적, 음원 분리 및 음성인식을 결합하여 잡음과 에코환경에서 강인하면서도 정확한 성능을 보여주었다. 이들 기술도 앞서 언급한 바와 같이 실생활에서 적용하기 위해서는 음성인식, 화자인식, 얼굴검출, 얼굴인식 기술들과 결합되어 좀 더 인간 친화적이어야 하며 여러 명이 존재하는 복잡한 환경에서 좋은 추적성능을 나타내야 한다. 이와 관련된 진보된 음원 추적 기술은 III장에서 기술되어질 것이다.

### 2. 화자인식 기술

화자식별 및 검증 기술은 로봇과 함께 생활하는 가족 구성원의 목소리를 듣고, 누구인가를 알아낼 수 있으며, 또한 발화자가 가족 구성원인지 아닌지의 여부도 알 수 있는 오디오 기반 인간-로봇상호작용 기술이다. 현재 화자인식 기술은 생체인식과 보안시스템에 관련된 전화망서비스나 인터넷보안에서 연구되어지고 있다. 주로 문장중속이나 문장제시형 화자인식방법이 보안시스템과 연관되어져 사용되고 있으며, 문장독립 화자인식의 경우에는 전화망과 같이 오랜 시간 동안 음성을 취득할 수 있는 환경에서 상용화되고 있다. 그러나, 지능형 로봇에서 음성인식 기술은 활발히 연구되고 있는 반면, 화자인식 기

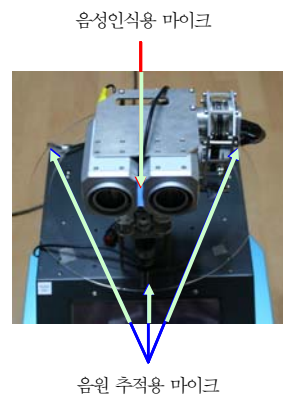
술은 아직 초보적인 수준에 불과하다. 로봇환경에서 화자인식 기술은 생체인식에서의 화자인식 기술과는 달리 여러 가지 고려되어야 할 문제점들이 있다. 첫번째로 로봇환경에서는 실시간으로 화자를 등록하는 온라인 화자등록 기능이 있어야 한다. 많은 문장을 통해 각 화자별 모델을 구축할 수도 있지만, 사용자는 불편함을 느끼기 때문에 적은 문장으로도 화자를 등록시킬 수 있도록 하는 방법이 사용되어야 한다. 이를 위해서는 일반화된 배경화자모델(UBM)을 생성하여 간편하게 온라인 화자 등록을 하는 방법이 주로 시도되고 있다. 또한, 아동을 위해서는 게임이나 노래와 같은 오락 및 교육 콘텐츠를 통한 등록방법도 고려해야 한다. 두번째로 로봇환경에서는 문장에 관계없이 화자를 식별해야 한다. 로봇과 대화 혹은 명령을 하는 동안 로봇은 발성자가 누구인지 어떠한 명령을 하는지를 파악해야 할 필요가 있다. 즉 음성인식 기술과 함께 사용되면 사용자에 맞는 맞춤형 서비스를 수행할 수가 있다. 세번째로 로봇환경에서는 거리와 잡음에 대해 강인한 특성을 가져야 한다. 화자인식은 근거리뿐만 아니라 원거리에서도 화자를 인식해야 하며, 잡음환경에 대해서도 강인한 알고리즘이 필요하다. 아직까지는 TV 잡음이나 라디오 잡음과 같이 사람의 음성이 혼합된 신호에 대해서는 연구가 좀더 필요한 실정이다. 네번째로 시간에 따라 변하는 특성을 보상할 수 있는 적응방법이 필요하다. 시나리오에서 일정주기마다 기존 화자모델을 적응방법을 통해 갱신해 주어야 한다. 이렇듯이 화자인식 기술은 비디오 정보를 이용하는 얼굴인식 기술과 함께 사용자인식 기술에서 중요한 요소가 된다. 음성인식 기술에 대한 전반적인 기술동향은 [11]에서 자세히 살펴볼 수 있다.

### Ⅲ. 오디오 기반 상호작용 기술발전

#### 1. 시청각 기반 음원 추적 기술

시청각 기반 음원 추적 기술은 II장에서 설명한 바와 같이 국내뿐만 아니라 국외에서도 많은 연구가

진행되고 있다. 그렇지만, 실제 가정과 같은 환경에서는 적용하기 어려운 문제점들이 있다. 기존의 연구에서는 박수소리나 일반적인 음성을 이용하거나, 아주 조용한 환경에서 로봇의 호칭을 사용하였다. 가정환경에서는 다양한 잡음이 존재하기 때문에 호출하지 않았음에도 불구하고 음원 추적을 수행하여 로봇이 회전하는 경우가 있다. 이러한 문제점을 해결하기 위해 음성인식을 함께 적용하여 로봇 호칭을 통해 호출한 음성에 대해서만 음원 추적을 수행하는 방법이 유용하다. 또한, 호출음성을 듣고 막연하게 회전하는 것보다 호출자가 누구인지 알고서 회전하는 것이 좀더 인간친화적인 방법이 될 수 있다. 이를 위해서는 화자인식 기술을 통해 호출자를 식별하는 기술이 필요하다. 게다가 기존의 방법에서는 음원 추적을 수행한 후에 단지 얼굴검출 기술만으로 호출자에게 다가가기 위해서는 많은 제약이 따른다. 특히 여러 명이 로봇 카메라의 범위 안에 보여진다면 특정 호출자에게 다가가기 어려운 문제점이 있다. 기존 연구에서는 얼굴영상이 가장 큰 사람에게 다가가거나 카메라 중심에 위치한 사람에게 다가가곤 하였다[3]. 이러한 문제점을 해결하기 위해 화자인식으로부터 얻어진 호출자의 정보를 가지고 얼굴 추적 및 다중 얼굴인식을 수행하여 호출자를 찾도록 하는 것이 타당하다. (그림 1)은 ETRI 지능형로봇연구단에서 개발된 연구용 로봇 플랫폼인 Wever-R2이다. 이 로봇은 서버/클라이언트 구조를 갖춘 네트워크 기반 지능형 로봇이다. 그림에서 보는 바와 같이 3

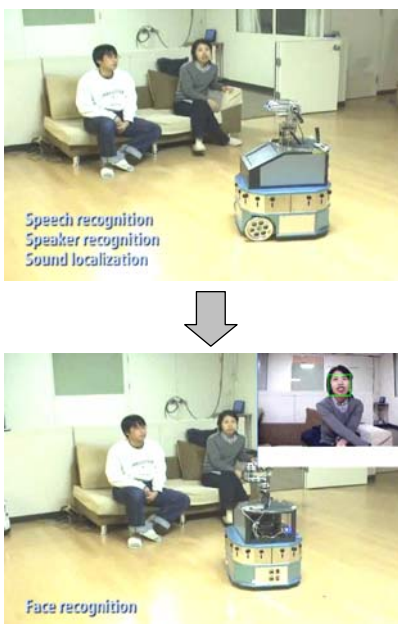


(그림 1) 연구용 로봇 플랫폼(Wever-R2)

개의 음원 추적용 마이크로폰과 1개의 음성인식용 마이크로폰을 가지고 있다. 여기서 음원 추적은 로봇 자체에서 수행되고, 음성인식은 네트워크를 통해 전송된 음성신호로부터 서버에서 음성인식을 수행하고 있다. 서버에서 호출음성으로 인식된 경우에만 음원 추적 결과를 사용해서 회전하도록 되어 있다. 현재 개발된 음원 추적 기술로는 마이크로폰간의 시간 지연 측정을 위해 상호상관방법을 사용하는 데 있어서 마이크로폰의 이득 특성이 다르기 때문에 신호간의 상호상관성을 잘못 측정하는 문제점이 있다. 이러한 문제점을 마이크로폰의 이득을 보정한 마이크로폰의 이득 특성에 강인한 음원 추적을 개발하였다. 또한, 음성신호의 현재 샘플과 이전 샘플과의 차로 새로운 신호를 만듦으로써 좀 더 정확한 시연지연을 측정할 수 있는 계차 기반의 음원 추적 기술도 함께 개발하였다. 그 밖의 GCC-PHAT 기반 음원 추적, 신호 보간법에 의한 음원 추적 기술 등이 개발되었다.

개발된 시청각 기반 음원 추적 기술을 요약하면 다음과 같은 특징을 가지고 있다.

- 로봇호칭 기반 음원 추적(웨버, 애니봇 등)



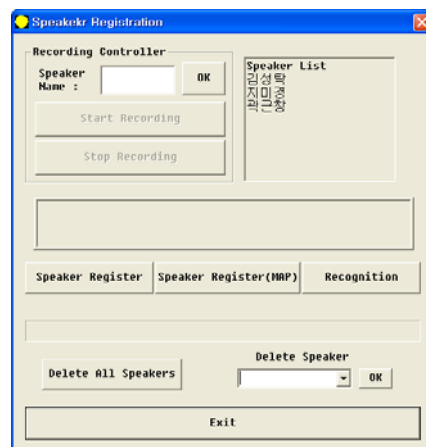
(그림 2) 시청각 기반 음원 추적 시나리오

- 음성인식을 통한 호출여부 인식
- 화자인식을 이용한 호출자 인식
- 얼굴검출을 통한 호출 후보자들의 얼굴검출
- 다중 얼굴인식을 통한 호출 후보자들의 얼굴인식
- 원거리 음원 추적(5m 이내)

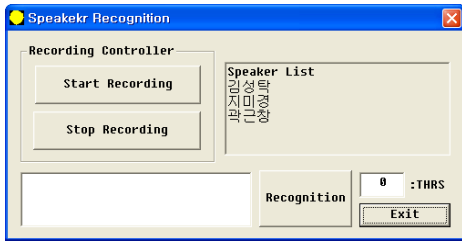
(그림 2)는 시청각 기반 음원 추적 기술의 시나리오를 보여주고 있다. 이 시나리오에서 음성인식, 음원 추적, 화자인식, 얼굴검출 및 인식 기술 등이 효과적으로 결합되어 좀 더 인간친화적인 시청각 기반 음원 추적 기술을 보여준다.

## 2. 문장독립 화자인식 기술

문장독립 화자인식 기술은 II장에서 언급한 것처럼 생체인식에서의 화자인식 기술과는 여러 가지 측면에서 다른 특징을 가지고 있다. 따라서 이러한 문제를 해결하기 위해 온라인 화자등록, 문장독립 화자식별 및 검증, 적응기법에 초점을 맞추어서 연구를 진행하였다. 먼저 (그림 1)의 Wever-R2에서 화자인식을 수행하면 음성인식과 마찬가지로 음성신호를 네트워크를 통해 전송한 다음 서버에서 화자인식을 수행한다. (그림 3)은 온라인 등록 및 화자모델 구축을 위한 GUI 환경을 보여주고 있다. 그림에서 보는 바와 같이 화자의 이름을 기입한 후 보여주는 문장들을 읽음으로써 화자를 실시간으로 등록할 수



(그림 3) 온라인 등록 및 화자모델 구축을 위한 GUI 환경



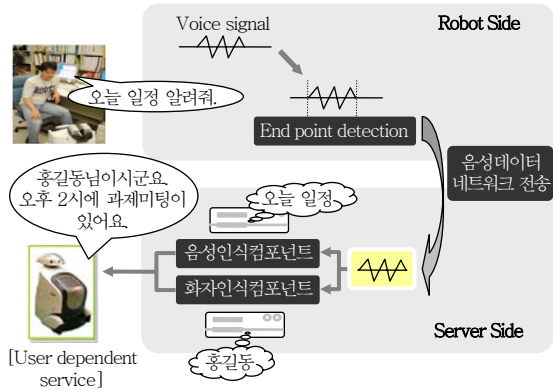
(그림 4) 화자식별 및 검증을 위한 GUI 환경

가 있다. 일반화된 배경화자모델을 구축하여 몇 문장을 발성함으로써 간편하게 화자를 등록할 수 있다 [12]. (그림 4)는 화자식별 및 검증단계로서 임의의 입력문장에 대해서 특징추출을 하고, 이전 등록된 화자의 모델과의 로그 우도 값을 계산하여 최대 값을 갖는 화자 모델을 선택함으로써 화자식별을 수행한다. 또한, 화자검증은 임계 값을 설정하고, 임계 값보다 작으면 거부하고, 임계 값보다 큰 결과를 얻으면 화자를 받아들인다. 여기서 임계 값은 등록자와 침입자의 로그 우도 값의 차이의 분포에 의해 설정된다. 실제환경에서 거리 변화, 문장길이 변화, 환경 변화에 따라 임계 값이 민감하게 작용하므로 설정하기 어려운 문제점이 있다. 현재 화자식별방법은 MFCC-GMM 방법과 MFCC-GMM-PCA 방법을 적용하고 있다. 화자검증방법으로는 화자혼잡도(SCR) 방법과 GMM-UBM 방법이 개발되었다.

개발된 문장독립 화자인식 기술을 요약하면 다음과 같은 특징을 가지고 있다.

- 간편한 온라인 화자등록(UBM을 통한 적용)
- 문장독립 화자식별
- 화자검증을 통한 가족/비가족 구별
- 시간변화에 따른 적응기능
- 음성인식과 화자인식을 결합한 맞춤형 서비스
- 로봇의 모든 방향에서 인식(다채널 보드)
- 원거리 화자인식(5m 이내)

(그림 5)와 (그림 6)은 각각 화자인식과 음성인식의 결합을 위한 개념도 및 시나리오를 보여주고 있다. 그림에서 보는 바와 같이 화자인식과 음성인식을 동시에 수행함으로써 로봇은 누가 무엇을 말하고 있는지를 알기 때문에 상세일정, 전화 메시지, 선호



(그림 5) 화자인식과 음성인식 결합을 위한 개념도



(그림 6) 화자인식과 음성인식 결합에 의한 시나리오

하는 TV 프로그램 및 음악, 동영상 이메일과 같이 그 화자에 맞는 맞춤형 서비스를 수행할 뿐만 아니라, 기존의 음성인식만 수행하는 것에 비해 좀 더 인간친화적인 방법이 될 수 있다.

## IV. 결론

본 고에서는 오디오 기반 인간로봇 상호작용 기술 가운데에서 음원 추적과 화자인식 기술의 연구동향을 살펴보고, 현재 ETRI 지능형로봇연구단에서 개발된 시청각 기반 음원 추적과 문장독립 화자인식 기술 및 시나리오를 다루었다. 현재 음원 추적의 연구방향은 서버/클라이언트 구조에서 호칭을 포함한 간단한 문장을 통해 시청각 기반 음원 추적을 수행하고, 다른 HRI 컴포넌트들과 통합하여 적용할 예정이며, 화자인식의 경우에는 네트워크와 로봇환경에 맞는 새로운 화자식별 및 검증방법뿐만 아니라 스코

어 정규화에 관한 연구를 진행할 예정이다. 최근에는 대기업 및 로봇관련 업체들이 인간의 일상생활 지원을 위한 고품질 서비스를 지원하는 개인서비스용 로봇을 비롯하여, 외부환경의 변화나 작업변경을 인식하고 스스로 상황을 판단하여 자율적으로 동작하거나 인간과 상호작용을 하는 지능형 로봇에 대한 연구개발 및 실용화에 대한 관심이 고조되어 있어 인간-로봇 간의 효과적인 상호작용 기술의 발전이 급속히 이루어질 것으로 전망된다.

## 약어 정리

|      |                               |
|------|-------------------------------|
| GCC  | Generalized Cross-Correlation |
| GMM  | Gaussian Mixture Model        |
| HCI  | Human-Computer Interaction    |
| HRI  | Human-Robot Interaction       |
| ITD  | Interaural Time Difference    |
| PCA  | Principal Component Analysis  |
| PHAT | Phase Transform               |
| SCR  | Speaker Confusion Rate        |
| UBM  | Universal Background Model    |

## 참고 문헌

[1] 특허청, 인간-로봇 상호작용기술(HRI), 2005.  
 [2] 이건상, 양성일, 권영현, 음성인식, 한양대학교 출판부, 2001.  
 [3] Jong-Suk Choi, Munsang Kim, and Hyun-Don Kim, "Probabilistic Speaker Localization in Noisy Environments by Audio-Visual Integration," *In Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS06)*, Beijing, Oct. 2006, pp.4704-4709.  
 [4] 박종옥, 나승유, 김진영, "휴머노이드 로봇을 위한 시청각 정보기반 음원 정위 시스템 구현," *음성과학논문지*, 제 11권, 제 4호, 2004, pp.29-42.  
 [5] Isao Hara, Futoshi Asano, Hideki Asoh, Jun Ogata,

and Naoyuki Ichimura, "Robust Speech Interface Based on Audio and Video Information Fusion for Humanoid HRP-2," *In Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS04)*, Sendai, Sep. 2004, pp.2404-2410.  
 [6] Jie Huang, Tadawute Supaongprapa, Ikutaka Terakura, Fuming Wang, Noboru Ohnishi, and Noboru Sugie, "A Model-based Sound Localization System and Its Application to Robot Navigation," *Robotics and Autonomous Systems*, Vol.27, No.4, 1999, pp.199-209.  
 [7] Jean-Marc Valin, Francois Michaud, Jean Rouat, and Dominic Letourneau, "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot," *In Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS03)*, Las Vegas, Oct. 2003, pp.1228-1233.  
 [8] Jean-Marc Valin, Francois Michaud, Brahim Hadjoui, and Jean Rouat, "Localization of Simultaneous Moving Sound Sources for Mobile Robot Using a Frequency-domain Steered Beamformer Approach," *In Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA04)*, New Orleans, Apr. 2004, pp.1033-1038.  
 [9] K. Nakadai, H.G. Okuno, and H. Kitano, "Real-time Sound Source Localization and Separation for Robot Audition," *In Proc. IEEE Int'l Conf. on Spoken Language Proc.*, 2002, pp.193-196.  
 [10] S. Yamamoto, Kazuhiro Nakadai, Mikio Nakano, Hiroshi Tsujino, Jean-Marc Valin, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno, "Real-time Robot Audition System That Recognizes Simultaneous Speech in the Real World," *In Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS 06)*, Beijing, Oct. 2006, pp.5333-5338.  
 [11] 이윤근, 박준, 김상훈, "음성 인터페이스기술," *전자통신 동향분석*, 제 20권, 제 5호, 2005. 10., pp.1-15.  
 [12] 지미경, 김성탁, 김희린, "지능형 서비스 로봇을 위한 문맥독립 화자인식 시스템," *로봇공학회 논문지*, 제1권, 제 2호, 2006, pp.151-157.