

CRE: 인간-로봇 상호작용을 위한 인지기반 행위제어 소프트웨어†

성균관대학교 이석한 · 백승민 · 박연출

1. 서 론

가정환경에서 인간과 공존하며 다양한 서비스를 제공하는 로봇의 실현으로 인간 삶의 질은 획기적으로 향상될 것이다. 그런데 인간과 로봇의 공존은 상호작용의 문제를 야기한다. 이는 조난현장의 구조 로봇을 원격으로 조작하는 것에서, 뇌파로 휠체어를 조작하는 인터페이스에 이르기까지 광범위한 영역에서 고려되는 문제이지만, 이 글에서는 가정이나 사무실과 같은 일상적인 생활공간에서 서비스를 제공하는 서비스 로봇과 관련한 인간-로봇 상호작용(Human-Robot Interaction, 이하 HRI) 문제에 초점을 둔다.

그림 1은 인간-로봇 상호작용 기술과 관련된 거시적 로드맵을 보여준다[1]. 서비스 로봇은 숙련된 작업에 대한 보조, 위험한 일이나 일상의 허드렛일에 대한 대체, 오락이나 교육, 인간과 로봇의 협동작업, 비서업무, 그리고 생활의 동반자에 이르기까지 다양한 수준의 응용 분야로 확대 될 것이다[1-3]. 단순한 작업에 대한 보조는 마치 가전제품을 조작하듯이, 리모콘, 핸드

폰, 웹 등의 도구를 통해 이루어질 수 있지만, 응용의 범위가 확대될수록, 인간과 인간의 상호작용에서 이루어지는 것과 같이, 음성이나 제스처, 표정 등과 같은 의사소통 수단에 대한 의존도와 요구가 점차 높아질 것이다. 향후에 인간 삶의 동반자로서의 로봇은 어린아이에서부터 노인에 이르기 까지 인지능력, 교육수준, 연령, 성별, 문화 등 모든 특성을 지닌 사용자들과 의사소통이 가능해져야 한다. 이는 단순한 명령에 대한 이해를 넘어서, 정서나 의도 그리고 사회적 맥락에 대한 이해까지도 요구된다[4, 5].

현재의 HRI 관련 연구들은 인간의 명령과 행위를 이해하는 기본적인 능력을 로봇에게 부여해 주는 음성 인식, 얼굴 표정 해석, 제스처 인식과 같은 요소기술에 집중하고 있다. 그러나 이들 요소기술들은 아직 많은 과제가 남아 있는 불완전한 기술들인 반면에, 가정이나 사무실과 같은 일상 환경은 초보적인 인지능력을 가지고 있는 로봇에게는 너무나도 다루기 어려운 복잡한 환경이다. 따라서 이러한 간격을 보완할 수 있는 연구가 시급하다. 그러므로 인간-로봇 상호 작용의 효과적인 구현은 음원추적, 얼굴인식, 얼굴검출, 표정인식, 제

† 본 논문은 정보통신부 지원 "Ubiquitous Robotic Companion(URC)" 과제의 수행결과로서 관계자 여러분께 감사의 글을 드립니다.

Technology	HRI: Verbal Pictorial Gesture Mental & Professional Emotional & Social Where- Face Speech Body Behavior & body Who- Face (frontal) Speaker Face (3D) Gait Dialogue What- Command Gesture Context understanding Why- Facial expression Behavior Platform: Mobile Robot Platform Humanoid Platform Robot Behavior Behavioral and Deliberate Hybrid Cognitive Engine Dependability/Safety: Soft Robot Self-Repair Self-Adaptation								From Navigation to Manipulation Higher Level of Autonomy Natural HRI Emphasis on Dependability and Robustness Power Supply Integration with Intelligent Space and Network Modularization and Standardization Cost < Price < Value
	Applications	Robot Companion Robot Secretary Human-Robot Cooperative Task Execution Personal and Public Entertainment and Tutoring Replacing Human Labor for Daily Chore and Unsafe Work Assisting Human for Skilled Work							
Related		PC Power 3G-80G 10G-200G 30G-600G 100G-1200G Broadband 10Mbps 20Mbps 100Mbps 300Mbps Batteries 200 WH/kg 300 WH/kg 400 WH/kg 500WH/kg Transducers Stereo Camera 3D Imaging Sensors Tactile Arrays Smart Actuators							
	1995 1998 2001 2004 2007 2010 2013 2016 2019								

그림 1 인간-로봇 상호작용 거시적 로드맵 (Macro Roadmap)

스처인식 등과 같은 각 요소기술의 개발만이 아니라, 이러한 요소 기술들을 통합, 보완하여 주어진 과제를 수행하는 기술의 개발을 동시에 추구해야 할 것이다.

이 논문에서는 서비스 로봇에서의 HRI 연구의 최신 동향을 살펴보고, HRI의 주요 이슈들과 인지적 관점에서의 HRI 실현을 위한 주요한 연구 과제들을 개관한다. 그리고 본론에서 HRI를 실현하기 위한 통합적 접근 방법으로 인지적 로봇 엔진을 제안하고 이를 호출자 식별 문제에 적용한 실험결과를 보인다.

2. 서비스 로봇에서의 HRI 기술 동향

본 장에서는 이제까지 개발된 로봇들과 현재 진행 중인 서비스 로봇 프로젝트를 기준으로 인간-로봇 상호작용 기술 동향을 소개한다. 그림 2는 다양한 서비스 로봇을 보여준다.

서비스 로봇 로드맵 상에서 가장 먼저 출현할 것으로 예상되거나 이미 상품화된 예로는 청소, 잔디깎이와 같은 단순한 작업보조 로봇이 있다. 이탈리아의 Ambrogio(6)와 미국 IRobot사의 Roomba(7)는 각각 자동으로 잔디를 손질하거나, 집안을 청소하기 위한 목적으로 설계된 서비스 로봇들이다. 청소로봇의 경우는 국내의 가전 및 로봇 회사들에 의해 경쟁적으로 제품화되고 있다. 이들 로봇은 버튼을 누르거나 숫자를 디스플레이하는 정도의 매우 단순한 상호작용만을 요구한다.

경비, 안내와 같은 원격조작이나 간단한 음성 및 화상 처리를 기반으로 한 서비스 로봇도 이미 많은 곳에서 개발을 마쳤다. 전시회장에서 관람객들을 안내하는 로봇으로는 CMU의 Minerva(8)와 KIST에 의해 만들어져 현대 박물관에서 안내를 수행하는 Jiny가 대표적인 예이다. 이러한 유형의 로봇들은 주로 대화형 디스플레이와 음성 명령어를 이용하여 관람객들과의 상호작용을 수행한다.

CMU의 Minerva는 입술, 눈썹 그리고 두 개의 카메라가 달린 얼굴을 가지고 있으며, 얼굴검출 기술을 이용하여 자신을 주시하는 사람을 확인할 수 있고, 기본 안내 멘트를 말할 뿐만 아니라 이를 변형해서 표현하는 것이 가능하다.

NEC의 PaPeRo(9, 10), 유진로보틱스의 아이로비(11)와 같은 로봇은 가정 내에서 간단한 음성 및 화상 인식을 통해 정보제공, 교육 등을 수행하는 상품화된 로봇의 예이다. PaPeRo는 동반자형의 개인용 로봇으로서, 음원 위치 추적, 음성 인식 및 합성, 얼굴 검출, 추적, 인식 기능을 내장하고 있다. PaPeRo는 얼굴을 검출하기 위해 우선 배경 영상과 움직이는 얼굴 영역의 차이를 이용하여, 머리 영역의 후보를 선정하고, 실

제 얼굴 여부를 검증하기 위해 스테레오 카메라로부터 얻어진 거리 정보를 이용한 후에 템플릿 매칭을 수행한다. 음원의 위치 추적을 위해서는 복수개의 마이크로폰에 도달한 음성신호의 상대적인 위상 차이가 이용된다.

외형이 사람이나 동물과 유사한 로봇들이 나타나면서 생명체의 감성까지 모사하는 연구들도 주목되기 시작했다. 감성의 인식과 표현은 더 자연스러운 상호작용의 필수 요소이기 때문이다. MIT의 KISMET은 감성엔진을 가지고 있고, 얼굴표정을 통해 감성 표현도 가능한 3세 정도의 아기의 반응을 모델로 만들어진 로봇이다(12).

KISMET 연구의 목적은 로봇에게 사회적인 상호작용 능력을 심어주는 데 있다.

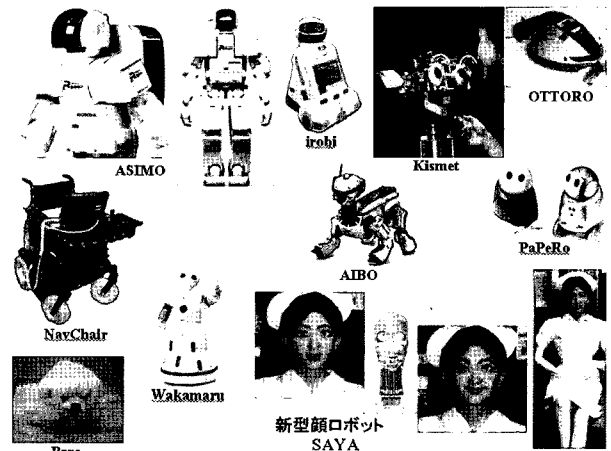


그림 2 다양한 서비스 로봇들

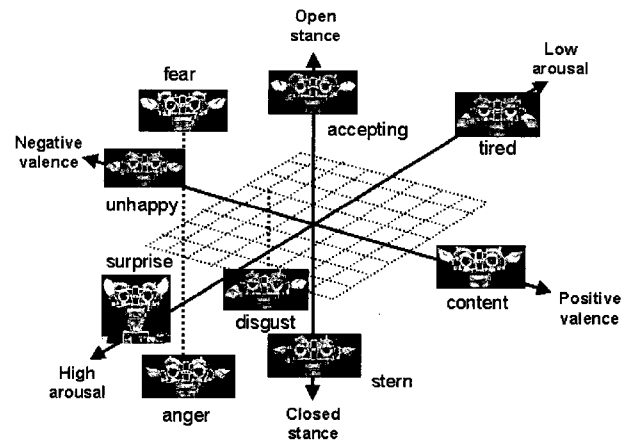


그림 3 KISMET의 감성모델

사람을 포함한 환경 인식과 감성 반응을 보이는 행동의 구현을 위해 심리학적 생물학적 연구들을 이용했으며, 그림 3과 같은 감성모델을 가지고 있다.

와세다의 WE 시리즈(13)도 KISMET과 유사하게 심리학적 연구결과를 바탕으로 내부적인 감성모델과 감성변화를 구현하였으며 감성에 따른 사람의 얼굴표정을 모사했다. 도쿄 이과대의 SAYA는 인공 피부와

근육을 모사한 구동장치를 이용하여 사람의 표정을 더 현실적으로 표현하는 로봇이다[14, 15]. 로봇에 감성을 부여하는 연구들은 로봇이 사람과 공존하면서 자연스러운 상호작용을 할 수 있도록 하고 사람에게 하나의 생명체로 인식되도록 한다.

서비스 로봇이 상호작용에서 실제 생명체나 동반자처럼 느껴지기 위해서는 감성뿐만 아니라 개성도 중요한 요소이다. 로봇의 개성은 미리 주어지기도 하지만, 사람과의 상호작용을 통해 변해가는 것이 자연스럽게 느껴진다. 실제 소니의 애완 강아지로봇 AIBO[16]는 주인과의 상호작용에 따라 성격이 변하도록 프로그램되어 있다. 일본 AIST에서 개발된 PARO는 이동성은 없으나 사람의 쓰다듬거나 만지는 행동에 반응하도록 만들어져, 사람의 정서적, 심리적 치료에 쓰인다[17].

로봇의 외형이 사람과 유사할수록 친근감을 줄 수 있기 때문에, KAIST의 AMI[18], 독일의 HERMES[19], Mitsubishi의 Wakamaru[20] 등과 같이 상반신이 휴머노이드 형태인 로봇도 많이 개발되고 있다. 기본적인 음성 인식, 합성뿐만 아니라 각종 센서와 스테레오 카메라를 이용한 환경인식과 사람처럼 두 팔을 이용한 제스처 표현이 가능하다. Wakamaru는 약 10,000개의 단어인식과 음성합성과 제스처를 이용한 표현이 가능하고, 전방향 카메라로 자신의 위치와 움직이는 사람의 위치를 확인하여 템플릿 매칭으로 얼굴을 인식하는 기능을 가지고 있다.

혼다의 Asimo[21], 소니의 Qrio[22]와 같이 이동 방식도 사람과 같은 이족 보행이 가능하며, 계단, 문턱도 이동할 수 있는 사람과의 공존에 한발 더 다가선 로봇도 개발되고 있다. Asimo는 50여개의 호출, 인사, 질문과 관련된 문장을 이해하고 반응할 수 있으며, 30여개의 신체 동작과 관련한 명령을 수행할 수 있다. 또한 얼굴인식, 움직이는 사람을 따라가는 기능, 손으로 지시하는 장소를 파악하는 등의 간단한 제스처 인식 기능을 갖고 있다.

이러한 휴머노이드 로봇들에게는 현재까지 개발된 음성, 비전, 촉각인식 등의 요소기술들을 응용한 멀티모달 상호작용 혹은 통합적 인지 기능이 필수적이다.

눈동자의 움직임을 감지하여 휠체어 로봇이 이동하거나, 뇌파와 신경신호를 감지하고 전기적으로 변환하여 로봇팔 등을 움직이는 등의 특수한 인간-로봇 상호작용 기술이 장애인 혹은 노인의 일상 생활을 돕기 위한 로봇들에게 적용된다. Michigan 대학의 Navbelt와 NavChair는 보행시에 장애물을 탐지하고, 경보음을 울릴 수 있도록 시각장애인을 보조하는 기능을 제공한다[23]. 최근에는 신체장애자에게 조작 능력을 제

공하기 위하여 로봇 머니플레이터를 제공하는 시도가 이루어지고 있으며, 더 나아가서 인체의 신경을 팔이나 다리의 손상된 부분을 대체하기 위한 로봇 제어 시스템과 연결하는 연구가 진행되고 있다. 그 외에 군사용, 의료용 로봇의 원격조작 및 힘 피드백이 되는 햅틱 인터페이스 개발도 여러 연구기관에서 이루어지고 있다.

3. 서비스 로봇의 행위제어를 위한 인지적 접근 방법

3.1 EU의 Cogniron Project

유럽연합에서 진행 중인 Cogniron 프로젝트의 HRI 부분 세부 연구과제 구성은 다음과 같다.

3.1.1 멀티 모달 담화(Multi-modal dialogues, UniBi)

멀티 모달 담화의 주된 연구 주제는 컴퓨터 비전, 음성언어 처리, 인간-컴퓨터 상호작용 등이며, 이중 컴퓨터 비전에 관한 연구는 VAMPIRE(Visual Active Memory Processes and Interactive REtrieval)와 물체 인식을 위한 다양한 모듈의 통합(Integrating Different Processing Paths for Object Recognition) 등으로 다시 나뉜다. VAMPIRE는 Active Memory를 사용하고 사람과의 상호작용을 통해 물체를 인식하는 기술이며, 사람이 바라보고 있는 지점을 인지하고 이를 물체 인식에 이용하는 경우가 일례이다(그림 4).

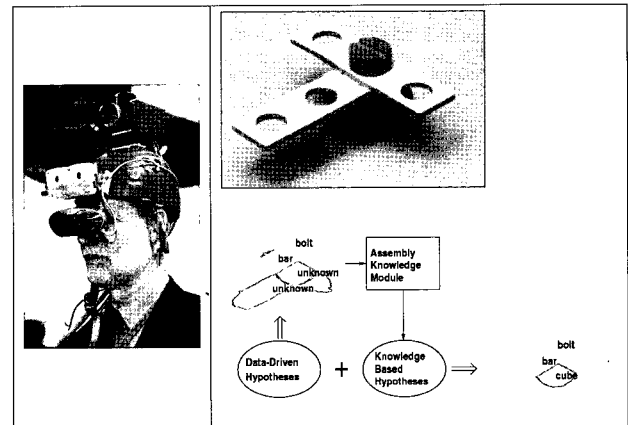


그림 4 VAMPIRE를 이용한 물체 인식(왼쪽), 물체 인식을 위한 각각의 처리 경로 결합(오른쪽)

다음으로 로봇 시스템을 위한 음성 언어 처리 분야는 Speech 시스템의 개발을 통해 영상 센서와의 융합을 시도하는 연구로서 통합을 위한 멀티모달 시스템의 구성도는 그림 5와 같다.

마지막으로 인간-컴퓨터 상호작용(Human Computer Interaction: HCI)에 대한 연구는 인간 검출(Human detecting), 제스처 검출 및 인식의 두 가지 핵심기술

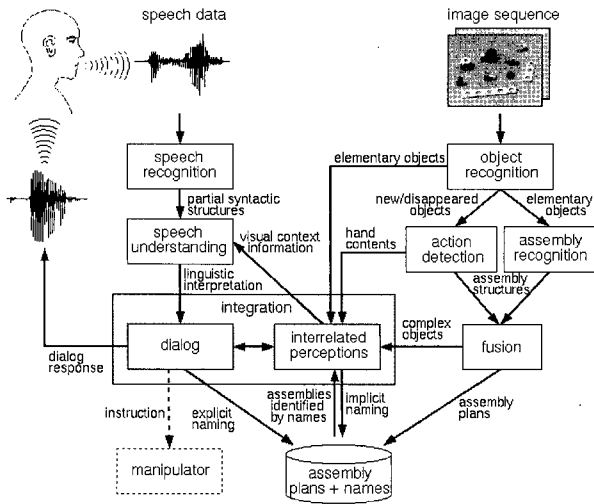


그림 5 멀티모달 시스템 구성도

의 개발로 나뉘어 진행된다. 이중 인간 검출은 레이저 스캐너를 이용하여 그림 6과 같이 대상 분할, 분류와 그룹화를 통해 다리를 찾아낸 후 사람과 대화를 시작할 수 있도록 하는 기술이며, 제스처 검출(Gesture detecting)은 사용자가 지시하는 손의 형태를 파악하여 지시하는 방향을 바라보는 행동에 관한 연구이다. 구체적인 과정은 피부 색상을 찾고 이를 칼만 필터를 사용하여 추적하는 순서로 진행되며 이때 행동을 인식하기 위한 알고리즘은 Viola와 Jones의 물체 인식 알고리즘을 사용한다. 또한 움직이면서 생성한 궤적이 어떤 움직임(예를 들어 책을 보는 모습, 음료수를 먹는 모습 등) 인지를 인지하는 방법으로는 Black과 Jepson의 조건부 확률밀도함수 전파 알고리즘(Conditional Density Propagation)을 사용한다.

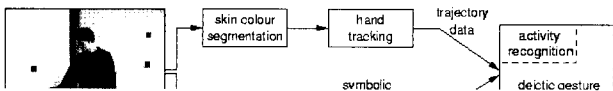
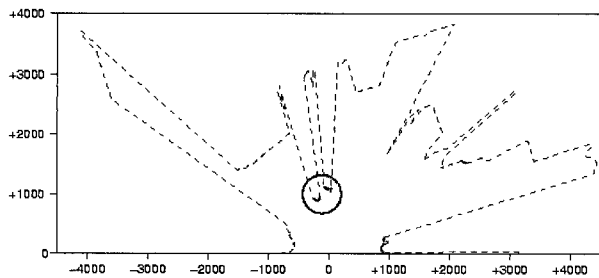


그림 6: 인간 검출(위) 및 제스처 인식(아래)

3.1.2 인간-로봇 상호작용(Human-Robot-Interaction)

인지 개념을 이용해 사람과 대화하며 상호 교류할 수 있는 로봇에 대한 연구가 수행되고 있다.

이 연구를 위해 보통 사람이 사용할 수 있도록 제작

된 높이 141cm의 BIRON[25]은 칼라 pan/tilt 카메라를 가장 위에 탑재하였고 터치스크린과 스테레오 마이크를 이용하여 사용자로부터 명령을 입력 받는다. 전방에 부착된 스피커를 이용하여 사용자와 대화하며 레이저 센서와 초음파 센서를 사용하여 물체인식 및 장애물 회피 동작을 한다.

제어 구조(Control Architecture)는 전형적인 3계층 혼합 제어 구조(3 layer Hybrid control architecture)이다. 가장 하위의 반사 계층(Reactive layer)을 사용하여 실시간성을 확보하였고 숙고계층(Deliberative Layer)은 복잡한 임무 수행 시 계획하고 이해하는 역할을 담당한다(그림 7).

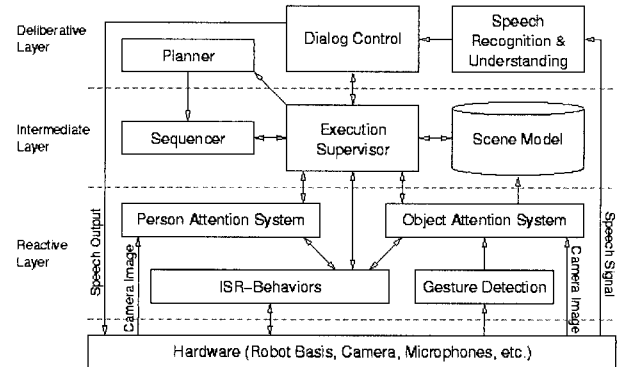
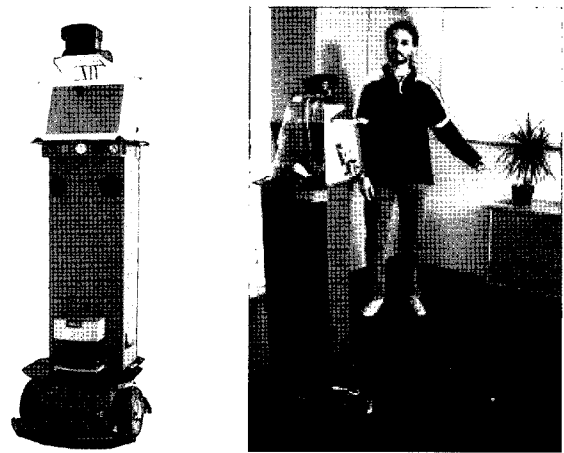


그림 7 BIRON(위), Cogniron의 3 Layer Architecture (아래)

3.2 MIT의 Cog

MIT CSAIL에서는 지능형 휴머노이드 로봇인 COG 및 감성지능 로봇 Kismet 등을 개발하여 인지/학습, 추론 등 각 요소 기술들과 이를 통합하는 지능통합 구조를 연구하고 있다[26].

Cog는 고차원적인 인지 기능을 시뮬레이션 하는 것이 아니라, 관찰, 모방 등을 통한 학습으로 새로운 기능을 습득하는 것이 가능한 통합 구조에 초점을 둔다.

현재 Cog 연구는 로봇이 사람, 환경과의 상호작용

을 사람처럼 자연스럽게 할 수 있는 지능의 구현과 이들이 사회적 기술을 학습을 통해 발전시키듯이 로봇도 발전 가능한 사회적 기술의 구현을 진행하고 있다.

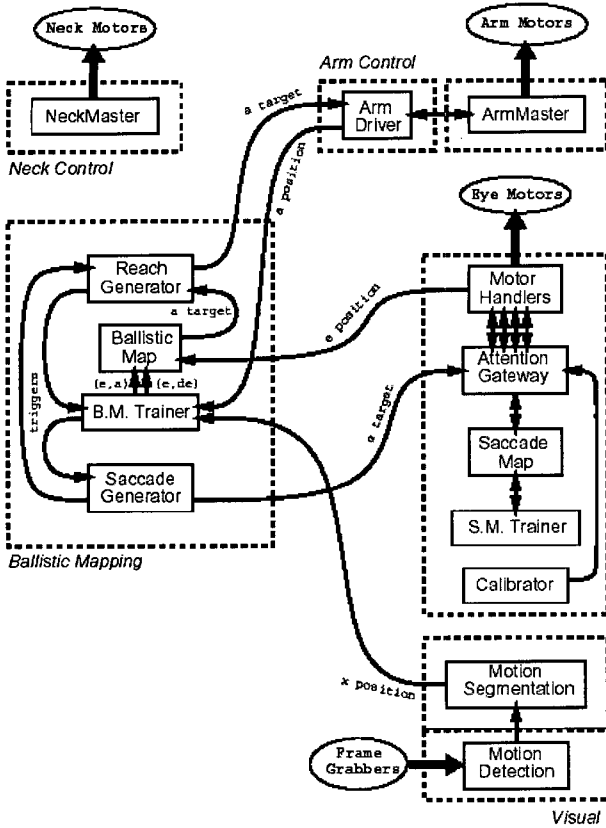


그림 8 Cog의 시스템 구조 연결 다이어그램

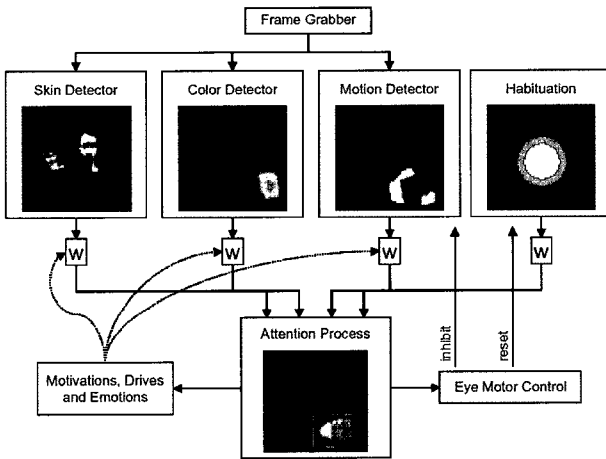


그림 9 Cog 로봇의 주의집중 과정을 나타내는 개념도. 살색 영역 탐지, 모션 탐지, 두드러진 색상 탐지 등의 저수준 특징 탐지 기능들이, 동기와 결합되어 눈과, 목의 운동과 연동된다.

3.3 기존 연구의 문제점

국내에서는 일반적으로 널리 알려진 반사계층(Reac-

tive Layer), 중간계층(Intermediate Layer), 숙고계층(Deliberative Layer)을 갖는 3계층구조의 로봇 제어 구조를 채택하고 있으며 지능형 서비스 로봇을 위한 독자적인 인지적 통합 프레임워크 연구는 미진한 상태이다.

EU의 Cogniron에서는 각 요소기술의 지능화와 멀티-모달 접근방식을 추구하고 있으나 로봇의 능동적인 행위를 고려하고 있지 못한 한계가 있다. 그리고 MIT의 Cog 프로젝트는 행위기반 로봇 제어구조를 통한 반사적인 행위(reactive behavior)의 구현에는 용이하나 실제 서비스와 같은 주어진 태스크를 수행하기에는 문제가 있다. 실제 지능형 서비스 로봇은 불완전한 환경인식 기능만을 갖추고 있더라도 능동적인 행위를 이용하여 임무 완수를 위한 정보의 불확실성을 줄여가는 인지적 능력이 필수적이라는 측면에서, 기존의 연구들은 로봇의 자연스러운 감지/추론/행동을 통합하는 프레임워크에는 적합하지 못하다. 따라서 이를 보완할 수 있는 새로운 통합 프레임워크에 대한 연구가 필요하다.

4. CRE(Cognitive Robotic Engine)

서비스 로봇의 자연스러운 인간-로봇 상호작용 실현을 위해서는 우선 시각, 청각 등의 개별 단서에 기반한 요소기술들을 조명 변화, 배경 잡음, 가려짐, 변형, 신호의 혼합 등과 같은 다양한 환경 요인들로부터 강인하게 동작하도록 하는 과제가 요구된다. 그러나 인간-로봇 상호작용은 다양한 감각 단서에 의존하면서 동시에 운동정보의 해석과 같은 하위 수준의 단서에서부터 상황의 이해나, 대화 상대에 대한 인지모형에 이르는 상위 수준의 정보가 통합된다는 점에서, 분명 개별 요소기술과는 다른 수준의 문제가 제기된다. 즉, 자연스러운 인간-로봇 상호작용을 실현하는 하나의 접근 방법으로서, 특정한 단위 요소 기술에 의존하는 것이 아니라, 각 개별 요소기술을 최대한 효과적으로 통합하여 주어진 과제를 해결하는 것에 초점을 둔 방법과 기술의 개발이 요구된다.

본 장에서는 HRI의 개별 요소기술을 통합하는 하나의 접근 방법으로서 CRE를 제안하고, 이를 서비스 로봇의 호출자 식별에 적용한 사례를 제시한다.

4.1 Conceptual Overview of CRE

연구실에서 개발되고 있는 로봇들이 실제의 복잡한 환경에서 서비스를 수행하기 어려운 이유 중의 하나는 로봇이 잘 정돈된 연구실 환경 이외의 다른 환경이나 수시로 변화되는 실제 환경에 잘 적응하지 못하기 때문이다. 예를 들자면, 조명이 변화되어 Vision 처리에

실패하거나, 외부 환경의 소음으로 인하여 음성 처리가 오동작 하는 것 등을 들 수 있다. 이러한 문제를 해결하기 위한 많은 연구들이 진행되고 있지만 아직 총체적인 해결 방법이 제시되고 있지는 못하다. 개별적인 모듈(음성인식 등)의 성능을 향상시키기 위한 연구는 많이 진행되어 왔으나 개별 모듈의 성능 향상이 어느 정도 이루어진다고 하더라도 개별 모듈로 실생활에서 요구되는 상황이해를 만족시키기에는 충분하지 않다.

복잡한 상황에서도 탁월한 수행을 보이는 사람의 인지와 행동이 로봇의 신뢰성 있는 지각 문제를 해결하기 위한 하나의 시사점이 될 수 있다. 사람은 적절한 인지와 행동을 통하여 목표를 이룬다. 로봇과 사람이 다른 중요한 특성 중의 하나는, 일반적인 로봇이 특별한 환경과 조건 하에서만 가능하며 인식성능에만 의존한 인지모듈을 사용하여 상황을 인식하려는 것과는 달리, 사람은 현재 자신이 사용하는 지각이 얼마나 불확실한지를 인지하고 각 인지모듈(오감)에서 들어오는 여러 가지 단서들을 종합하여 판단하는 상위 수준에서의 인지를 하는 것이다. CRE는 위에서 말한 것처럼 여러 가지 단서들을 종합하여 판단하는 신뢰성 있는 인간의 인지 방법, 판단 방법을 다양한 인지모듈 들을 가지고 있는 Service Robot System에 어떻게 실제적으로 확립하여 구현 할 것인가에 중점을 둔다.

복잡하고 시끄러운 파티장에서 누군가가 자신을 호출하고 피호출자가 그것을 인지하는 과정을 살펴 보자. 호출자가 피호출자를 시끄러운 파티장 같은 환경에서 호출 한다면 그 호출음은 파티장의 여러 가지 소리와 혼합되어 들리는 굉장히 불확실한 소리(위치에 대한 불확실, 호출했다는 믿음에 대한 불확실성을 내포)가 될 것이다. 호출함과 동시에 피호출자는 호출자가 어디에 있는지 확인하기 위한 일련의 Mission이 발생하게 될 것이다.

이때 호출로 발생한 Mission은 호출자가 자신을 어디에서 호출했다고 하는 것이 어떤 충분한 수준에 도달 할 때까지(즉 호출자를 찾았다고 확신할 때까지) 계속 피호출자의 의식에 남아 있게 된다(물론 Mission fail의 경우에는 Cancel도 가능하다. 그러나 여기에서는 Mission Fail을 고려하지 않고 설명한다). 위의 상황에서 Mission을 달성하여야만 피호출자의 의식에서 호출자를 찾는 Mission은 사라지게 되는 것이다.

위에서의 시끄러운 환경에서의 호출음처럼 불확실하거나 부족한 단서들이 Mission을 발생시켰지만 현재 가지고 있는 단서들로는 발생한 Mission을 달성할 수 없다면 Mission을 달성시키기 위해 가능한 방법으로 는 부족한 정보를 좀 더 수집하거나, 모호한 정보의 확

신도를 높이는 쪽으로의 연산을 통하여 불확실성을 줄여나가는 방법이 필요하다. Mission에 대한 불확실성을 감소시키는 방법은 능동적인 방법과 수동적인 방법으로 나눌 수 있다. 정보를 좀 더 수집하기 위하여 좀 더 기다려 볼 수도 있고, 피호출자 스스로 주변을 찾아 보며 호출자를 찾아 볼 수도 있을 것이다.

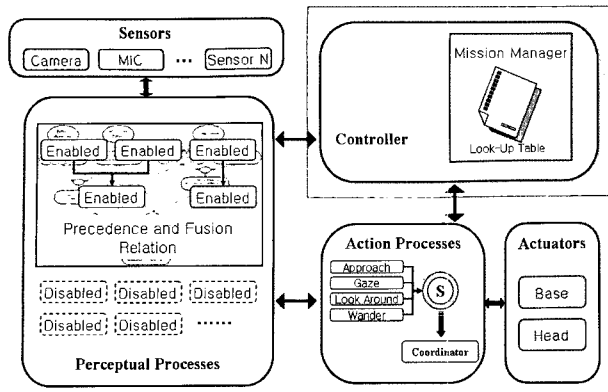
어떠한 단서들의 조합이나 개별적인 단서 하나로 특정한 Mission이 발생되고, 현재의 단서들로는 발생한 Mission을 달성할 수 없다고 판단 할 때 Mission 달성을 위한 단서 수집을 하게 되며 수집된 정보들을 계속적으로 융합하고 가공하면서 호출자를 찾게 될 것이다.

인간의 경우처럼 로봇에서는 이러한 추가적인 정보의 수집 및 가공을 위하여, 기본적인 Sensing unit부터 상위개념의 Recognition unit까지의 추상적이고 복잡한 여러 가지 수준의 Perception unit로 이루어져 있는(Building Block을 구성하는 Unit의 개수, Packet의 형식, Processing 특성들은 미리 정의 되었다고 가정) Perceptual Building Block들 사이에서 자연스럽게 정보가 교환되고 가공되는 비동기적이며 병렬적인 Perceptual Processing 이 일어나게 되는 것이다.

좀 더 자세히 말하면, 음향이나 영상인식 블록의 비동기적이며 동시발생적인 데이터의 흐름에 의해 발생하는 Calling Hand Gesture나 Calling Facial Expression 같은 음향, 영상 모듈이 결합된 복합적인 모듈로부터 들어오는 단서들을 충분히 모으게 될 것이다. 로봇의 센서로부터 Perceptual Building Block으로 연결되는 경로들은 매우 많은 수 일 것이다. 이 중에서 Mission을 달성하기 위한 최적의 경로를 선택하여 판단하게 된다. 여기서의 최적경로라 함은 목적하는 바에 따라 달라지겠지만 달성시간이나 자원의 할당 등과 같은 일련의 자원을 고려하여 경로를 선택하게 되는 것이다. 또한 Uncertainty를 감소시키기 위하여 컴포넌트의 성능이 충분하지 않다면, 좀더 정보를 획득하거나, 불확실한 정보의 구체화를 위하여 Action을 선택하는 능동적인 행위를 하게 된다.

위에서 기술한 사람의 자연스러운 인지과정처럼 로봇이 하듯이 Component의 인식성능 극대화로 인한 인지가 아닌, 현재 가지고 있는 Component상호간의 자연스러운 통합을 통해서 신뢰할만한 판단을 할 수 있는 제어 구조를 구현하는 것이 Cognitive Robotic Engine(CRE)의 목표이다.

그림 10은 CRE의 전체적인 구조에 대해서 나타내고 있다.



< Overall Architecture of Cognitive Robotic Engine >

그림 10 CRE 시스템의 기본적인 전체 구조

CRE의 기본적인 구조는 크게 Perceptual Part, Control Part, Action Part의 세 부분으로 나눌 수 있다. 각 부분에 대해서 좀 더 세부적으로 살펴보면 다음과 같다.

4.1.1 Perceptual Part

Perceptual Process는 인지의 기본적인 Building Block으로 정의된다. Perceptual System을 위한 Building Block에는 많은 수의 Perceptual Process가 존재할 수 있으며 Perceptual Process들은 Precedence상에서의 제약이 있기 전에는 독립적인 것으로 간주된다.

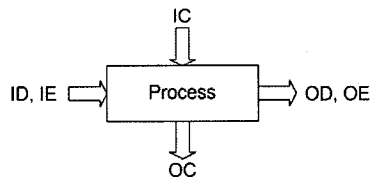


그림 11 Perceptual process 에서의 입력과 출력과의 관계

표 1 Perceptual Process의 입출력 관계

항목	설명
입력	IC 프로세스의 활성화/비활성화를 위한 제어 입력
	ID 선행된 프로세스로부터의 입력
	IE ID와 관련한 불확실성 값
출력	OC OE로부터 일어날 수 있는 Action의 목록
	OD 프로세스 시간과 관련한 프로세스의 출력
	OE OD 출력과 관련한 불확실성 값

적절하게 표현을 하기 위해 Perceptual Process를 인간에 비유하자면 오감의 기능 및 처리를 하는 곳이 된다. 시각과 청각 등의 감각 정보를 수집하여 가공하는 곳이다. Perceptual Process의 구성을 보면, 먼저 외부환경들이나 정보들이 가장 처음으로 입력되는 가

장 첫 번째 단계의 Sensor부와 Sensor에 의해서 수집된, 그리고 가공된 단서들의 관계를 나타내는 Precedence Relations, 서로 다른 정보들을 융합하여 새로운 단서들을 만들거나 불확실성을 높이는데 사용되는 Fusion Relation으로 구성된다.

Precedence Relation 과 Fusion Relation의 설명은 다음과 같다.

가. Precedence Relation

만일 하나나 그 이상의 Perceptual Process에서 나오는 Output이 다른 Process로의 Input이 된다면, 그런 관계를 갖는 둘 이상의 Perceptual Process들은 Precedence Relation을 갖는다고 정의한다.

다음 그림 12는 호출자 인식 문제를 해결하기 위해 정의된 Perceptual Process의 Precedence Relation을 나타내고 있다.

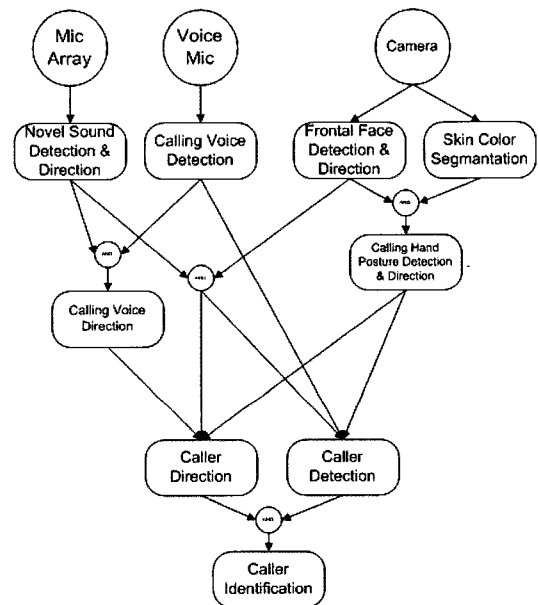


그림 12 Precedence Relation for Caller Identification

나. Fusion Relation

각 단서들이 개별적으로 존재 할 때의 Certainty와 단서들이 함께 존재 할 때의 Certainty는 달라지게 된다. 서로 다른 Process들에서 출력되는 단서들이 결합할 때 Mission을 달성하기 위한 Certainty를 높여 주게 된다면 이러한 관계가 있는 단서들을 Fusion Relation 이 있다고 이야기 한다.

(1) Control Part

각 Process로부터 들어오거나 또는 들어온 정보들을 사용하여 최적 패스를 예측하거나 Uncertainty를 줄이기 위해 적합한 Action을 결정하는 역할을 한다.

Control 모듈은 Perceptual Process에 존재하는

Precedence Relation 과 Fusion Relation을 똑같이 가지고 있다. 실제로 센서에서 들어온 정보를 사용하여 다른 Process들과의 관계를 예측하거나(실제 Perceptual Process 상에서는 실제 Perceptual Process에서 연산되어 출력되는 결과를 가지고 연산을 하게 되지만 Controller 에서는 존재하는 정보를 통하여 정보가 존재하지 않는 다른 Perceptual Process 들의 출력 및 관계까지 연산하게 된다.) 로봇이 Mission을 가장 효율적으로 달성하기 위한 최적의 경로와 방법을 계산해 낸다. 또한 각 Process에서 Uncertainty를 줄이기 위해 정보를 더 수집하는 능동적인 방안으로써 후보로 내놓은 Action 중에서 가장 적합한 Action을 결정해서 실행하게 하는 역할을 한다.

(2) Action Part

로봇의 Uncertainty를 줄이기 위한 능동적인 방법인 Behavior들이 정의 되어 있는 곳이다. Action Behavior, Action Process로 구성 되어 있다.

표 2는 Action Behavior에 대한 정의이다.

표 2 Action Behavior의 정의

Sub Behavior	정의
Gazing (GA)	관심 대상에 대한 지속적인 Head 움직임
Look around (LA)	Random searching을 위한 Head 움직임
Heading (HE)	주어진 방향에 대한 Head 움직임
Wandering (WA)	일정한 속도로 돌아다니기
Approaching (AP)	목표물로 접근 후 일정 거리를 유지
Turning (TU)	의도된 방향으로 돌
Speech Action (SA)	사람과 상호작용하기 위해 로봇이 말함
Searching (SE)	Wandering behavior + Look around behavior
Attentive Approaching (AA)	Gazing + Approaching

Action Process는 각각의 구동부와 관련해 Action Behavior를 생성해낸다. 우리는 첫째로 Action Process 들을 서로 Mutually exclusive한지 그렇지 않은지에 따라 구분했다. 그 이유는 Action Process들 중에 중복되는 자원을 사용하는 것들은 동시에 수행될 수 없기 때문이다.

(3) Mission

CRE에서의 Mission 이라 함은 로봇이 달성 하고자 하는 최종 목표를 의미한다. Mission을 달성하고 나면 이에 해당되는 Perceptual Process나 Action은 종료된다. Mission의 완료는 또 다른 Mission을 Invoke시킬 수 있으며, 동시에 여러 가지 Mission이 Invoke(Multi Mission)될 수도 있다.

4.2 구현을 위한 Hardware/Software 구조의 설계

이 절에서는 실제로 로봇에 구현된 하드웨어와 소프트웨어 구현방법에 대하여 기술한다.

4.2.1 구현된 Perceptual Process의 세부 정의

표 3에서는 그림 11에서 보여준 각각의 Perceptual Process 들에 대한 세부적인 정의와 설명을 보여준다.

표 3 Caller Identification을 위하여 구현된 Perceptual Process의 정의

Novel Sound Detection & Direction	
DEFINITION	When the sound volume exceeds the threshold, estimates the direction of source
SOURCE	Mic Array
INPUT	Raw data of sound
OUTPUT	Direction and certainty of novel sound. Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Calling Voice Detection Caller Direction Caller Detection
Calling Voice Detection	
DEFINITION	Recognize voice commands which this module knows
SOURCE	(Voice) Mic Array
INPUT	Raw data of voice sound
OUTPUT	Certainty of voice. Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Calling Voice Direction Caller Detection
Frontal Face Detection & Direction	
DEFINITION	Finds face region by image feature
SOURCE	Camera
INPUT	Raw image from Camera
OUTPUT	Coordinate, size and certainty of detected face. Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Calling Hand Posture Detection & Direction Caller Direction

Skin Color Segmentation	
DEFINITION	Distinguishes skin region by RGB condition and makes others black in image
SOURCE	Camera
INPUT	Raw image from Camera
OUTPUT	Skin color segmented image. Most probable direction and certainty that callers exist in. Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Calling Hand Posture Detection & Direction
Calling Hand Posture Detection & Direction	
DEFINITION	Estimates calling hand by skin color in face adjacent area
SOURCE	Frontal Face Detection & Direction Skin Color Segmentation
INPUT	Coordinate and size of detected face Skin segmented image
OUTPUT	Direction, distance and certainty of caller Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Caller Direction Caller Detection
Calling Voice Direction	
DEFINITION	Finds voice direction with voice and noble sound direction
SOURCE	Novel Sound Detection & Direction Calling Voice Detection
INPUT	Calling voice and Novel sound direction
OUTPUT	Direction and certainty of callers Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Caller Direction
Caller Direction	
DEFINITION	Finds caller direction by putting together sound and image direction
SOURCE	Novel Sound Detection & Direction Calling Voice Direction Frontal Face Detection & Direction Calling Hand Posture Detection & Direction
INPUT	Direction information and certainties from calling voice and novel sound. Face location and certainty
OUTPUT	Location of callers and certainty Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Caller Identification
Caller Detection	
DEFINITION	Judges the certainty that callers exist
SOURCE	Novel Sound Detection & Direction Calling Voice Direction Frontal Face Detection & Direction Calling Hand Posture Detection & Direction.

INPUT	Certainty of calling voice Certainty of calling hand posture
OUTPUT	Certainty that callers exist Candidate of action behavior and certainty improvement expected. Processing Time
DESTINATION	Caller Identification
Caller Identification	
DEFINITION	Integrates certainties from pre-step modules and identifies callers existence and location
SOURCE	Caller Direction Caller Detection
INPUT	Certainties of Caller Direction and Caller Detection
OUTPUT	Certainty of Caller Identification Candidate of action behavior. Processing Time.
DESTINATION	N/A

4.1.2 CRE-BOT System

아래에서 보여지는 그림 13은 서로 다른 컴퓨터를 사용하고 멀티 쓰레드를 사용하여 Asynchronous and Concurrent 한 Perceptual Process를 구현한 CRE-BOT System 의 전체 구조를 나타내는 그림이다.

Caller Identification을 예로 들어서 간단한 동작 원리를 설명해 보면 다음과 같다.

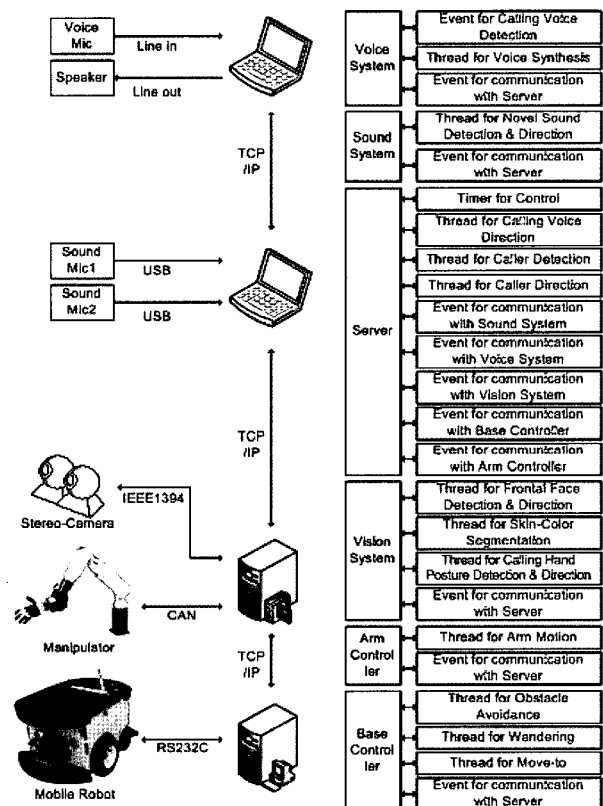


그림 13 CRE-BOT System/A DESTINATION

그림상에서 첫 번째 컴퓨터는 음성 및 소리 인식을 위한 시스템이며 로봇이 호출되어지는 경우 Calling Voice Detection Process를 사용하여 분석하게 된다. 그리고 그 결과를 Server로 보고하게 된다.

각 Perceptual Process들은 출력되는 정보를 다음 Process들로 보내기도 하지만 동시에 Server에 있는 Controller로도 보내게 된다 이는 각 Perceptual Process에서 일어나는 출력의 전파와 동시에 Mission을 달성하기 위한 최적의 경로를 예측과 함께 필요한 경우 Arm과 Base를 구동하는 Action Part 컴퓨터에게 일련의 Action Behavior를 요구하여 Certainty를 높이기 위한 능동적인 Action을 발생하게 한다.

5. 실험결과

본 장에서는 Caller Identification Mission에 CRE의 개념을 도입하여 Mission이 주어졌을 때 자신이 현재 보유한 Component만을 사용하여 어떠한 방식으로 불확실한 상황을 극복하며 어떻게 Mission으로 수렴해 가는지에 대해 실험하고 그 결과를 제시 하였다.

사람이 로봇을 호출할 때 여러 조건에 따라 다양한 상황이 연출될 수 있으나 구현된 CRE의 평가를 위하여 우리는 두 가지 상황만을 가정하였다. 첫 째는 호출자가 로봇의 앞에 위치하는 경우이고, 다른 하나는 호출자가 로봇의 옆 또는 뒤에 존재하는 경우이다.

5.1 호출자가 로봇의 정면 좌측에서 호출한 경우

그림 14는 호출자가 로봇의 정면에서 호출한 경우의 상황을 보여주고 있다. 이 상황에서는 영상 정보만을 가지고 충분히 호출자를 식별해 낼 수 있으므로 시간이 지남에 따라 자연스럽게 호출자에게 다가가는 것을 목표로 삼았다.

그림 15와 그림 16은 그림 14의 상황에 따른 실제 실험의 결과를 보여주고 있다.

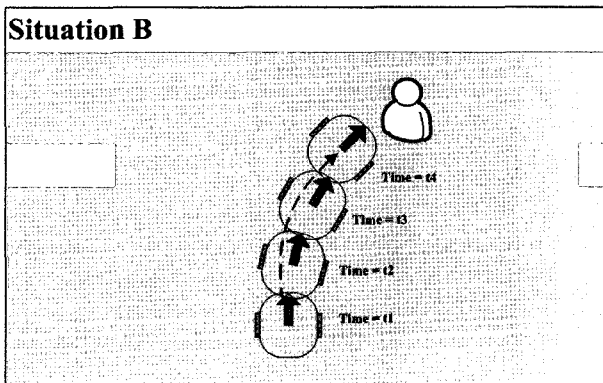


그림 14 호출자가 로봇의 정면 좌측에서 호출한 경우

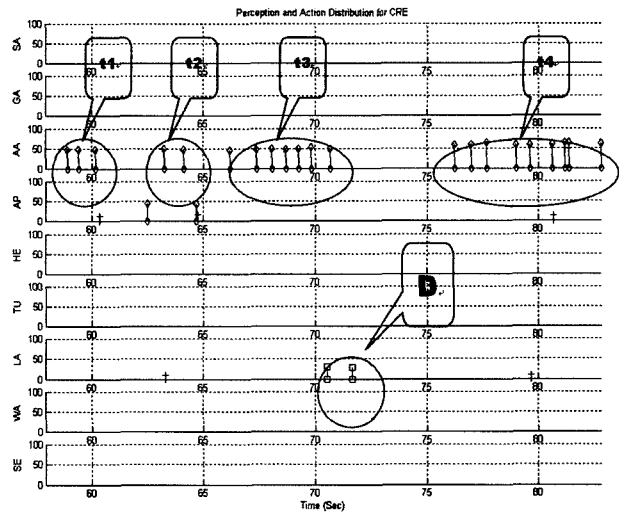


그림 15 호출자가 로봇의 정면 좌측에서 호출한 경우 Perception & Action Distribution Graph

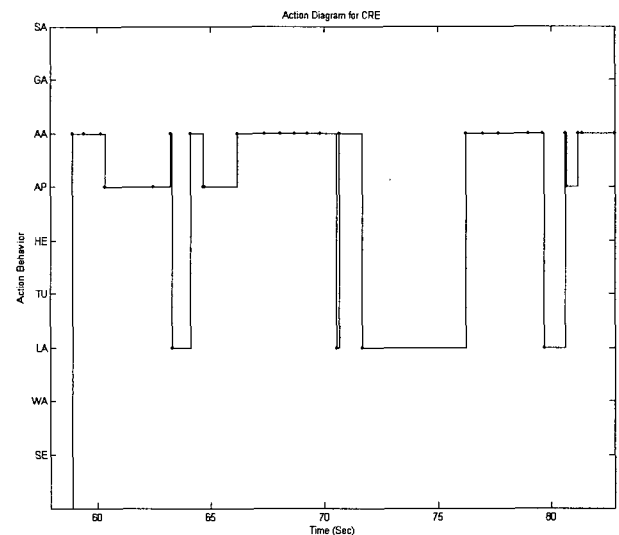


그림 16 호출자가 로봇의 정면 좌측에서 호출한 경우 Action Diagram

그림 16에서는 호출자가 로봇의 정면에서 FD에 의해서 Caller Identification Mission 이 Invoke 된 후 Caller 라고 추정되는 사람의 Certainty를 높여 Mission을 달성하기 위해 수행한 일련의 행동을 보여준다. 그림 20에서의 D는 로봇 Body의 자체적인 소음에 의해서 Invoke 된 것으로 부분적으로는 잘못된 결과라고 볼 수도 있지만 전체적인 로봇의 행위에 크게 영향을 미치지 못함을 그림 16에서 확인할 수 있다. 로봇 자체적인 노이즈에 의해서 잠시 주춤 했던 로봇은 다시 일정 시간 후에 원래의 목표로 수렴함을 볼 수 있다. 또한 로봇의 행위 중 가끔씩 Caller가 손을 들어 줌으로써 HD에 의한 또 다른 행위 유발에 대해서 실험할 수 있었다.

5.2 호출자가 로봇의 우측에서 호출한 경우

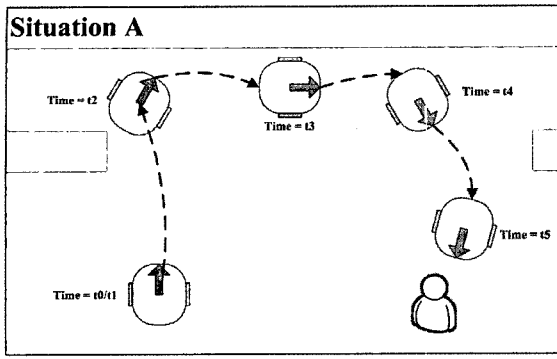


그림 17 호출자가 로봇의 우측에서 호출한 경우

그림 17은 그림 14에서의 상황과 달리 로봇이 처음 위치에서 호출자를 보지 못하고 있는 경우 호출에 의해서 호출자에게 다가가는 상황을 보여주고 있다. 처음에 영상 정보를 통해서 호출자를 찾아낼 수 없지만 호출의 음성이나 일정 크기 이상의 소리에 의해서 호출자가 있는지에 대한 Mission을 Invoke하고 소리의 방향을 찾다가 영상정보에서 호출자를 찾고 앞으로 다가갈 수 있다는 것이 이 상황의 목표가 되었다.

그림 18과 그림 19는 그림 17의 상황에 따른 실험의 결과를 보여주고 있다.

그림 18에서의 t1시점에서 로봇의 시야에서 벗어나 우측에 대기하고 있던 호출자가 소리로 Caller Identification Mission 을 Invoke 시켰음을 볼 수 있다. 그 후 로봇이 Mission 을 달성하기 위해 Invoke 되는 행위와 behavior의 전환 과정이 그림 19에 나타나 있고 최종적으로 그림 18의 t4 시점에서 호출자의 영상 정보를 획득하게 되었다. 결국 호출자에게 다가가는 데 성공을 했지만 한 편으로는 로봇자체의 소음이 많이 발생하여 예상치 않은 NS들이 많이 발생하였다.

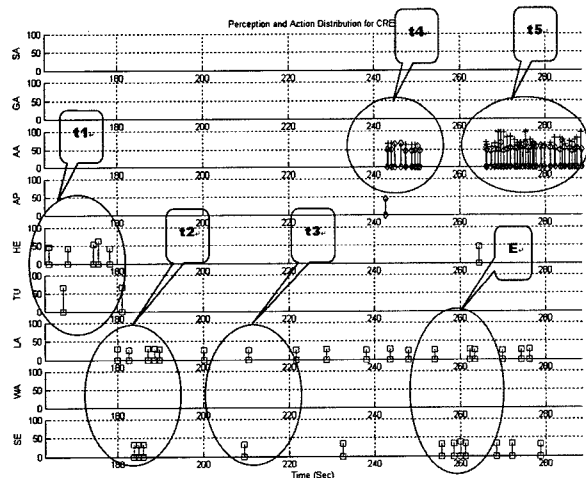


그림 18 호출자가 로봇의 우측에서 호출한 경우
Perception & Action Distribution Graph

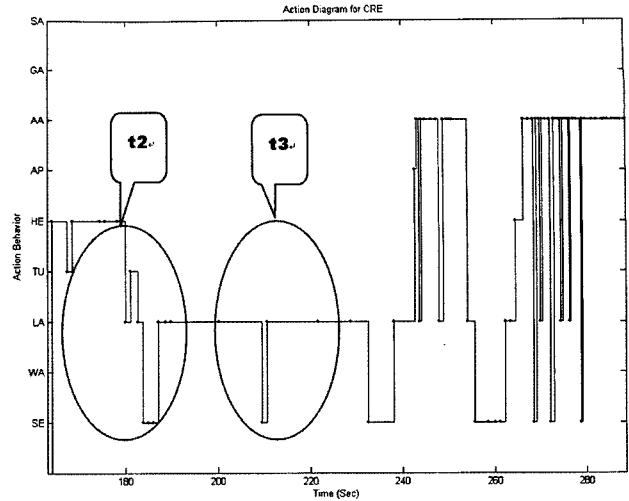


그림 19 호출자가 로봇의 우측에서 호출한 경우
Action Diagram

위에서 일어난 일련의 과정들을 시간에 따라 요약하여 설명해보면 다음과 같다.

두 가지 경우 모두 만약 비전 감지 모듈이 호출자를 찾지 못할 경우 로봇은 SE, WA, LA 행위를 사용하여 호출자를 찾고자 시도한다. 그리고 만약 비전과 사운드 감지 모듈이 적절한 호출자의 방향을 검출한다면 로봇은 호출자를 HE와 TU 행위를 사용하여 찾을 것이다. 또한 로봇은 미션의 certainty를 높이기 위하여 AA, AP 그리고 GA 행위를 통해 호출자 쪽으로 접근하게 된다.

우리는 다양한 상황에서 여러 번 로봇 시스템을 테스트하였으며, 그리고 대부분의 경우 미션은 위의 결과처럼 적당한 시간의 간격을 두고 수행/완료되었다.

5. 맺음말

종래의 behavior based 로봇은 동적인 실제 환경에서, 깨지기 쉬운 가정에 의존하는 추론 과정을 거치지 않고, 즉각적 감지-행동 반응을 통하여 충돌회피, 코너 탈출 등의 문제를 해결함으로써 로봇 행동의 강인성/자연성을 부여하는데 성공적인 결과를 얻었다. 그 결과 최근에는, 상위에서 목표 수행에 관한 계획을 추론을 통해 수립하는 decision layer와 하위에서 예상하지 못한 실제 상황에 대하여 즉각적 감지-행동 반응을 통해 대처하는 behavior layer가 존재하고, 이 두 layer의 중간에서 coordination layer가 조정하는 역할을 수행하는 세 계층의 hybrid 로봇 제어구조가 많은 로봇 플랫폼에 채택되고 있다.

본 연구에서는 로봇 행동의 강인성/자연성을 부여하는데 성공적이었던 behavior 개념을 즉각적 감지-행동 반응에만 한정하여 적용하지 않고, 불확실하고 불완

전한 감지 환경에서 여러 센서와 이들 센서 정보로부터 제공되는 여러 증거를 자연적으로 융합하고, 또 더 필요한 증거를 행동을 통하여 수집하며, 이를 통하여 주어진 미션을 계획하는 보다 상위 차원의 로봇 인지 영역으로 확대하는 방안을 제안하였다. 또한 이러한 확장behavior 인지 개념을 통하여 보다 강인하고 자연적이며 동시에 미션 수행에 효율적인 로봇을 실현할 수 있음을 실제 로봇에 적용한 실험을 통하여 보였다. 본 연구에서 제안하는 CRE(Cognitive Robotic Engine)는 종래의 3 layer architecture의 제한을 뛰어넘을 수 있는 기틀을 마련하는 우리 고유의 독창적인 로봇 architecture라 볼 수 있다.

지금까지는 로봇 인지 엔진을 구성하고 기초적인 효용성을 다양한 플랫폼(PowerBot, Wever, IRobi)을 통하여 검증하는 연구를 진행하여 왔다. 향후에는 로봇 시스템의 자원 제약을 고려한 정보 처리 경로의 판단과 행위의 영향에 따른 감지 정보 변화를 고려한 행위 선택 방안 등에 대해 집중적으로 연구가 진행되어야 한다.

6. 요약

서비스 로봇의 지각과 행동에서의 신뢰성은 각 개별 요소기술의 완벽함으로부터 얻어지는 것이 아니라 비록 성능이 불완전한 요소기술들을 사용하더라도 각 요소기술들의 정보를 융합하고 능동적인 행위를 발생시키는 시스템적인 접근방식으로부터 얻어진다. 그러므로 본 연구에서는 자연스러운 HRI 구현을 위하여 “로봇 인지 엔진”이라는 독창적인 통합 구조를 제안하였으며, 비동기/병렬 정보처리 구조, 다양한 지각정보들의 융합, 행위 결함을 위한 능동적인 지각정보 수집의 특징을 갖는다.

참고문헌

[1] S. H. Lee, “The Emergence of Intelligent Service Robotics,” UN-ECE, Oct., 20, 2004.
 [2] Technology roadmap: Robot, 산업자원부 한국 산업기술 평가원, 2001.
 [3] ‘IT 신성장동력’ 발전전략, 정보통신부, 2003.
 [4] E. Rogers and R. R. Murphy, Human-Robot Interaction: Final Report for DARPA/NSF Study on Human-Robot Interaction, <http://www.csc.calpoly.edu/~erogers/HRI/HRI-report-final.html>
 [5] T. Fong, I. Nourbakhsh, K. Dautenhahn,

“A survey of socially interactive robots,” Robotics and Autonomous Systems, Vol. 42, pp.143-166, 2003.

[6] <http://www.ambrogiorobot.com/>
 [7] <http://www.irobot.com/>
 [8] <http://www-2.cs.cmu.edu/~minerva/>
 [9] <http://www.incx.nec.co.jp/robot/>
 [10] A. Kobayashi, I. Kume, A. Ueno, Y. Kono, M. Kidode, “A Robot Programming Model for Meditating Between Familiarity-Oriented Behaviors and Environment-Oriented Behaviors,” The 7th World Multi-Conference on Systemics, Cybernetics and Informatics, pp.295-302, July, 2003.
 [11] <http://www.yujinrobot.com/>
 [12] <http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>
 [13] <http://www.takanishi.mech.waseda.ac.jp/eyes/>
 [14] <http://koba0005.me.kagu.sut.ac.jp/newsinfo.html>
 [15] http://koba0005.me.kagu.sut.ac.jp/SAYA_2004_March.pdf.
 [16] <http://www.sony.net/Products/aibo/>
 [17] <http://www.mel.go.jp/soshiki/robot/biorobo/shibata/shibata.html>
 [18] http://mind.kaist.ac.kr/3_re/HumanRobot/HumanRobot.htm
 [19] <http://www.unibw-muenchen.de/hermes/>
 [20] <http://www.sdia.or.jp/mhikobe-e/products/etc/robot.html>
 [21] <http://asimo.honda.com>
 [22] <http://www.sony.net/SonyInfo/QRIO/>
 [23] http://www.engin.umich.edu/research/mrl/00MoRob_19.html.
 [24] R. C. Arkin, Behavior-Based Robotics, The MIT Press, 1998.
 [25] <http://www.techfak.uni-bielefeld.de/ags/ai/projects/BIRON/welcome.html>
 [26] <http://groups.csail.mit.edu/lbr/humanoid-robotics-group/cog/>

이 석 한



1972 서울대학교 공과대학(학사)
1974 서울대학교 공과대학(석사)
1982 Purdue University, West
Lafayette, Electrical
Engineering(박사)
1983 Univ. of Southern California
조교수
1998 삼성종합기술원 System &
Control Sector 연구소장 전무

2003 성균관대학교 정보통신공학부 교수
관심분야: Robotics, 자동화, 지능시스템, Mechatronics, 정보
저장, 처리 및 Mobile System을 위한 Micro 및
Nano Systems(MEMS/NEMS)연구
E-mail : Lsh@ece.skku.ac.kr

백 승 민



1996 성균관대학교 전자공학과(학사)
1998 성균관대학교 전기전자 및 컴퓨터
공학부(석사)
2002 성균관대학교 전기전자 및 컴퓨터
공학부(박사)
2002 한국생산기술연구원 위촉연구원
2003 일본 나고야대학교 박사후연구원
2004 성균관대학교 ISRC 연구교수
관심분야: 로봇아키텍처, 로봇비전,
자율주행 및 환경지도 작성
E-mail : smbak@ece.skku.ac.kr

박 연 출



1999 숭실대학교 컴퓨터학과(석사)
2004 숭실대학교 컴퓨터학과(박사)
2000 (주)애니큐브 연구개발팀 팀장
2003 (주)엠아이엔디 선임연구원
2004 (주)포도시스템즈 CTO
2005 성균관대학교 ISRC Post-Doc
관심분야: 로봇비전, 이미지프로세싱, 로봇
자가치유, 로봇인지엔진
E-mail : fearhope@gmail.com
