

음성 및 제스처를 이용한 멀티 모달 명령어 인식 시스템

김정현, 노용완, 권형준, 홍광석
성균관대학교 정보통신공학부

Multi-Modal Instruction Recognition System using Speech and Gesture

Jung-Hyun Kim, Yong-Wan Rho, Hyung-Joon Kwon and Kwang-Seok Hong
School of Information and Communication Engineering, Sungkyunkwan University

요약

휴대용 단말기의 소형화 및 지능화와 더불어 차세대 PC 기반의 유비쿼터스 컴퓨팅에 대한 관심이 높아짐에 따라 최근에는 펜이나 음성 입력 멀티미디어 등 여러 가지 대화 모드를 구비한 멀티 모달 상호작용 (Multi-Modal Interaction : MMI)에 대한 연구가 활발히 진행되고 있다. 따라서, 본 논문에서는 잡음 환경에서의 명확한 의사 전달 및 휴대용 단말기에서의 음성-제스처 통합 인식을 위한 인터페이스의 연구를 목적으로 Voice-XML과 Wearable Personal Station(WPS) 기반의 음성 및 내장형 수화 인식기를 통합한 멀티 모달 명령어 인식 시스템 (Multi-Modal Instruction Recognition System : MMIRS)을 제안하고 구현한다. 제안되어진 MMIRS는 한국 표준 수화 (The Korean Standard Sign Language : KSSL)에 상응하는 문장 및 단어 단위의 명령어 인식 모델에 대하여 음성뿐만 아니라 화자의 수화 제스처 명령어를 함께 인식하고 사용함에 따라 잡음 환경에서도 규정된 명령어 모델에 대한 인식 성능의 향상을 기대할 수 있다. MMIRS의 인식 성능을 평가하기 위하여, 15인의 피험자가 62개의 문장형 인식 모델과 104개의 단어 인식 모델에 대하여 음성과 수화 제스처를 연속적으로 표현하고, 이를 인식함에 있어 개별 명령어 인식기 및 MMIRS의 평균 인식율을 비교하고 분석하였으며 MMIRS는 문장형 명령어 인식모델에 대하여 잡음환경에서는 93.45%, 비잡음환경에서는 95.26%의 평균 인식율을 나타내었다.

Keywords : HCI, Haptics, Multi-Modal Recognition, Gesture Recognition, Embedded and Ubiquitous System

I. 서론

인간과 컴퓨터 상호작용 (Human Computer Interaction : HCI)은 정보통신 산업사회의 컴퓨터 하드웨어 및 소프트웨어 관련 시스템을 연구, 개발함에 있어서 생리적, 지능적, 감성적 특성 등의 인간 요소를 고려한 사용자 중심의 설계 및 구현을 추구하여 사용자 편의성과 수용성이 우수한 인간 최적합의 컴퓨터 시스템을 구축하기 위한 제학문적 연구로 정의할 수 있다. 즉, 컴퓨터 시스템 설계시 인간의 개별적 요소를 고려하여 시스템의 효율성을 높여 사용자가 배우기 쉽고, 상용시 불편함을 최소화하여 만족스러운 작업을 할 수 있도록 구성하는 것이며 이를 위하여 최근의 컴퓨터의 운영체계가 텍스트 위주의 명령에서 WIMP (Windows, Icon, Menu, and Pointer) 인터페이스와 같은 GUI(Graphical User Interface) 방식으로 바뀌고 있고 인간의 다양한 기능을 사용하는 Multi-Modal

Interface (MMI) 시스템으로 발전하고 있다. 이러한 MMI를 위한 대표적인 사례연구로서, Siska Fitriane 연구팀은 'Eliza question-answering system' 기반의 "Multi-Modal Communication System"[1]을 제안하였고 Mark Barnard 연구팀에서는 축구 게임의 분석을 위한 "Multi-Modal Audio-Visual Event Recognition"[2]을 소개하였다. 또한 저자 Martin Fuchs는 일반적인 대화에 있어 구문론적인 부분과 의미론적인 부분의 분별을 위한 "Multi-modal Dialogue System의 구조적 체계"[3]를 기술하였다. 그러나, 이러한 기존 연구의 대부분은 음성-시각 또는 음성-감정의 통합에 연구의 중점을 두었고 데스크탑 PC 및 유선통신망을 기반으로 인간의 자연적 감각 요소 중 하나만을 인식하고 표현하는 유니모달 (Uni-Modal) 인식시스템을 구현하는데 초점을 두고 있다. 또한 멀티모달을 위한 구성 요소로서 제스처 인식 및 시각적 정보의 획득을 위하여 영상 기술을 이용하였으며 이에 따른 복잡한 계산 알고리즘의 필요성 및 제스처 대

이터의 부정확성 그리고 카메라의 성능 및 주변 환경의 변화에 따른 인식 결과의 다형성과 같은 몇 가지 제약사항 및 문제점을 가지고 있다. 무엇보다도 인간의 가장 보편적이며 자연적인 표현수단인 음성과 제스처의 인식에 있어 화자의 정보 및 주변의 잡음 환경을 전혀 고려하지 않았고 멀티 모달 인식의 관점에서 음성-제스처를 포함하는 두 감각 기관의 통합에 대한 연구는 미비한 수준으로 머물고 있다. 즉, 화자의 제스처 정보를 인식하고 표현하려는 시도가 진행 중으로, 잡음이 심한 환경에서 음성 신호는 크게 영향을 받지만 햅틱 장치를 이용한 화자의 제스처 신호는 잡음에 전혀 영향을 받지 않기 때문에 이들의 정보를 함께 분석하고 사용함으로써 주어진 환경에 대한 최적의 인식 성능을 보장할 수 있다. 따라서, 본 논문에서는 잡음환경에서의 명확한 의사 전달 및 휴대용 단말기에서의 음성-제스처 통합 인식을 위한 사용자 인터페이스 관련 연구를 목적으로 Voice-XML과 WPS 기반의 음성 및 내장형 수화 인식기를 통합한 멀티모달 명령어 인식 시스템을 제안하고 구현한다. 제안되어진 MMIRS는 잡음 환경을 분석하고 측정함으로써 잡음의 정도에 따라 개별 언어 인식기에 서로 다른 가중치를 부여함으로써 주어진 환경에서 보다 명확하고 자연적인 대화 요소를 제공할 수 있으며 인터넷과 무선통신망을 기반으로 규정된 명령어 및 언어 정보를 인식하고 표현함으로써 보다 광범위한 정보의 교환이 가능하다.

본 논문은 구성은 다음과 같다. 제 2절에서는 WPS기반의 내장형 수화 인식기의 설계 및 구현에 대하여 기술하고 제3에서는 Voice-XML을 이용한 웹 기반의 음성 인식 및 합성 시스템과 음성-제스처 통합 기반의 MMIRS의 설계 및 구현에 대하여 표현하며, 개별 언어 인식 시스템과 MMIRS의 인식 실험 및 성능 평가는 제 4절에서 주어진다. 마지막으로 제 5 절에서는 본 연구의 대한 결론 및 향후 과제를 요약한다.

II. WPS기반의 내장형 수화 인식기

2.1 수화 제스처의 규정 및 제스처 입력모델

한국표준수화(KSSL)는 국내 청각 및 언어장애인에 의해 사용되어지는 시각적이며 공간적인 언어로서 손으로 표현되는 언어이며 음성언어가 청각으로 이해되고 음성으로 표현되는 청각-음성체계에 반하여 수화는 시각적으로 이해되고 손 운동으로 표현되는 시각-운동체계로 분류할 수 있다. 하지만 일반인에게 있어 수화의 완벽한 학습 및 자연적인 해석은 매우 어려울 뿐만 아니라 전문적인 해석과 번역을 위해선 많은 학습 시간이 요구되어지며 청각장애인에 의해 표현되어지는 모든 대화 요소를 인식하고 해석하는 것은 불가능하다. 따라서, 본 논문에서는 WPS 기반의 내장형 수화 인식기를 구현하기 위하여 "한국어 표준 수화 교본[4]"을 분석하고 수화소의 분류 기

준에 따라 25가지의 기본 수화 제스처를 선정하였고 수화 제스처에 필요한 관련 수형(手形) 제스처를 손의 형태, 굽힘 그리고 뒤틀림 등의 정도에 따라 23가지 손 제스처로 분류하여 이들의 조합과 표현에 따라 여행 정보 안내와 관련되어진 62개의 문장형 인식 모델과 104개의 단어 인식 모델을 선정하였으며 수화 제스처에 필요한 관련 손 제스처의 예는 그림 1로 표현하였다.

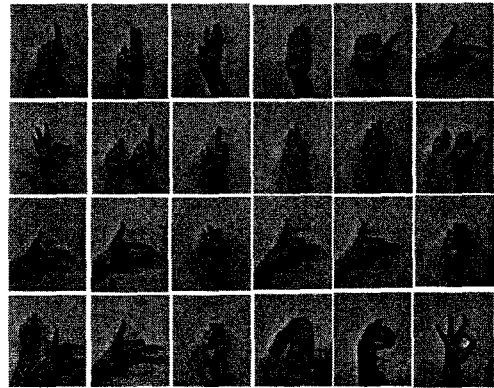


그림 1. 수화 제스처에 필요한 관련 손 제스처의 예

제스처 인식 및 표현을 위한 기존의 연구에서는 일반적으로 제스처 데이터의 획득을 위하여 시각 인식 기술 기반의 영상 기술을 이용하였으며 이러한 영상 기술의 활용은 정확한 제스처 정보를 획득할 수 없을 뿐만 아니라 주변 환경의 변화, 특히 배경 및 조명 조건의 변화 그리고 카메라의 성능 등은 제스처 인식 성능에 매우 큰 영향을 미친다. 따라서, 본 논문에서는 수화 제스처를 획득하기 위한 장치로 햅틱 응용 분야에서 주요한 입력장치로 사용되어지고 있는 5DT사의 무선 장갑 장치와 모션 추적기(Fastrak®)를 이용하였다. 무선 장갑 장치는 Fiber-Optic Flex Sensor를 사용하여 사용자의 손가락 굽힘 정도를 파악하고 이로부터 데이터를 획득할 수 있는 기본적인 제스처 인식장치로서 Pitch & Roll 센서를 자체 내장하고 있으므로 손목의 굽힘과 뒤틀림에 대한 측정도 별다른 장비 없이 가능한 특징을 가진다. 장갑장치의 출력데이터는 9bytes로 header, f1, f2, f3, f4, f5, pitch, roll, checksum의 구조를 가지며 손가락 움직임 정보는 f1=엄지, f2=인지, f3=중지, f4=약지, f5=소지의 순서로 출력되어지고 각각의 데이터 영역은 0에서 255까지 설정되어, 0에 가까운 값이면 손가락의 상태는 펴진 상태이고 255에 가까운 값이면 굽힌 상태를 나타낸다. 또한 Fastrak®은 3차원 공간에서 사용자의 수화 제스처를 실시간으로 측정하는 고정밀의 모션 추적 시스템으로서 공간상에서 움직이는 작은 수신기들의 움직임을 실시간으로 측정하는데 일반적으로 6자유도로 정의되는 위치값(X, Y, Z축에 대한 좌표값)과 방위값(X, Y, Z축에 대한 회전값)으로 환산하여 직렬 blue-tooth를 통하여 차세대 착용형 PC 플랫폼(embedded i.MX21 board)으로 전송해준다[5], [6]. 수화 제스처의 획득과 처리를 위한 제스처 입력 모듈의 개

요 및 구조적 모델은 그림 2와 같다.

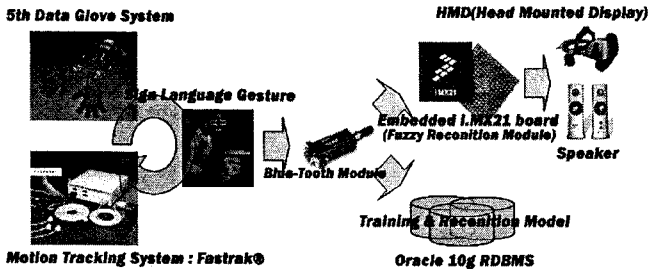


그림 2. 수화 제스처 입력 모델

입력된 사용자의 동적 수화 제스처 데이터는 수화 제스처 인식 알고리즘이 내장된 WPS (embedded i.MX21 board) 및 수화 제스처의 클러스터링 및 분류를 위한 데이터베이스 서버로 전송되어진다. WPS로 전송된 제스처 데이터는 의미 있는 제스처 인식을 위한 Fuzzy 인식모델의 입력변수로 사용되어지며, 데이터베이스 서버로 전송된 수화 제스처 데이터는 RDBMS의 SQL analytic function에 의해 의미 있는 제스처 인식모델을 형성한다.

2.2 데이터-클러스터링 : 수화 인식모델

K-means 클러스터링, QT (Quality Threshold) 클러스터링, Fuzzy c-means 클러스터링, Self-Organizing Map (SOM)과 같은 통계적 클러스터링 알고리즘들은 기존의 unsupervised 패턴 인식 및 기계 훈련, 데이터마이닝, 이미지 분석 등의 분야에서 보편적으로 이용되어져왔다. 그러나 이러한 기법들의 적용에 있어서는 다차원적 특징에 따른 복잡한 수학적 계산 알고리즘이 필요하고 분산처리 환경에서는 적용하기가 매우 힘들며 패턴(데이터)의 크기에 따라 계산비용이 결정되어지고 메모리 교환과 할당의 최소화를 위한 여러 가지 제약사항 및 문제점들을 고려하여야만 한다[7], [8], [9]. 따라서 본 논문에서는 입력되어진 수화 제스처 데이터에 대하여 의미 있는 유효 제스처 구간과 동작간의 천이구간을 분류하고 효율적인 분석 및 클러스터링을 위하여 RDBMS(Relational Database Management Systems)를 채택하였고 SQL (Structured

Query Language)기반의 analytic function을 적용함으로써 보다 합리적이고 향상되어진 수화 제스처 학습 및 인식모델을 형성할 수 있었다. 수화 제스처의 효율적인 클러스터링을 위한 RDBMS 클러스터링 모듈의 구성 및 규칙은 그림 3과 같다.

1. Difference Between Preceding Average(preceding 3 and 1) and Current Row Value
 2. Decide validity : Preceding Average - Current Value = 82.17 = 35

Average Between 3 Preceding and 1 Preceding from Current Row : 52

1. Difference Between Preceding Average and Current Row Value > 5
 2. 'X'(invalidity Record) check

RowNum	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	VALIDITY
22	73	86	256	255	255	52.98	14.93	-4.78	56.42	-5.71			X
23	81	-8	255	255	255	82.98	-11.97	-2.13	52.23	-4.95			X
24	12	15	154	255	255	82.55	-1.98	6.01	56.62	-1.76			X
25	20	10	0	255	255	81.69	-28.05	-1.69	55.21	-1.89			X
26	7	0	255	255	255	82.81	-1.86	-1.82	49.25	-6.23			
27	0	0	255	254	255	82.78	-8.77	-1.75	51.58	-6.62			
28	0	0	253	254	255	49.87	-15.18	-2.12	48.97	-6.27			
29	0	0	256	255	254	48.71	-9.43	-1.92	52.86	-4.89			
30	0	0	255	256	255	81.09	-8.75	-1.75	47.93	-1.94			

그림 3. RDBMS에 의한 클러스터링 및 인식모델의 구현

데이터의 현재 행과 과거 3개의 행의 평균값을 비교하여 그 차이가 5이상이면 천이구간의 데이터로 간주하고 대상 데이터의 다섯 손가락 값 중 하나의 손가락이라도 5 이상의 차이가 날 경우 천이 구간의 데이터로 간주한다. 또한 연속된 천이구간과 유효구간의 구분을 위하여 위의 과정에서 유효하다고 판정된 구간일 지라도 연속적인 값이 5 이상이 차이하지 않을 경우 천이 구간으로 간주한다.

2.3 Fuzzy Max-Min 합성을 이용한 수화 인식 모듈

수화 제스처의 보다 정확한 인식과 WPS 기반의 내장형 수화 인식 시스템의 구현을 위하여 본 논문에서는 Fuzzy Max-Min 합성 및 추론 기법을 적용하였다. 또한 임베디드 시스템의 제에 있어 보다 쉽게 표현이 가능한 사다리꼴 모양의 fuzzy 소속 함수를 적용함으로써 다양한 수화 제스처 인식 모델에 대하여 보다 빠른 인식 시간과 효율적인 제스처 데이터의 처리가 가능하였다.

제안되어진 Fuzzy Max-Min 합성은 수식(1) [10]로 정의되어지며, Fuzzy Max-Min 수화 인식모델의 합성과

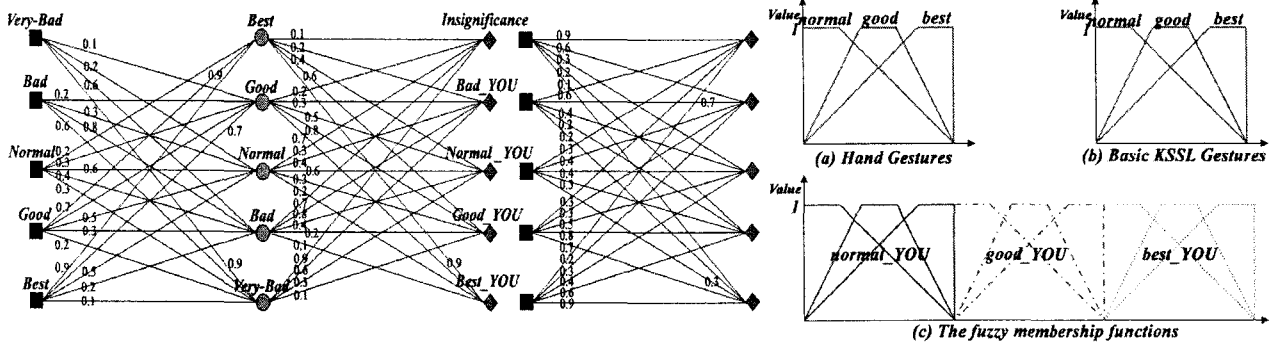


그림 4. Fuzzy Max-Min 합성과 fuzzy 소속함수

추론 규칙에 필요한 가중치, 그리고 명령어 인식모델 “당신”에 해당하는 fuzzy 소속 함수를 그림 4로 표현하였다.

$$For(x, y) \in A \times B, (y, z) \in B \times C,$$

$$\mu_{SR}(x, z) = \text{Max}_y \{ \text{Min}(\mu_R(x, y), \mu_S(y, z)) \} \quad (1)$$

III. 멀티모달 명령어 인식 시스템

3.1 Voice-XML 기반의 웹 음성 인식 및 합성

Voice-XML(VXML)은 웹기반에서 음성 인터페이스를 지원하고 개발할 수 있도록 구성된 마크업 언어로서 사용자와 서비스 시스템간의 실제 음성 흐름을 제어하고 VXML문서 형식은 XML에 내포된 태그형식을 갖는 텍스트 파일로서 이해 및 작성 용이하며 음성 포털 구축 시 유지보수 측면에서 유리하다는 장점을 가진다. W3C의 표준화에 따른 VXML 2.0의 구조 및 구성 요소는 그림 5로 표현되어진다[11].

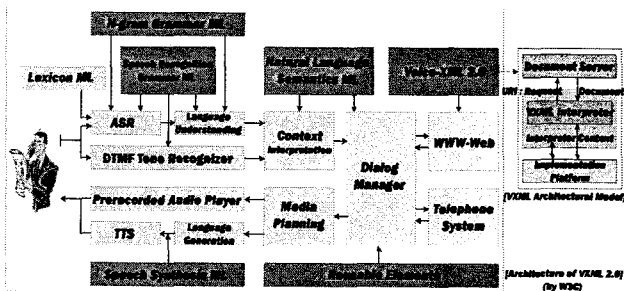


그림 5. VXML 2.0의 구조적 모델 및 구성 요소

VXML 구성요소 및 실행과정은 다음과 같다 : ① Document Server(web server)는 VXML Interpreter인 클라이언트 어플리케이션으로부터 요청을 받는다. ② 서버는 VXML interpreter에게 응답으로서 VXML Document를 제공한다. ③ VXML Interpreter context는 VXML interpreter와 동등하게 사용자 입력을 모니터링한다. ④ Implementation platform은 VXML interpreter context와 VXML interpreter에 의해 통제된다. 또한, 웹기반의 음성인식 및 합성을 위하여 본 논문에서는 ASR-engine으로서 KT의 'HUVOIS-TTS'를 이용하였으며 'HUVOIS-TTS'는 시각-청각 장애인을 위한 음성 인식 및 합성 소프트웨어로 제공되어진다.

3.2 음성 및 제스처 인식 기반의 Multi-Modal Interaction

현재 W3C의 멀티 모달 인터랙션 프레임워크 표준안을 기반으로 멀티 모달 인터페이스를 이용한 응용시스템을 구현하기 위해 요구되어지는 기본적인 요소는 크게 입력 요소, 출력 요소, 상호작용 관리기(또는 대화관리시스템)로 분류할 수 있다[11]. 입력요소는 다시 인식, 해석, 통합 모듈로 세분화되어지며 이를 통해 다중 감각 기관 및 인

식 대상에 대한 인식결과는 해석되어지고 통합되어진다. 본 논문에서는 인식의 대상을 인간의 오감 요소 중 음성과 제스처를 입력 요소로 규정하고 개별 감각 요소에 의한 명령어 인식 모델을 웹과 WPS를 기반으로 인식하며 “멀티 모달 서버”에서 통합, 해석되어진다. 해석되어진 인식 결과는 음성, 제스처, 포인팅 디바이스 등의 정보를 통합시켜 인터랙션 관리기로 전달되어지며 인터랙션 관리기로부터 사용자에게 전달할 정보가 입력이 되면 어떤 모드로 출력할 것인지를 결정하게 된다. 즉 음성, 그래픽 등을 결정하게 된다. 이때 출력 모드가 하나로 결정이 되고 만약에 주 모드로 출력이 불가능한 경우에는 보충 모드가 출력으로 결정된다. 특별히, 본 연구에서는 차세대 PC 플랫폼으로서의 WPS를 기반으로 제스처를 인식하고 통합되어진 인식결과를 WPS에서 음성과 시각적 요소로 표현하기 위해서 HMD (Head-Mounted Display)를 포함하는 확장된 출력모듈을 구현하였다. 그림 6에서는 본 연구에서의 음성-제스처기반의 MMI의 기본 개념과 구조를 도식화하였다.

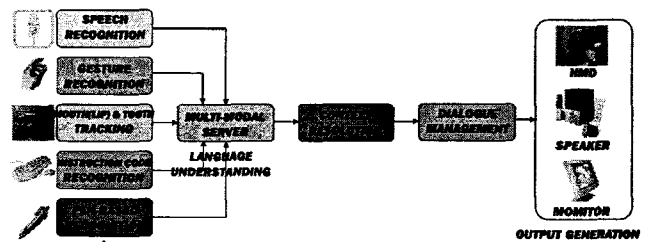


그림 6. Multi-Modal Interaction (MMI)의 구조 및 구성요소

3.3 음성 및 제스처의 통합을 위한 MMIRS의 아키텍처

본 논문에서 제안되어진 음성 및 제스처의 통합을 위한 MMIRS의 아키텍처는 그림 7로 표현되어진다.

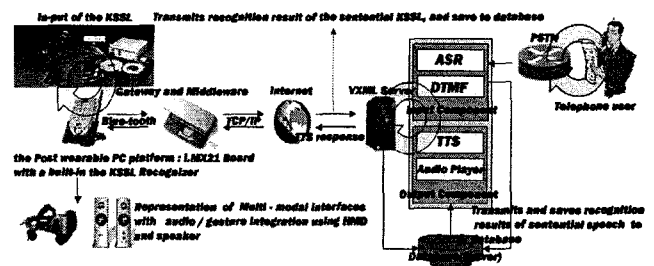


그림 7. 음성 및 제스처의 통합을 위한 MMIRS의 아키텍처
사용자는 전화 단말기와 무선 햅틱 장치를 이용하여 규정된 명령어 인식 모델을 음성과 수화 제스처로 입력하게 되고 입력된 개별 감각 명령어는 수화인식기가 내장된 WPS와 웹기반의 음성인식을 위한 VXML Server로 전송되어지며 입력된 명령어에 대하여 인식 과정을 수행하게 된다. 개별 감각 요소의 인식 결과는 인식 결과의 통합과 해석을 위한 MMI 데이터베이스로 전달되어지는데, 이 과정에서는 블루투스 통신을 기반으로 하는 WPS

의 TCP/IP 통신을 위한 미들웨어 및 스마트게이트가 요구되어진다. MMI 데이터베이스에서 통합된 개별 감각 요소의 인식 결과는 규정된 명령어 인식 모델과의 비교를 통해 주어진 상황에 가장 적합한 멀티모달 인식 결과를 사용자에게 음성과 시각적 표현 요소(그래픽)로 제공한다. 즉, 잡음환경에서는 음성 인식기의 성능이 매우 저하되는 특징을 가지므로 잡음 특성과는 전혀 무관한 제스처 인식 결과를, 비잡음환경에서는 인간의 보편적 의사교환 수단인 음성 인식 결과를 사용자에게 제공한다. 그림 8은 음성과 제스처 인식 기반의 MMIRS의 명령어 인식 및 표현을 위한 전체 흐름도를 보여주며, 아래의 소스코드는 이러한 비교 과정 및 최종 출력결과를 사용자에게 제공하기 위한 내부 웹 로직을 부분적으로 기술하였다.

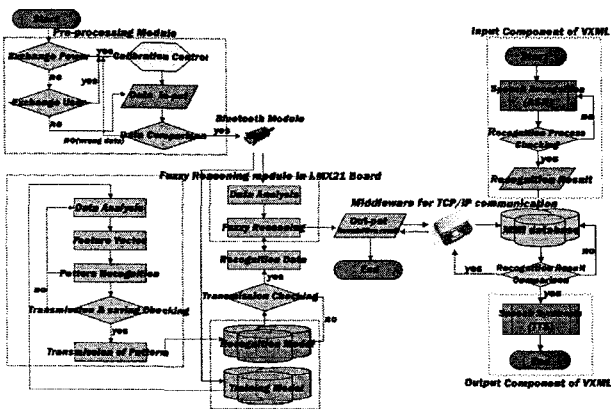


그림 8. 음성과 제스처 인식 기반의 MMIRS의 순서

```
<?xml version="2.0" encoding="EUC-KR"?>
<vxml version="2.0" >
<%

if rs("Gesture") <> "0" and rs("Voice") <> "0" then
conn.Execute("update play set Checking=1")

Set rs2 = conn.Execute("select * from data where
Contents=' " & rs("Gesture") & "'")

if not rs2.eof then
conn.Execute("update play set Result=1")
conn.Execute("update play set Contents=' " & rs("Gesture") & "'")
:
end if

Set rs3 = conn.Execute("select * from data where Contents=' " & rs("Voice") & "'")
if not rs3.eof then
conn.Execute("update play set Result=1")
conn.Execute("update play set Contents=' " & rs("Voice") & "'")
:
end if
end if

if Result_value <> "ok" then
conn.Execute("update play set Result=0")
conn.Execute("update play set Contents=' " & "Not Matching" & "'")
end if

%>
</vxml>
```

IV. 실험 및 결과

제안되어진 MMIRS의 실험 및 성능 평가를 위한 실험 환경은 다음과 같다. 본 논문에서는 수화 제스처의 획득을 위한 무선 햅틱 장치로서 5DT 사의 data glove와 fastrak®를 채택하였고 내장형 수화 인식기의 구현을 위한 H/W 플랫폼은 임베디드 Linux기반의 WPS (i.MX21 test board)를 적용하였으며 웹기반의 음성 인식 및 합성을 위하여 웹기반 음성 인터페이스 VXML을 도입하였다.

본 논문에서는 수화 제스처 입력 모듈과 내장형 수화 인식기(WPS)간의 이격거리를 반경 10M의 타원 형태로 구성하고 햅틱 장치를 착용 후 규정된 위치로 이동하면서 인식 모델에 상응하는 수화 제스처를 취함과 동시에 휴대용 단말기의 헤드셋을 이용하여 규정된 명령어 인식 모델을 발생하게 된다. 15인의 피험자가 62개의 문장형 인식 모델과 104개의 단어 인식 모델에 대하여 잡음과 비잡음 환경에서 연속적으로 10회의 반복실험을 수행하였고 주어진 상황에 따른 개별 명령어 인식기 및 MMIRS의 평균 인식율을 비교, 분석하였다. 표 1은 잡음과 비잡음 환경 하에서의 규정된 인식 모델에 대한 개별 명령어 인식기 및 MMIRS의 평균 인식율을 보여주며 그림 9~12를 통해 실험결과를 도식화 하였다.

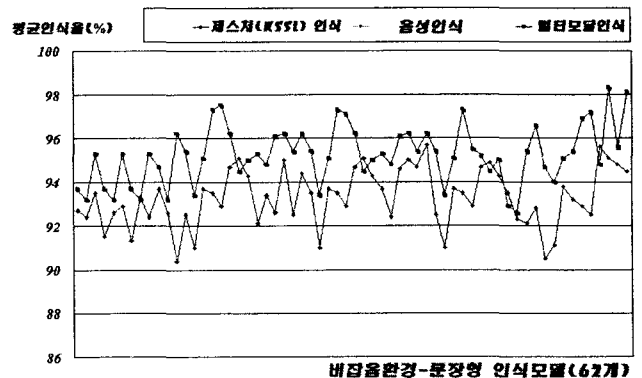


그림 9. 비잡음환경에서 문장형 모델의 인식율

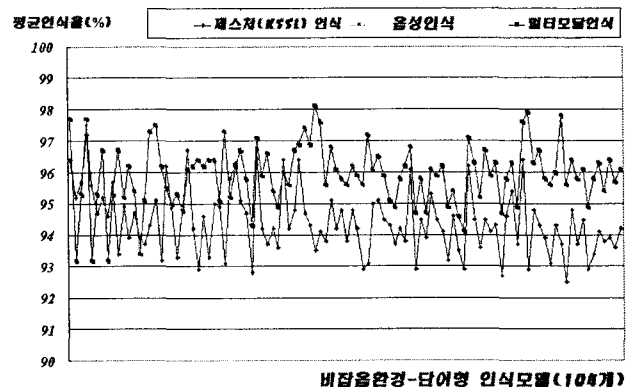


그림 10. 비잡음환경에서 단어형 모델의 인식율

표 1. 개별 감각 명령어 인식기 및 MMIRS의 평균 인식율

		개별 감각 인식기의 평균 인식율		MMIRS의 평균 인식율
		음성(%)	음성(%)	수화 음성(%)
잡음환경	문장형 인식모델 (62개)	93.45	86.09	93.45
	단어형 인식모델 (104개)	94.50	87.15	94.50
비잡음환경	문장형 인식모델 (62개)	93.29	95.26	95.26
	단어형 인식모델 (104개)	94.41	95.93	95.93

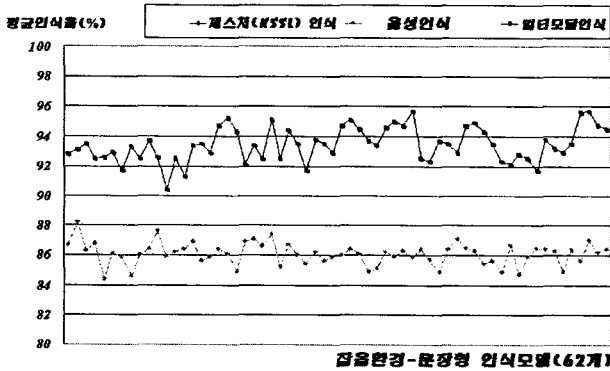


그림 11. 잡음환경에서 문장형 모델의 인식율

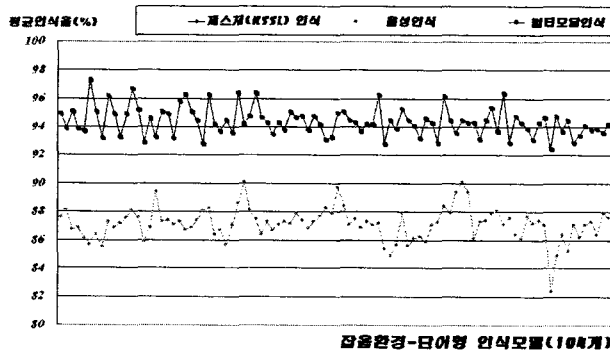


그림 12. 잡음환경에서 단어형 모델의 인식율

V. 결론 및 향후과제

인간의 오감 요소 중 음성과 제스처에 의한 의사 표현은 가장 보편적이며 일반화되어진 의사 전달 및 의사 결정 수단으로 사용되어져 왔다. 근래에 있어 멀티 모달 상호작용(Multi-Modal Interaction) 및 차세대 PC를 위한 오감 인식 기술에 대한 관심과 연구가 활발히 추진되어짐에 따라 제안되어진 두 감각 요소는 단순한 의사 교환의 수단이 아닌 기존의 마우스나 키보드를 대체할 수 있는 새로운 입력 요소로 연구의 범위가 확대되어지는 경향을 보이고 있다. 따라서, 본 논문에서는 WPS기반의 내장형 수화 인식기와 VXML을 이용한 웹기반의 음성 인식 시스템을 통합한 멀티 모달 명령어 인식 시스템을 제안하고 구현하였다.

그러나 음성과 제스처의 통합 인식 및 표현에 있어 인간의 발성과 제스처의 표현은 서로 다른 시간적 차이를 가지게 되며 이로 인한 동기화의 문제는 매우 중요하다. 그리고 이러한 입력 동기화 문제의 해결을 위하여 지속

적인 연구가 요구되어지며 향후 인간의 오감 요소 중 보다 많은 감각 요소를 인식하고 통합함으로써 보다 향상된 HCI의 구현과 더불어 인간의 오감요소를 이용한 사용자 인증 기술의 설계 및 구현을 향후 과제로 남기고자 한다.

감사의글

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 지원 사업 (IITA-2005-C1090-0501-0019) 및 2006년도 두뇌한국21사업의 연구 결과로 수행되었음.

참고문헌

- [1] Siska Fitrianie. et al.: A Multi-modal Eliza Using Natural Language Processing and Emotion Recognition. TSD 2003. LNAI, Vol. 2807.(2003)
- [2] Mark Barnard. et al.: Multi-Modal Audio-Visual Event Recognition for Football Analysis. IEEE XI11 Workshop on Neural Networks for Signal Processing, IEEE Workshop Proceedings (2003) 469-478
- [3] Martin Fuchs. et al.:Architecture of Multi-modal Dialogue System. TSD 2000, LNAI, Vol.1902, (2000)
- [4] S.-G.Kim.: Korean Standard Sign Language Tutor, 1st, Osung Publishing Company, Seoul (2000)
- [5] J.-H.Kim. et al.: Hand Gesture Recognition System using Fuzzy Algorithm and RDBMS for Post PC. FSKD2005. Lecture Notes in Artificial Intelligence, Vol. 3614. Springer-Verlag, Berlin Heidelberg New York (2005) 170-175
- [6] 5DT Data Glove 5 Manual : <http://www.5dt.com>
- [7] Richard O. Duda, Peter E. Hart, David G. Stork.: Pattern Classification, 2nd, Wiley, New York (2001)
- [8] Dietrich Paulus and Joachim Hornegger.: Applied Pattern Recognition, 2nd, Vieweg (1998)
- [9] J. Schuermann.: Pattern Classification: A Unified View of Statistical and Neural Approaches, Wiley&Sons (1996)
- [10] W. B. Vasantha kandasamy.: Smaranda Fuzzy Algebra. American Research Press, Seattle (2003)
- [11] Scott McGlashan et al.: Voice Extensible Markup Language (VoiceXML) Version 2.0. W3C Recommendation, <http://www.w3.org> (1992)