
스마트 홈 환경에서 사용자 상황정보 기반의 음성 인식 시스템 개발

Development of Speech Recognition System based on User Context Information in Smart Home Environment

김중훈*, 심재호*, 송창우*, 이정현**
인하대학교 컴퓨터정보공학과*, 인하대학교 컴퓨터정보공학부**

Jong-Hun Kim(jhkim@hci.inha.ac.kr)*, Jae-Ho Sim(whitelazy@gmail.com)*,
Chang-Woo Song(whrma@hci.inha.ac.kr)*, Jung-Hyun Lee(jhlee@inha.ac.kr)**

요약

현재 높은 인식성능을 보이고 있는 대용량의 음성인식 시스템의 대부분은 고립 단어 음성인식 시스템이다. 이러한 시스템의 인식범위를 늘리기 위해서는 검색 하려는 단어수를 늘려야 한다. 하지만 검색하려는 단어수가 늘어남에 따라 시스템의 속도 및 인식 성능이 저하되는 문제점이 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 스마트 홈 환경에서 음성인식 성능에 영향을 주는 상황정보를 정의하고 관성 센서와 RFID(Radio Frequency Identification)를 사용한 사용자 위치 추정 방법을 제안한다. 또한 음성인식 시스템의 상황정보에 의한 단어모델 도메인을 구축하여 기존의 시스템보다 높은 성능을 보이는 음성인식 시스템을 개발 한다. 스마트 홈 환경에서 본 연구에서 제안된 음성 인식 시스템이 인식률의 저하 없이 동작하는 것을 확인하였다.

■ 중심어 : | 음성인식 | 위치 추정 | 유비쿼터스 컴퓨팅 |

Abstract

Most speech recognition systems that have a large capacity and high recognition rates are isolated word speech recognition systems. In order to extend the scope of recognition, it is necessary to increase the number of words that are to be searched. However, it shows a problem that exhibits a decrease in the system performance according to the increase in the number of words. This paper defines the context information that affects speech recognition in a ubiquitous environment to solve such a problem and develops user localization method using inertial sensor and RFID. Also, we develop a new speech recognition system that demonstrates better performances than the existing system by establishing a word model domain of a speech recognition system by context information. This system shows operation without decrease of recognition rate in smart home environment.

■ keyword : | Speech Recognition | Localization | Ubiquitous Computing |

1. 서 론

정보통신 기술의 발달로 인간과 기계사이의 인터페

이스의 필요성이 증가하였다. 특히, 스마트 홈 환경에서 다양한 디바이스들과 가장 자연스러운 의사소통과 가장 편리한 인터페이스를 충족시키기 위해서는 음성 인

* 본 논문은 인하대학교의 지원에 의하여 연구되었습니다.

접수번호 : #071226-002

접수일자 : 2007년 12월 26일

심사완료일 : 2008년 1월 21일

교신저자 : 이정현 : jhlee@inha.ac.kr

식 기술이 필수적으로 필요하다. 음성 인식 기술은 인간사이의 의사소통 수단인 음성 속에 포함된 언어 정보와 음향 정보를 추출하고 추출된 정보를 저장 및 훈련하여 음성정보의 의미를 기계가 이해하도록 하는 것을 말한다.

음성인식 형태는 구분 발성된 단어를 인식하는 고립 단어 인식과 연속 발음을 인식하는 연속 음성인식, 더 나아가서는 대화체 음성을 대상으로 하는 음성이해의 형태로 발전하고 있다. 이러한 음성인식기술의 궁극적인 목표는 모든 화자의 음성을 모든 환경에서 인식하는 것이다.

고립단어 음성인식기 중에서 상용화되었던 대표적인 시스템은 Dragon System의 Voice Scribe 1000 Dragon Dictate, IBM의 Voice Command, Texas Inst.의 Speech Command 등이 있다. 이 시스템들은 대략 1000 단어 내외의 단어를 인식할 수 있다. 현재는 IBM의 Tangora 시스템과 NEC의 영어, 일본어 고립단어 음성 인식 시스템이 있다. 각각 2만 개와 1천 800개의 단어 규모로 95% 이상의 인식 성능을 보인다. 이러한 고립단어 음성인식 시스템은 인식대상이 되는 단어수가 증가하면 인식 속도 및 인식률이 급격히 저하되는 문제점이 있다.

따라서 본 논문에서는 음성인식 서비스를 받는 사용자의 상황정보를 사용하여 스마트 홈 환경에서 높은 인식률을 보이는 시스템을 설계한다. 음성 인식 시스템의 성능에 영향을 주는 요인을 상황정보로 구성하고 온톨로지(Ontology)를 사용하여 정의 한다. 노이즈를 측정하고 RFID(Radio Frequency Identification)를 통해 사용자 정보를 획득한다. 또한 RFID를 관성센서와 함께 사용하여 사용자의 위치정보를 획득한다. 본 시스템은 실시간으로 상황정보를 획득하고 획득된 정보를 애플리케이션에 제공하기 위해 유비쿼터스 미들웨어 중 하나인 OSGi(Open Service Gateway Initiative)를 기반으로 설계한다. 본 시스템의 음성 인식 알고리즘은 HMM(Hidden Markov Model)을 사용하였으며, 상황정보에 따른 HMM 도메인을 구성하여, 고립 단어 음성 인식기의 단점을 극복하고 성능을 향상시켰다. 시스템에 등록되어있는 여러 사용자를 대상으로 시스템 평가 결과 스마트 홈 환경에서 높은 음성인식 성공률을 보였다.

본 논문의 구성은 먼저 2장에서 사용자 위치추정 방법을 제시하고 3장에서 음성인식을 위한 상황정보 모델 및 음성인식 단어모델 도메인을 구성한다. 4장에서 음성 인식 시스템을 설계하고 5장에서는 시스템 평가를 수행한다. 마지막으로 6장에서 결론을 맺고 향후 연구방향을 제시한다.

II. 관성 기반 사용자 위치추정

관성센서를 사용하여 사용자의 위치를 추정하는 방법은 사용자의 이동한 거리를 누적해서 더하여 현재의 사용자의 위치를 측정하는 방식이다. 이러한 방식은 사용자가 움직이기 시작한 처음의 위치를 알아야 사용자의 위치를 추정할 수 있고 장시간 운용 시 센서의 측정 오차에 의해서 오차가 누적되는 문제점이 있다. 따라서 본 연구에서는 오차가 누적되는 문제를 해결하기 위해서 RFID를 사용한 근접 위치 측정법을 관성센서와 함께 사용하는 위치 추정 방법을 개발한다. 위치 추정 관리자에 의해 확인된 사용자의 위치는 음성 인식 시스템의 도메인 분류에 사용된다.

1. 관성 항법 시스템

스트랩다운 관성 센서는 관성을 측정하는 장치로 가속도계 및 각속도계(Gyroscope)로 이루어진다. 가속도계 및 각속도계는 3축으로 구성되며, 각각 X, Y, Z축으로의 가속도와 회전 각속도를 측정한다. 이 장에서는 관성 항법 시스템의 좌표계, 초기정렬 및 위치 계산 알고리즘에 대해 설명한다[1][2][3].

관성 센서에서 측정되는 정보는 이동 객체가 어느 방향으로 얼마만큼 힘을 받았는지에 대한 관성 정보이다. 이를 실제로 활용하기 위하여 항법 좌표계로의 변환이 필요하다.

- 관성 좌표계

지구 중심을 원점으로 하고 지구의 자전과는 관계없이 X, Y축은 적도 평면상에 위치하고, Z축은 지구의 회전축과 일치하는 좌표계이다.

- 지구 좌표계

지구 중심을 원점으로 하고 경도와 적도가 만나는 점을 X축으로, 북극을 향하는 축을 Z축으로, X축에서 반시계 방향으로 90도 회전한 방향을 Y축으로 하는 좌표계이다.

- 항법 좌표계

이동 객체의 중심을 원점으로 하고 X, Y, Z 세 축을 북(N), 동(E), 수직 아래(D)로 삼는 NED 지역 좌표계이다.

- 동체 좌표계

관성센서가 부착된 동체의 좌표계로서 전방을 X축, 오른쪽 방향을 Y축, 동체의 수직 아래방향을 Z축으로 한다.

관성 항법 시스템에서는 초기정렬이 필요 하다. 초기 정렬이란 이동 객체가 위치 추정을 시작하기 전에 정지 상태에서 이동 객체의 자세를 구하는 것이다. 정지 상태에서의 동체 가속도 f^b 는 다음과 같이 측정 된다.

$$f^b = [f_x \ f_y \ f_z] = C_b^n f^n = \begin{bmatrix} \sin\theta g^n \\ -\sin\phi \ \cos\theta g^n \\ -\cos\phi \ \cos\theta g^n \end{bmatrix} \quad (1)$$

여기서 정지상태의 가속도는 $f^n = [0 \ 0 \ -g^n]^T$ 이고, g^n 는 중력가속도의 크기를 나타내며, ϕ, θ, ψ 는 각각 자세각의 롤(roll), 피치(pitch), 요(yaw) 각이다. 이를 통하여 초기 좌표변환 행렬 $C_b^n(0)$ 을 구한다.

$$C_b^n = \begin{bmatrix} c(\theta)c(\psi) & s(\phi)s(\theta)c(\psi) - c(\phi)s(\psi) & c(\phi)s(\theta)c(\psi) - c(\phi)s(\psi) \\ c(\theta)s(\psi) & s(\phi)s(\theta)s(\psi) - c(\phi)c(\psi) & c(\phi)s(\theta)s(\psi) - s(\phi)c(\psi) \\ -s(\theta) & s(\phi)c(\theta) & c(\phi)c(\theta) \end{bmatrix} \quad (2)$$

$c = \cos \quad s = \sin$

동체 좌표계에서 측정된 선형 가속도 f^b 와 초기 정렬을 통해 얻은 좌표 변환 행렬 C_b^n 을 이용하여 항법 좌표계에서 가속도 f^n 으로 변환시킨 다음 항법 방정식을 계산하여 항체의 위치와 속도를 얻는다. 속도를 계산하기 위한 방정식은 다음과 같다.

$$v^n = C_n^b f^b - (2w_{ie}^n + w_{en}^n) \times v^n g^n \quad (3)$$

여기서 $v^n = [(v_N \ v_E \ v_D)]^T$ 는 항법 좌표계에서 표현된 속도 성분이다. w_{ie}^n 과 w_{en}^n 는 각각 항법 좌표계에서 표현된 항법좌표계의 관성좌표계에 대한 각속도와 지구 고정 좌표계의 관성 좌표계에 대한 각속도를 나타내며, g^n 은 항법 좌표계에서의 중력 가속도를 나타낸다. (식 3) 에서 속도와 위치를 구하는 과정은 이동 객체의 운동으로 인해 발생하는 Coriolis 항과 중력항을 보정하고 적분하여 얻는다. w_{ie}^n 과 w_{en}^n 는 다음과 같이 얻을 수 있다.

$$w_{ie}^n = [\Omega \cos L \ 0 \ -\Omega \sin L] \quad (4)$$

$$w_{en}^n = [\rho_N \ \rho_E \ \rho_D] = [l \cos L - L \ l \sin L] \\ = \left[\frac{v_E}{(R_t + h)} - \frac{v_N}{(R_m + h)} - \frac{v_E \tan L}{(R_t + h)} \right] \quad (5)$$

$$R_m = \frac{R_0(1 - e^2)}{(1 - e^2 \sin^2 L)^{3/2}}, \quad R_t = \frac{R_0}{(1 - e^2 \sin^2 L)^{1/2}} \quad (6)$$

여기서 L 은 좌표계에서 X축의 값, l 은 Y축, h 는 Z축의 값이다. R_0 는 지구타원체의 적도반경, e 는 지구 이심률, Ω 는 지구 자전각속도를 나타낸다. 위의 (식 3) 에서 구해진 속도 v^n 을 이용하여 항체의 위치인 L, l, h 좌표 값을 다음과 같이 계산하여 얻을 수 있다.

$$L = \int_0^t \frac{v_n}{(R_m + h)} dr + L(0) \\ l = \int_0^t \frac{v_E}{(R_t + \cos L)} dr + l(0) \quad (7) \\ h = \int_0^t (-v_D) dr + h(0)$$

2. RFID를 이용한 위치 초기화

본 논문에서 사용하는 관성센서를 통한 위치 추정기법은 사용자의 초기 위치를 기준으로 관성센서로부터 얻은 데이터를 통하여 이동한 방향과 거리를 계산하고 이를 초기위치에 더하여 위치를 계산한다. 따라서 초기 위치를 구하고 이를 원점으로 설정해 주어야 상대적인 위치가 아닌 현실 공간에 대응하는 절대적인 위치를 얻

고 활용할 수 있다. 새로운 위치로 현재 위치를 초기화하므로 오차도 초기화 되어 오차 누적 문제 또한 초기화 된다. 위치를 초기화하기 위해 RFID를 이용한 근접 위치 측정법을 사용한다.

RFID는 사람이나 사물을 식별할 수 있는 정보를 무선 주파수를 이용하여 접촉 없이도 정보를 전달하는 기술을 말한다. RFID의 목적은 재난재해관리, 생산관리, 물류시스템, 재고관리 등을 목적으로 개발되었다. 그러나 비접촉식으로 정보를 저장하고 전달하는 특징으로 경로 추적, 도난 방지 등의 분야에도 활용되고 있다. RFID를 활용한 근접 측정 방식의 원리는 이동 객체에 부착된 태그가 RFID 리더에 감지되면 감지된 위치를 현재의 위치로 삼는다. 태그가 감지된 시간순서에 따라 이동 객체가 이동한 경로의 파악 및 방향 등을 추측할 수 있다. 고정적인 위치 정보를 얻을 수 있다는 장점이 있으나 정확도를 높이기 위해서는 설치된 태그의 밀집도가 높아야 한다. 하지만 태그의 밀집도가 높아지면 반대로 서로간의 간섭으로 인하여 태그의 인식률이 떨어진다[4].

[그림 1]에서는 본 논문에서 제안하는 위치 초기화 방법에 대해 보여준다. RFID 리더를 사용하기 위해 초기화 명령을 수행한 후 태그가 리더에 인식되기를 기다린다. 리더에 태그가 인식된다면 인식된 태그의 ID를 태그의 위치를 저장해 놓은 테이블에서 검색한다. 만약 해당하는 ID가 테이블 내에 존재한다면 저장되어있는 ID의 위치를 가지고 온 뒤에 이를 현재의 위치 값으로 설정한다.

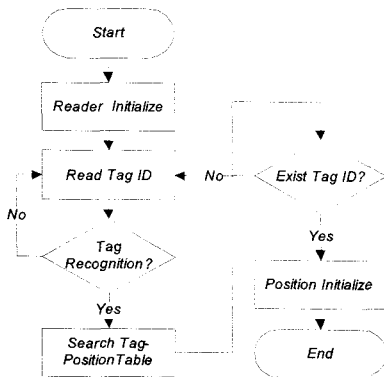


그림 1. RFID 위치초기화 방법

3. 관성 측정 유닛

관성 측정 유닛은 [그림 2]와 같이 가속도 센서, 각속도 센서, RFID 리더, 제어부(Control Unit) 그리고 통신부(Communication Unit)로 구성된다.

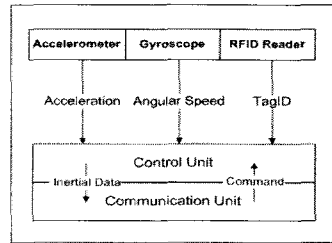


그림 2. 관성 측정 유닛 구조도

가속도 센서와 각속도 센서는 각각 X, Y, Z축에 대응하기 위해 각 3축으로 구성되며, 각 센서에서 측정되는 가속도와 각속도 그리고 위치초기화를 위한 RFID 리더는 제어부에게 명령을 전달 받고 RFID 태그의 정보를 수집하여 통신부를 통하여 무선으로 서버로 전송된다.

III. 상황정보 모델 및 음성 단어모델 도메인 구성

보통의 음성인식 시스템에서는 단어모델 데이터베이스만을 사용하여 음성인식을 수행하지만 본 시스템에서는 유비쿼터스 컴퓨팅에서 사용되는 상황 정보 모델을 음성 인식의 환경정보로 사용하여 음성 인식 성능을 향상시켰다. 상황 인식 시스템의 핵심 기능 중의 하나는 상황 인식 응용에서 이용가능한 사용자의 주변 환경에 관한 정보를 표현하기 위한 형식적인 상황 정보 모델 및 이를 관리하고 이용하는 방법을 제공하는데 있다[5]. 음성 인식에서 상황정보를 사용하기 위해서 보통의 상황 인식 컴퓨팅 환경에서 이용되는 다양한 상황 정보들 중에서 음성 인식에 영향을 주는 정보만을 선택하여 상황정보를 구성하고 정의 한다. 이렇게 정의된 상황정보를 기반으로 단어모델 도메인을 구축하여 음성 인식에 사용한다.

1. 음성인식 상황정보 모델

상황정보 기반의 음성 인식 시스템을 위한 상황정보의 구성은 사용자 정보(성별, 나이), 노이즈, 오브젝트, 그리고 위치 정보로 구성된다.

사용자 정보, 노이즈, 오브젝트 그리고 위치 정보는 미리 온톨로지(Ontology)로 정의하고 센서로부터 데이터를 입력받는다. 노이즈 정보는 무선 통신이 가능한 센서로부터 OSGi 프레임워크와의 통신으로 데이터를 주고받는다. 사용자 정보 및 위치정보는 사용자가 지니고 있는 관성 측정 유닛을 통해 실시간 지그비(Zigbee) 통신으로 획득 된다.

아래의 [표 1]에서는 상황정보 구성을 온톨로지 모 델링하기위해 성별은 남자/여자, 연령은 5등급, 노이즈는 낮음/보통/높음으로 정의한다. 노이즈 정보는 사용자로부터 발화된 음성의 정규화에 사용된다. [표 2]에서는 오브젝트 구성을 가정에서 대표적으로 사용하는 여섯 가지의 전자제품으로 구성하고 음성 인식 서비스가 제공되는 공간을 홈으로 가정하여 사용자가 위치할 수 있는 지역을 Balcony, Bathroom, Bedroom, Guestroom1, Guestroom2, Kitchen, 그리고 Livingroom 으로 제한한다.

표 1. 사용자 정보 및 노이즈 정보

Sex	Age		Noise	
class	num.	class	Num.(dB)	class
Male	0~7	Infant	20~39	Low
	8~11	Child		
	12~17	Young Adult	40~59	Normal
Female	18~61	Adult	60~	High
	62~	Old Adult		

표 2. 사용자 위치 및 사물 정보

Object	Location
class	class
Computer	Balcony
Television	Bathroom
Radio	Bedroom
Refrigerator	Guestroom1,2
Washing machine	Kitchen
Electric Lamp	Livingroom

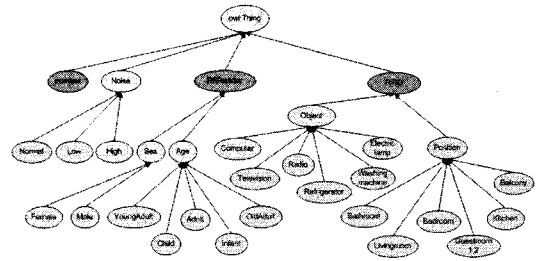


그림 3. 음성 인식 시스템을 위한 상황정보 온톨로지

본 상황정보 기반의 음성 인식 시스템에서는 상황의 정확한 표현, 다양한 관계 기술 등 명시적 규정을 위하여 시맨틱 웹(Semantic Web)에서 사용되는 온톨로지 로 상황을 정의 한다. 특히, 도메인 구성 및 음성 인식에 사용될 상황정보는 온톨로지 언어인 OWL(Web Ontology Language)로 구성한다. [그림 3]은 본 시스템의 OWL로 작성된 온톨로지 클래스의 구성 및 계층관계를 보여준다.

2. 음성인식 단어모델

음성인식에 사용되는 가장 일반적인 방법은 패턴 인식에 의한 방법으로 템플릿 기반의 패턴 매칭 방법을 이용하는 DTW(Dynamic Time Warping)와 통계적 패턴 인식을 이용한 HMM(Hidden Markov Model)이 있다. HMM은 수학적인 배경에서 개발된 알고리즘으로 1975년 음성 신호 처리 분야에 도입되어 고립단어 인식에서부터 대화체 음성 인식까지 폭 넓게 사용되는 알고리즘이다. 또한 신경 회로망을 이용한 방법도 음성 인식에 이용되나 많은 계산량 등의 문제로 최근에는 HMM과 결합한 형태를 많이 취하고 있다.

본 연구에서는 음성인식 알고리즘으로 패턴인식 방법인 HMM을 도메인에 따라 분류하여 사용한다. HMM 단어모델의 학습방법으로는 Baum-Welch를 사용한다. 또한 Vitervi 알고리즘을 사용하여 HMM 단어 모델의 확률 값을 계산한다.

HMM 단어모델의 파라미터는 상태간의 천이확률, 상태에 종속된 출력확률, 상태의 초기 존재확률로 구성 된다. HMM 단어모델의 파라미터를 간단하게 표현하면 (식 8)과 같다. A 는 상태천이 확률 분포, B 는 관측심볼 확률 분포 이고, π 는 초기상태 분포이다.

$$\lambda = \langle A, B, \pi \rangle \quad (8)$$

본 논문에서는 위의 (식 8)을 표현하기 위해 [그림 4]와 같이 5개의 상태를 가지는 Bakis Model의 변형된 형태를 사용한다. HMM 단어모델 Topology의 특징은 다섯 개의 STATE를 갖고 First-Order-Markov Chain 모델이며, 각 STATE마다 자기 이동 전위 및 다음 상태 전위 확률을 가진다. 또한 마지막 STATE는 DUMMY STATE로 이동 전위 확률을 갖지 않고 각 STATE는 512개의 OBSERVATION SYMBOL 확률을 가진다.

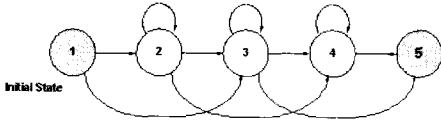


그림 4. 사용자 기반 유사도 측정 개념도

[그림 5]는 컴퓨터 활용 단어중 하나인 “aljip” HMM의 내부 확률 값을 보여준다.

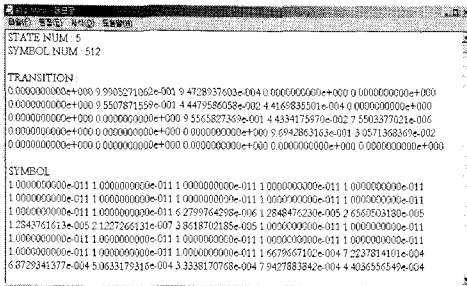


그림 5. 5 state와 512 symbol을 가지는 단어모델

3. 음성인식 단어 도메인

본 음성 인식 시스템에서는 크게 세 개(상위, 중위, 하위)층의 도메인을 구성한다.

상위 층으로, 발생자의 정보, 발생위치에 따라 HMM 도메인을 구성 한다. 사용자에게 부착된 관측 측정 유닛과 집안 곳곳에 부착되어있는 RFID를 통해 발생자의 정보와 위치를 파악한다. 상위 계층의 HMM 도메인은 크게 사용자의 성별, 나이에 따라 구성하고 하위로 사

용자의 위치로 구성한다.

중위 층으로, 관측열 길이에 따른 도메인을 구성한다. 본 음성 인식 시스템에서 사용하는 단어모델의 경우 발화 단위와 인식 단위가 동일하기 때문에 발화 길이에 따라서 생성되는 관측열의 길이도 변하게 된다. 따라서 동일한 HMM 토폴로지를 발화 길이가 상이한 고립단어에 동일하게 적용할 경우 인식을 저하가 발생하며, 인식 단어가 증가함에 따라 인식 속도가 낮아지게 된다.

하위 층으로, 음성인식 시스템의 성능 향상을 위해서 상황 정보 도메인, 관측열 도메인 구성과 함께 음절 개수 도메인을 구성한다. 음성의 모음부는 자음부와 달리 주기적인 특성을 가지고 있다. 따라서 모음부의 신뢰성 있는 검출을 통하여 음절 개수 도메인을 구성, 음성인식 시스템의 성능을 향상 시킬 수 있었으며, 향후 모음인식을 통한 음소 단위의 음성인식 시스템 개발이 가능하게 된다. 음절개수는 고립단어의 주파수, 포먼트 특징 추출 자료의 분석을 통해 만들어진다.

IV. 음성 인식 시스템 설계

본 장에서는 2장과 3장에서 제시한 사용자 위치추정 및 단어 도메인을 바탕으로 자바 기반의 OSGi(Open Service Gateway Initiative), 프레임워크 기반의 스마트 홈 환경에서 높은 성능을 보이는 음성 인식 시스템을 설계하고 구현한다.

본 시스템에 사용되는 OSGi는 가정 정보 기기 및 보안 시스템과 같은 인터넷 장비의 표준 연결 방법을 위해 OSGi 단체가 제안한 산업체 표준안을 말한다[5]. 개방형 자바 임베디드 서버인 JES 기반의 게이트웨이 소프트웨어 플랫폼 응용 소프트웨어 등에 전혀 구애 받지 않고 보안 기능이 우수한 멀티 서비스를 장치나 설비에 서비스 할 수 있는 기능이 있다. 특히, 블루투스, HAVi, 홈 PNA, 홈 RF, USB, VESA 등 다양한 유무선 네트워크 기술을 수용하는 개방형 네트워크 기술이다.

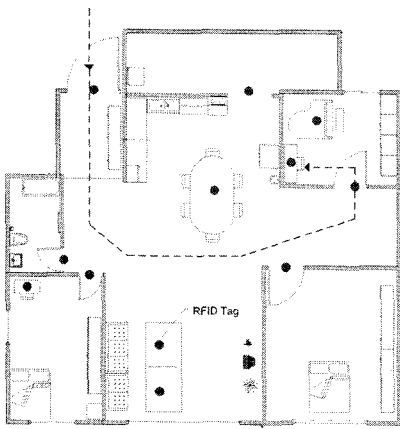


그림 7. 실험 공간 구성도

사용자 위치 추정 실험은 [그림 7]과 같이 사용자가 현관문을 통해 실내 공간에 진입하여 RFID 태그를 통하여 초기위치를 설정한 후 이동하면서 가속도와 각속도 데이터 및 태그의 ID를 수집하여 위치 계산 모듈로 수집된 데이터를 전송한다. 위치 계산 모듈은 수집된 데이터를 기반으로 사용자의 위치를 계산 한다.

[그림 8]에서는 OSGi상에서 각 번들이 동작하는 모습을 보인다. 각 번들은 관성 측정 유닛으로부터 들어온 데이터를 수집하는 관성 측정 유닛 Driver 번들과 태그의 위치를 검색하고, 검색된 위치를 현재의 위치로 설정하여 초기화를 수행하는 RFID 태그 관리자 번들, 그리고 위치 계산을 수행 하는 위치 계산 번들 등이 있다.

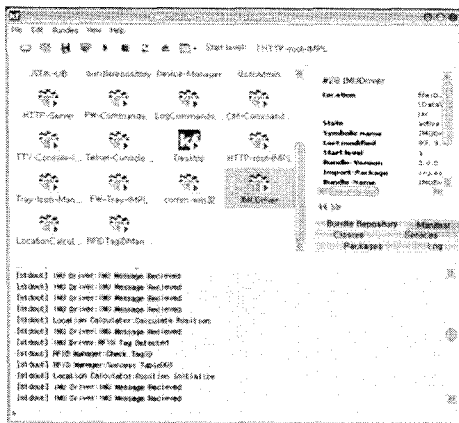


그림 8. OSGi 기반의 데이터 수집 번들

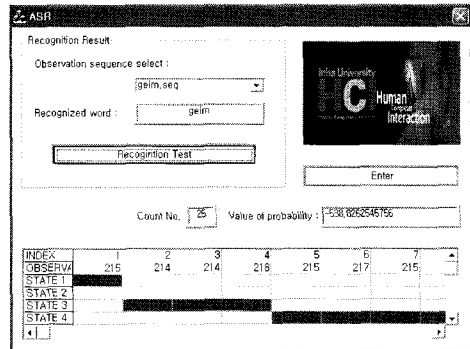


그림 9. HMM 상태 천이와 확률값에 따른 단어 검출

음성 인식 실험에서는 우선 컴퓨터 활용단어 25개를 대상으로 HMM과 인식 알고리즘의 정확성을 테스트 한다. [그림 9]는 HMM 훈련에 사용된 관측열중 하나를 대상으로 25개의 HMM 단어모델 중에서 가장 확률이 높은 모델을 찾는 실험을 보여준다.

음성 인식 시스템의 효율성을 실험하기 위해 화자 3명이 홈 환경에서 녹음한 컴퓨터 및 가전제품 활용 단어 100개를 대상으로 스마트 홈 공간에서 실험을 수행하였다. 실험 데이터는 16kHz로 샘플링 되고, 16bit로 A/D 변환하였다.

[그림 10]은 노이즈 정보에 따라서 발화 음성을 정규화 하였을 때의 인식률을 보여준다. 실험은 화자 1명이 노이즈가 Low, Normal, High일 때 100단어를 발화하여 인식률을 측정하였다. 노이즈가 적을 때는 인식률 차이가 거의 없었지만, 노이즈가 많을 때는 16%의 큰 인식률의 차이를 보였다.

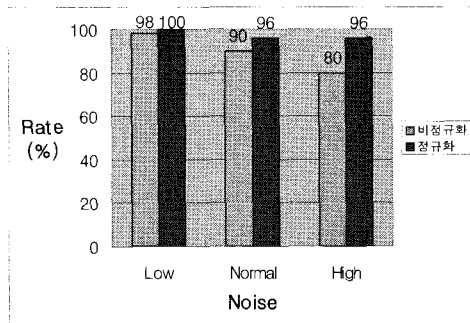


그림 10. 노이즈 정보의 사용에 따른 인식률

[그림 11]은 계층별로 도메인을 구성하였을 때에 인식 성공률 차이를 보여준다. 화자 3명이 스마트 홈 환경에서 위치를 변경하며 도메인 구성에 따라 각각 100단어씩 화자 한명 당 총 400단어를 발화하여 인식률을 측정하였다.

실험 수행결과 HMM 도메인을 구성하지 않았을 때 평균적으로 95%정도의 인식 성공률을 가지는 시스템이 사용자 정보와 위치 정보에 따른 HMM 도메인을 구성하여 2%의 인식률 향상을 보였다. 또한 관측열에 따른 도메인과 모음 개수에 따른 도메인을 추가로 구성하여 각각 1% 정도 인식률이 증가됨을 확인 하였다. 결과적으로 3개 층으로 도메인을 구성할 때 가장 인식률이 높은 것을 확인할 수 있다. [그림 10]과 [그림 11]의 결과에 따라서 잡음 환경에서는 능동적으로 노이즈 정보를 획득하여 잡음 환경에 강인한 시스템을 개발할 수 있다. 본 연구에서와 같이 사용자 위치정보와 사용자정보를 기반으로 단어모델 도메인을 구성할 경우, 인식률이 향상되고 화자독립의 대응량 고립단어 음성 인식기의 개발이 가능함을 보였다.

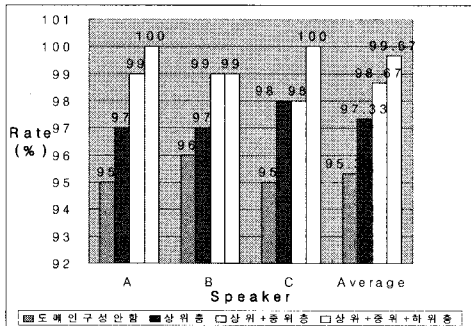


그림 11. HMM 도메인 구성에 따른 인식률

VI. 결론

기존의 애플리케이션 또는 가전제품의 컨트롤을 목적으로 상용화 되어있는 고립 단어 음성 인식 시스템은 화자의 음성만을 사용하기 때문에 제한적인 환경에서만 높은 인식 성능을 보인다. 이는 사용자의 발화상태와 발화 목적을 음성 인식 시스템이 파악하지 못하기

때문이다. 또한 단어모델을 사용하기 때문에 인식 단어가 증가하게 되면 단어모델 탐색 시간이 증가되고 인식률이 낮아지게 되는 단점을 가지고 있다.

본 논문에서는 스마트 홈 환경에서 발화자의 신원정보를 획득하고 발화 목적을 위치와 근처의 사물을 통해 파악하여 음성인식기의 성능을 향상시킨 시스템을 설계 및 구현 한다. 이를 위해 발화자의 상황정보를 성별, 나이, 잡음, 오브젝트, 사용자 위치로 구성하고 온톨로지로 정의한다. 사용자의 위치 정보는 RFID와 관성 센서를 함께 사용하여 장애물에 제약 없이 지속적으로 추정할 수 있는 시스템을 개발하여 사용한다. 상황정보에 적합한 음성을 인식하기 위해 본 시스템은 위치와 사용자정보 그리고 단어특성에 따른 단어모델 도메인을 구성한다. OSGi 기반의 상황인식 관리자에서 능동적으로 획득된 상황 정보는 음성 인식을 위한 단어모델 도메인 선택의 중요한 정보가 된다. 본 시스템의 성능평가 결과 스마트 홈 환경의 어느 위치에서도 기존의 단어모델 도메인을 사용하지 않는 음성인식 시스템 보다 높은 인식 성공률을 보였다.

향후 연구로는, 발화자의 의도를 정확히 파악할 수 있도록 다양한 센서를 통해 사용자의 상황을 자세히 모델링하고 이를 음성 인식 시스템에 적용하겠다. 또한 노이즈와 성별, 나이에 대한 음성 DB를 추가로 보완하여 지금의 시스템보다 잡음환경에 강인하고 성별과 나이에 제약 받지 않는 시스템을 개발하겠다.

참고 문헌

- [1] 김성룡, 스트랩다운 관성항법시스템의 초기정렬 알고리즘 구현, 석사학위 논문, 인하대학교, 2000(2).
- [2] 박명훈, 심현민, 이용혁, 홍승홍, “관성센서를 이용한 노약자의 개인항법시스템의 설계 및 구현,” 대한전자공학회, 하계종합학술대회, 제27권, 제1호, pp.1245-1248, 2004(6).
- [3] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*, IET, 2004.

[4] 문승욱, 지용관, 박장현, "정적 RFID 수동태그와 이동로봇의 상대위치인식에 대한 기법연구", 한국정밀공학회, 추계학술대회 논문요약집, 2005.

[5] L. Gong, "A Software Architecture for Open Service Gateways," IEEE Internet Computing, Vol.5, No.1, pp.64-70, 2001.

[6] P. Dobrev, D. Famolari, C. Kurzke, and B. A. Miller, "Device and Service Discovery in Home Networks with OSCi," IEEE Communications Magazine, Vol.40, Issue8, pp.86-92, Aug. 2002.

[7] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition," in Proc., IEEE, Vol.77, No.2, pp.257-286, Feb. 1989.

[8] Mark Weiser, "The Computer for the Twenty-first Century." Scientific American, Vol.265, No.3, pp.94-104, 1991.

[9] P. J. Brown, J. D. Bovey, and X. Chen, "Context-Aware Application: From the Laboratory to the Marketplace," IEEE Personal Communication, pp.58-64, 1997.

[10] P. Bellavista, A. Corradi, and C. Stefanelli, "Mobile Agent Middleware for Mobile Computing," IEEE Computer, Vol.34, No.3, Mar. 2001.

[11] Liu, T. and Martonosi, M., "Impala: A Middleware System for Managing Autonomic, Parallel Sensor Systems," ACM SIGPLAN Symp. Principles and Practice of Parallel Programming, June. 2003.

[12] <http://www.w3.org/2004/OWL/>

[13] Strang, T. and Linnhoff-Popien, C., "A Context Modeling Survey." UbiComp 1st International Workshop on Advanced Context Modelling, Reasoning and Management, Nottingham, pp.34-41, 2004.

[14] T. Gu, H. K. Pung, and D. Q. Zhang, "An Ontology-based Context Model in Intelligent

Environments," Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference, pp.270-275, 2004.

저 자 소 개

김 중 훈(Jong-Hun Kim)

정회원

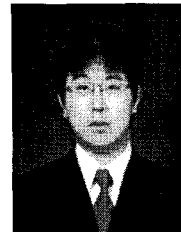


- 2001년 2월 : 인천대학교 물리학과(학사)
- 2003년 2월 : 인하대학교 전자계산공학과(공학석사)
- 2007년 8월 : 인하대학교 컴퓨터정보공학과 박사수료

<관심분야> : 음성인식, 임베디드/유비쿼터스 시스템, 데이터마이닝, 인공지능

심 재 호(Jae-Ho Sim)

준회원

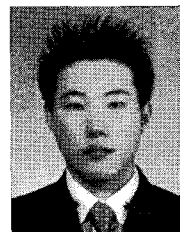


- 2006년 2월 : 인하대학교 컴퓨터공학과(학사)
- 2006년 ~ 현재 : 인하대학교 컴퓨터정보공학과 석사과정

<관심분야> : 임베디드시스템, USN, 홈네트워크

송 창 우(Chang-Woo Song)

정회원



- 2004년 : 한국교육개발원 학점은행제 컴퓨터공학 전공(학사)
- 2007년 2월 : 인하대학교 전자계산공학과(공학석사)
- 2007년 4월 ~ 현재 : 인하대학교 컴퓨터정보공학과 박사과정

<관심분야> : 임베디드/유비쿼터스 시스템, 상황인식, 데이터마이닝

이 정 현(Jung-Hyun Lee)

정회원



- 1977년 : 인하대학교 전자공학과 (공학사)
- 1980년 : 인하대학교 대학원 전자 공학과(공학석사)
- 1988년 : 인하대학교 대학원 전자 공학과(공학박사)

- 1979년 ~ 1981년 : 한국전자기술 연구소 시스템 연구원
- 1984년 ~ 1989년 : 경기대학교 전자계산학과 교수
- 1989년 ~ 현재 : 인하대학교 컴퓨터공학부 교수

<관심분야> : 자연어처리, HCI, 정보검색, 컴퓨터구조