논문 2006-43CI-6-13

심리음향 모델을 이용한 무선 음성인식 시스템

(Wireless Speech Recognition System using Psychoacoustic Model)

노 진 수*, 이 강 현**

(Jin Soo NOH and Kang Hyeon RHEE)

요 약

본 논문에서는 무선 음성 센서를 사용하여 스위치 제어나 생체신호 인증과 같은 유비쿼터스 센서 네트워크 응용 서비스를 지원하기 위한 음성인식 시스템을 구현하였다. 제안된 시스템은 무선 음성센서와 심리음향 모델을 이용한 음성인식 알고리즘과 에러정정을 위한 LDPC(Low Density Parity Check) 모듈로 구성된다. 제안된 음성인식 알고리즘은 센서의 소비 에너지를 효율적으로 사용하기 위하여 호스트 컴퓨터에 삽입되며, 음성인식의 정확도를 향상시키기 위하여 전방향 에러정정 알고리즘을 사용하였다. 또한, 효율적으로 무선채널의 잡음을 제거하고 무선채널 에러를 정정하기위하여 실험 환경과 실험 계수를 최적화 하였다. 결과적으로, 센서와 음원 사이의 거리가 1.0m 이하 일 때 FAR 0.126%와 FRR 7.5%를 얻었다.

Abstract

In this paper, we implement a speech recognition system to support ubiquitous sensor network application services such as switch control, authentication, etc. using wireless audio sensors. The proposed system is consist of the wireless audio sensor, the speech recognition algorithm using psychoacoustic model and LDPC(low density parity check) for correcting errors. The proposed speech recognition system is inserted in a HOST PC to use the sensor energy effectively and to improve the accuracy of speech recognition, a FEC(Forward Error Correction) system is used. Also, we optimized the simulation coefficient and test environment to effectively remove the wireless channel noises and correcting wireless channel errors. As a result, when the distance between sensor and the source of voice is less then 1.0m, FAR and FRR are 0.126% and 7.5% respectively.

Keywords : Speech recognition, Wireless audio sensor, LDPC, Psychoacoustic model

I. 서 론

무어의 법칙에 힘입어, 우리는 수많은 컴퓨터에 둘러싸여 있다. 여기에 최근 정보통신 기술의 비약적인 발전은 기존의 계산기로서의 컴퓨터가 아닌 정보단말기로의 컴퓨터로 발전하여 더욱더 우리의 생활에 밀접한 영향을 주고 있다. 이러한 기술의 진보는 유비쿼터스 컴퓨팅이라는 새로운 정보통신 혁명을 야기하게 되었고, 이런 사회발전의 흐름과 환경을 인간 친화적으로 만들고자하는 인간의 끊임없는 욕구가 맞물려 무선 센서 네트워크 (WSN: Wireless Sensor Network)의 필요성이 제기되고 있다.

무선 센서 네트워크란 센서가 달려 있어 센싱이 가능하고 센싱된 정보를 가공할 수 있는 프로세서가 달려 있으며 이를 전송할 수 있는 무선 송수신기를 갖춘 소형장치, 즉, 센서 노드로 구성된 네트워크를 의미하며, 기존의 네트워크와 다르게 의사소통의 수단이 아니라 환경에 대한 정보를 수집하는 것을 그 목적으로 한다. 이에 따라, 인간, 사물 그리고 컴퓨터가 유기적으로 연계되어 다양하고 편리한 서비스를 제공해 주는 유비쿼터스 컴퓨팅 환경에서, 외부 환경의 감지와 제어 기능을 수행하는 센서 네트워크 기술이 최근 활발히 연구되고 있다[1,2].

무선통신기술과 전자디바이스기술의 발전으로 말미암아 저가격, 저전력, 다기능 센서 노드로 구성된 무선 센서 네트워크에 대한 관심이 급격히 고조되고 있다[3]. 무선 센서 네트워크는 기존에 구축된 센서네트워크를 무선네

* 학생회원, ** 평생회원, 조선대학교 전자공학과
(Dept. of Electronic Engineering, Chosun University)
접수일자: 2006년9월21일, 수정완료일: 2006년10월30일

트워크로 대체 하는 기술로서, 각 센서노드는 센서에 의한 센싱, 센싱된 데이터의 처리, 멀티 홉을 통한 네트워킹 등의 기능을 가지고 있으며, 이는 기존의 전통적 의미의 센서에서 정보처리 능력의 향상을 의미한다[4]. 이러한 정보처리 능력의 향상에 힘입어 WSN의 응용 분야는 군대 정보의 센싱과 추적, 환경 모니터링, 환자 감시 그리고 스마트 환경 등의 분야에 적용되어지고 있다.

현재 센서 네트워크는 하드웨어와 소프트웨어 플랫폼의 개발과 응용분야의 발굴을 위하여 학계와 산업체에서 많은 연구가 이루어지고 있으며, 대표적인 센서 네트워크 연구 그룹으로는 센서노드용 하드웨어인 MICA[5]와 운영체계인 TinyOS[6]를 개발한 버클리 대학과 상업용 응용 시스템에 필요한 연구를 수행하는 Intel[7]을 들 수 있다.

본 논문에서는 음성 신호를 센싱할 수 있는 무선 음성 인식 센서를 구현하였으며, 구현된 센서를 통하여 센싱된 음성신호를 인식할 수 있는 알고리즘을 제안하였다. 설계된 무선 음성인식 센서를 동작시키기 위하여 UStar-2400[8]을 기본 플랫폼으로 사용하였다. UStar-2400는 ChipCon CC2420 RF 모듈을 사용하여 2.4GHz ZigBee 통신이 가능하며, 저 전력 프로세서인 Atmel Atmega128L 프로세서가 장착되어있다. 또한 음성 인식의 정확도를 향상시키기 위하여 심리음향 모델[9]을 이용한 음성 인식 알고리즘과 무선 채널 잡음에 강인성을 가지기 위하여 LDPC[10] 모듈을 설계하여 삽입하였다.

본 논문에서는 제안된 음성 인식 시스템의 성능 평가를 위해 FAR과 FRR을 측정하였으며, LDPC 모듈의 무선 환경에서 발생되어지는 외부 잡음에 대한 강인성을 측정하였다. 이를 위해 Ⅱ장에서는 USN 환경에서 이루어지는 생체 인식 시스템에 대하여 알아보고, Ⅲ장에서는 본 논문에서 음성 인식 알고리즘으로 사용된 심리음향 모델과 무선채널에서 잡음에 강인성을 가지는 LDPC에 대하여 설명하겠다. Ⅳ장에서는 음성인식 센서 및 제안 알고리즘의 H/W와 S/W적인 구현 방법을 기술한다. 그리고 Ⅴ장에서 제안된 알고리즘의 성능 측정 및 결과 검토를 하고 마지막 Ⅵ장에서 결론과 향후 연구방향에 대해 고찰한다.

Ⅱ. 이론적 배경

1. 심리음향 모델 (Psychoacoustic Model)

심리음향 모델은 무선 음성인식 센서로부터 전송되어 오는 오디오 데이터를 입력받아 본 논문에서 제안된 음성 인식 알고리즘에 필요한 정보를 제공해주는 역할을

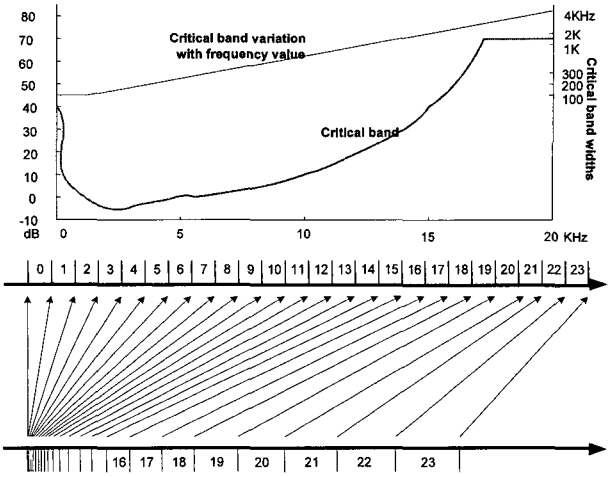

그림 1. 절대 가청 주파수
Fig. 1. Absolute threshold frequency.

표 1. 심리음향 모델 (14단계)
Table 1. Psychoacoustic Model (14steps).

순서	과 정
1	입력 샘플 값 재 구성
2	FFT 계산
3	예측된 값의 계산
4	비 예측 값의 측정
5	에너지와 각각 분학된 비 예측 값의 계산
6	에너지와 비 예측 값의 컨볼루션 실행
7	각각의 분할된 음조의 계산
8	각각의 분할된 SNR의 계산
9	각각의 분할된 Power Ratio의 계산
10	분할된 실제적인 에너지 쓰레스홀드의 계산
11	Quiet 상태에서의 쓰레스홀드와 pre-echo 조절
12	PE (Perceptual Entropy) 계산
13	블록 형태의 결정
14	각 SWB에서의 1/SMR의 계산

한다. 사람의 귀가 인지할 수 있는 신호 대역은 20Hz에서 20KHz 이며, 이를 절대 가청 주파수(Absolute threshold) 라고 하며, 아래 식 (1)과 같이 표현되고, 그림 1과 같은 값을 가진다.

$$ATH(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-33)^2} + 10^{-3}(f/1000)^4 \quad (1)$$

$ATH(f)$는 ATH안의 dB 값을 나타내고, f는 주파수 값이다. 이 주파수는 낮은 영역과 높은 영역에서 큰 값을 갖고 중간 대역(1KHz-5KHz)에서 낮은 값을 갖는다. 이것은 중간 대역에 음성의 중요한 정보가 들어 있다는 것을 의미한다. 위 곡선에서 가장 낮은 지점은 싸인 레벨의 ±1

LSB와 사운드 압력이 동일하다. 사람의 귀는 충분히 높은 주파수의 순간적인 소리는 일정시간이상 유지를 하지 못한다. 이 영역에서 특정 주파수에 따라 듣는 특성이 달라지며, 신호크기가 일정 수준 이상에 이를 때까지 그 소리를 듣지 못한다. 심리음향모델은 신호에너지에 의해 마스킹되는 최대왜곡에너지를 계산하는데 이 에너지를 쓰레스홀드(threshold)라고하며 입력신호를 주파수와 위상으로 분석한다. 입력된 오디오 데이터는 출력단에서 long 블록과 short블록으로 구분된다. 출력의 long블록, short 블록 안의 데이터 값들은 에너지 밀도, 채널대역폭, 채널수로 구성되어 있으며 이 값들을 이용하여 음성인식을 진행한다. 심리음향모델의 전체 14단계는 표 1과 같다.

2. LDPC

LDPC 부호는 최근에 가장 주목 받는 오류정정부호로 1960년대 초 Gallager에 의해 제안된 부호로, 패리티 검사 행렬에 0이 아닌 원소의 수가 부호의 길이에 비해 현저히 적게 존재하는 부호로 정의되며 섀논(Shannon) 한계에 가장 근접하는 오류정정부호로서, 터보부호와 더불어 제4세대 이동통신시스템에 활용될 수 있는 매우 우수한 오류정정부호로 평가되고 있다.
식 (2)은 LDPC 부호화 과정을 나타낸다.

$$H=(A_p^{-1}\cdot A)\bmod 2=[I\ \ A_2]$$
$$G=\left(\frac{A_2}{I}\right) \qquad (2)$$
$$c=(G\cdot m)\bmod 2$$

A, H : 패리티 체크 행렬
A_p^{-1} : 역피봇 행렬
G : 생성 행렬
m : 전송 메시지
c : 부호어

패리티 체크 행렬 A와 H를 이용하여 생성 행렬 G를 만든 다음 G와 메시지 m을 사용하여 부호어 c가 생성되어진다.

식 (3)는 LDPC 복호화 과정을 나타낸다.

$$q_n(x)=\alpha P(c_n=x|r_n)\Pi_{m\in n}P(z_m=0|c_n=x,r) \qquad (3)$$

$q_n(x)$: 가상 사후 확률
$\alpha P(c_n=x|r_n)$: 내부 확률
$\Pi_{m\in n}P(z_m=0|c_n=x,r)$: 외부 확률
z_m : 패리티 체크 비트s
c_n : 부호어
n, m : 행렬 인덱스
r : 전체 수신된 데이터

전송 채널을 통과한 부호어 c는 잡음 및 에러 성분이 첨가되어 r의 형태로 수신되고 수신된 신호는 식 (3)에서 채널의 사후확률 계산을 통하여 복호된다.

III. 제안된 알고리즘과 시스템 구현

본 논문에서는 USS-2400 ISP, USS-2400, 그리고 본 논문에서 설계된 음성인식 센서 모듈을 사용하여 건물 출입자를 인증할 수 있는 음성 인식 모듈을 제안하였다. 제안된 시스템의 전체 블록도는 그림 2와 같다.

센서를 통하여 수집된 음성 신호는 FFT를 사용하여 주파수 성분으로 변환시킨 후 심리음향 모델을 적용시켜 음성신호를 인증하는 데 유효한 24개의 음성 특성값을 추출한다. 추출된 데이터는 HOST PC에 저장된 데이터베이스와의 상관관계를 검사하여 출입자를 인증하는 시스템이며 전체 시스템은 JAVA기반으로 설계되었다.

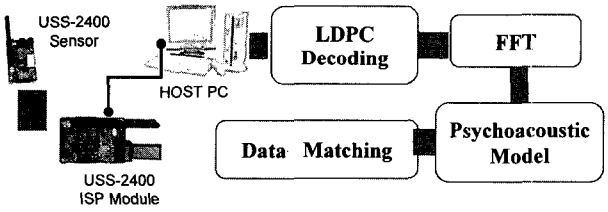

그림 2. 제안된 전체 블록도
Fig. 2. The proposed total block diagram.

1. USS-2400

USS-2400 ISP는 HOST PC와 USS-2400을 연결하기 위해 필요한 보드로 구성은 HOST PC와 연결하기 위해 병렬 케이블과 직렬 케이블 단자가 있고, USS-2400에 다운로드하기 위한 Download Port, POWER S/W, RESET PUSH BUTTON S/W, 상태 LED가 있다. 또한 5V 전압을 입력 받아 3.3V로 변환하여 사용하며, 직렬 케이블을 통하여 USS-2400을 디버깅할 수 있다.

USS-2400의 MCU는 Atmega128L, RF 모듈은 Chipcon의 CC2420을 사용하였다. 그림 3은 USS-2400의 블록도이다.

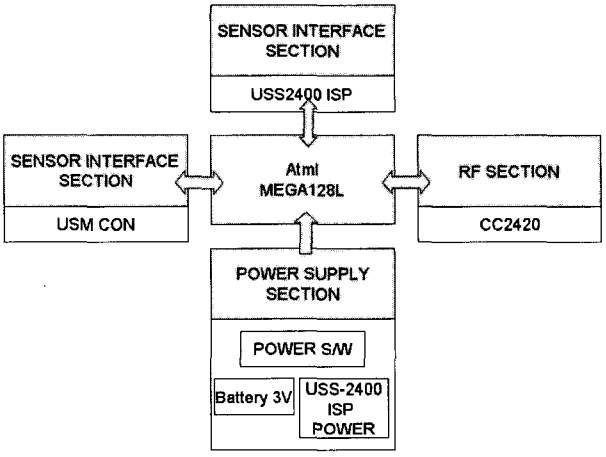

그림 3. USS-2400 블록도
Fig. 3. USS-2400 block diagram.

2. 제안된 무선 오디오 센서

그림 4는 본 논문에서 설계한 무선 오디오신호 전송 센서의 블록도이다. 전체 구성은 USS-2400의 데이터 전송 모듈, ADC, 오디오 센서, LDPC 생성 알고리즘으로 구현되어 있다.

LDPC 블록은 식 (2)에서 생성된 생성행렬 값과 ADC 블록을 통과한 디지털 신호를 이용하여 생성하였

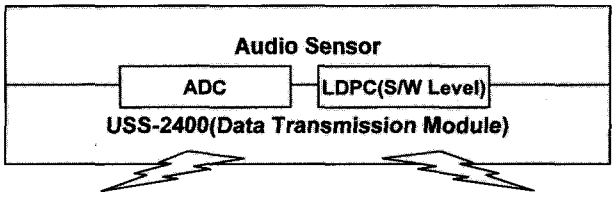

그림 4. 제안된 무선 오디오 센서 블록도
Fig. 4. The proposed wireless audio sensor block diagram.

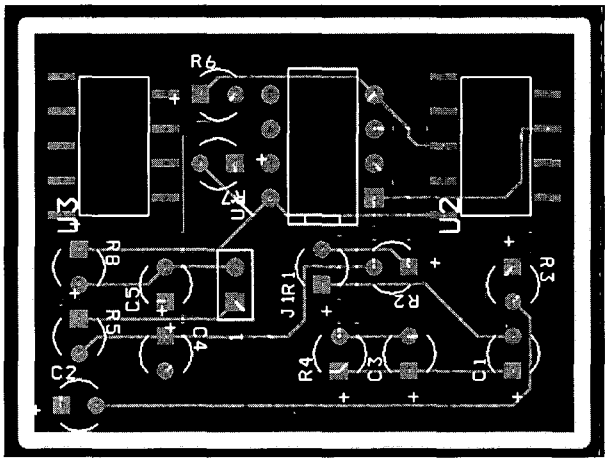

그림 5. 무선 오디오 센서의 PCB
Fig. 5. The PCB of wireless audio sensor.

으며 복호기는 JAVA 프로그램 상에서 구현하였다. 그림 5는 본 논문에서 설계한 PCB 보드이다.

3. 음성인식 알고리즘

음성인식은 FFT와 심리음향 모델을 사용하여 구현하였다. 음성 인식 대상자가 자기 이름을 말하면 센서를 통해서 HOST PC에 전달되고, 전달된 신호를 FFT 처리한 후 심리음향 모델을 적용하여 사람마다 가지는 고유 특성 값 24개를 추출한 다음 DB에 저장되어 있는 값과 비교하여 인증 여부를 결정한다. 사용된 심리 음향 모델은 식 (1)을 사용하였으며 표 2는 고유 특성 값을 추출하기 위해 사용한 채널의 주파수 범위 값이다.

표 2. 채널의 주파수 범위
Table 2. Frequency ranges of channels.

채널 번호	주파수 [Hz]	채널 번호	주파수 [Hz]	채널 번호	주파수 [Hz]
1	100	9	1,270	17	4,400
2	200	10	1,480	18	5,300
3	300	11	1,720	19	6,400
4	400	12	2,000	20	7,700
5	510	13	2,320	21	9,500
6	630	14	2,700	22	12,000
7	770	15	3,150	23	15,500
8	920	16	3,700	24	End

IV. 실험 및 결과 검토

USS-2400 ISP를 통하여 수집된 음성 신호는 JAVA로 구현된 오실로스코프 상에서 확인 가능하며 본 논문에서는 그림 6과 같이 실험실에 4개의 Audio Sensor를 배치하여 음성 인식 실험을 진행하였다.

ISP 모듈에서 수집된 데이터는 HOST PC로 전송이 되며 JAVA로 구현된 LDPC 디코더 블록을 통과한 후 오실로스코프 상에서 4개의 채널을 통해 들어오는 데이터를 확인할 수 있으며, 최종적으로 JAVA 프로그램 상에서 음성인식 시스템이 동작한다. 음성 인식 결과 본인이 확인되면 ISP 모듈을 통하여 ACK 신호가 전송되며 Audio Sensor에 있는 LED가 동작되게 된다.

센서의 동작은 주변에서 발생하는 소리가 일정 레벨 이하일 때는 슬립 모드로 동작하다가 일정 레벨 이상의 소리 신호에 대해서만 신호를 수집하는 형태로 동작한

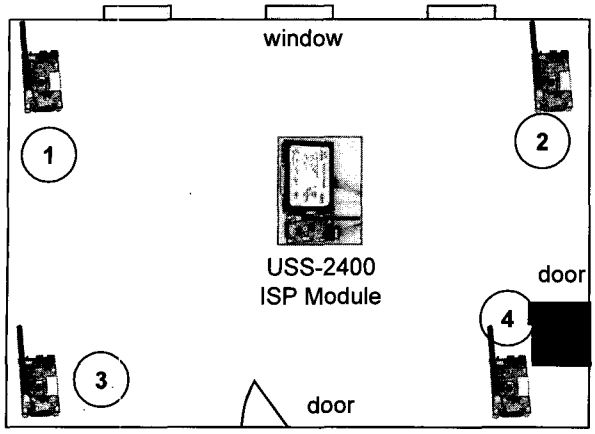

그림 6. 센서 배치도
Fig. 6. Deployment of sensor.

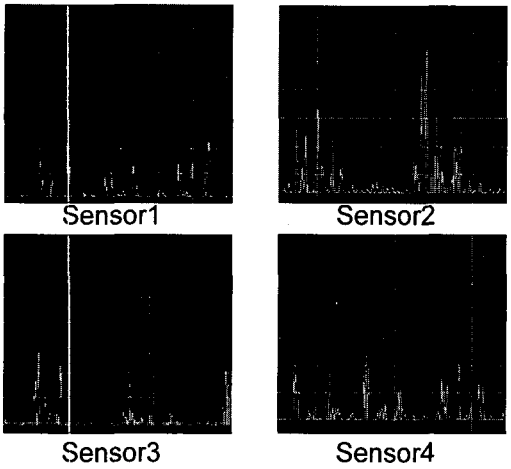

그림 7. 센서로부터 수집된 음성 신호
Fig. 7. The collected voice signal from sensors.

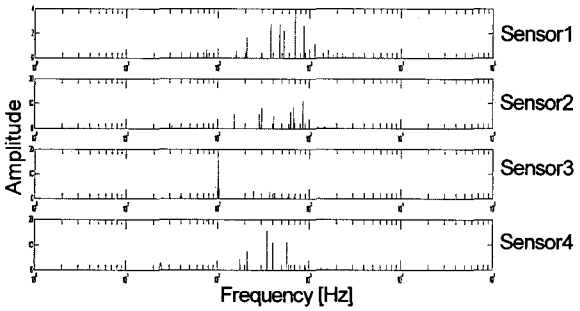

그림 8. 심리음향 모델을 사용하여 음성 신호로부터 추출된 특성 값
Fig. 8. The extracted value using psychoacoustic model from voice signal.

다. 그림 7은 LDPC 복호 후 오실로스코프에서 보여 지는 오디오 신호이며, 그림 8은 음성신호로부터 추출된 음성 특성값이다.

각각의 센서에서 수집된 음성신호로부터 계산된 24

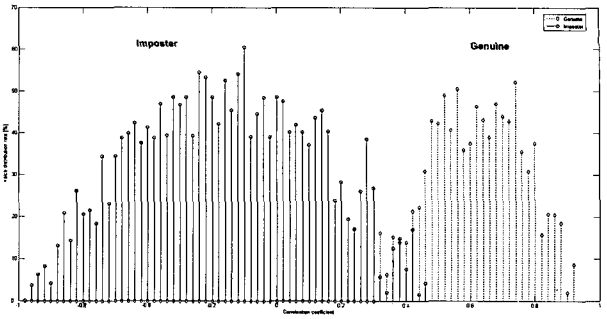

그림 9. 음성의 상관계수
Fig. 9. The correlation coefficient of voice.

개의 특성값과 데이터베이스에 저장된 특성값의 유사도를 비교하여 본인 인증 여부를 결정한다. 유사도의 비교는 식 (4)의 상관계수를 사용하였다.

$$k = \frac{\frac{1}{n}\sum_{m=1}^{n}(a_m - \bar{a})(b_m - \bar{b})}{\sigma_a \sigma_b} \tag{4}$$

k : coefficient of correlation(,$-1 \le k \le 1$)

$\bar{a}, \bar{b}$: average

σ_a, σ_b : standard deviation

그림 9는 센서를 통하여 입력된 음성을 본인 대 타인 매칭 시뮬레이션 한 결과 생성된 상관계수이다. 이 결과로부터 음성인식 임계값(Critical value)을 0.40으로 설정하여 실험을 진행하였다.

본 논문의 성능 평가에는 타인수락오류율(FAR : False Acceptance Ratio)과 본인거부오류율(FRR : False Rejection Ratio)을 사용하였다. 여기서 타인수락오류율은 입력 음성을 데이터베이스의 데이터와 비교했을 때 동일인의 음성이 아님에도 불구하고 동일인의 음성으로 잘못 판정한 비율이고, 본인거부오류율은 동일인의 음성임에도 불구하고 동일인의 음성이 아니라고 잘못 판정한 비율이다. 본 논문에서는 타인수락오류율(FAR)과 본인거부오류율(FRR)로서 식 (5)를 사용하였다.

$$FAR = \frac{\gamma_a}{\gamma_b} \qquad FRR = \frac{\delta_a}{\delta_b} \tag{5}$$

γ_a : 타인이 본인으로 오인식된 횟수

γ_b : 본인대타인 매칭 횟수

δ_a : 본인이 타인으로 오인식된 횟수

δ_b : 본인대본인의 매칭 횟수

표 3. FAR와 GAR 성능
Table 3. The FAR and GAR performances.

Critical Value	γ_a	FAR[%]	δ_a	FRR[%]
0.38	10	0.035	2	5.0
0.42	25	0.109	2	5.0
0.40	29	0.126	3	7.5
0.42	49	0.213	6	15.0

음성 인증 과정에서 임계값 범위 이내에 들어가는 데이터가 두개 이상 일 때에는 다시 한번 음성을 획득하여 음성을 재인식하는 구조를 가지도록 설계하였다.

표 3은 임계값의 변화에 따른 시스템의 FAR과 FRR을 나타낸다.

표 3에서와 같이 임계값이 0.40 일 때 FAR과 FRR로 각각 0.126[%], 7.5[%]를 얻었다.

또한, 센서와 음성발원지의 거리에 따라 음성인식률에 차이가 발생됨을 확인할 수 있었다. 실험결과 음성발원지로부터 1.0m 이내에 음성 인식 센서가 있을 경우에는 인식률에 큰 차이가 없으나 1.5m 이상부터는 인식률이 급격하게 떨어짐을 확인할 수 있었다.

V. 결 론

유비쿼터스 환경에서 휴먼 인터페이스 기술 중 하나인 음성인식 기술의 필요성이 증가하고 있다. 이에 따라 본 논문에서는 무선 음성인식 센서를 이용하여 출입자의 음성을 인식할 수 있는 시스템을 설계하였다. 하드웨어는 USS-2400 ISP, USS-2400, 그리고 본 논문에서 설계한 무선 음선인식 센서와 채널잡음을 줄이기 위하여 설계한 LDPC 부/복호기로 구성되며, 소프트웨어는 심리음향모델, FFT, 그리고 데이터의 인증 알고리즘으로 구성되어 있다.

본 논문에서 구현된 시스템은 1.0m 이내의 거리에서 FAR 0.126%, FRR 7.5%의 성능 즉, 92.5%의 인식률을 가진다. 하지만 1.5m 이상의 거리에서는 센서와 음원 사이의 거리에 비례하여 성능이 급격히 떨어짐을 확인할 수 있었다.

또한, 본 논문에서 구현된 시스템은 대부분의 처리를 베이스 스테이션에 의존하기 때문에 센서노트의 개수가 증가할수록 베이스 스테이션에 오버헤드가 커지는 단점이 존재하며, 또한 센서와 음원과의 거리에 비례하여 성능이 급격히 떨어지는 문제점을 내포하고 있다. 앞으로는 이러한 문제점을 해결할 수 있는 방향의 추가적 연구와 무선 영상 센서를 이용한 영상인식 시스템에 대한 연구가 이루어져야 할 것이다.

참 고 문 헌

[1] H. Wang, D. Estrin, and L. Girod, "Preprocessing in a tiered sensor network for habitat monitoring," EURASIP JASP special issue of sensor networks, pp. 392-401, 2003.

[2] S. Shukla, N. Bulusu, and S. Jha, "Cane-toad monitoring in kakadu national park using wireless sensor networks," in Proceedings of APAN, Cairns, Australia, July 2004.

[3] Paramvir Bahl and Venkata Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," Proc. of IEEE INFOCOM, vol. 2, p.p. 775-784. March 2000.

[4] Nissanka B., Priyantha, Anit Chakraborty, Hari Balakrishnan, "The cricket location-support system," Proc. of MOBICOM 2000, p.p.32-43, Boston, MA, Aug. 2000, ACM, ACM Press.

[5] Ian D. Chakeres and Luke Klein-Berndt, "AODVjr, AODV Simplified," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 6, no. 3, Jul. 2002, p.p.100-101.

[6] E. H. Callaway, "Wireless Sensor Networks Architectures and Protocols," Auebach, 2003.

[7] MICA2, http://www.xcross.com

[8] TinyOS, http://webs.cs.berkeley.edu.

[9] Intel, http://www.intel.com/research/exploratory/wireless-sensors.htm#sensornetwork.

[10] Ustart-2400, http://www.huins.com

[11] "ISO/IEC MPEG-2 Advanced Audio Coding 382(N-1)"-Presented at the 101st Convention 1996 November 8-11 Los Angeles, California, AN AUDIO ENGINEERING SOCIETY PREPRINT, 1996.

[12] R.G. Gallager, "Low-density parity-check codes," IRE Trans. Inform. Theory, vol. IT-8, pp. 21-28, Jan. 1962.

저 자 소 개

노 진 수(학생회원)
2002년 조선대학교 전자공학과 학사졸업.
2004년 조선대학교 전자공학과 석사졸업.
2006년 조선대학교 전자공학과 박사과정.
<주관심분야 : UWB, 생체인식, 영상신호처리, Ubiquitous Sensor Network>

이 강 현(평생회원)-교신저자
1979년, 1981년 조선대학교 전자공학과 공학사 및 석사
1991년 아주대학교 대학원 공학박사
1977년~현재 조선대학교 교수
1991년, 1994년 미 스탠포드대 CRC 협동연구원.
1996년 호주 시드니대 SEDAL 객원교수
2000년~현재 한국 멀티미디어기술사협회 이사
2002년 영국 런던대 객원 교수
2002년 대한전자공학회 멀티미디어연구회전문위원장
2003년 한국 인터넷 방송/TV 학회 부회장
2003년~현재 대한전자공학회 상임이사
2005년~현재 조선대학교 RIS 사업단장
<주관심분야 : 멀티미디어 시스템설계, Ubiquitous convergence>