

u-Green City 구현을 위한 상황인지기반 지능형 음성인식 시스템

Intelligent Speech Recognition System based on Situation Awareness for u-Green City

조 영 임*, 장 성 순
(Young Im Cho and Sung Soon Jang)

Abstract: Green IT based u-City means that u-City having Green IT concept. If we adopt the situation awareness or not, the processing of Green IT may be reduced. For example, if we recognize a lot of speech sound on CCTV in u-City environment, it takes a lot of processing time and cost. However, if we want recognize emergency sound on CCTV, it takes a few reduced processing cost. So, for detecting emergency state dynamically through CCTV, we propose our advanced speech recognition system. For the purpose of that, we adopt HMM (Hidden Markov Model) for feature extraction. Also, we adopt Wiener filter technique for noise elimination in many information coming from on CCTV in u-City environment.

Keywords: situation awareness, HMM, wiener filter, MFCC

I. 서론

2008년 등장한 그린 IT의 개념은 다음 그림1에서와 같은 일반적인 개념을 의미한다[1] 즉, IT 시스템 설계의 그린화, IT 시스템 사용의 그린화, IT 시스템 생산의 그린화, IT 시스템 폐기의 그린화 등 4가지 방향으로 그린 IT 정책을 실현해 나가고 있다. 즉, IT를 환경적 접근방법으로의 그린화를 말하는 것이며, 이것이 일반적인 그린 IT에 대한 접근방법이다.

그러나 2008년 발표된 국가정보화 기본계획보고서에 의하며, 그린 IT의 개념은 휴머니티(humanity), 지능(intelligence), 상생(win-win)의 3요소를 가장 핵심요소로 하고 있다[2]. 그린 IT란 IT기반의 친환경(green) 기술(technology)을 포함하는 신조어로 2008년에 새롭게 등장한 개념으로, 전 세계적으로 대규모 데이터처리시스템들이 늘어나면서 미래 IT 기술 선도과 시장선점에서 구체적 대안으로 제시되고 있다. 현재 그린 IT의 응용영역이 환경 분야에서부터 전 사회영역으로 확산되고 있는 추세이다. 예를 들면, 환경 감시 및 관리 체제, 정부기관 설비시설 오토메이션, 대체 에너지 활용, 중앙 난방장치, 자택근무, 온라인 서비스, 폐기물 관리, 정책 이행 비용 등이 응용영역이 될 수 있다. 위의 3가지 개념은 그린 IT의 환경적 접근법 뿐 아니라 개념적 그린 IT의 실천전략이 매우 중요하다는 것을 의미한다. 따라서 본 논문에서는 위의 개념적 그린 IT의 3가지 핵심요소를 구현한 시스템으로서의 상황인지기반 지능형음성인식 시스템을 구현하고자 한다. 즉, 가장 인간적인 휴머니티 기반 컴퓨터를 지능적으로 구현하면서 상호 윈윈(win-win)할 수 있는 지능형 음성인식 시스템 구현에 관한 것이다.

이러한 개념을 발전시킨 동기는 다음과 같다. 유비쿼터스 시대의 도래와 함께, 각종 정보기기들의 소형화 등 홈네트워크 시장의 확대로 인하여 음성 기술은 이들 장비간의 제어를 가장 효율적으로 할 수 있는 인터페이스로서 점점 중요성이 높아지고 있다[3,4].

현재 CCTV 환경은 기본적으로 주변에 소음이 생길 수밖에 없는 잡음환경에 처해 있으며, 응급 상황이 일어 날 수 있는 동적 환경을 기반으로 운영되고 있다. 그러나 항상 관리자가 모니터링을 할 수 있다는 장점이 있기 때문에 보안의 중요한 도구로 사용되고 있다.

그러나 이러한 보안기술 측면에서 CCTV상에서의 영상 분석이 많이 연구되고 있는데, 이러한 영상을 분석한 연구에서의 문제점은 첫째, CCTV 카메라의 영상인식이 가지고 있는 많은 기술적 문제들, 특히 기상 변화, 그림자 등 조명의 변화에 따른 오인식과 같은 문제점이 발생한다는 점이다[5]. 둘째, 어두운 밤이나 혹은 화면상으로 구분이 불가능한 응급 상황 발생 시에 이를 확인하기 어렵다는 점이다. 따라서 보다 효과적인 응급 상황 대처를 위해 상황인지를 통한 음성인식 기술을 이용하여 보안성 강화를 고려한 연구들이 병행되어 연구되고 있다.

그러나 CCTV 상에서의 응급시 음성을 확인하는 지금까지의 연구들은 홈 네트워크라는 환경 내에서 연구가 주로

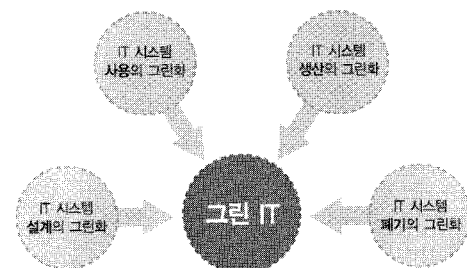


그림 1. 그린 IT의 기본개념.
Fig. 1. The concept of green IT.

* 책임저자(Corresponding Author)

논문접수: 2009. 9. 10., 수정: 2009. 10. 5., 채택확정: 2009. 10. 30.

조영임, 장성순: 수원대학교 컴퓨터학과

(ycho@suwon.ac.kr/veteranus@gmail.com)

※ 본 논문은 경기도지역협력연구센터 사업의 일환으로 수행되었음 [GRRC 수원 2009-B3, u-City 보안감시기술협력센터].

이루어지고 있으나[6], 본 연구에서는 음성인터페이스 기술을 사용하여, CCTV 환경에서 음성인식이 가능한 연구로 확장 연구하여 응급 상황에 대처할 수 있는 시스템을 개발하고자 한다. 즉, 홈 네트워크 환경처럼 정적이고 제한적인 환경이 아닌 동적이고 잡음의 영향이 큰 상황에서의 u-City 구현을 위해 상황인지를 통한 음성인식이 가능한 시스템구현에 관해 연구하고자 한다.

그런 IT기반 u-City란 위에서 설명한 그런 IT의 주요 개념을 u-City에 적용시킨 것으로, 상황인지[7]를 하지 않았을 경우와 상황인지를 통한 음성인식에서 컴퓨터 프로세스상에서의 절감효과를 가져올 수 있다는 것이다. 예를 들면, CCTV상의 많은 음성정보를 모두 인식하려면 수많은 프로세싱 모듈과 프로세싱 타입이 필요하지만, 상황인지를 통해 위급정보만을 인지하여 프로세싱 하였을 경우에는 상당한 프로세싱 비용을 줄일 수 있을 것으로 예상된다. 본 논문의 연구 결과물은 향후 CCTV가 있는 실제 환경에 적용되어 응용 및 활용하고자 한다. 따라서 본 논문에서는 II 장에서 일반적인 음성인식의 주요 기술을 서술하며, III 장에서는 그런 IT 개념을 갖는 상황인지기반 음성 인식의 주요 알고리즘을 소개하고, IV 장에서는 다양한 시뮬레이션을 통한 결과를 설명하고, V 장에서 결론을 맺고자 한다.

II. 일반적인 음성인식 시스템의 기술

기본적으로 음성인식 시스템은 그림 2에서와 같이 총 6 단계에 걸쳐 구성된다. 1단계는 음성신호를 전기신호로 변환하여 디지털화하여 전송하는 음성입력 단계이며, 2단계는 주위 잡음을 제거하고 음성신호를 분리하여 음성이 있는 구간을 찾아내게 되는 전처리 단계이다. 3단계는 음성인지 모델을 통하여 음성인식에 유용한 특징을 뽑아내는 특징추출 단계이며, 4단계는 음성 인식 훈련 과정으로 표준 패턴 DB를 생성하는 단계이다. 5단계는 미리 생성된 기준패턴과 입력되는 음성을 비교하여 가장 비슷한 것을 인식결과로 결정하는 음향모델 단계인 탐색과정이다. 마지막으로 이러한 인식결과를 원하는 응용에 적용하여 사용자 인터페이스 기술을 이용하게 되는 단계이다.

강인한 음성인식을 위하여 음성 신호의 입력 후 잡음의 제거를 위해 Wiener 필터를 사용하며 그림 2의 ②전처리 단계에서 적용된다. Wiener 필터의 성능을 좌우하는 요소는 잡음구간검출과 동적 잡음으로부터의 음성 분리이다. 잡음 구간 검출은 매 프레임별 정확한 잡음 모델 추정을 위한 것으로 통계기반 VAD (Voice Activity Detection) 기법을 적용하여 성능을 향상시켰다. 모델기반 Wiener 필터의 구성도

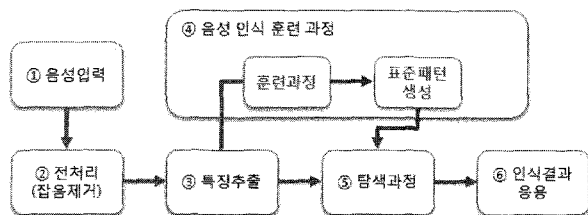


그림 2. 일반적인 음성인식 시스템 구조.
Fig. 2. The general structure of speech recognition system.

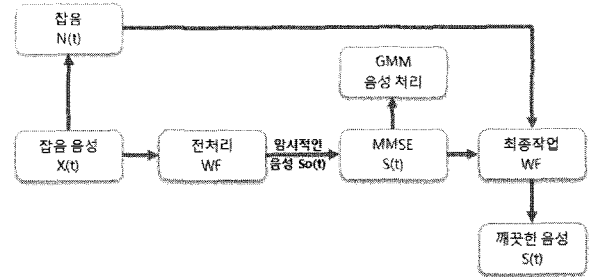


그림 3. 모델 기반 Wiener 필터의 구성도.
Fig. 3. The structure of model based Wiener filter.

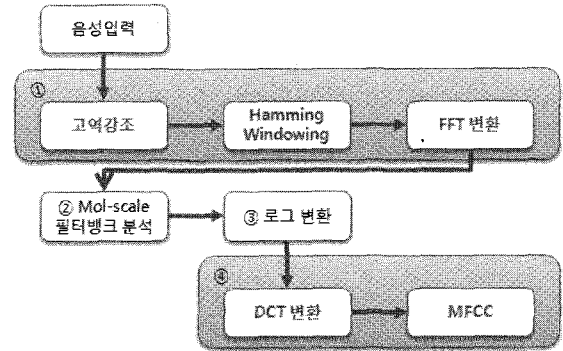


그림 4. MFCC 과정.
Fig. 4. The process of MFCC.

는 그림 3과 같다[5,6].

일반적인 모델 기반 Wiener 필터의 구현 과정에서 Wiener 필터의 최적화 방식을 사용하는데, 이것은 비대칭 윈도우 사용으로 인한 잡음 제거 알고리즘에서 소용되는 시간의 최소화를 이루는 것이다[8,9].

잡음이 제거된 입력신호로부터 음성인식에 유용한 특징을 추출시에는 MFCC (Mel Frequency Cepstral Coefficient)을 사용하며 그림 2의 ③ 특징추출단계에서 적용된다. 이 알고리즘은 사람의 귀가 주파수 변화에 반응하게 되는 양상이 선형적이지 않고 로그스케일과 비슷한 멜(mel) 스케일을 따르는 청각적 특성을 반영한 캡스트럼 계수 추출 방법이다. MFCC의 처리 과정은 그림 4와 같다.

- ① 분석구간의 음성 신호에 푸리에(fourier) 변환을 취하여 스펙트럼을 구한다.
- ② Mel 스케일에 맞춘 삼각 필터뱅크를 대응시켜 각 밴드에서의 크기의 합을 취한다.
- ③ 필터뱅크 출력값에 로그를 취한다.
- ④ 로그를 취한 필터뱅크 값에 이산 코사인 변환 (DCT, Discrete Cosine Transform)을 하여 최종 MFCC를 구한다.

현재 대부분의 인식 시스템은 음성신호의 통계적 특성을 나타내기 위해서 HMM (Hidden Markov Model)을 사용하며 그림 2의 ④ 음성인식훈련과정과 ⑤ 탐색과정에서 사용되는 방법이다. 음성 인식에서 주도적인 역할을 해온 DTW (Dynamic Time Warping) 알고리즘의 단점을 보완할 수 있는 방법으로 알려져 있다. HMM은 관측이 불가능한 과정을 관측이 가능한 다른 과정을 통해 추정하는 이중 확률 처리로서 음성과 같이 다변성이 발생과정을 알 수 없는 처리를

표현하는데 적당한 방법이다. HMM은 안정된 구간이나 특이한 구간의 판별, 구간에 따른 연속적 변화 특성의 묘사 및 각 구간에 대한 공통적인 단 구간 모델을 효과적으로 처리할 수 있다.

Markov 특성은 시간 영역의 사건들에 대해 과거와 현재의 사건이 주어 졌을 때 현재 사건의 조건 확률 밀도(conditional probability density)가 가장 최근의 j개 사건에 영향을 받는다는 현상을 말하며 이러한 특성을 만족시키는 과정이 Markov 과정이라고 하고 다음과 같이 모델링된다.

$$\begin{aligned} P[q_t = S_i | q_{t-1} = S_j, q_{t-2} = S_k, \dots] = \\ P[q_t = S_i | q_{t-1} = S_j] \end{aligned} \quad (1)$$

여기서 위 식의 우측 항은 시행 횟수에 대해 독립적이므로 전이 확률 a_{ij} 는

$$a_{ij} = P[q_t = S_i | q_{t-1} = S_j], 1 \leq i, j \leq N \quad (2)$$

으로 표시할 수 있고

$$a_{ij} \geq 0, \sum_j a_{ij} = 1 \quad (3)$$

의 특성을 갖고 있기 때문에 일반적인 확률적 제한에 따른다.

이와 같이 HMM은 유한개의 상태를 갖는 Markov 모델과 출력 분포의 집합으로 특성 지을 수 있다. Markov chain에서 출력 분포의 파라미터가 음성 신호의 변이성을 모델링하는 반면, 전이 파라미터는 시간의 변화를 모델링 한다. 이러한 2가지 타입의 변화성으로 음성의 성질을 특성 지을 수 있다[10]. 학습 과정에서 forward-backward 알고리즘은 대량의 데이터로부터 자동적이고 효율적으로 전이 파라미터와 출력 파라미터 값을 추정한다. 이 알고리즘은 E-M(Estimate-Maximize) 알고리즘의 한 예로서 명확한 수학적 기반을 가지고 있고, 복잡도(complexity)가 입력 음성의 길이에 선형적이라는 장점을 가지고 있다[11]. HMM은 관측 불가능한 프로세스를 관측 가능한 출력을 내는 프로세스로

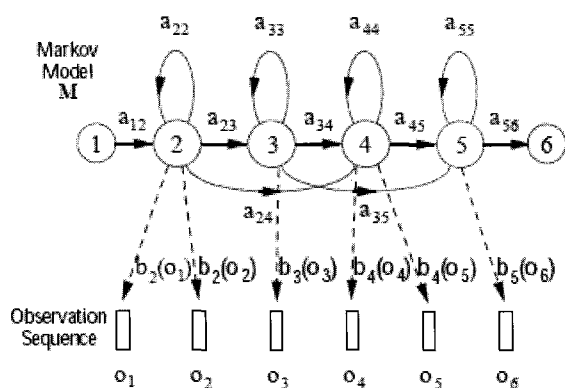


그림 5. HMM 알고리즘 인식과정.

Fig. 5. The recognition process of HMM algorithm.

추정하는 방식으로, 초기상태확률(initial state probability), 상태전이확률(state transition probability), 그리고 관측확률(observation probability)로 나타내어진다. 일반적으로 HMM은 흔히 위의 3개 파라미터로 표시된다.

ECHOS (Easy Compact Hanguk Object-orienter Speech recognizer)는 개발된 객체지향 음성인식기인 ezCSR을 바탕으로 오류 수정, 기능 추가를 거쳐서 개발된 한국어 음성인식 플랫폼이며, 잡음제거, ESTI 특징 추출, N-Best 탐색, Word graph 생성이 가능한 음성 디코더이다. 한글 처리가 가능하며 객체 기반의 구조를 가지고 있으며, 음성인식 DB를 구축한 HTK와의 호환이 가능하다[12].

또한, 음성의 인식을 향상 위한 음성인식 후처리 방법에 대해서는 많은 연구들이 진행되고 있으며, 이 중 오류 패턴비교, 어휘 의미 패턴[13]에 대해 간단히 기술한다.

오류 패턴 비교는 미리 구축된 오류 패턴 데이터를 이용하여, 오류를 수정하는 후처리 하는 방식이다. 음성의 오인식은 일정한 유형을 가지고 발생하며 무작위로 발생하는 오인식 유형은 거의 없다고 가정한다. 따라서 훈련 데이터를 통해 미리 오인식 패턴을 모은 후 이를 오류 패턴 데이터베이스로 구축하여 후처리에 이용한다.

어휘 의미 패턴의 경우에는 단어의 의미적 정보를 고려한 후처리를 위해 LSP (Lexico-Semantic Pattern)을 제안하였다. 연속 음성인식에서 발화된 문장을 단어별로 각각 어휘 및 의미정보를 포함한 특정 범주 정보로 대치하여 연결한 스트링이며, 후처리에 사용될 LSP는 훈련 데이터를 통하여 데이터로 미리 구성된다.

III. 상황인지기반 음성인식 알고리즘

본 논문의 목적은 일반적인 대어휘 연속음성인식이기보다는 u-Green City 내에서 지능적 도시 구현을 위한 응급(위급) 환경을 인지하는 음성인식 알고리즘 개발하는데 있다.

일반적으로 음성인식에 사용되는 탐색 알고리즘 중 플랫폼 lexicon과 lexical 트리를 기반으로 하여[14], 빠른 인식속도를 요구하는 상황인지가 가능하도록 설계하고자 한다.

제안하는 음성인식 시스템의 전체구조는 그림 6과 같이 구성되며 이러한 음성인식 시스템을 이용하여 응급상황 발생시에 효과적인 대처가 가능하도록 CCTV 환경을 구축할 것이다. 이 시스템의 각 과정은 다음과 같은 과정을 수행하게 된다. 자세한 내용은 다음 절에서 설명하고자 한다.

1. 음성인식 DB구축

제안하는 음성인식 시스템을 위해 먼저, HTK (Hidden

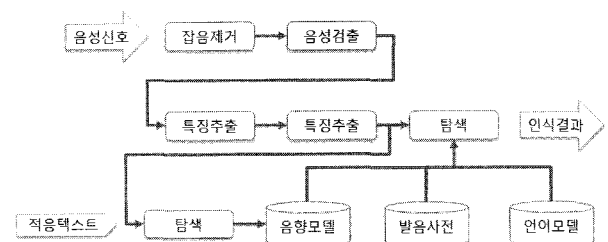


그림 6. ECHOS의 구조.

Fig. 6. The structure of ECHOS.

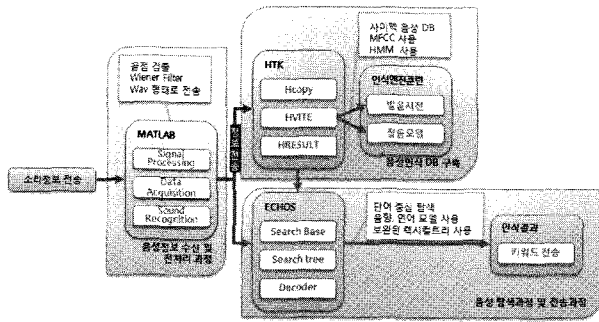


그림 7. 음성인식 시스템 전체구조.
Fig. 7. The overview of proposed speech recognition system.

Markov Model Toolkit)[11]을 사용하여 음성인식 DB를 구축하며, 사용되는 음성 DB는 500여명의 발성, 16kHz, 16Bit의 데이터를 가지고 구축하였으며, 기본 음성인식 DB와 함께, 응급상황에 따른 음성인식 DB를 구축하였다.

2. 음성정보 수신 및 전처리 과정

CCTV에서 음성이 수신이 되면 MATLAB[15]에서 전처리 과정을 거치게 된다. Data acquisition 모듈과 signal processing 모듈을 이용하여, 다음과 같은 과정을 수행한다.

- ① 음성의 발화구간을 추출한다.
- ② 추출된 음성 구간에서 특징벡터를 추출한다.
- ③ 추출된 특징벡터를 Wiener 필터 과정을 거친다.

u-City 환경 내에서 상황인지를 위한 음성 정보의 처리 과정에서 불필요한 정보가 포함된 음성 정보의 처리를 위하여 Wiener 필터 과정을 거치지만, 이러한 과정으로 말미암아 시간 지연이 생기다면 목적에 부합되지 못하기 때문에, 이러한 문제점을 해결하기 위하여, Wiener 필터의 최적화설계를 통하여, 시간 지연과 잡음 제거의 성능간의 관계를 고려하였다.

비대칭 윈도우의 적용으로 잡음 제거 시에 소요되는 시간을 최소화하기 위한 것이다. 일반적으로 잡음의 통계량은 그 자체로 정상적이라고 판단이 가능하지만, 음성 정보의 통계량과 비교한다면 차이를 알 수 있을 정도로 비정상성을 확인할 수 있으므로, 그 차이를 구분하여 Wiener 필터의 최적화를 고려하였다.

$$h(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{P_1}\right), & 0 \leq n < n_0 \\ \cos\left(\frac{2\pi(n-n_0)}{P_2}\right), & n_0 \leq n < N \end{cases} \quad (4)$$

식 (1)에서 P_1, P_2 는 비대칭 창함수의 왼쪽 및 오른쪽 부분을 나타내기 위한 주기값이며, n_0 및 N 은 최대치가 존재하는 위치 및 창함수 전체의 길이를 나타낸다.

잡음제거 후의 의미 있는 신호의 양은 실제 신호와 저주파 필터 후 신호 사이 포락의 상관관계 값을 통해서 얻어 내어 음성 신호에 있는 감정을 구분하여 탐색이 필요여부를 구분하는데 사용된다[16]. 이 때 각 감정의 구분은 [17]의 논리에 따라 그림 8과 같이 구분되며, 본 논문에서는 이 주파수 영역에 따라 '분노'에 해당하는 인식 주파수 대역만을 상황인지 하고자 한다. 음성으로서의 의미가 있는 주파

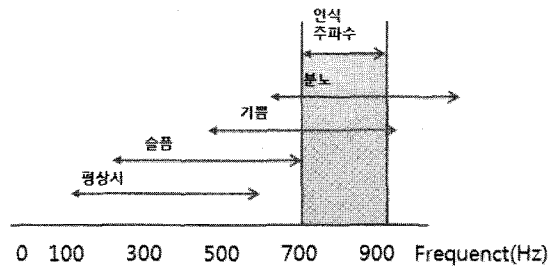


그림 8. 음성신호의 주파수 영역.
Fig. 8. The frequency range of speech signal.

수 영역 중 감정의 인식이 구분어 용이한 대역의 주파수만을 고려하였다.

3. 음성 탐색과정 및 전송과정

구축된 음성인식 DB를 바탕으로 탐색과정은 그림 9와 같으며, 응급상황 인식 및 탐지에 사용되게 된다. 이 때 기본적으로 음소를 기반으로 단어를 인식하고자 DB를 구축하는 것을 기본으로 구축하였다.

음향모델을 중심으로 하여 단어(keyword)를 탐지하는 것을 우선으로 선정하였으며, 이는 다른 연구[6]를 통해서 확인되었으며, 이러한 단어 중심의 인식 시스템에서는 플랫폼, 렉시컬 트리[14]를 이용한다. 렉시컬 트리는 메모리 사용은 효율적이지만, 언어모델 확률값의 적용 지연과 단어 간 모델링 구현의 복잡성이 존재하기 때문에, 트리 복사 알고리즘[6]을 구현하였다. 그리고 본 논문에서는 렉시컬 트리가 가지고 있는 단점을 보완하기 위해 단일 음소로 이루어진 단어에 대해서는 렉시컬 트리를 구성할 때, 다음 그림 10에서와 같이 별도의 병렬적인 구조를 갖도록 설계하여 문제점을 해결하고자 한다.

각 단계를 거쳐 최종적으로 검색 후 나온 결과는 텍스트 형태의 정보로 사용자 인터페이스에 전송된다.

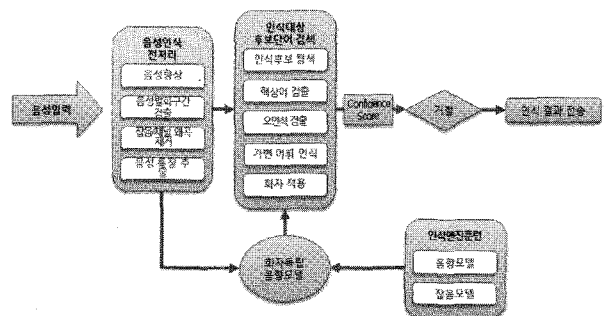


그림 9. 음성인식 부분 구조.
Fig. 9. The proposed speech recognition part of system.

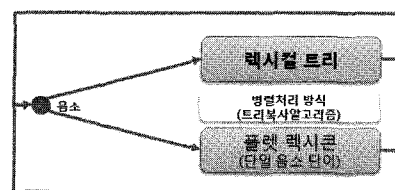


그림 10. 렉시컬 트리에 의한 문제점 해결 전략.
Fig. 10. The solution strategy of lexical tree problem.

IV. 실험결과

음성인식에서의 성능을 평가하기 위해 SITEC에서 제작한 스피치 단어 음성 데이터베이스를 사용하였으며, 16kHz/16bit로 녹음되어있으며, 음향 모델 학습에 500명 인원의 음성을 사용하였다. 추후 비교 음성은 16kHz/16bit의 형태로 마이크 혹은 CCTV에서 전송되는 데이터를 기준으로 사용하였다.

Wiener 필터 적용 전에 이러한 과정을 거쳐야만 응답속도에서 결과를 얻을 수 있기 때문에, 이런 음성 감지를 위한 에너지량의 체크는 필수적으로 필요하다.

이렇게 저장된 음성정보들만을 이용하여, Wiener 필터를, MATLAB에서 구현하여 그림 9에서와 같이 음성 정보에 포함된 잡음을 제거하는 과정을 통하여, 정확한 인식효과를 가져 올 수 있게 하였다.

이와 같이 필터링 전 음성 정보들의 불필요한 부분, 즉 인식 결과에 큰 영향을 끼치지 못하는 음성의 부분이나 소리가 발생 시에 일어나는 각종 잡음, 혹은 CCTV에서 정보를 전송하면서 생기는 각종 잡음들을 효과적으로 필터링을 하여, 인식률의 상승효과를 가져 올 수 있는 결과를 확인할 수 있다(그림 11).

본 논문에서 사용한 HMM의 음성신호와의 관계는 그림 12의 상태와 같이 한 대응관계를 나타낼 수 있다. 처리과정은 다음과 같다.

- ① S₁ 상태에서 x₁,x₂,x₃가 발생되며, S₂상태에서는 x₄,x₅, S₃ 상태에서 x₆,x₇이 대응된다.
- ② Frame shift 시간 간격에 MFCC를 이용하여 하나의 특징벡터가 음성신호로부터 추출한다.
- ③ 천이확률은 상태 S_i에서 S_j로 천이한다.
- ④ 관측확률밀도함수는 b_j(x_i)는 상태 S_j에서 특징벡터 x_i를 관측할 확률이다.

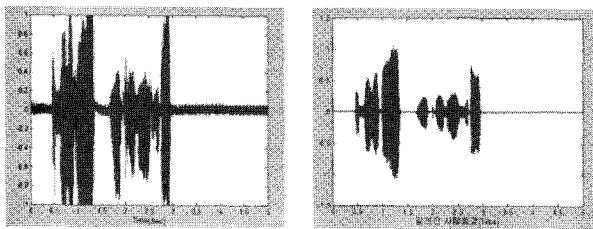


그림 11. Wiener 필터링 적용 전후 결과 비교.
Fig. 11. The before and after results of Wiener filter adoption.

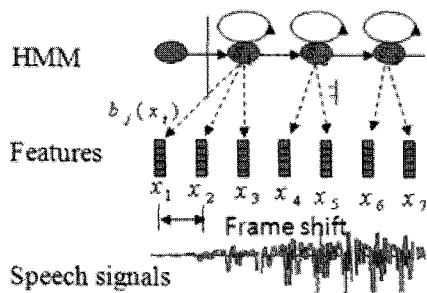


그림 12. HMM와 음성신호의 대응관계.
Fig. 12. The relation between HMM and speech signal.

표 1. 제안하는 시스템의 인식 결과.

Table 1. The recognition result of proposed system.

탐색방법	단어간 모델 사용	단어 인식률 (%)	인식 시간 (sec/문장)
정방향	X	77.2	5.4
정방향 + 역방향	X	80.1	6.3
정방향	O	88.9	21.0
정방향 + 역방향	O	90.0	22.1

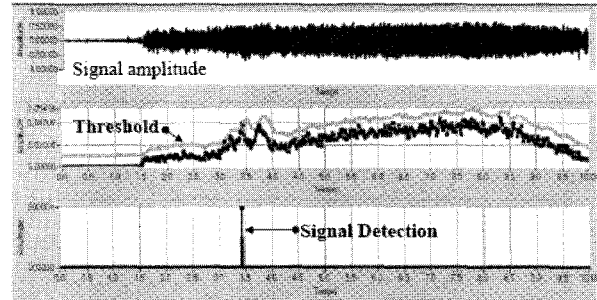


그림 13. 상황인지를 위한 음성신호의 인식의 접근 방법.
Fig. 13. The approach of speech signal recognition for situation awareness.

이처럼 좌에서 우로의 천이확률만을 갖는 HMM에서 천이확률은 각 상태에서 머무는 프레임수와 다음 상태로 이동하는 프레임의 수의 비로부터 구할 수 있다.

구축된 음성인식 DB를 이용하여 ECHOS를 이용하여 two-pass 탐색 시에 정방향 탐색(bigram)시와 정방향과 역방향(trigram) 시와의 비교결과, 다음 표 1과 같이 단어 모델의 사용 여부가 인식률과 인식 시간 모두에 큰 영향을 미치고 있음을 알 수 있었다. 즉, 탐색하는 단어들의 연관관계를 고려하는 단어 모델을 사용하는 경우 그렇지 않는 경우보다 보다 나은 인식률을 가져오는 것을 알 수 있었다.

또한, 이와 같은 결과를 바탕으로 HMM의 기본적인 탐색방법(좌에서 우로)을 따르는 정방향 탐색 시보다는 음소들 간의 관계를 고려하는 역방향 탐색이 단어 인식률이 8% 이상 상승으로 단어의 인식 시에 음소들 간의 관계를 고려하여 시스템을 구축하여 탐색하는 것이 정확한 인식률을 얻을 수 있음을 알 수 있었다. 그러나 음성인식 DB에서 가장 일치하는 단어열을 찾는 확률이 높아지는 것이 결과적으로는 탐색 시간의 증가도 가져오고 있음을 알 수 있었다.

CCTV에서 전송되는 모든 소리를 음성인식 탐색 구간에서 이용하지 않고, 3.2절에 언급한 듯이, 감정부분이 탐지되는 구간만을 감지를 하여, 그림 13과 같은 상황인지를 위한 프로세싱 타임의 감소와 함께 보다 빠른 상황 대처가 가능하게 된다.

따라서 음색인식율과 인식시간의 상관관계로부터 보다 효과적인 음성인식 시스템을 구축하려면 효과적인 음향모델과 언어모델을 도입하여 인식률과 인식시간의 감소를 가져오는 방향으로의 연구가 선행되어야 할 것이다.

V. 결론

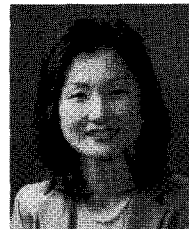
상황인지는 여러 가지로 정의될 수 있으나, 본 논문에서는 상황을 주위환경이나 처해있는 상황을 의미하는 용어로 정의하여 연구를 수행하였다. 따라서 본 논문에서는 CCTV 등에서 들어오는 음성정보를 위급한 사운드와 위급하지 않은 사운드로 상황을 인지하여, 위급한 사운드에 대해 음성인식을 좀더 빠르고 정확하게 처리될 수 있도록 함으로써 그린 u-City를 구현하고자 한 점이 특징이다. 따라서 본 논문에서는 응급상황용 음성인식 DB를 새로이 구축하여 음성 DB를 이중으로 구성하여, 각 검색을 병렬로 처리하여, 긴급을 요하는 처리를 빠르게 감지하여 보다 뛰어난 결과를 가져올 수 있었다. 물론, 응급상황용 음성인식DB의 구축에 따라 그 결과가 다르게 나올 수 있는 문제점을 고려하여, 추후 연계되는 연구를 통하여 기 구축된 음성인식 DB를 자율 신경망 알고리즘을 이용하여, 자체적으로 정보를 추가하여 보다 많은 변수가 일어나는 환경에서 그 감지율을 높이고자 하는 연구가 계속 되어야 할 것이다.

참고문헌

- [1] “그린IT 활용: 원칙과 실천,” 한국소프트웨어진흥원 정책연구센터 연구보고서, pp.64-81, 2008.
- [2] 국가정보화 기본계획, 정보화추진위원회, 2008.
- [3] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, “Toward conversational human-computer interaction,” *AI Magazine*, vol.22, no. 4, pp. 27-37, 2001.
- [4] H. Kruegle, “CCTV Surveillance: Analog and digital video practices and technology,” *Elsevier*, pp. 227-239, 2007.
- [5] 유장희, 문기영, 조현숙, “지능형 영상보안 기술현황 및 동향,” 전자통신동향분석, vol. 23, no. 4, pp. 80-89, 2008.
- [6] M. Vacher, S. Jean-François, S. Chaillol, D. Istrate, and V. Popescu, “Speech and Sound Use in a Remote Monitoring System for Health Care,” *LNAI* 4188, vol.4188, pp.711-718, 2006.
- [7] 임신영, 허재두, “상황인식 컴퓨팅 응용기술동향,” 전자통신동향분석, 제19권 제5호, 2004.
- [8] 정상배, 이성득, 한민수, “VoIP 환경에서의 잡음제거를 위한 최적화된 위너 필터,” 대한음성학회 논문지, vol.

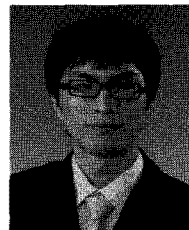
64, pp. 105-119, 2007.

- [9] ETSI ES 300 726, “Digital cellular telecommunications system; Enhanced Full Rate(EFR) speech transcoding (GSC06.60),” pp.20, 1997.
- [10] 박재홍, 이광호, 안동순, “HMM에 기반 음성인식을 위한 Toolkit의 구성요소,” 한국정보과학회 논문지, vol. 26, no. 1, pp. 472-473, 1999.
- [11] 이윤근, 박준, 김상훈, “음성인터페이스 기술,” 전자통신동향분석, 제20권 제5호, pp. 1-15, 2005.
- [12] 한국어 음성 인식 공통 플랫폼(ECHOS), http://www.sitec.or.kr/kongji_show.asp?num=111.
- [13] 송원문, 김명원, “문맥 및 사용 패턴 정보를 이용한 음성인식의 성능 개선,” 정보처리학회 논문지, 제13권 제5호, pp. 553-560, 2006.
- [14] 권오욱, 권석봉, 윤성락, 장규철, “한국어 음성인식 플랫폼(ECHOS)의 개선 및 평가,” 대한음성학회 논문지, vol. 59, pp. 53-68, 2006.
- [15] The MathWork, Getting started guide, 2009.
- [16] 김은호, 현경학, 곽윤근, “음성의 특정 주파수 범위를 이용한 잡음환경에서의 감정인식,” 한국정밀공학학회지, vol. 23, no. 5, pp. 68-76, 2006.
- [17] 정병욱, 천성표, 김연태, 김성신, “음성신호를 이용한 감정인식, 한국지능시스템학회 논문지,” vol. 18, no. 4, pp. 494-500, 2008.



조영인

1988년 고려대 컴퓨터학과 졸업. 1990년 동 대학원 석사. 1994년 동 대학원 박사. 현재 수원대학교 컴퓨터학과 교수. 관심분야는 인공지능, 에이전트시스템, 음성인식, 유비쿼터스 시스템.



장성순

2008년 수원대 컴퓨터학과 졸업. 2009년 동 대학원 석사과정. 관심분야는 인공지능, 에이전트시스템, 정보검색, 음성인식, 유비쿼터스 시스템.