

논문 2004-41CI-2-1

# 자동차 텔레매틱스용 내장형 음성 HMI 시스템

## (The Human-Machine Interface System with the Embedded Speech recognition for the telematics of the automobiles)

권 오 일\*

(Oh-II Kwon)

## 요 약

자동차 텔레매틱스 용 음성 HMI(Human Machine Interface) 기술은 차량 내 음성정보기술 활용을 위하여 차량 잡음환경에 강인한 내장형 음성 기술을 통합한 음성 HMI 기반 텔레매틱스 용 DSP 시스템의 개발을 포함한다. 개발된 내장형 음성 인식엔진을 바탕으로 통합 시험을 위한 자동차 텔레매틱스 용 DSP 시스템 구현 개발을 수행하는 본 논문은 자동차용 음성 HMI의 요소 기술을 통합하는 기술 개발로 자동차용 음성 HMI 기술 개발에 중심이 되는 연구이다.

## Abstract

In this paper, we implement the Digital Signal Processing System based on Human Machine Interface technology for the telematics with embedded noise-robust speech recognition engine and develop the communication system which can be applied to the automobile information center through the human-machine interface technology. Through the embedded speech recognition engine, we can develop the total DSP system based on Human Machine Interface for the telematics in order to test the total system and also the total telematics services.

**Keywords :** Human Machine Interface, Digital Signal Processing, Global Positioning System, Telematics

## I. 서 론

텔레매틱스<sup>[1]</sup> 및 이를 기반으로 하는 차량용 정보서비스는 자동차 운행 중에 운전자와 외부환경과의 의사소통을 도와주는 장치 및 시스템과 차량 및 운전자 맞춤형 서비스로 정의된다. 구체적으로 보면 GPS(Global Positioning System)와 무선통신망 기술을 이용한 위치 정보를 바탕으로 자동차와 운전자, 그리고 탑승자에게 교통정보, 응급상황 시 대처, 원격차량진단 등의 자동차

영역에 특화된 차량 정보 서비스와 인터넷 이용(금융거래, 뉴스, 이메일, 엔터테인먼트 등) 등 기존 Web 기반 정보 서비스를 이동 중인 차량에서도 가능하게 하는 차량 내장형 단말기와 운영체제 및 차량 영역 특화 정보 서비스를 말한다.

그러나, 이동 중인 다른 상태와는 달리, 운전이라는 특성 상, 손은 핸들에 눈은 전방을 주시해야하므로, 다른 사람과 통화를 하거나, 웹상의 정보를 검색하거나, e-메일의 송수신 등의 정보 교환 및 접근을 위하여 손과 눈을 이용할 수 없다. 이에 따라 자동차 운행 중에 정보의 접근을 위하여 손과 눈을 배제한 편리한 운전자용 정보 접근 인터페이스가 필수적으로 요구되며, 이를 위한 가장 자연스러운 인터페이스가 바로 음성인 것이다. 지능형 자동차에서 가장 중요한 것은 안전 운전과

\* 종신회원, 현대오토넷주식회사  
(Hyundaiatonet co., Ltd)

※ 본 연구는 산업자원부 중기거점 출연금 기술 개발 지원사업에 의하여 연구 되었음  
접수일자 : 2002년9월5일, 수정완료일 : 2004년2월16일

자동 운전, 그리고 다양한 정보의 운전 중 접근이며, 이를 위한 운전자용 정보 서비스 인터페이스로 가장 자연스럽게 경쟁력 있는 수단은 다름 아닌 음성인 것이다. 따라서, 자동차용 음성 HMI 기술은 향후 자동차 산업의 경쟁력에 있어서 매우 중요한 기술이며, 이의 핵심이 바로 자동차용 음성 인식/합성 기술이라고 할 수 있다.

자동차 텔레매틱스의 상업적 시도가 선진국에서는 2001년부터 시작되고 있다. 대표적인 예가 GM에 의한 OnStar(<http://www.onstar.com>), Ford와 Qualcomm에 의한 Wingcast(<http://www.wingcast.com>)를 들 수 있다. 국내에서도 텔레매틱스를 기반으로 하는 차량용 정보 센터의 구축을 현대/기아 자동차와 LG 텔레콤, 대우 자동차와 한국통신프리텔, 삼성자동차와 SK텔레콤이 각기 시작하고 있다. 차량 내장형 텔레매틱스는 현대오트넷이 현재 상업용 제품을 생산하고 있으며, 향후, 현대자동차의 차량정보센터와 연동한 텔레매틱스를 생산할 계획으로 있다. 음성 HMI 기반의 차량용 텔레매틱스 및 차량정보 센터는 OnStar에서는 2001년 하반기부터 Virtual Advisor란 개념으로 시작하고 있다. 그러나, 기존 전화망 기반IVR(interactive Voice Response) 시스템으로 구축하여, 차량의 잡음 환경과 특성을 반영하지 못하여 음성 인식 성능이 사용자의 요구에 미치지 못하고, 운전자가 대화형으로 접근하기에 매우 불편한 VUI(Voice User Interface)를 지원하고 있다. 음성 HMI 기반 차량 내장형 텔레매틱스는 연구가 진행 중이다. IDC, Dain Rauscher Wessels 등과 같은 산업 및 시장 리서치 전문기업의 보고서에 따르면, 이상에서 언급한 국내외 기술 보유 회사들이 차량 특성을 반영한 음성 HMI 기술의 필요성을 강조하고 있으며 활발한 연구 개발을 진행하고 있다.

## II. 본 론

본 논문은 차량 텔레매틱스 용 음성 HMI 전체 통합 시스템 개발 및 자동차용 음성 정보 서비스를 위한 효과적인 음성활용 기술의 개발이 목표이다.(그림1. 참조)

차량 내의 각종 센서, 도로, 교통관계 센터, 메시징 센터, GIS, 그리고 기타 정보원(날씨, 식당/호텔 예약) 등에서 발생하는 다양하고 복잡한 정보를 음성을 통하여 운전자(또는 가정)에게 전달하고 운전자(사용자)가 이들 정보를 음성으로 브라우징하고 차량의 각종 기기를 음성으로 제어할 수 있는 전체 프레임을 개발하는 것이 목표이다.

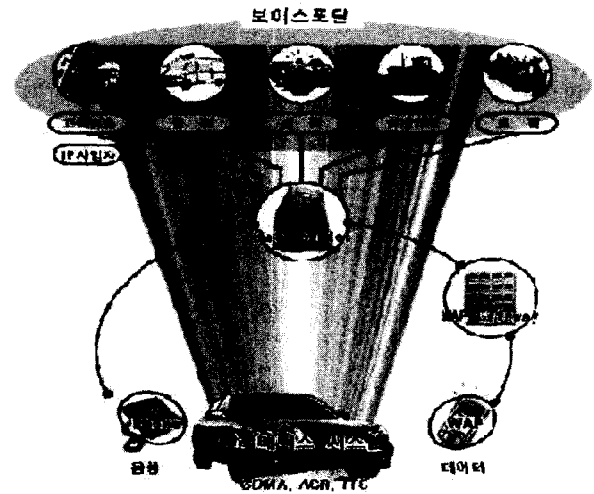


그림 1. 음성정보 서비스 전체 시스템.  
Fig. 1. The total system of SIT service.

### 2.1 차량 내장형 텔레매틱스용 음성 HMI 통합 시스템 개발

#### 2.1.1 차량 내장형 텔레매틱스 용 음성 HMI 보드 개발

기존 텔레매틱스 및 차량내 각종 기기를 음성으로 제어할 수 있도록, 텔레매틱스와 연동할 음성 HMI 모듈을 DSP<sup>[2]</sup> 보드로 개발한다. 차량 내장형 음성 기술<sup>[3]</sup> (잡음 처리 기술<sup>[10]</sup>, 내장형 음성 인식/합성 기술<sup>[4][5][7]</sup>, 내장형 음성 미들웨어 기술)의 결과물을 통합하는 테스트 시스템을 구축하고, 통합 시험 및 보완 개발의 결과물은 차량 내장형 음성 텔레매틱스 시스템으로 완성한다.

#### 2.1.2 텔레매틱스용 음성 HMI 시스템 보드 구현

음성 HMI 시스템 보드는 TI 사의 부동소수점 연산이 가능한 TMS320C6711C DSP 칩을 사용하였고, 32MB의 SDRAM과 32MB의 Flash 메모리를 사용하여 구현하였다.(그림 2.참조) 또한 음성 입출력 인터페이스 부분에는 실시간 음성인식엔진 구현을 위하여 타이머 인터럽트 혹은 DMA를 사용하여 일정한 간격으로 데이터를 샘플링하는 방안을 구현하였다. 이때 사용하는 코덱은 11.025 KHz 샘플링 레이트와 16비트 양자화 레벨을 지원하며 이를 위한 별도의 라이브러리 함수를 구현하였다.

그리고 향후 음성 응용 application 프로그램을 용이하게 포팅하게 하기 위하여, 상위의 응용 프로그램 개발자가 사용하기 편하도록 H/W 자원을 접근하는 기능을 가진 함수를 구현하였다. 응용 프로그램 개발자는 H/W에 대한 깊은 전문 지식이 없어도 프로그램을 설계할 수 있도록 ANSI-C 혹은 DOS-C 수준의 사용하기에 편

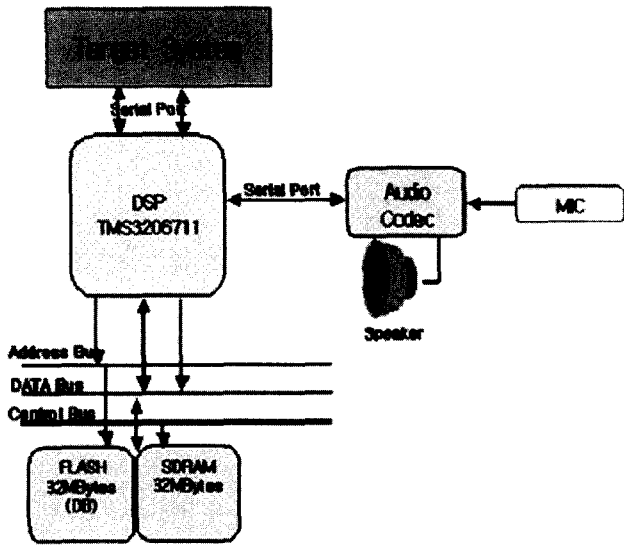


그림 2. 텔레매틱스용 음성 HMI 보드 블록도.  
Fig. 2. The block diagram of HMI telematics board.

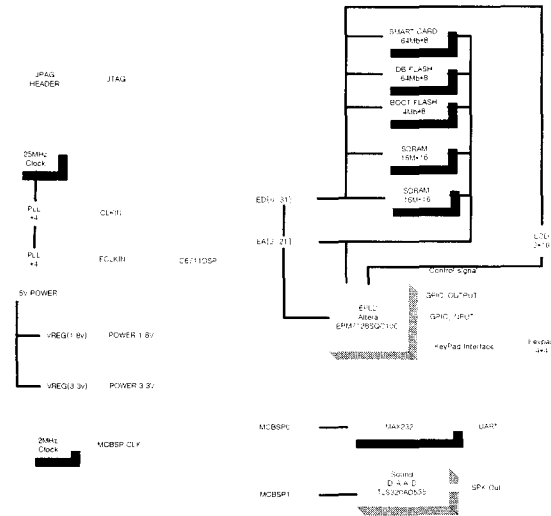


그림 3. 시스템 블록도.  
Fig. 3. The block diagram of the system.

안한 라이브러리 함수를 개발하였다. 그리고 외부 텔레매틱스 시스템과의 연동을 위하여 MAX232 칩을 이용하여 시리얼 통신 부분을 구현하였다.

리 함수를 구성하였다.

2.1.3. 상세 시스템 블록도

다음은 DSP 상세 시스템 블록도(그림 3 참조) 이다.

2.1.4. 입출력 장치

2.1.4.1 한글 LCD 출력 장치

조합형 한글 출력을 지원하기 위하여 제품화된 한글 출력용 LCD를 사용한다. 디지털 I/O를 이용하여 내부 S/W RS-232 프로토콜을 설계하고 이를 이용해 LCD를 제어한다.

2.1.4.2. RS-485, RS-232 직렬통신

각종 임베디드 시스템과 별도의 시스템과의 통신을 위한 직렬통신을 지원한다.

2.1.4.3 디지털 I/O 장치

제품 개발을 위한 에뮬레이터용 I/O를 지원하며 보드 구성을 위한 PLL 에뮬레이터를 지원한다. 제품 개발을 위한 에뮬레이터용 I/O를 지원하며 보드 구성을 위한 PLL 에뮬레이터를 지원한다.

2.1.4.4. 음성 입출력 장치

실시간 음성인식 구현을 위하여 타이머 인터럽트 혹은 DMA를 사용하여 일정한 간격으로 데이터를 샘플링하는 방안을 모색하였다. 이때 사용하는 코텍은 11KHz 샘플링레이트를 지원하며 이를 위한 별도의 라이브러

2.1.5 라이브러리 함수 개발

상위의 응용 프로그램 개발자가 사용하기 편하도록 H/W 자원을 접근하는 기능을 함수로 지원한다. 응용 프로그램 개발자는 H/W에 대한 깊은 전문 지식이 없어도 프로그램을 설계할 수 있도록 ANSI-C 혹은 DOS-C 수준의 사용하기에 편한 라이브러리 함수를 제공한다.

2.1.6. DSP를 이용한 음성인식 프레임워크 (그림 4 참조)

2.1.7 차량 제어용 음성 HMI 시스템 인터페이스 구현

차량 내부 제어를 위하여 외부 시스템과의 인터페이스 회로를 그림 5. 과 같이 구현하였다. 차량 내부의 여러 가지 모듈, 즉 에어컨, 파워 윈도우, AV 시스템, 실내 등등의 모듈 제어를 위하여 개발하였다. 외부 모듈과의 통신 프로토콜은 PPM 코드 방식으로 구현하였다. 실제 제작된 DSP 보드는 그림 6와 같다

2.1.8 내장형 음성인식 엔진

2.1.8 .1 음향 모델 최소화

Embedded 시스템용 음향 모델은 기존의 CHMM의 저장 메모리 문제를 해결하고, 동시에 인식 성능의 저하를 최소로 하도록 구성하는 것을 기본으로 한다. 이러한

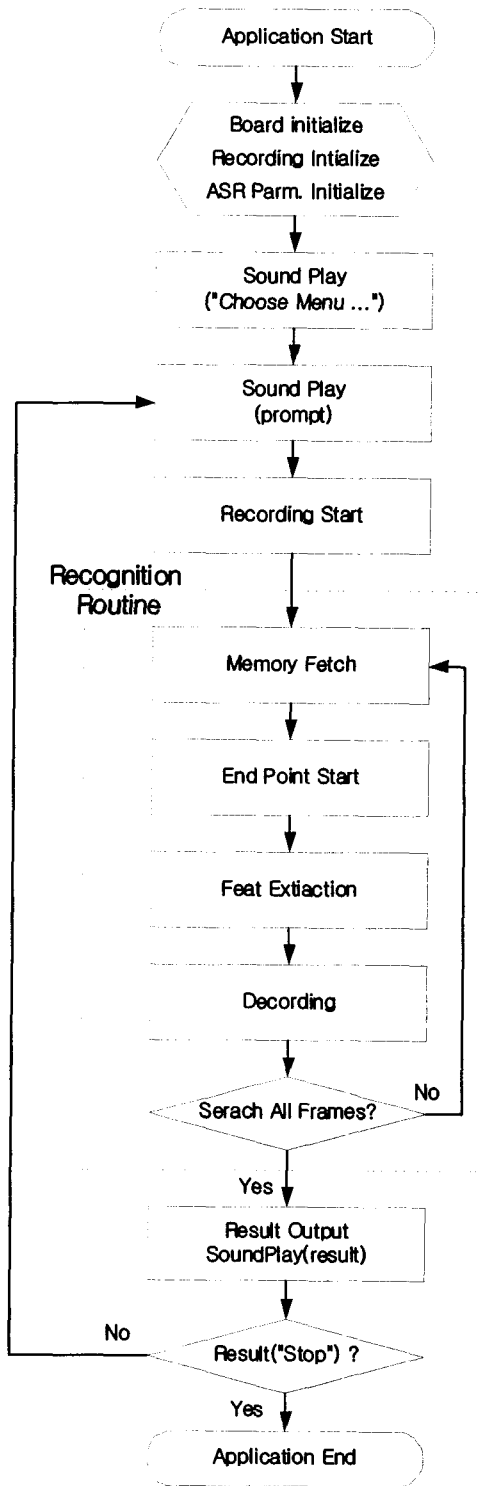


그림 4. DSP를 이용한 음성인식 프레임워크.  
Fig. 4. The flowchart of speech recognition.

문제 해결을 위해 반 연속HMM(SCHMM) 모델링 기법을 도입하였으며, tied-mixture 모델링 방법을 통하여 구현하였다. 또한 문맥 종속 모델(triphone)을 적용하여 앞,뒤 문맥에 따른 조음효과를 최대한 반영하도록 하였

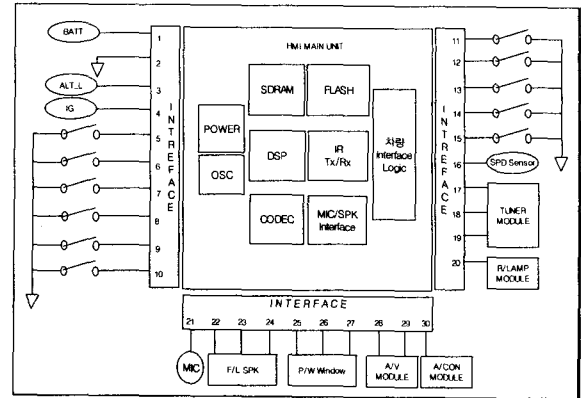


그림 5. 음성 HMI 시스템 인터페이스 회로도.  
Fig. 5. The interface circuit of HMI system.

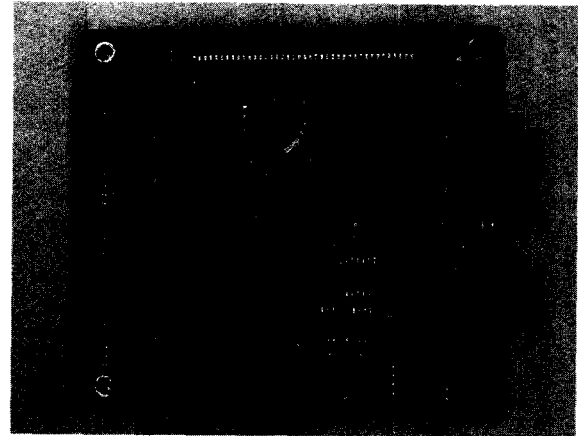


그림 6. 음성 HMI 시스템 보드.  
Fig. 6. The system of Speech HMI board.

다. 이렇게 구성된 음향 모델은 CHMM 모델 크기의 20% 정도의 크기를 가지며 인식 성능은 1% 미만의 범위로 줄어 든 것을 확인할 수 있다.

2.1.8 .2 반연속 HMM (SCHMM)

SCHMM은 이산 HMM(DHMM)과 연속 HMM (CHMM)의 절충된 형태의 HMM 출력 분포를 가진다<sup>[6][8][9]</sup>. DHMM은 벡터 양자화(VQ)을 통해 입력 벡터에 대한 codebook을 형성하고 그 codebook의 중심 값이 가장 가까운 하나의 인덱스를 HMM의 입력으로 결정하고 HMM의 state에서 각 index에 대한 이산 출력 확률의 분포를 갖는 형태의 HMM이다. CHMM은 VQ 과정 없이 입력 vector 그 자체에 대한 연속적인 확률 분포를 HMM의 state에서 가지는 형태의 HMM을 말한다. 따라서, DHMM은 확률 값의 계산이 용이한 반면 VQ과정에서 생기는 오차에 의해 모델의 정확성이 다소 떨어지는 단점이 있다. 또한 CHMM은 연속 확률 분포를 추정하여 state의 출력 확률 분포로 삼지만 Gaussian

mixture에 대한 연산의 복잡성이 크고, 모델의 저장 공간이 매우 커지는 단점을 가진다. 이 두 가지 형태의 HMM의 절충된 형태인 SCHMM은 VQ를 통해 각 codeword에 대한 Gaussian parameter (mean, variance)를 가지는 Gaussian pool 형태의 codebook을 형성하고 각 HMM의 state들은 그 codeword들의 weight만을 가진 형태로 state 출력 확률 분포는 Gaussian pool의 weighted sum 형태를 가지게 된다.

$$b_s(\vec{x}_t) = \sum_{i=1}^L b_s(O_i) N(\vec{x}_t; \mu_i, \Sigma_i) \quad (1)$$

식 (1)은 SCHMM의 state s에서 벡터  $\vec{x}_t$ 에 대한 출력 확률 값을 나타낸다. 여기서 L은 codebook의 크기이며,  $b_s(O_i)$ 는 벡터  $\vec{x}_t$ 의 이산 확률 값으로 weight를 나타내며  $N(\bullet)$ 은 Gaussian 분포를 나타낸다.

2.1.8 .3 Tied-mixture HMM modeling 기법

SCHMM의 codebook 생성을 위해 tied-mixture HMM modeling 기법을 적용 한다. Tied-mixture modeling 기법은 기존의 CHMM의 Gaussian mixture들을 이용하여 새로운 Gaussian pool을 형성하는 codebook을 생성하는 방법이다. CHMM은 각 state 별로 Gaussian mixture를 가지고 있고, mixture의 weight들을 가지고 있는 형태이다. Tied-mixture HMM modeling에서는 존재하는 모든HMM의 Gaussian mixture들을 하나의 pool에 모아서 설정한 크기의 codebook이 되도록 한다. 이 때, 존재하는 Gaussian의 개수가 codebook의 크기보다 작은 경우는 variance가 가장 큰 Gaussian부터 mean deviation을 통한 분할을 통하여 부족한 개수만큼 codebook 크기를 채우며, Gaussian의 개수가 codebook의 크기보다 많으면 weight가 작은 것부터 제거하는 방법으로 codebook을 형성한다. 이렇게 형성된 codebook을 각 HMM의 state들은 식 (1)과 같은 형태로 공유하며 각 codeword에 대한 codebook 크기 만큼의 weight array를 가진다. 그림 7.는 Tied-mixture HMM modeling 기법의 기본 개념도이다.

HMM parameter re-estimation 과정에서는 codebook의 각 codeword들의 mean과 variance, 그리고 각 state의 weight array를 re-estimate 하고, codeword들이 converge하면 weight array만 re-estimate한다.

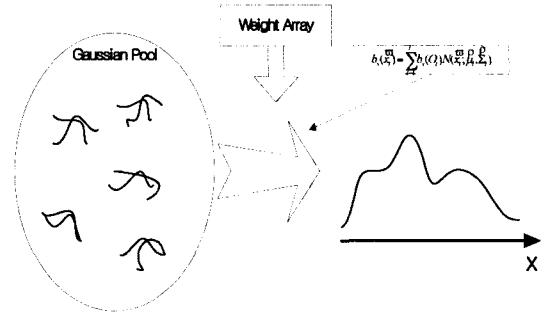


그림 7. HMI 공유 mixture 모델.  
Fig. 7. Tied-mixture HMM modeling.

가) Tied-mixture HMM의 구성

Tied-mixture HMM은 크게 세 가지의 parameter definition으로 구성된다.

(1) Gaussian Pool

각 Gaussian의 이름이 정의 되어 있으며 이름은 stream 정보와 codeword의 일련 번호를 포함한다. 예를 들어 첫번째 stream의 50번째 codeword에 해당하는 Gaussian이면 "TM\_1\_50"의 이름으로 정의 된다. 또한 mean vector와 covariance matrix가 정의 되어있다.

(2) State Macro Definition

State tying이 이루어져 있는 모델의 경우 state macro가 정의 되어 있으며 이름은 중심 음소의 이름과 해당 state 번호 그리고 class 이름으로 구성된다. 각 state macro에 해당하는 weight array가 정의 되어 있어서 실질 state를 구성하는 output distribution은 이 weight와 codeword들의 weighted sum으로 구성되게 되며, tying이 이루어진 state들의 공유 state parameter가 된다.

(3) HMM macro definition

실질적인 HMM의 정의를 한다. 여기서는 HMM의 state에 따라, stream에 따라 HMM의 parameter가 정의 되며, mixture 또는 state가 tying 되어 있는 경우는 이미 정의된 mixture 또는state macro의 이름을 link함으로써 중복 definition을 방지하고 모델의 저장 크기를 줄인다.

나) Reduced SCHMM

위에서 언급한 바와 같이 codeword들에 대한 weighted sum으로 state 출력 확률 분포를 계산하면 실질적인 state당 mixture 수가 매우 많게 되고, 계산 시간

또한 증가하게 된다. 따라서, 일정 기준 이하의 값을 가지는 weight는 고려하지 않는 reduced SCHMM 기법을 적용 하였다. Reduced HMM 기법에 관한 많은 논문들이 발표 되었다. 본 논문에서는 fixed threshold removing 알고리즘을 적용 하였다. 이 알고리즘은 고정된 기준치 이하의 weight를 0 으로 padding하고 계산 루틴에서 0 인 weight를 제외하는 방법이다. 설정한 threshold는  $2 \times 10^{-5}$  이다.

다) Context-dependent 음향 모델링

문맥에 따른 조음효과를 충분히 반영하기 위해 context window가  $\pm 1$  인 triphone을 사용하였다. 즉, 바로 앞과 뒤의 문맥에 따라 변하는 발성을 다르게 모델링 함으로써 모델의 변별력을 높였다. 그러나 triphone에 의한 모든 문맥을 고려하면 각 문맥에 해당하는 훈련데이터의 부족 현상과, 모델 definition의 증가로 인해 embedded system에 부적절한 모델 구성이 된다. 따라서 state tying 기법을 적용하였다. State tying 기법은 CHMM에서의 기법과 동일하며, unseen context 문제 해결을 위해 tree based clustering 알고리즘을 이용하였다.

따라서 HMM definition file은 이중의 parameter tying으로 이루어져 있다. state는 Gaussian pool의 Gaussian을 공유하고, triphone HMM들은 이 state를 tying하여 공유하고 있는 형태이다.

라) Tied-mixture HMM model 성능 평가

모델의 훈련은full training된 CHMM으로부터 tied-mixture Gaussian pool을 생성한 후, 5회의 codebook 및 weight를 re-estimation 하고, 5회의 weight re-estimation을 추가한 과정을 거쳐 이루어졌다.

성능 평가에 사용된 음성 데이터베이스는 공통 음성 데이터베이스 PBW(Phonetically Balanced Words)452 로써, 남성 화자 38명, 여성 화자 32명의 각 2회 발성으로 구성되어 있다. HMM의 구조는 3 state, 음소단위 모델이고, 남자 26명, 여자 24명이 각 2회 발음한 발성으로 훈련하였다. 모델 초기화 과정에서 4500개의 발성을 수작업 labeling을 한 bootstrapping 데이터가 사용되었다. 인식 테스트는 훈련에 포함되지 않은 남자 12명, 여자 8명이 각 2회 발음한 발성이 사용되었다. 성능평가의 결과는 <표 1> 과 같다.

표 1. Embedded system의 음향 모델의 크기 및 성능. Table 1. The size and performance of the acoustic model.

	인식률		모델크기	
	CHMM	SCHMM	CHMM	SCHMM
CI(256 Codebook Size)	94.06 %	93.43 %	2,498k B	1,010 kB
CD(16 Gaussian Mx)	99.04 %	98.55 %	21,082 kB	5,726 kB

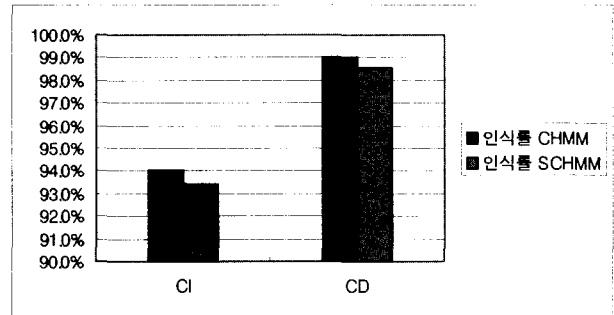


그림 8. Embedded system의 음향 모델의 성능 비교 그래프.

Fig. 8. The comparison of the embedded acoustic model.

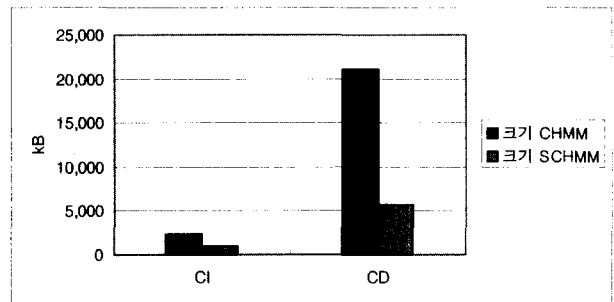


그림 9. Embedded system의 음향 모델의 크기 비교 그래프.

Fig. 9. The comparison of the embedded acoustic model.

III. 결론

텔레매틱스 및 차량내 각종 기기를 음성으로 제어할 수 있도록, 텔레매틱스와 연동할 음성 HMI 모듈을 DSP 보드로 개발하였다. 그리고, 차량 내장형 음성 기술과 차량 정보 센터의 결과물을 통합하는 시스템 보드를 구현하였고, 차량 내부 및 외부의 정보원으로부터 차량정보센터, 텔레매틱스를 통한 음성 정보 서비스의 효과적인 구조를 설계하여 자동차용 음성정보기술을 기반으로 한 음성 HMI 시스템을 개발하였다.

음성 HMI 시스템 보드는 TI 사의 부동소수점 연산이

가능한 TMS320C6711 DSP 칩을 사용하였고, 32MB의 SDRAM과 32MB의 Flash 메모리를 사용하여 구현하였다. 또한 음성 입출력 인터페이스 부분에는 실시간 음성인식엔진 구현을 위하여 타이머 인터럽트 혹은 DMA를 사용하여 일정한 간격으로 데이터를 샘플링하는 방안을 구현하였다. 이때 사용하는 코덱은 11.025KHz 샘플링 레이트와 16비트 양자화 레벨을 지원하며 이를 위한 별도의 라이브러리 함수를 구현하였다.

음성 HMI 시스템 보드에 포팅한 Embedded 시스템용 음향 모델은 기존의 CHMM의 저장 메모리 문제를 해결하고, 동시에 인식 성능의 저하를 최소화 하도록 구현하였다. 이러한 인식 성능 저하 문제 해결을 위해 반연속 HMM(SCHMM) 모델링 기법을 도입하였으며, tied-mixture 모델링 방법을 통하여 구현하였다. 또한 문맥 종속 모델(triphone)을 적용하여 앞뒤 문맥에 따른 조음효과를 최대한 반영하도록 하였다. 이렇게 구성된 음향 모델은 CHMM 모델 크기의 20% 정도의 크기를 가지며 인식 성능은 1% 미만의 범위로 줄어 든 것을 확인할 수 있다.

그리고 향후 음성 응용 application 프로그램을 용이하게 포팅하게 하기 위하여, 상위의 응용 프로그램 개발자가 사용하기 편하도록 H/W 자원을 접근하는 기능을 가진 함수를 구현하였다. 응용 프로그램 개발자는 H/W에 대한 깊은 전문 지식이 없어도 프로그램을 설계할 수 있도록 ANSI-C 혹은 DOS-C 수준의 사용하기에 편한 라이브러리 함수를 개발하였다. 그리고 외부 텔레메틱스 시스템과의 연동을 위하여 MAX232 칩을 이용하여 시리얼 통신 부분을 구현하였다.

## 참 고 문 헌

- [1] International Telematics Exhibition & Conference, 한국커머스넷, 2002.
- [2] L. R. Rabiner, R. W. Schafer, "Digital Processing Of Speech Signals", Prentice Hall, 1978.
- [3] L. R. Rabiner, B. H. Jung, "Fundamentals Of Speech Recognition, PTR Prentice Hall, 1993.
- [4] C. H. Lee., F. K. Soong and K. K. Paliwai, Automatic Speech And Speaker Recognition, Kluwer Academic Publishers, 1996.
- [5] Richard Sproat, Multilingual Text-to-Speech Synthesis: The Bell Labs Approach, Kluwer Academic Publishers, 1998.
- [6] Manfred R. Schroeder, Computer Speech, Springer, 1999.
- [7] Jonathan Allen, M. Sharon Hunnicutt, Dennis Klatt, From text to speech, Cambridge University Press, 1987.
- [8] Michael D. Riley, Speech Time-Frequency Representation, Kluwer Academic Publishers, 1989.
- [9] Daniel Jurafsky, James H. Martin, Speech & Language Processing, Prentice Hall, 2000.
- [10] D. K.Seo, et. al., "Endpoint detection in the car noise environment for speech recognition", The Journal of the Acoust. Soc. of Korea. Vol. 17, No. 1 pp.43-79,1998.

---

저 자 소 개

---



권 오 일(중신회원)

1991년 고려대학교 전자공학과  
학사

1993년 고려대학교 전자공학과  
석사

1996년 고려대학교 전자공학과  
박사

2000년 현대전자산업(주) 차장

현재 (주)현대오토넷 책임연구원