

# 자동차 환경내의 음성인식 자동 평가 플랫폼 연구

이 성 재\*, 강 선 미°

## A Study of Automatic Evaluation Platform for Speech Recognition Engine in the Vehicle Environment

Seong-Jae Lee\*, Sun-Mee Kang°

### 요 약

주행 중 차량내의 음성인터페이스 에서 음성인식기의 성능은 가장 중요한 부분이다. 본 논문은 차량내 음성인식기의 성능 평가를 자동화하기 위한 플랫폼의 개발에 대한 것이다. 개발된 플랫폼은 주 프로그램, 중계 프로그램 데이터베이스 관리, 통계산출 모듈로 구성된다. 성능 평가에 있어 실제 차량의 주행 조건을 고려한 시뮬레이션 환경이 구축되었고, 미리 녹음된 주행 노이즈와 발화자의 목소리를 마이크를 통해 입력하여 실험하였다. 실험 결과 제안하는 플랫폼에서 얻어진 음성인식 결과의 유효성이 입증되었다. 제안한 플랫폼으로 사용자는 음성인식의 자동화와 인식결과의 효율적인 관리 및 통계산출을 함으로서 차량 음성인식기의 평가를 효과적으로 진행할 수 있다.

**Key Words** : 차량 음성 인터페이스, 음성 인식기, 음성인식 자동 평가 플랫폼  
vehicle speech interface, speech recognizer, automatic evaluation platform

### ABSTRACT

The performance of the speech recognition engine is one of the most critical elements of the in-vehicle speech recognition interface. The objective of this paper is to develop an automated platform for running performance tests on the in-vehicle speech recognition engine. The developed platform comprise of main program, agent program, database management module, and statistical analysis module. A simulation environment for performance tests which mimics the real driving situations was constructed, and it was tested by applying pre-recorded driving noises and a speaker's voice as inputs. As a result, the validity of the results from the speech recognition tests was proved. The users will be able to perform the performance tests for the in-vehicle speech recognition engine effectively through the proposed platform.

### I. 서 론

현재 보급되고 있는 음성인식 기술은 로봇, 가전 제품 등의 다양한 분야에서 활용되고 있다. 특히 주행중 내비게이션, 오디오, 공조기 등 편의장치의 조작성이 제한된 자동차 환경에서 음성인식 기술의 중요성과 관련제품 시장의 규모는 날이 갈수록 커지고 있다. 그러나 자동차 환경에서의 음성인식은

엔진음, 공조음 같은 차량 모델별 내부 잡음과 노면 마찰음, 풍절음과 같은 외부 잡음 때문에 음성인식기 개발과 안정적인 상용화에 어려움이 있다.

자동차 환경에서 음성인식을 수행하는 장치 중 AVN(Audio Video Navigation)이 있다. 이것은 차량의 센터페시아에 위치하며, 사용자가 내비게이션, 라디오, 공조장치 등 차량의 전반적인 기능을 음성으로 조작할 수 있게 하는 기능을 탑재하고 있다.

\* 주저자 : 서경대학교 전자공학과 Intelligent Human & Computer Interface 연구실, sj3452@cyworld.com, 준회원

° 교신저자 : 서경대학교 전자공학과 Intelligent Human & Computer Interface 연구실, smkang@skuniv.ac.kr, 종신회원

논문번호 : KICS2012-05-250, 접수일자 : 2012년 5월 14일, 최종논문접수일자 : 2012년 6월 27일

상기와 같은 AVN의 상용화를 위하여 차량별, 화자별 등 각종 조건을 고려한 음성인식기의 성능을 평가하는 방법은 발화자가 직접 차량에 탑승하여 AVN에 육성으로 명령을 내리거나, HATS(Head and Torso Simulator)를 이용하여 사람 대신 녹음된 음성을 발화하여 인식 결과를 기록하는 방법이 있다<sup>1)</sup>. 이와 같은 방식은 음성인식기 시험 전과정을 사람이 직접 차를 주행하며 각각 발화자별로 진행하기 때문에 시간 및 비용의 문제가 발생한다. 또한 시험한 결과를 일일이 전산입력을 해야 하기 때문에 시험 결과의 사후 관리에도 어려움이 있다.

본 논문에서는 기존의 시험 방법을 개선하여 자동차 환경내의 음성인식기의 평가 과정을 실제 차량을 운행하지 않고도 가상 시뮬레이션을 통하여 수행하는 것을 목적으로 하는 시스템 구성에 대하여 제안하였다. 이를 위하여 주 프로그램(Main Program)은 음성인식기를 능동적으로 제어하기 위하여 설계된 중계프로그램을 통해 AVN과 실시간으로 기체취한 잡음과 녹음된 음성 명령어 데이터를 주고받아 음성인식기 성능을 자동으로 평가 가능하게 하며, 이를 통해 수집된 음성인식 결과들을 자동으로 온라인 데이터베이스화 하여 효율적으로 관리한다. 또한 발화자, 주행환경, 차량모델 등 어떠한 요인이 음성인식기의 성능에 영향을 미치는지 알아보기 위한 통계 분석을 실시하는 모듈을 제공하여 평가에 있어 최소한의 기회비용으로 높은 효과를 볼 수 있게 한다.

본 논문의 구성은 제 2장에서 제안한 시스템의 전체 구성에 대하여, 제 3장에서는 시스템 설계 및 구현에 대하여 기술하고, 제 4장에서 실험 및 결과 그리고 제 5장에서 결론을 맺는다.

## II. 전체 시스템 구성

본 논문에서 제안하는 전체 시스템은 크게 인식 명령어와 주행 노이즈 관리 및 재생과 음성인식기를 제어하는 주 프로그램과 음성인식기와 주 프로그램간의 통신을 담당하는 중계 프로그램으로 구성되었다. 시스템의 구성도는 그림 1 과 같다.

주 프로그램은 MICROSOFT社의 C# .NET 플랫폼에서 개발되었으며, PC에 Standalone Software 형태로 설치된다. 구성은 음성인식기의 제어를 담당하는 제어부와 사용자의 음성 명령어를 스피커를 통해 출력하는 출력부, 시험이 완료된 결과를 효율적으로 관리하기 위한 데이터베이스 관리 모듈과

저장된 데이터를 기반으로 한 통계 모듈이 있다.

중계 프로그램은 AVN 보드내에 S/W 형태로 설치되며 주 프로그램에서 송출되는 음성인식기 제어 신호를 수신하고, 주 프로그램으로 음성인식 결과를 전송하는 송 수신부와 수신한 신호를 해석하여 음성인식기를 제어하는 제어부로 나누어진다.

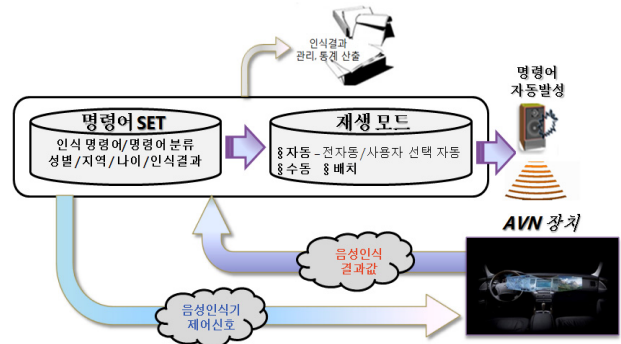


그림 1. 제안한 전체 시스템 구성도  
Fig. 1. Overall structure of the proposed system

음성인식기 평가를 자동화하기 위하여 사용자는 발화자의 명령어, 성별, 나이 등의 정보가 수록된 파일을 주 프로그램으로 불러온 뒤, 중계프로그램을 통해 AVN을 가동하여 음성인식기를 인식대기 상태로 진입하게 한다. 이후 발화 시작을 알리는 신호를 전송하여 AVN의 마이크를 개방하고, 스피커를 통해 미리 녹음된 음성인식 명령어를 출력 한다. 이때 출력된 음성은 마이크를 통해 음성인식기에 입력된다. 음성의 출력이 종료됨과 동시에 음성인식기에 종료 신호를 전달하고 인식기로부터 얻어진 음성인식 결과(인식, 오인식, 인식불가)는 다시 중계프로그램을 통하여 주 프로그램으로 전송된다. 자동모드의 경우 상기 전 과정을 명령어 SET 파일, 혹은 배치 파일 목록의 마지막 항목까지 자동 반복한다.

## III. 시스템 설계 및 구현

### 3.1. 주 프로그램

주 프로그램은 각 모듈과 데이터들을 제어하고 관리하는 기능을 담당하며 전체 시스템에서 중추적 역할을 수행한다. 동작순서는 발화자의 이름, 나이, 지역과 음성인식 명령어가 포함된 데이터를 불러온 후 사용자의 선택에 따라 다양한 옵션을 지정하여 음성인식기 평가 준비를 완료한다. 사용자가 평가 시작 명령을 내리면 저장된 명령어를 재생하고 중계 프로그램으로 부터 인식결과 문자열을 수신 받아 저장한다. 상기과정을 그림 2 에 나타내었다.

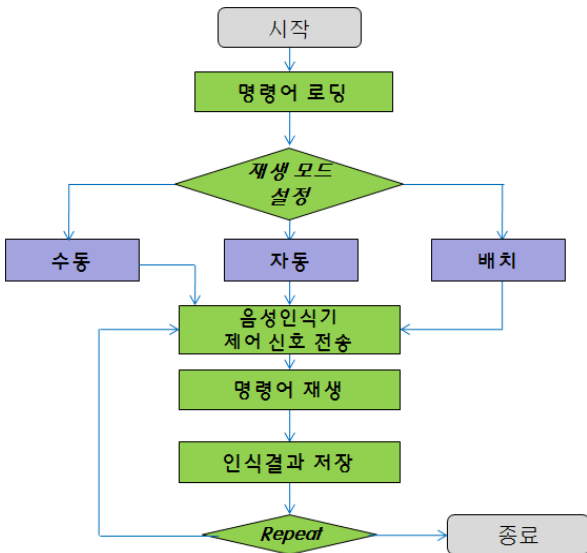


그림 2. 주 프로그램의 동작 순서도  
Fig. 2. Flow chart of main program

3.2. 데이터베이스 관리

음성인식 평가 자동화 과정을 거치면 주 프로그램의 데이터를 저장하는 관리모듈에 명령어 인식결과가 기록된다. 본 논문에서는 음성인식 명령어 SET 파일의 항목들을 대상으로 한 Client/Server 기반의 온라인 데이터베이스를 구축하여 발화자의 인적사항 및 명령어와 음성인식기로 부터 전송된 인식결과를 저장한다. 또한 별도의 데이터베이스 관리 모듈을 추가하여 인식결과 목록을 서버에 저장하고 사용자가 원하는 조건에 적합한 데이터를 검색할 수 있게 한다. 이는 네트워크만 연결된 환경이라면 장소와 PC, 노트북, 스마트폰 등의 접속장비에 구애받지 않고 명령어 SET 및 결과 파일을 체계적으로 관리할 수 있다. 또한 본 논문에서 개발된 주 프로그램과 통계분석 틀이 서버에 저장된 데이터베이스를 공유함으로써 효율성을 극대화한다.

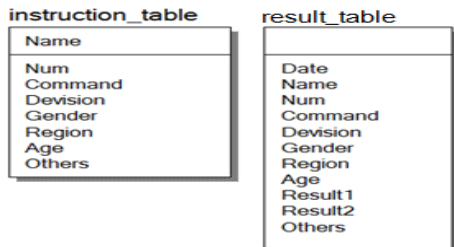


그림 3. 구축된 데이터베이스 테이블 스키마  
Fig. 3. Database table schema in the system

그림 3에 도시된 구축된 데이터베이스의 테이블은 발화자의 정보와 명령어가 저장된 Instruction과 인식된 결과를 저장하는 Result 테이블로 구성된다.

3.3. 통계 분석

자동 인식실험 결과를 데이터베이스화하여 저장된 데이터는 단순히 저장 및 열람뿐만 아니라 인식 성공, 오인식, 인식실패와 같은 인식결과의 비율 산출과 화자의 성별, 거주지역과 같은 요인이 인식률에 어떻게 영향을 미치는지 알아보기 위한 통계 분석에 사용된다. 이를 위하여 본 논문의 통계 분석 모듈에서는 분석방법을 기초분석과 교차분석으로 분류하였다.

기초분석은 전체 항목 또는 성별 및 연령 등의 전체 항목에 대한 각각의 인식률을 산출한다. 그림 4는 기초분석 방법을 통해 산출한 주행속도별 인식률을 나타낸 것으로서, 좌측 상단의 수치해석 테이블은 각각의 항목에 대한 개수와 전체항목 중 차지하는 비율을 나타낸다. 하단의 그래프 중 ‘First’와 ‘Second’ 항목은 각각 1차와 2차 시도 인식 실험에서 음성인식이 성공했음을 의미하며, 나머지 항목들은 인식에 실패한 경우들을 나타낸다. 두개의 그래프 축 중 좌측은 정차시를, 우측은 120km/h 주행상황을 나타낸다. 이를 통하여 주행속도가 올라갈수록 인식률이 저하됨을 알 수 있다.

교차분석은 상기에 열거된 성별 및 연령과 같은 각종 요소들이 인식률에 영향을 미치는 정도를 산출하는 방법으로서, 통계학에서 사용되는 ANOVA (Analysis of variance)<sup>1)</sup> 분석법을 적용하였다.

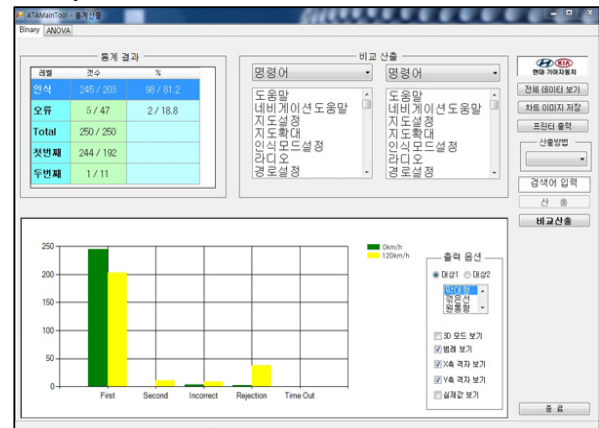


그림 4. 통계모듈을 통해 분석된 차량속도별 음성인식률  
Fig. 4. Result in speech recognition test using the statistical analyzer module

1) 두 개 이상 다수의 집단을 비교하고자 할 때 집단내의 분산, 총 평균과 각 집단의 평균의 차이에 의해 생긴 집단 간 분산의 비교를 통해 만들어진 F분포를 이용하여 가설검정을 수행함

### 3.4. 중계 프로그램

중계 프로그램은 주 프로그램과 음성인식기 간의 중계자 역할을 한다. 주 프로그램의 상태 및 정보는 중계 프로그램을 통해 음성인식기에 전송되며, 음성인식기는 주 프로그램에게 자신의 상태 및 음성인식 결과 값을 전송한다. 전체 시스템에서의 역할은 다음과 같다.

첫째, 주 프로그램에서 보내는 신호를 통하여 음성인식기를 제어한다. 예를 들어 주 프로그램에서 종료 신호를 전송하면 중계 프로그램은 음성인식기를 강제 종료시킨다. 이는 스피커를 통한 화자의 명령어 발성이 종료되는 즉시 신호를 전달함으로써 시험 속도를 높이기 위함이다.

둘째, 음성인식기가 스피커를 통해 발화된 명령어를 인식한 결과를 주 프로그램에게 전송하는 역할을 한다. 주 프로그램에서는 중계프로그램으로 부터 수신된 결과를 원래의 명령어와 비교하여 인식 성공여부를 판별한다.

주 프로그램과 중계 프로그램간의 통신에 사용되는 신호들은 표 1과 같다.

표 1. 중계 프로그램과 주 프로그램간의 통신 인터페이스  
Table 1. Communication interface between main and agent program

전달 신호	의 미	분 류
0	중계 프로그램 시작	음성인식기 상태
1	인식 프로세스 시작	
2	인식 프로세스 종료	
3	인식명령어 재생	
6	다이얼로그 시작	
7	다이얼로그 종료	음성인식 결과
rejection	음성인식 실패	
4	음성인식 시간 초과	
result	음성인식 결과 값	

## IV. 실험 및 결과

본 논문에서 구현한 시스템의 검증은 제안한 시스템으로 수행한 음성인식기의 인식 결과가 실제 차량에서 수행된 인식 결과와 유사한 결과를 나타내는지 확인하는 것이다. 이를 위하여 PCM wave 파일로 녹음된 발화자의 음성명령어 데이터는 (주)미디어젠 으로부터 제공받은 것으로서 남성과 여성을 동등한 비율로 구성하였다. 음성인식 결과에 영향을 미치는 차량 내부의 잡음 데이터는 (주)현대모

비스에서 제공받은 ‘YF 소나타 하이브리드’ 모델의 것을 사용하였다. 잡음 데이터는 정차, 60km/h, 120km/h 주행상황으로 구성하였다. 본 논문에서 사용한 인식 단어는 현대자동차 AVN 에서 사용되는 AV, BT, HELP, NAVI, ETC 명령어로 구분되는 50개의 명령어를 사용하였다.

음성인식 모듈의 언어모델은 AVN 명령어에 적합하게 학습된 ETRI 에서 제공된 낭독체 데이터베이스를 사용하였으며, 음성인식기는 Carnegie Mellon 대학에서 개발한 sphinx모델을 이용하여 음성인식을 수행하였다.

실험 방법은 노이즈 녹음에 사용된 모델의 차량 주행을 하며 현장의 노이즈를 직접 소음측정기로 음량측정 후, 16bit로 양자화된 녹음 노이즈와 명령어 발화 목소리를 표 2의 측정된 결과 값과 동일하게 음량이 설정된 스피커를 사용하여 재생하였다. 실험에서 음성과 노이즈의 입력은 AKG社 에서 제작한 지향성 마이크를 사용하였으며, USB 타입 Audio Interface 카드를 통하여 PC로 전송하였다. 음성 출력은 BEHRITONE社의 스튜디오 모니터용 스피커를 사용하였다. 이후 각 발화자당 50개의 단어를 3가지의 주행상황 별로 5번씩 총 750번 재생하여 그 결과의 평균을 계산하였다.

표 2. 측정된 자동차 주행 소음량 (단위 : dB)  
Table 2. Noise level measurement inside the vehicle (Unit : dB)

주행속도 측정시기	정차	60km/h	120km/h
	1차	49.4	70.0
2차	55.0	70.6	73.7
3차	53.0	69.4	75.3
평균	52.46	70.0	74.0

그림 5는 본 논문에서 제안한 시스템을 사용하여 측정된 주행상황 및 화자별 음성인식 성공률의 결과를 나타낸다. 주행 속도가 올라갈수록 음성인식 성공률은 낮아졌으며, 최고 인식률 면에서 정차시 99.2%를 나타내었으며, 60km/h와 120km/h 주행이 각각 96%와 87.6%를 기록했다. 전체적으로 화자별 인식률 감소 정도가 유사하여 가상 시뮬레이션 환경이 신뢰성이 있음을 보여준다.

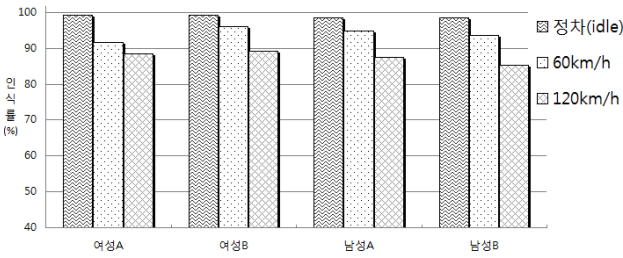


그림 5. 제안한 시스템의 음성인식 성공률 결과  
Fig. 5. Success rates in speech recognition test using the proposed system

### V. 결 론

본 논문은 차량내 음성인식기 성능 향상을 위한 자동 테스트 플랫폼 개발에 있어 음성인식 명령어 SET와 전체 모듈을 관리하는 주 프로그램, 음성인식기와 주 프로그램간의 음성 및 데이터 통신을 수행하는 중계 프로그램, 효율적인 데이터 관리를 위한 데이터베이스 관리 툴, 음성인식 데이터를 토대로 인식률에 영향을 미치는 요소를 파악하기 위한 통계분석 툴의 개발에 대한 것이다.

본 논문에서 구현한 시스템의 검증을 위하여 실제 차량과 유사한 환경을 조성하여 시험을 진행하였으며, AVN에서 사용하는 명령어와 각 주행 속도별 노이즈를 각각의 모니터 스피커를 통하여 재생하였다. 실험 결과 평균 인식률은 정차시 98.8%, 60km/h 주행시 94%, 120km/h 주행시 87.6%를 나타내었다. 또한 오류가 자주 발생하는 명령어 패턴이 존재함을 확인하였다.

본 논문의 시스템 개발로 인하여 차량내 AVN의 음성인식기의 테스트 방법에 있어 사람이 차량에 탑승하여 직접 음성인식 명령어를 발화하여 수동으로 인식 혹은 오인식 여부를 확인해야했던 기존 방식과 달리, 전체 시험 과정을 자동화하고 음성명령어와 주변노이즈를 시뮬레이터로 생성하여 유사한 결과를 얻을 수 있게 되었다. 뿐만 아니라 인식결과를 토대로 자동 통계분석을 실시하여 어떠한 요소가 음성인식기 성능에 영향을 미치는지도 알 수 있어 차량 음성인식기의 성능 향상에 큰 기여를 할 것으로 기대한다.

향후 과제로서 노면, 날씨 등의 더욱 다양한 노이즈 환경 조건을 추가하여 현실성이 증가된 평가를 진행해야 할 것이다.

### References

- [1] JinWoo Yang, SoonHyob Kim, "A study on Speech Recognition in a Running Automobile", *The Journal of the Acoustical Society of Korea* pp.3-8 ISSN 1225-4428, 2000.
- [2] KwangHyun Lee, "Comparison of Speech Intelligibility & Performance of Speech Recognition in Real Driving Environments", *The Joournal of Korean Society of Speech Sciences* pp.99-110 ISSN 2005-8063, 2004.
- [3] JongGeon Kim, YounSeok Choi, NaeYoung Lee, "Unbalanced ANOVA for Testing Shape Variability in Statistical Shape Analysis", *The Korean Journal of Applied Statistics*, p317-323, 2010.
- [4] 강성훈, "Head and Torso Simulator를 이용한 음향측정 및 음장재생"(SeongHoon Kang, "Sound measurement and acoustic field playback using Head and Torso Simulator"), *The Joournal of Korean Society of Speech Sciences* pp.80-85 ISSN 1225-4428, 1992
- [5] P.Damaske, "Head-related two channel stereophony with loudspeaker reproduction." *J.Acoust. Soc. Am.* 52,pp 1109-1115, 1971.
- [6] Takeshi Yamadam, "Performance Estimation of Speech Recognition System Under Noise Conditions Using objective Quality Measures and Artificial Voice", *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO.6, NOVEMBER 2006.*
- [7] CheolJun Hwang, DeokSoo Kim, HyunYeol Chung, "Performance Evaluation of Multimodal Speech Recognition System by On-Line Tests", *The Journal of Korea Information and Communications Society*, 1998.
- [8] Nicolas Scheffer, Luciana Ferrer, "THE SRI NIST 2010 SPEAKER RECOGNITION EVALUATION SYSTEM", *IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague, 2011.*

**이 성 재 (Seong-Jae Lee)**



2011년 2월 서경대학교 컴퓨  
터과학과 졸업

2011년 3월~현재 서경대학교  
전자컴퓨터공학과 석사과정  
<관심분야> 음성신호처리, 패  
턴인식

**강 선 미 (Sun-Mee Kang)**



1981년 2월 고려대학교 전자  
공학과 졸업

1988년5월 Erlangen-Nurnberg  
대학교 전기전자 공학과  
Diplom

1992년 8월 고려대학교 전자  
공학과 박사

1997년 3월~현재 서경대학교 전자공학과 교수  
<관심분야> 음성신호처리, 영상신호처리, 패턴인식