자동차용 음성 DB 구축 시스템 개발

Database Collection System for the Automotive Environment

권 오 일*
Ohil Kwon

ABSTRACT

We collect the Korean Database which can be trained for the speech recognition engine in an automotive environment. We describe the overall trends of the Korean database collections in this paper and suggest a database collection method for the speech recognition system of the car-kit and explain several conditions in collecting the database in the automotive environments.

Finally, we expain an effective method of the Korean database collection in the automobile and the results of the database colletions, and the devised softwares used for the collection of the database.

Keywords: Database, TTS, PBW, DAT, Segmentation, Validation

1. 서 론

화자 독립 음성 인식기를 개발하는데 있어서 인식기를 개발하는 것만큼이나 중요하고 필수적인 작업으로 인식 및 학습에 사용되는 음성 데이터베이스를 구축하는 것으로 음성 데이터베이스는 음성 인식기를 훈련시키는데 필요할 뿐만 아니라 구현된 음성 인식기의 오류를 찾는데도 매우 유용하다. 음성 데이터베이스를 얼마나 잘 구축했는가에 따라서 인식기의 성능이 좌우되며, 음성 데이터베이스와 음성 인식기는 매우 밀접한 관계를 갖는다고 말할 수 있다.

음성 인식기의 구현 과정 중 가장 먼저 할 일은 적용 분야를 결정하는 일이다. 그리고 이를 바탕으로 음성 인식에서 사용될 최적의 기본 단위(subword, 예를 들어 단어, 음절, 음소)[1][6]를 선정한 후 발성 목록에 이 기본 단위(만일 기본 단위가 음소라면 각 음소들)를 고르게 포함되도록 한다(PBW; Phonetically Balanced Words).[2][3][4] 또한 가급적 발성 문장과 단어 수가 적을수록 좋다. 왜냐하면 음성을 제공하는 사람의 수고를 줄일 뿐 아니라, 데이터베이스 구축시간, 메모리, 처리시간, 훈련시간 등을 줄일 수 있기 대문이다. 이 내용을 모두 준비한 후 계획에 따라 음성 녹음작업에 들어가면 된다. 그런데 한 가지 참고할 사항은 항상

* (주) 현대오토넷

여분의 발성을 더 받아야 한다는 것이다.

음성의 각 부분에 대응하는 음절, 혹은 음소 기호를 할당하는 것을 '레이블링'이라고 한다. 레이블링 단위는 단어, 문장 등도 가능하며, 음소보다 더 작은 단위를 이용할 수도 있다.[5][6][7] 그리고 음성 파형에서 음소의 경계를 검출하는 일을 세그멘테이션(segmentaion)이라고 한다. 일반적으로 2 개 이상의 음소가 연결되어 있을 경우 세그멘테이션 작업은 음성, 음운학적인 지식이 요구된다.[8][9] 이런 작업은 보통 스펙트로그램(spectrogram), 에너지, 영교차율 등을 참고해 이루어진다. 따라서 세그멘테이션 작업은 음성 데이터베이스 구축에 있어서 가장 힘든 부분이며, 가장 중요한 부분이기도 하다.

본 논문에서 구축한 음성 데이터베이스는 자동차용 음성 인식기 개발을 위한 것으로 자동차 환경에서 주로 사용하는 몇 개의 명령어를 음성 데이터베이스 구축의 기본 단위로 하였다. 또한 '세그멘테이션'의 의미를 달리 설정하였다. 자동차 환경 하에서 한 화자가 전체 인식 단어를 일정 시간에 발음하도록 하여 하나의 음성 파일로 녹음하였기 때문에 이를 단어 단위로 분할하는 것을 '세그멘테이션'으로 명명하였다. 본 논문에서는 국내에서 구축되어 있는 음성 데이터베이스의 현황과 자동차용 음성 인식기 개발을 위한 음성 데이터베이스 구축 방법, 그리고 수집 환경에 대해서 기술하였다. 마지막으로 음성 데이터베이스 구축의 결과와 차후 음성 데이터베이스 구축시 보완할 점 및 앞으로의 향후 계획에 대해서 기술하였다.

2. 국내 음성DB 현황

음성 데이터베이스는 음성 인식 및 합성 등 음성 신호처리에 있어서 매우 중요한 요소이다. 음성 인식과 같이 데이터로부터 훈련하여 시스템을 구성하는 경우 주의 깊게 디자인하고 검증된 양질의 대량 음성 데이터베이스가 시스템의 성능을 좌우한다. 또한 음성 합성에서는 음성 데이터베이스에 따라서 합성 음질이 좌우된다. 그러나 이러한 음성 데이터베이스를 제작하는 것은 많은 노력과 비용이 소요된다. 우리말 음성 DB의 경우 공통적으로 사용할 수 있는 데이터베이스가 소수인 실정이다. 또한 표준화 등 여러 가지 문제점을 안고 있다.

음성 언어 코퍼스에 관한 중요성을 일찍부터 인식한 ETRI는 1980년대 중반부터 지속적으로 음성DB를 구축하여 왔다. 특히 단음절, 단독 숫자음, 연결숫자음, 각종 기능제어명령어, 음성밸런스 단어(PBW) 445 단어, POW (Phonetically Optimized Word) 3,848 단어, 호텔예약 테스크의 문장음성(낭독음성), 스케쥴링 테스크의 모의대화음성(자유발화), 발화의 자유도에 따른 다단계음성DB 등이 구성되어 지속적인 양의 확대와 레이블링이 계속되고 있으며 최근에는 음성 번역연구를 위하여 자유발화 음성 자료의 확보에도 노력을 기울이고 있다.

과기부의 지원으로 SERI, KAIST와 함께 원광대가 주관하여 공동이용을 목적으로 한 4연숫자, 단독숫자, 단문, PBW 452 단어, PBS(Phonetically Balanced Sentence) 589 문장 등의 음성DB를 구축한 바 있다.

한국통신의 멀티미디어연구소에서는 기관의 특성상 주로 전화음성코퍼스를 주 대상으로 하고 있다. 즉, 전화회선을 통한 단어 및 문장음성 코퍼스를 구축하여 자체 인식 시스템 개발에 사용중이다. 최근에는 다양한 발성 및 회선 환경을 고려한 음성, 전화음성자동수집시스템

에 의한 대량의 전화음성, 그리고 연속음성인식시스템 구축을 위한 열차표 예약 테스크의 연속음성 등의 코퍼스가 구축되었다.

한국과학기술원(KAIST)에서는 공동이용을 목적으로한 3,000 단어 규모의 무역관련 문장음성, 가변길이 연결숫자음, Phonetically Balanced 75고립단어, 지역명 관련 500 단어 등이 작성되었다.

3. 자동차용 음성DB 구축 시스템

자동차용 음성 데이터베이스는 자동차 주행시 발생할 수 있는 모든 상황, 예를 들어 녹음하는 사람의 심리상태나 성별, 나이, 방언과 같은 인적 요소와 주행 속도, 주변 소음, 기상상태, 도로상태와 같은 환경 요소들을 모두 고려한 음성 샘플의 수집 조건들을 설정한 후 음성을 녹음 수집하여 구축한 데이터베이스를 의미한다. 자동차 환경의 음성 데이터베이스를 효율적으로 구축함으로써 음성 인식기의 개발을 위한 전처리 과정에서의 음성 및 소음 분석에 사용할 수 있는 자료로서 활용할 수 있다.

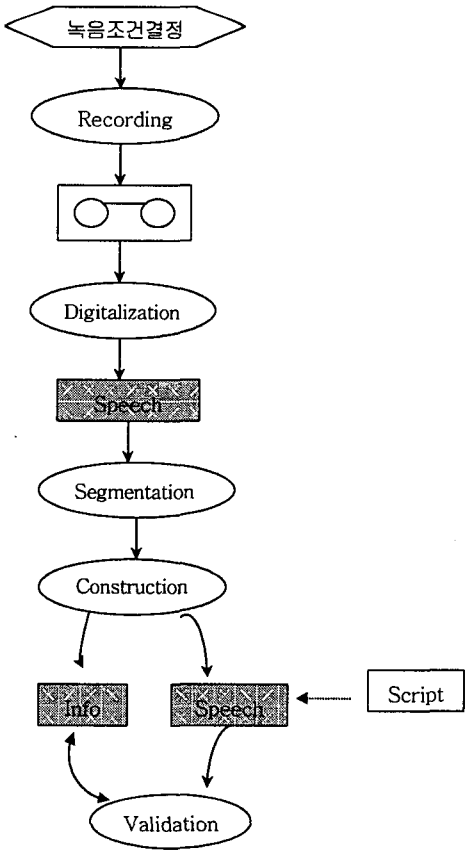


그림 1. 음성 데이터베이스 구축 흐름도

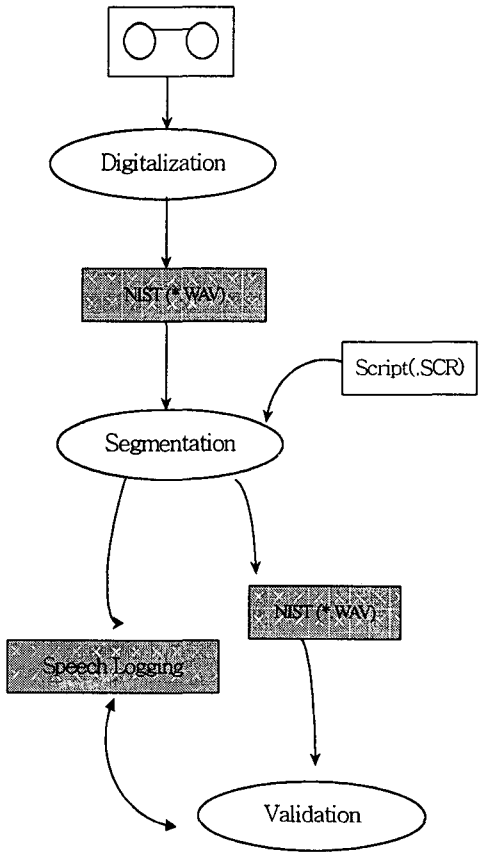


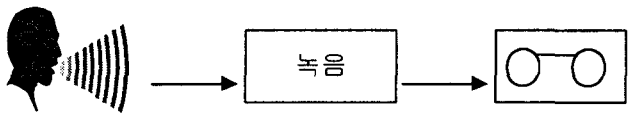
그림 2. 음성 데이터베이스 데이터 흐름도

음성 데이터베이스 구축 시스템의 전체적인 구조는 녹음 작업(Recording), 디지털 변저장 작업(Digitalization), 단어 단위별 분할 작업(Segmentation), 음성 데이터베이스 구축 작업 (Construction), 웨이브 파일의 등급을 결정하는 확인 작업(Validation)으로 구성되어 있다. 본 논문에서 개발된 음성 데이터베이스 구축 소프트웨어는 세 개의 프로그램으로, DAT(Digital Audio Tape Recorder)에 녹음된 음성을 웨이브 파일로 변환 저장하는 Digitalization 프로그램과 전체 인식 단어가 저장된 파일을 단어 단위로 분할하고, 음성 데이터베이스를 구축하는 Segmentation 프로그램, 웨이브 파일의 확인 작업을 하기 위한 Validation 프로그램이다.

음성 데이터베이스는 트리 형태로 구성되어 있으며, 웨이브 파일이 저장된 SPEECH 디렉토리와 웨이브 파일에 대한 여러 가지 정보가 기록된 정보 파일이 저장된 DATA 디렉토리로 구성되어 있다. 그림 1과 그림 2는 음성 데이터 베이스의 구축 흐름도와 데이터 흐름도를 나타낸 것이고, 그림 3은 음성 데이터베이스의 구조를 보여주고 있다.

3.1 녹음

자동차 환경에서 발생할 수 있는 인적, 환경적 요소들을 모두 고려한 녹음 환경 테이블을 만든 다음 각 화자마다 세 가지 녹음 환경을 설정하여 음성을 DAT(Digital Audio Tape recorder)에 녹음한다.



3.2 디지털화

디지털화 작업은 화자 단위로 전체 인식 단어가 DAT에 녹음된("세션"이라 함) 음성을 디지털로 변환시켜 웨이브 파일로 컴퓨터에 저장하며, Digitalization 프로그램을 사용한다. 저장 가능한 파일은 세 가지 형태가 가능하고, 본 시스템에서는 NIST 웨이브 파일 형태를 사용하고 있다.

3.3 분할

분할 작업은 세션 단위로 저장된 웨이브 파일을 단어 단위로 나누는 작업이며 Segmentation 프로그램을 사용한다. 분할 작업의 결과로 SGL (Segmentation Logging) 파일이 생성되며, 음성 데이터베이스를 구축할 때 사용되는 정보 파일로서, 분할에 관련된 정보가 저장되어 있다.

3.4 음성DB 구축

디지털화 작업으로 생성되는 웨이브 파일을 스크립트(Script) 파일과 웨이브 파일에 연결된 SGL파일의 분할 정보를 이용하여 웨이브 파일에 대한 정보를 기록한 SPL파일과, 단어별 웨이브 파일들을 생성한다. 스크립트 파일은 음성 데이터베이스 구축을 위해 시스템 및 녹음,

전반적인 사항이 기술되어 있는 파일이다. 음성 데이터베이스는 Segmentation 프로그램에서 지원해 주는 기능으로 자동으로 구축된다. 음성 데이터베이스의 구조는 그림 3과 같이 트리(Tree) 형태이며, 특징은 "Localize"하다는 것이다. 이것은 분할 작업이 끝난 웨이브 파일이 음성 데이터베이스 구조의 어느 부분에도 삽입이 가능하며 삭제도 간단하다는 것을 의미한다. 또한 본 시스템으로 구축된 음성 데이터베이스는 웨이브 파일들에 대해 모든 정보가 저장되어 있는 정보 파일(SPL파일)이 연계되어 있는 특징을 가지고 있다. 음성 데이터베이스는 크게 두 개의 디렉토리로 구성되어 있는데 DATA와 SPEECH이다.

DATA디렉토리는 아래에 기술된 예제와 같이 스크립트 번호, 장치 번호, 그룹 번호의 하부 디렉토리로 구성되며, 그룹 번호 디렉토리 내에 SPL파일들이 기록된다. 스크립트 파일은 음성 데이터베이스 구축시 사용되는 파일로 본 논문에서는 30 개 단어와 50 개 단어의 음성 데이터베이스 구축을 위해 두 개의 파일을 만들었다. 장치(Unit) 번호는 많은 시간을 요하는

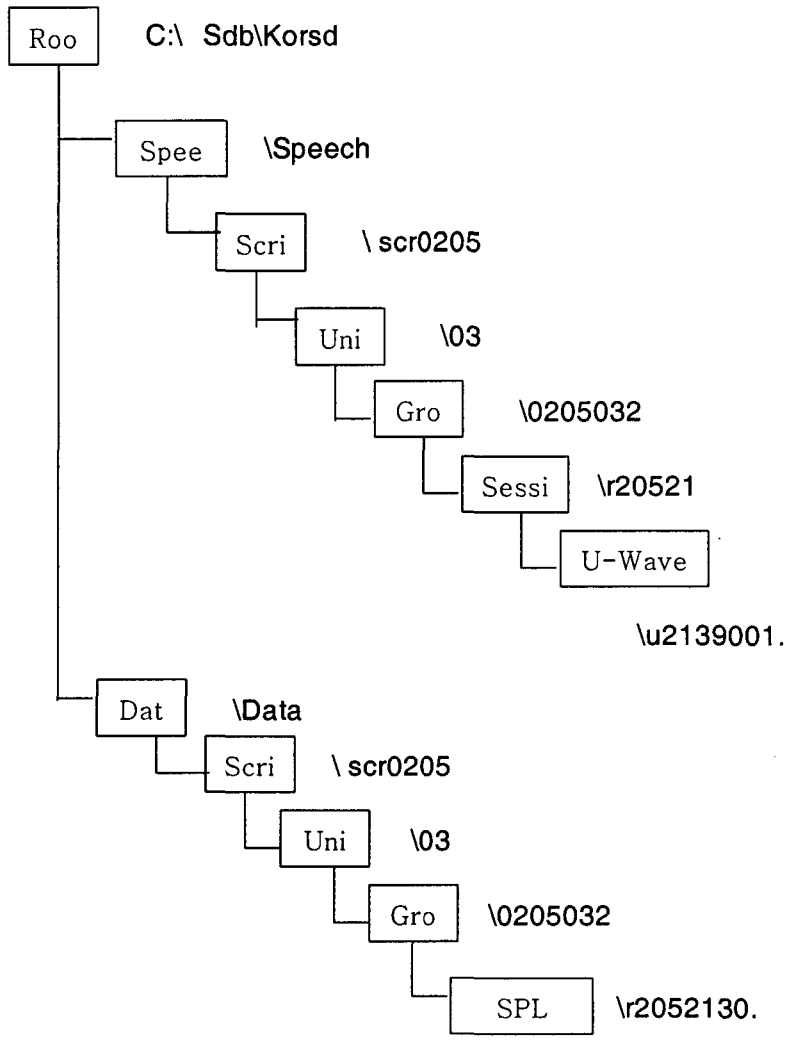


그림 3. 음성 데이터베이스의 디렉토리 구조

확인 작업을 여러 사람이 여러 대의 컴퓨터에서 작업하기 때문에 이를 구별하기 위한 것이다. 그룹(Group)은 많은 양의 데이터를 분류하기 위한 것이며 한 개 그룹은 100 개의 단위로 구성된다. 그룹 번호 하부 디렉토리에는 한 세션마다 한 개의 SPL(Speech Logging) 파일이 생성되며, 파일의 형태, 크기 및 웨이브 파일에 대한 여러 가지 정보가 저장된다.

SPEECH 디렉토리에는 DATA디렉토리와 마찬가지로 스크립트 번호, 장치 번호, 그룹 번호의 하부 디렉토리로 구성되어 있고, 그룹 번호 디렉토리 아래에 세션 번호 디렉토리가 존재하며, 이 디렉토리 아래에 단어 단위의 웨이브 파일들이 저장되어 있다. 세션(Session)은 한 명의 화자가 주어진 수집 조건의 환경 하에서 녹음한 것을 의미한다.

3.5 확인

확인 작업은 음성 데이터베이스로 구축된 각각의 웨이브 파일에 대해 어떤 소음이 포함되어 있는지 음성에 대한 평가를 하며, 음성의 앞뒤에 약 200 msec 묵음 구간이 있도록 끝점 검출(End point Detection)을 한다. 또한 화자에 대한 여러 가지 정보, 예를 들어 이름, 나이, 지역, 녹음 환경, 자동차 모델, 참고 사항을 기록한다.

음성의 등급 평가는 먼저 소음(Noise)이 포함되어 있는지 유무를 검사한 다음, 그 소음이 어떤 종류이며 어느 정도인지를 아래에 기술한 기준에 의하여 평가한 후 등급 (A~E)을 결정한다. 웨이브 파일이 A와 B등급이면 인식 또는 학습하기에 좋다고 평가할 수 있으며, C등급은 보통이고, D와 E등급은 사용하기에 나쁘다고 할 수 있다.

소음은 NOI, SND, SPC, UTT, DST의 다섯 가지로 분류한다. NOI(noise)는 소음이 명확하게 들릴 때, SND(sounds)는 주변 소음이 포함되어 있을 때, SPC(speech not speaker)는 화자 이외의 사람에게서 발생하는 소음이 있을 때, UTT(utterance)는 화자가 단어를 발음할 때 생기는 소음 및 오류가 발생할 때, DST(distortion)는 발음에 클리핑(cliping)과 같이 왜곡이 있을 때를 의미한다.

음성의 등급 평가 기준

A: 소음이 거의 없는 이상적인 음성

B: 음성이 비교적 좋으며, 아래와 같은 소음이 허용된다.
소음이 적음/ 낮은 기침/ 명확하지 않은 에코/ 숨소리/ 천천히 발음함.
(음성에 위의 소음들이 복합적으로 나타날 수도 있다.)

C: 음성의 질이 보통일 때
B에서 정한 기준보다 더 크게 들리는 소음
음성의 앞뒤 묵음 구간이 100 msec보다 짧고, 10 msec보다 길 때

D: 순간 소음이 음성을 초과할 때
발음이 올바르고 이해할 수 있을 정도
앞뒤 묵음 구간이 10 msec보다 짧을 때
녹음 상태가 나쁠 때

E: 묵음 구간이 0 msec보다 적을 때
인식 단어와 다른 단어일 때

발음을 이해할 수 없을 때

소음의 분류 기준

SND (sounds): 주변 소음
주변 교통 상황에서 발생하는 소리
물체와 부딪치는 소리(책상, 의자 등)
사이렌이나 경적 소리
발자국 소리, 문 노크 소리,
딸까닥 소리

SPC (speech): 화자 이외의 다른 요인에서 발생하는 소음
라디오, 호출 또는 휴대폰 소리/잡담/아기 우는 소리

UTT (utterance): 화자가 발생시키는 소음
기침소리
더듬거릴 때(hesitation)
숨소리(breath)
과도한 분절이 있을 때
목소리에 떨림이 있을 때
숨을 꼴깍이는 소리
음성을 우물거리며 발음할 때(swallowing)
발음에 이상이 있거나 두 개 이상의 표준 발음이 존재할 때

DST (distortion): 목소리가 반사됨(echo)-사무실 환경에서
클리핑(clipping)이 일어날 때

NOI (noise): 소음이 명확하게 들릴 때
녹음환경과 관련 없는 배경소음이 계속해서 발생할 때

4. 음성 수집 환경

음성 샘플은 화자 800 명에 대해 인식 단어 80 개를 주행중인 자동차 환경과 사무실 환경에서 녹음 수집하였다. 녹음 화자는 서울, 경기, 부산, 대구, 대전, 광주, 그리고 강원 지역을 대상으로 800 명을 선택하였으며 지역별, 연령별, 지역별, 성별 분포가 고르도록 하였다.

자동차 환경에서는 450 명에 대해 주행 속도, 엔진과 팬의 가동 유무, 주행 도로를 기준으로 하여 구분된 16 개의 환경 조건(C001~C016)을 갖는 주행 환경 테이블에서 정지 상태(C001~C008)에서 한 번, 저속 상태(C009~C012)에서 한 번, 그리고 고속 상태(C013~C016)에서 한 번씩 녹음하여 한 화자당 세 번씩 녹음하였다. 사무실 환경에서는 350 명에 대하여 한 번씩 녹음하였다.

자동차 환경에서는 기본적으로 네 개의 창문이 모두 닫혀 있고, 라디오 및 기타 소음이 생길 수 있는 장치는 꺼져 있는 상태에서 녹음을 한다. 사무실 환경에서는 사무실에서 발생

할 수 있는 잡담이나 걸어가는 소리, 전화벨, 문 여닫는 소리, 컴퓨터, 라디오, 에어컨 소리 등 기타 소음이 포함되지 않도록 하며, 사무실에서 음성이 울리지 않도록 방음이 잘 되어 있는 곳에서 녹음을 하여 소음이 실제 음성 신호를 왜곡시키지 않도록 주의한다.

4.1 음성 샘플의 분포

1) 지역 분포

지역	서울	경기	부산	대구	대전	광주	강원	합계
인원	250	100	100	100	100	100	50	800

2) 연령 분포

연령	20 대	30 대	40 대	50 대 이상
비율	20%	30%	30%	20%

3) 성별 분포

남	여
50%	50%

4.2 주행 환경 테이블

자동차 환경에서의 녹음 환경－속도, 엔진, 팬의 가동유무, 도로－을 16 가지 조건을 갖는 테이블로 만들어 음성을 수집한다.

	속도	엔진	팬	도로
001	0	Off	Off	도심
002	0	Off	Off	주차
003	0	Off	On	도심
004	0	Off	On	주차
005	0	On	Off	도심
006	0	On	Off	주차
007	0	On	On	도심
008	0	On	On	주차
009	40	On	Off	도심
010	40	On	Off	도심
011	60	On	On	도시근교
012	60	On	On	도시근교
013	80	On	Off	일반도로
014	80	On	Off	일반도로
015	100	On	On	고속도로
016	100	On	On	고속도로

C: 자동차 환경 녹음 세션의 조건을 정의하기 위한 참조 번호
속도: 도로 상황을 고려한 주행 속도
엔진: OFF, ON (=Running)
팬: 보통 세기에서 off, on
예를 들어 4 단계 스위치가 있다면 두 번째 것을 선택한다.
도로: 녹음하는 동안의 차의 위치
주차: 소음이 없는 조용한 곳
도심, 도시근교, 일반도로, 고속도로: 교통 소음이 있는 곳

4.3 화자 및 녹음 환경 기록

음성 샘플에 대한 정보를 기록하기 위해 녹음에 들어가기 전에 화자의 이름이나 성별, 나이, 지역, 녹음 차량, 녹음 환경 조건을 시트에 미리 적는다. 그리고 녹음 도중에 예외 상황이 발생할 때도 마찬가지로 시트에 기록한다. 화자에 대한 정보나 녹음 환경에서 발생하는 상황을 기록한 시트의 내용은 확인(Validation) 작업할 때 정보를 삽입하며, SPL파일에 저장된다.

자동차 환경에서의 예외 상황이란 자동차가 지하도를 지나거나 사이렌이나 경적 소리와 같은 외부 소음이 발생했을 때, 도로의 과속방지턱에 심하게 차량이 흔들려서 소음이 발생할 때, 도로가 콘크리트인 경우, 그리고 비가 오거나, 와이퍼를 작동시켰거나 깜박이등을 작동시켰을 때 등 주행시 발생하는 여러 가지 상황을 의미한다. 사무실 환경에서의 예외 상황은 빈 사무실에서 녹음했을 때 울림(Echo)이 발생했을 때, 자동차 환경에서 발생할 수 없는 이상 소음, 예를 들어 프린터 소리가 나거나 전화벨이 울리거나 발자국 소리, 문 여닫는 소리가 발생할 때 등을 의미한다.

<u>확인 작업시 기록하는 사항</u>
이름: 한글로 기입
화자번호: 고유번호를 갖는 일정 형식을 갖는다.
예) 서울-C-25-45 : 자동차환경에서 서울 25 번째 사람이 녹음한 것으로 테이프 45에 저장함.
성별: 남/여
나이: 20 대/30 대/40 대/50 대 이상
지역: 서울/ 경기/ 대구/ 부산/ 대전/ 광주/ 강원
자동차: 브랜드 및 모델
환경: 자동차 환경(1)/ 사무실 환경(2)
녹음 조건: C001~016 중의 한가지 조건
참고사항: 기상상태, 와이퍼 작동 유무, 도로 노면 상태로 인한 소음 등

4.4 파일 형식

본 음성 데이터베이스 구축 시스템에서는 녹음 음성은 NIST (National Instruments Standard Tecchnology) 형태의 웨이브 파일로 저장되며, 분할 작업이 끝나면 SGL (Segmentation Logging) 파일이 생성된다. 음성 데이터베이스가 구축되면 단어별 NIST 형태 웨이브 파일과

한 세션에 대해 한 개의 SPL (SPeech Logging) 파일이 생성된다. 분할 작업 및 확인 작업에 사용되는 스크립트 파일은 음성 데이터베이스를 구축하기 위한 시스템에 대한 전반적인 사항 및 녹음 정보가 기술되어 있는데, 이 파일은 음성 데이터베이스 구축하기 전에 미리 설계되어야 하고, ASCII 형태로 저장한다.

4.4.1 NIST 웨이브 파일

NIST웨이브 파일은 National Instruments사에서 고안한 파일 형태로 원래 UNIX상에서 운영하기 위해 설계되었다. NIST 형식 웨이브 파일의 특징은 1,024 바이트 크기의 ASCII 헤더를 가지고 있으며 객체 지향적인 특징을 가지고 있다.

4.4.2 SGL 파일

SGL (Segmentation Loggig) 파일은 분할 작업 이후에 생성되는 ASCII파일로, 한 세션에 한 개의 SGL파일이 만들어진다. SGL파일은 각 웨이브 파일과 연결되어 있고, 단어당 분할 정보를 가지고 있다. SGL파일은 3 개의 섹션 — [System] 섹션, [Segment] 섹션, [Session] 섹션 — 으로 구성되어 있다.

[System] 섹션: 시스템 구성에 대한 기술
코딩 형태, 주파수, 파일 형식, 저장 방식, 샘플 수, 웨이브 파일 이름
[Segment] 섹션: 분할 시작과 끝 위치 및 레이블링 기술
[Session] 섹션: 녹음 정보에 대한 사항 기술, 녹음 날짜 및 시간, 전체 샘플수, 실제 분할 샘플수, 기타

4.4.3 SPL파일

SPL (Speech Logging)파일은 음성 데이터베이스 구축의 결과로 한 세션에 한 개씩 생성되는 아스키 파일이다. SPL파일은 5 개의 섹션-[System] 섹션, [Record states] 섹션, [Validation states] 섹션, [Session] 섹션, [Info states] 섹션-으로 구성되어 있다.

[System] 섹션: 시스템 구성에 대한 기술
스크립트 번호, 주파수, 코딩 형태, 저장 방식
[Record states] 섹션: 음성 녹음 이후의 발음에 대한 기술
일련 번호, 레이블링, 헤더 크기(1,024 바이트), 헤더가 포함된 파일 크기, 인식 단어별 웨이브 파일명, 분할 정보가 들어있는 SGL파일명, R-Wav 파일 내에서의 시작과 끝 위치
[Record states]
1=2;;창문올려;1024;44196;U0501001.WAV;spk0501.sgl;1024;44196;….
30=2;;아니오;;1024;32604;U0501030.WAV;spk0501.sgl;1099602;1131182;

Header	Wav 1	Wav 2	Wav 3	…	Wav 30
1024	44196	86170	127344	1099602	1131182

R-Wav 파일 구조의 예 : r0010501.wav

[Validation states] 섹션: 확인 작업 이후의 발음에 대한 기술
일련 번호, 화자가 실제로 발음한 단어의 레이블링,
R-Wav 파일 내에서의 시작과 끝위치,
파일의 등급 (A~E) 표시
소음 분류(NOI, SND, SPC, UTT, DST) 표시 (유: 1, 무: 0)
인식 단어별 웨이브 파일의 이름,
묵음구간(약 200 msec) 분할 결과로 생긴 파일 내의 상대적 시작과 끝 위치
소음의 구체적인 기록

[Validation states]
1=창문올려;1024;44196;QUA:C;NOI:0;SND:0;SPC:0;UTT:0;DST:0;U0501001.WAV;1024;44196;micnoise;;;
….
30=아니오;1099602;1131182;QUA:A;NOI:0;SND:0;SPC:0;UTT:0;DST:0;U0501030.WAV;1024;32604;;;;
[Session] 섹션: 정보와 관련된 레코드 세션(record session) 기술
[Info states] 섹션: 화자에 대한 정보 기록
화자의 나이, 성별, 이름, 녹음 환경, 생활 지역 등을 기록한다.

4.4.4 스크립트 파일

스크립트(Script) 파일은 음성 데이터베이스 구축시 (DSDR 프로그램으로 SDB를 구축할 때) 필요한 정보를 미리 기록한 파일로 시스템, 화자, 녹음에 관련된 정보가 포함되어 있다. 스크립트 파일은 음성 데이터베이스 구축할 때와 타스크(Task)를 만들어 확인 작업을 할 때 사용된다. 스크립트 파일은 세 개의 섹션 — [System] 섹션, [Info] 섹션, [Record] 섹션 — 으로 구성되어 있으며, 그 특징은 다음과 같다.

4.4.4.1 스크립트 파일의 특징

아스키 파일
객체 지향적
계층적 (Hierarchical)
계속적인 추가 가능
데이터 오브젝트 타입
섹션 (Sections) - 라인 (Lines) - 필드 (Fields) - 리스트 (Lists)

4.4.4.2 스크립트 파일의 구성

스크립트 파일은 3 개의 섹션으로 구성되어 있는데, system, Info, record section으로 구성되어 있다.

[System] section: 시스템 구성(Configuration)에 대한 기술
Platform : Sounblaster 16 bit / NI DSP 2200
Sampling rate
Coding type : Linear / Mu-Law / A-Law PCM
Field delimiter : Default: ' ; '
Environment memo (선택)

[Info] section: 화자에 대한 정보 기술
이름
나이
성별 : 남 / 여
태어난 지역 및 생활 지역(사투리)
기타 정보 (예: 자동차 소유 유무 및 모델)
녹음 환경 : 자동차, 사무실
참고 사항

[Info] 섹션의 기술 형태:
MASK:
0;EDIT; <Info item name>; <Text>
<Info item name>: 영어로 아이템 기술
<Text>: 아이템을 화면에 표시할 때의 텍스트

[Record states] section : 인식 단어(Utterance)에 대한 기술

[record section] 섹션의 기술 형태:
MASK:
0; ; <Message content>;;;
0: 메시지 상태 지시자
<Message content>: 사용자에게 보여지는 텍스트

2; <Label>; <Text>; ; <Max record duration>
2: 발음 상태 구별자 (Utterance state identifier)
<Label>: 인식 단어 발음 레이블 (Utterance label)
<Text>: 화자에게 보여지는 단어
<Max. rec. duration>: 발음이 허용되는 시간

5. 결 론

자동차용 음성 인식기 개발을 위한 음성 데이터베이스는 서울, 경기, 부산, 대구, 대전, 광주, 강원지역 800 명의 인원을 대상으로 80 개의 단어에 대하여 자동차 환경에서 450 명, 사무실 환경에서 350 명 녹음하였다. 음성 데이터베이스는 PC에서의 작업을 쉽게 하기 위해 30 개 단어와 50 개 단어로 나누어 구축하였다.

자동차 음성 DB를 구축하기 위하여 자동화 프로그램인 Digitalization, Segmentation, Validation 시스템을 개발하였고, 또한 효율적으로 자동차용 DB를 저장하기 위하여 최적의 파일 트리 구조를 개발하였다.

그리고 자동차 주행조건과 주변환경을 세부 항목별로 설정하여 각각의 차량의 주변 환경을 상세히 포함할 수 있도록 하여 DB를 구축하였다.

참 고 문 헌

[1] 신지영. 2001. *말소리의 이해.* 한국문화사.
[2] 현대오토넷. 2002. *무제한 한국어 음성합성 시스템 개발에 관한 연구.*
[3] Sanders, Eric. 1995. *Using Probabilistic Methods to Predict Phrase Boundaries for a Text-to-Speech System.* MS Thesis of University of Nijmegen.
[4] Charniak, E. 1994. *Statistical Language Learning.* MIT Press.
[5] Rabiner, L. R. & Juang, B. H. 1986. "An Introduction to Hidden Markov Models." *IEEE Acoustic Speech and Signal Processing Magazine.*
[6] 서봉수. 2001. *가변어휘 음성인식기 구현 및 탐색 시간 단축 알고리듬 비교.* 석사학위논문, 전남대학교 전자공학과.
[7] Young, Steve, Julian Odell, Dave Ollason, Valtcho Valtchev & Phil Woodland. 1997. *The HTK Book (for HTK Version 2.1).* Entropic Cambridge Research Laboratory Ltd, Cambridge, UK.
[8] Knill, K. M., M. J. F. Gales & S. J. Young. 1996. "Use of Gaussian Selection in Large Vocabulary Continuous Speech Recognition Using HMMs." *IEEE ICSLP '96.*
[9] 김동화. 1999. *연속 음성인식을 위한 향상된 결정트리 기반 상태공유 기법 연구.* 박사학위논문, 부산대학교 전자계산학과.

접수일자: 2002. 7. 17.
게재결정: 2002. 8. 30.

▲ 권오일
서울시 강남구 역삼동 823-21 대공빌딩7층 (우: 135-080)
(주) 현대오토넷 차장
Tel: +82-2-3016-9457 Fax: +82-2-3016-9460
E-mail: koi@haco.co.kr