

# 한국어 표준 전화 음성 데이터 베이스 구축

Construction of the Standard Telephone Speech Database in Korean

김 중 교

전북대학교 공과대학 전기,전자,제어공학부

전화:0652-270-2402 FAX:0652-270-240

E-mail:ckkim@moak.chonbuk.ac.kr

## 요약

음성처리 연구자들이 연구 수행을 위해서 최초로 해야 할 일은 음성학적 범주가 풍부한, 많은 사람이 발성한 음성 데이터의 확보이다. 그러나 음성 데이터를 확보하는 데는 많은 시간과 노력이 소요되므로 음성 DB의 구축이 필요하다. 따라서 본 연구에서는 한국어 전화 음성 인식 기술 연구시 시스템의 객관적 성능평가와 알고리즘 개발등에 모든 연구 개발자들이 공통으로 사용할 수 있는 표준 음성 데이터 베이스를 구축하여 국내 음성처리 연구의 활성화를 꾀할수 있도록 하였다.

## 1. 서론

우리말의 공학적인 응용을 위해서는 그 기반이 되는 요소 기술로써 음성인식 및 합성으로 대표되는 음성처리기술과 언어이해 및 기계 번역으로 대표되는 언어처리기술의 연구가 필요하다. 이러한 음성 및 언어 처리 기술의 연구를 위해 가장 먼저 확보되어야 할 것이 음성, 언어 및 각종 사전DB 등 국어 정보 베이스이다. 이들의 체계적인 조기 확보 여하에 따라 음성 및 언어 처리 연구의 성패를 좌우한다고 해도 과언은 아니다. 특히 우리말을 대상으로 한 음성 DB(음성 코퍼스 또는 음성사전)는 음성언어 연구의 기본으로서 개발 초기부터 확보되어야 할 연구 자원이다. 따라서 한국어 전화 음성 인식 기술 연구시 시스템의 객관적 성능평가와 알고리즘 개발등에 모든 연구 개발자들이 공통으로 사용할 수 있는 표준 음성 데이터 베이스를 구축하고 전화 음성 인식 시스템의 실용화를 위해 성능 영향을 미치는 각종 요인을 분석 및 개선 방안의 제안을 동시에 수행한다.

## 2. 연구의 필요성

음성을 맨머신(Man-Machine) 인터페이스의 수단으로 활용하기 위한 음성정보처리 연구는 관련 기술의 진보에 따라 한정 어휘의 인식 및 합성 시스템들을 실용화에 접어들고 있고, 최근들어 무한정 어휘를 대상으로 하는 시스템들이 연구개발되고 있다. 특히 무한정 어휘 인식시스템 개발에는 음소 인식 기술이 필수적이지만, 연속음성중의 음소는 발성자에 따른 개인차는 물론이고 전후에 발성되는 음소의 영향에 의한 조음 결합에 따라 그 특성이 크게 변한다. 이러한 개인차 및 조음 결합의 현상을 분석하기 위해서는 많은 사람이 발성한 다양한 음성 데이터가 필요하다. 또한 시스템의 객관적인 평가를 위해서 표준적인 음성데이터가 필요하다. 이 음성데이터는 다중다양(성별, 연령, 방언, 발성자수 등)한 것이 필요하다. 지금까지는 각 연구자가 필요에 따라 음성데이터를 만들어 보관하고 이용해 왔다. 음성 연구가 진보되어 감에 따라 처리 가능한 데이터 수는 많아져 가고, 따라서 준비해야 할 데이터 량도 대폭적으로 증가되었다. 최근에는 음성인식의 경우, HMM이나 bigram/trigram등 언어 모델로 대표되는 통계적 수법의 발달에 따라 대량의 음성데이터가 시스템의 학습에 필요하게 되었다. 한편 음성정보처리시스템의 연구 개발을 위해서는 분석, 합성 인식의 각종 알고리즘을 적절하게 비교 평가할 필요가 있지만 이를 위한 방법으로는 현재까지는 공통 음성 데이터를 이용하여 알고리즘을 수행하고 그 결과를 비교하는 방법 이외에는 알려져 있지 않다. 따라서 공동으로 이용 가능한 각종 대량의 음성 데이터를 수록, 보관, 공개하는 것은 연구 개발 과정에서의 이용 및 인식장치의 성능평가 면에서 필요하다.

합성의 경우에도 지금까지 다이폰, 반음절 등 각종 단위에 의한 접속방식이 주류를 이루고 있고 최근에는 대량의 음성 DB로부터 임의의 길이의 음성 부분을 골라내어 접속함으로써 좋은 합성품질을 얻고 있다. 이를 위해서는 잘 정비된 대형의 음성 DB가 필요하다. 또한 인식 및 합성 알고리즘의 개발을 위해서는 다양한 환경의 음성언어학적 분석이 필연적으로 요구되는데 이를 위해서는 음성DB의 중요성은 크다.

### 3. 연구 개발의 내용 및 범위

한국어 전화 음성 인식 기술 연구시 시스템의 객관적 성능 평가와 알고리즘 개발 등에 모든 연구 개발자들이 공통으로 사용할 수 있는 표준 음성 데이터 베이스를 구축하였다.

#### 3.1 전화 음성 자동 수집을 위한 플랫폼 구성 및 수집 절차 연구

전화망 음성을 자동으로 수집하기 위하여 DIALOGIC사에 의해 개발된 D/4x voice board를 사용하였다. 이 보드는 다채널을 동시에 이용할 수 있어 대규모 전화음성 수집에 적당하다. 음성 데이터는 ADPCM과 PCM( $\mu$ -law, A-law) 방식에 의해 프로그램 상에서 선택적으로 변환하여 저장할 수 있다. 본 연구에서는 8bit  $\mu$ -law PCM 방식을 이용하였으며 자동 수집된 음성 데이터는 하드 디스크에 저장되고, 저장된 데이터는 본 연구실에서 개발한 S/W tool을 이용하여 segmentation 및 labeling 작업을 하여 최종적으로는 ASCII 데이터 포맷으로 CD-ROM에 저장하였다. 그림 3.1은 전화음성 자동수집 장치의 구성도를 나타낸 것이다.

#### 3.2 대규모 전화 음성 DB 구축

전화 음성 DB 구축을 위한 수집대상 어휘로는 한국 통신에서 제시한 Phoneme Balanced Word(PBW) 400개, Phoneme Balanced

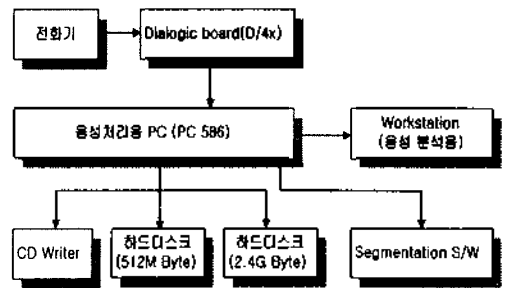


그림 3.1 전화 음성 자동 수집 장치 구성도

Sentence(PBS) 76개를 선정하였으며, 이 어휘의 양은 다른 필수 어휘를 추가하지 않더라도 한번에 발성하기에는 많은 양이다. 따라서, 이들 대상 어휘를 요소로 갖는 pool을 만들어, 이 pool에서 1회 발성분의 어휘를 무작위로 선택하여 발성 용지를 작성하는 과정이 필요하다. pool의 각 요소는 선택되는 빈도가 균형을 이루어야 하며, 일반적으로 1회 발성 어휘 수는 약 50개 정도로 하는 것이 타당하다.

본 연구에서의 발성화자 선정기준은 화자의 성별(남,여), 발성자의 주거 지역(서울·경기, 충북·충남, 전북·전남, 경북·경남), 화자의 연령 등의 균형을 고려하여 총 208명의 음성 데이터를 수집하였다.

표 3.1, 표 3.2 그리고 표 3.3은 전화음성 수집에 참여한 발성화자들에 대한 지역별, 연령별 학력별 통계이다.

표 3.1 전화 음성 화자 통계(지역별)

	남 자	여 자	계
서 울	15명	13명	13.5%
경 기	5명	5명	4.8%
충 북	11명	12명	11.1%
충 남	14명	12명	12.5%
전 북	24명	15명	18.8%
전 남	10명	10명	9.6%
경 북	23명	1명	11.5%
경 남	21명	17명	18.2%
계	123명 (59.1%)	85명 (40.9%)	100%

표 3.2 전화 음성 화자 통계(연령별)

	남 자	여 자	계
10세-19세	7명	13명	9.6%
20세-24세	65명	46명	53.4%
25세-29세	34명	17명	24.5%
30세-39세	13명	3명	7.7%
40세 이상	4명	6명	4.8%
계	123명 (59.1%)	85명 (40.9%)	100%

표 3.3 전화 음성 화자 통계(학력별)

	남 자	여 자	계
고졸이하	2명	6명	3.8%
고졸	10명	28명	18.3%
전문대졸	10명	7명	8.2%
대졸	60명	28명	42.3%
대학원	41명	16명	27.4%
계	123명 (59.1%)	85명 (40.9%)	100%

#### 4. 연구결과

Dialogic board를 이용한 전화음성 자동 수집플랫폼 구축하여 음성 데이터를 수집하였으며, 수집된 데이터를 segment하기 위해 segmentation S/W IPSS를 개발하였다.

또한 지역별, 연령별 균형을 고려한 전화음성 데이터는 다음과 같다.

- 서울·경기, 충남·충북, 전남·전북의 4개 지역(총 208명)
- 208명 \* 80단어(PBW) 수집
- 208명 \* 30문장(PBS) 수집

위와 같이 수집된 전화음성 데이터는 8bit  $\mu$ -law PCM방식에서 ASCII 데이터 포맷으로 변환하였으며 이를 시제품화하기 위하여 CD-ROM으로 제작하였다.

#### 5. 결론

본 연구에서는 회선 특징의 변화, 발성자의 개인차, 지역차 등을 고루 흡수할 수 있는 대규모의 전화 음성을 온라인으로 자동 수집하는 방안을 검토하여 자동음성수집시스템을 구축하였다. 그리고 자동음성수집시스템을 이용하여 지역별, 연령별로 변동요인을 흡수할 수 있는 전화 음성 DB를 구축하였다. 수집대상의 어휘로는 한국 통신에서 제시한 Phoneme Balanced Words(PBW) 400개, Phoneme Balanced Sentences(PBS) 76개를 선정하였으며, 선정된 발성화자 수는 208명으로 지역적 균형을 최대한 이루도록 하였다.(서울·경기, 충남·충북, 전남·전북, 경남·경북의 4개 지역에서 각각 50명 정도)

2.소속

- ① 전북대학교 전자공학과
- ② 전화  
(0652) 72-1177
- ③ Fax  
(0652) 270-2400
- ④ E-mail  
ckkim@moak.chonbuk.ac.kr

3. 현재 수행중인 연구

- ① 연구과제명  
음성인식을 위한 실시간 끝점 검출기의 구현
- ② 연구비 지원기관  
서울대 반도체 공동연구소
- ③ 연구개발비  
1,200만원
- ④ 연구기간  
97.8~98.7

4. 주요 연구실적(최근 5년간)

(1) 첫번째

- ① 연구제목  
RF-Modulator 및 Tuner의 전기적 성능 개선
- ② 연구비 지원기관  
대우전자부품(주)
- ③ 연구 개발비  
1,500만원
- ④ 연구기간  
92.12~93.12

(2)두번째

- ① 연구제목  
RF-Modulator의 전기적 성능 개선

- ② 연구비 지원기관  
대우전자부품(주)

- ③ 연구 개발비  
1,600만원

- ④ 연구기간  
94.3~95.2

(3) 세번째

- ① 연구제목  
한국어 표준음성 데이터베이스 구축

- ② 연구비 지원기관  
한국통신

- ③ 연구 개발비  
3,835만원

- ④ 연구기간  
94.4~94.12

(4) 네번째

- ① 연구제목  
차세대 지능망 서비스를 위한 음성인식에 관한 연구

- ② 연구비 지원기관  
한국전자통신연구소

- ③ 연구 개발비  
2,000만원

- ④ 연구기간  
94.3~94.12

(5) 다섯번째

- ① 연구제목  
한국어 표준 전화음성 데이터베이스 구축

- ② 연구비 지원기관  
한국통신

- ③ 연구 개발비  
4,825만원

- ④ 연구기간  
95.5~95.12

(6) 여성번제

① 연구제목

음성을 이용한 차세대 운전자지원시스템 개발

② 연구비 지원기관

전북대학교 자동차신기술 기술연구소

③ 연구 개발비

2,000만원

④ 연구기간

96.12~96.11

(7) 일곱 번제

① 연구제목

한국어 표준 전화음성 데이터베이스 구축

② 연구비 지원기관

한국통신

③ 연구 개발비

4,366만원

④ 연구기간

96.5~96.12

(8) 여덟번제

① 연구제목

단말장치 기술수준 수치의 타당성 검증 연구

② 연구비 지원기관

한국전자통신연구원

③ 연구 개발비

1,500만원

④ 연구기간

97.3~97.11

(9) 아홉번제

① 연구제목

음성인식을 위한 실시간 끝점 검출기의 구현

② 연구비 지원기관

서울대 반도체 공동연구소

③ 연구 개발비

1,200만원

④ 연구기간

97.8~98.7

5. 보유 장비

(1) 기자재명

Portable DSP Board

HP Laser Printer

Image Scanner

Axil 311, Axil 220

DAT LINK

Mixer

Xwaves\*

Portable DAT

DAT Reader

TMS 320 DSP Board

Pentium 9대

Pentium Pro 1대

6. 향후 연구계획 및 기술 교류

제목: 자동차 잡음하에서의 인식기술 확보를 위한 기술 연구

: 전체 기술의 초점은 자동차에서의 음성 인식기의 성능 향상이다.