

삼성종합기술원의 음성 DB 구축 현황

김 상 동

삼성종합기술원, 기반기술연구소, 음성연구실

Trend on the Speech Database of SAMSUNG Advanced Institute of Technology (SAIT)

S. R. Kim

Speech Technology Lab., Core Technology Research Center, SAIT

요 약

본 논문에서는 삼성종합기술원의 음성 인식, 합성 분야의 음성 데이터 베이스 구축 현황과 향후 연구 방향에 대하여 기술한다. 삼성종합기술원에서는 1989년 한국어 문음 변환기술 개발을 시작하여 그동안 남성음, 여성음 합성 시스템을 발표하였고, 최근에는 시각장애자용 컴퓨터를 개발하여 전국 13개 시각장애자 학교에 기증한 바 있다. 음성인식 분야는 100단어 내외의 소용량, 화자속속 시스템을 개발하여 키폰용 음성인식 다이얼 장치를 실용화하였다.

약 5년여에 걸친 연구 결과 자체적으로 구축하게된 음성 DB는 크게 남,여 합성용 DB와 인식용 DB로 요약할 수 있다. 이러한 경험을 바탕으로 향후 국내외 대학, 연구소 등과 공동연구를 통해 상품화 수준의 문음 변환기술과 대응량, 화자독립 음성인식 시스템을 개발하고자 한다. 궁극적으로는 휴대용 통역기의 요소 기술을 확보하여 제한된 영역에서 자동 통역기를 상품화하는데 이바지할 계획이다.

1. 서 론

삼성종합기술원에서는 1989년부터 한국어 문음 변환기술을 연구하여, 그 결과를 발표한 바 있다[1, 2, 3]. 한편, 1994년에는 문음 변환기술의 실용화 전 단계로서 시각장애자용 컴퓨터를 개발하여 전국 13개 장애자 학교에 무상으로 기증하였으며, 한국 전자전람회 (KE SHOW) 등 다수의 국내 전시회에 출품하여, 상품화 가능성을 확인하였다. 현재까지 문음 변환기술과 관련하여 남성음 DB와 여성음 DB를 각각

구축하였다. 한편, 음성인식과 관련하여서는 소용량 인식 시스템을 대상으로 DB를 구축하였으며, 소음환경 하에서의 인식실험을 위한 DB도 구축하였다.

본 논문에서는 개발된 음성 DB를 중심으로 DB의 규격과 생성 방법, 그리고 향후 연구 방향에 대하여 기술한다.

2. 음성합성 DB

합성 DB를 구축하기 위해서 먼저 녹음 방법을 설명한다. 무제한 음성합성을 위해서는 미리 정해진 합성단위를 기준으로 한글의 모든 발음 가능한 단위를 녹음해야 한다. 그러므로 사전에 모든 합성단위를 포함하는 발성 리스트를 만든다. 이렇게 작성된 발성 리스트를 화자가 발음하고 이것을 녹음한다. 녹음시에는 Reel Tape Recorder (모델명 : TEAC-X300) 와 마이크로폰 (모델명 : AKG D330 BT)를 사용하였다. 녹음환경은 조용한 사무실 환경이며, 특별히 주변 잡음이 발생하지 않도록 하였다. 발음한 음성 신호는 SPARC10 워크스테이션에 내장된 A/D변환기를 이용하여 8kHz, 16 bit 선형 PCM 형태로 양자화하였다. 이러한 방법으로 획득한 음성 데이터는 본 연구실에서 자체 개발하여 X 윈도우 환경에서 편리하게 사용할 수 있는 음성 분석 Tool을 이용하여 합성용 DB로 구축되었다.

• 여성음

- * 화자 : 성우, 연령 30대
- * Encoding : 10차 LSP(Line Spectrum Pair)
 - ⇒ 프레임 길이 : 10msec
 - ⇒ 압축률 : 8bytes/frame

- * 종류 : diphone, triphone 등 15종류
- * 갯수 : 7975개
- * Size : 약 2Mbytes

• 남성음

- * 화자 : 아나운서, 연령 50대
- * Encoding : 10차 LSP(Line Spectrum Pair)
 - ⇒ 프레임 길이 : 10msec
 - ⇒ 압축률 : 8bytes/frame
- * 종류 : diphone, triphone 등 7종류
- * 갯수 : 2715개
- * Size : 약 600Kbytes

3. 음성인식 DB

음성인식 연구는 현재까지 소용량, 고립어 인식 시스템의 상용화에 중점을 두어 왔으나, 앞으로는 소음 환경을 포함한 대용량, 연속어 인식 시스템의 연구를 진행하려한다 이와 관련하여 연속어 숫자음 DB와 자동차 소음환경에서의 인식 DB를 구축하였다. 녹음 방법 및 A/D 과정은 합성 DB의 구축과 같다

• 명령어 DB

- * 정해진 명령어 30개 단어
- * 남자 10명, 3회 반복
- * 여자 10명, 3회 반복
- * 환경 : 조용한 사무실

• 숫자음 DB

- * 성별 : 남자, 여자
- * 지역 : 10개 도시 (서울, 부산, 대구, 목포, 수원, 전주, 대전, 강릉, 청주, 제주)
- * 연령 : 10대, 20대, 30대, 40대, 50대, 60대, 각각 10명
- * 단어 : 무작위 1~7자리 숫자 연속음 100단어
- * 환경 : 조용한 사무실

• 소음 환경 DB

- * 자동차 운전관련 1000단어
- * 남자 2명, 5회 반복
- * 여자 2명, 5회 반복
- * 환경 : 자동차 소음 환경

4. 향후 연구 방향

• 음성 합성

음성 DB와 관련하여 음성 합성기술을 분류한다면, LPC 계열의 파라메타를 사용하는 파라메타 부호화법과 파형 자체를 사용하는 파형 부호화법으로 나눌 수 있다. 파라메타 부호화법이 음성 합성기술에 손쉽게 적용될 수 있지만, 파형 부호화법에 비하여 명료성이 떨어지는 단점이 있다. 한편, 파형 부호화법은 자연스럽게 기본주파수를 변경하기가 어려운 단점이 있다 결국, 고음질 음성합성을 실현하기 위해서는 두가지 방법의 단점을 보완하고, 장점을 살려 나가야 한다고 판단된다. 그러므로 본 연구실에서는 파형 부호화 방식을 사용하면서 기본주파수를 자연스럽게 변경할 수 있는 고음질 음성합성 방식을 연구할 계획이다.

• 음성인식

무제한 음성인식은 음성합성에 비하여 상용화하기에 어려운 분야라고 생각된다. 그러나 무제한 음성 인식이라는 목표를 실현하기 위해서 연구하는 과정에 생기는 중간 결과물만으로도 상용화가 가능한 부분이 많으리라고 생각된다. 1994년 개발한 키폰용 음성인식 다이얼 장치가 그 좋은 사례이다. 앞으로 연속 숫자음 인식등을 활용하여 휴대 전화기용 음성인식 장치 등을 연구할 계획이다.

참고문헌

[1] 하남규, 공병구 외, "Development on korean text to speech system." Korea-Japan joint symposium on acoustics, 1991

[2] 김상룡, 김정수, "형태소 해석을 이용한 합성 음성의 음운 및 운율 처리", 전자 공학회지 20권 5호, pp.508-515, 1993

[3] 공병구, 김상룡, 김정수, "이질음 접속에 의한 음질 저하 및 극복 대책 연구", 음성통신 및 신호처리 워크샵 논문집, pp.279-284, 1993