

## 청각장애인을 위한 자동 자막 시스템

박현근<sup>○</sup>, 이희석\*, 이상문\*

<sup>○</sup>충주대학교 컴퓨터정보공학과

e-mail: 0329484@cju.ac.kr, magozer@naver.com, smlee@cju.ac.kr

# A Study on The Automatic Caption System for Hearing Impaired Person

Hyon Gun Park<sup>○</sup>, Hee Suk Lee\*, Sang Moon Lee\*

<sup>○</sup>Dept. of Computer Information Engineering, Chungju Nat'l University

### ● 요약 ●

장애인도 비장애인과 정보접근 기회를 동일하게 가질 권리가 있다. 하지만 청각장애인은 의사소통 수단이나 정보획득의 수단으로부터 소외됨으로써 다양한 사회적 참여에서 배제되기 일쑤였다. 따라서 이 논문에서는 청각장애인을 위한 음성인식을 이용한 자동 자막 시스템을 제안한다. 이 논문에서 제안하는 시스템은 음성 특징 파라미터 추출과 DTW에 의한 음성인식을 통해 음성을 추출하여 화면에 자막으로 나타나게 하는 시스템이다.

키워드: 청각장애인(hearing impaired person), 음성인식(voice recognition), 자막(caption)

## I. 서론

21세기 정보화 사회에서는 인터넷과 TV를 통해 정보에 접근하고 있다. 현재는 인터넷을 통해 많은 정보를 얻고 있지만, TV를 통한 정보접근도 인터넷에서 얻는 정보 못지않게 중요한 역할을 하고 있다. 하지만 청각장애인의 방송매체 접근 및 이용 문제는 소홀히 다루어지고 있다. 여러 가지 이유로 후천적 장애인의 수가 점점 증가하고 있는 시점에서 청각장애인의 정보 접근권의 소외와 배제는 청각장애인들의 총체적인 삶의 질과 만족도에 부정적인 영향을 미치는 중요한 요인으로 작용할 수 있다. 정보화 사회에서 정보접촉의 권리를 보장 받지 못한다면 장애인과 비장애인과의 정보격차는 더욱 심화될 것이다. 따라서 본 논문에서는 청각장애인을 위한 음성인식 기술을 이용하여 자동으로 자막서비스가 되는 시스템을 제안하고자 한다.

## II. 관련 연구

최근의 연구경향은 2가지로 분류할 수 있다. 첫째는, LPC(Linear Predictive Coding)계수 및 DTW(Dynamic Time Warping)에 의한 음성인식 방법으로, 음성의 특징 파라미터를 추출하기 위해서 LPC계

수를 사용하고 있다. LPC는 사람의 발성기관을 하나의 필터로 가정하고 그 필터의 계수를 음성의 특징 벡터로 사용하는 방법이다[1, 2].

동일인이 동일한 단어를 발성해도 단어의 시간적 길이가 변화한다. 이를 표준 패턴과 단순히 비교하면 시간축이 고르지 않기 때문에 오류나 인식 불능이 생기게 된다. 이 문제를 비교적 고속이면서 효과적으로 처리할 수 있는 방법이 동적 프로그래밍에 기반한 시간축의 비선형 신축에 의한 정합법(DTW)이다[1, 3].

또 다른 것으로는 자막파일 생성기법으로, 이 방법에서는 음성인식을 통하여 추출된 자막데이터는 규격화된 자막파일로 변환된다. 자막파일의 규격인 SAMI 파일은 HTML 기반의 자막파일이다. 추출된 자막을 완전한 단어나 문장의 형태로 배열하기 위해서는 상황에 따라서 단위 혹은 일정 길이로 연결하는 작업이 필요하다. TV화면에 나타나는 자막의 행이나 열의 수를 결정하기 위한 기준이 된다[4].

## III. 시스템의 설계 및 평가

이 시스템은 동영상 방송프로그램에서 사용자에게 요청에 의해 동영상에서 출력되는 음성을 입력받아 음성을 문자로 변환 후 저장하는 방법으로 그림1과 같은 처리과정을 거친다.



그림 1. 시스템 처리 순서  
Fig. 1. System Processes Sequence

사용자에게 자막으로 전달하기 위하여 앞서 사례분석한 기법인 LPC, DTW 등의 기법을 통하여 저장된 음성을 인식한다. 그런 다음 이 정보를 문자로 변환처리하여 자막으로 출력하게 된다. 그리고 사용자에게 정확한 자막을 제공하기 위하여 1회 반복하여 검증하게 된다.

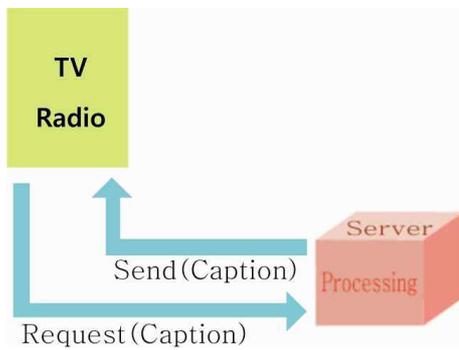


그림 2. 시스템 구조  
Fig. 2. System Architecture

현재 이 시스템은 실제 생방송에 적용하기에는 방송사와의 협조 등 절차적인 문제로 인해 완전한 시제품으로 적용하기에는 문제가 있다. 이런 문제를 해결하기 위한 방법으로 우선적으로 인터넷 개인방송을 이용하는 체제를 설계하였다.

하지만 이 방법은 실제 TV와의 H/W 사양을 맞출 수가 없기에 결과를 비교할 수 없었다. 또한 현재의 상태로는 인식속도 또한 2~3초 정도로 다소 느린 편이며, 이 연구에서 설계하여 제시한 시스템이 보다 높은 호환성을 획득하기 위해서 보다 적합한 최적화

과정이 요구된다고 하겠다. 그리고 이 시스템에 입력되는 자료는 연구 환경상 다양한 환경에서 자료를 이용한 테스트라 할 수 없으므로, 향후 다양한 환경에서 시스템을 테스트하여 문제를 도출하고 해결하는 업그레이드 과정이 필요하다.

#### IV. 결론

정보가 곧 가치이자 기회인 정보화 사회에서 시각장애인들은 사회적 차원의 특별한 조치와 추가적인 서비스가 없는 자유롭게 정보에 접근하여 활용하는 것이 거의 불가능하다. 그럼에도 불구하고 현재 우리나라의 방송환경에서는 청각장애인의 접근은 미흡하기 짝이 없는 실정이다.

따라서 이 논문에서 제안하는 시스템설계의 결과는 청각장애인들이 비장애인과 동일한 환경에서 TV를 통한 정보습득의 기회를 제공받게 될 것을 목표로 하고 있다. 드라마, 시사, 교양, 뉴스 등 다양한 TV프로그램의 동영상상을 이 논문에서 제안하는 시스템으로 자막을 추출하여 자막을 제공함으로써 비록 실시간 방송이 아닌 동영상방송이지만 청각장애인도 다양한 TV 정보매체를 접할 수 있는 기회가 생길 것이다. 또한, 저작권의 문제로 인해 방송 동영상의 확보가 어려운 점이 있다. 따라서 방송국에서는 청각장애인을 위한 청각장애인 동영상상의 배포가 필요하다.

향후 연구과제로는 청각장애인들을 위한 동영상이 무단으로 사용되는 것을 방지하기 위하여 자막 삽입 후 음성데이터가 삭제되는 추가 기능의 연구와 음성인식률을 높여 실시간으로 방송에 자막을 삽입할 수 있는 연구가 진행되어야 할 것이다.

#### 참고문헌

- [1] 이상문, “음성 인식 기술을 이용한 휴먼 인터페이스시스템”, 한국멀티미디어학회 춘계학술발표대회 논문집, pp.395-398, 2009. 5.
- [2] 장성열, “선형예측계수와 RBF 신경망을 이용한 화자인식”, 퍼지및지능시스템학회논문지, Vol.10, No.1, pp.16-22, 2000. 2.
- [3] 정영규, “성대신호 기반의 명령어인식기를 위한 특징벡터 연구”, 한국정보과학회 논문지B, Vol.34, No.3, pp.226-234, 2007. 3.
- [4] 김정연, “디지털방송 자막데이터를 이용한 멀티미디어 응용서비스 연구”, 방송공학회논문지, Vol.14, No.4, 2009.