음성인식 기반의 자동 프롬프터 시스템

김길연*, 김진우**
베스티안 파트너스*, KBS 기술연구소**

Auto-Scrolling Prompter System using Speech Recognition Technology

Kilyoun Kim*, Jinwoo Kim**
Bestian Partners*, KBS(Korean Broadcasting System)**
kykim@b-p.co.kr*, starseeker@kbs.co.kr**

Abstract

A prompter software is used, behind the camera, to scroll the script for a TV narrator. So far it has been manually operated by an assistant, who scrolls the caption following narrator's speech. Automating this procedure using a speech recognition technology has been investigated in this project. The developed auto-scrolling software was tested in offline and online, which shows performance good enough to replace an existing prompter software. This paper describes the whole development process and concerns to be cared.

I. 서론

프롬프터는 방송 출연자를 보조하는 부가적인 방송 도구 중 하나로 녹화할 대본을 카메라 뒤쪽에 표시해준다. 출연자는 이 프롬프터를 이용하여 카메라를 계속해서 주시하며 긴 문장의 메시지를 시청자들에게 용이하게 전달할 수 있다.

현재의 프롬프터는 출연자의 발성에 맞추어 수동으로 스크롤을 제어하고 있으며, 이는 많은 양의 메시지를 전달함에 있어서 적지 않은 노동력을 필요로 한다. 이러한 수동 제어 방식을 자동화하기 위한 기술 개발은 앞으로 많은 방송 제작에 효과적으로 활용될 충분한 기술적 가치가 있으며, 차후 방송제작 과정의 전반적인 자동화에 기반 기술로써 응용될 수 있다.

방송에 활용되는 프롬프터를 자동으로 제어하기 위해서는 출연자의 음성을 인식하여 대본에서의 진행정도를 자동으로 파악하는 음성인식 기술과, 프롬프터의 스크롤 및 기타 기능을 제어하고 표출되는 대본 문서의 크기 및 형식 변환을 가능하게 하는 소프트웨어 개발 기술이 필요하다.

본 논문에서는 이러한 다양한 기술들을 바탕으로 출연자의 음성에 따라 자동으로 스크롤이 가능한 음성인식 기반 자동스크롤 프롬프터 소프트웨어의 개발 결과를 정리한다.

2장에서는 방송용 프롬프터에 음성인식을 적용한 방법론과 인식률을 살펴보고, 3장에서 후처리 모듈을 통한 성능향상과정을 제시한다. 4장에서는 개발된 프롬프터 S/W를 소개하고 5장에서 결론을 맺는다.

II. 음성인식 시스템

2.1 연속어 음성인식 시스템

방송 출연자의 발성은 "안방불켜" 와 같은 단어의 발성이 아니라 "화재가 발생하여 많은 사람이 다쳤습니다." 와 같이 연속된 문장의 발성이다. 따라서 발성이 이루어지는 도중에 중간 인식결과를 표시해 주는 연속어 음성인식을 수행해야 한다. [그림1] 에서와 같이 각 어절 사이의 휴지기에 현재까지의 인식결과를 제공해야 한다.

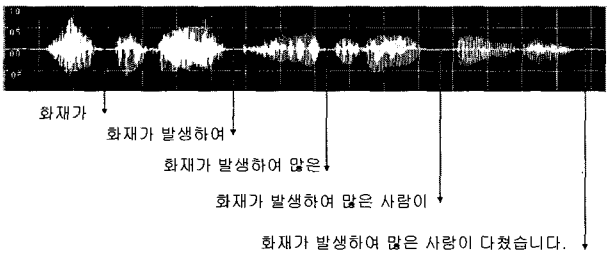

그림 1. 연속음성인식 시스템

음성인식 시스템의 세부 Configuration은 다음과 같다.

항목	설정
Feature Set	MFCC_E, Delta, Acc. (39차)
Sampling Rate	16KHz, 16bit
Frame Size	40ms frame, 20ms shift
Grammar	FSN Grammar, Cross-word

2.2 방송 프롬프터를 위한 음성인식 문법

방송 멘트를 음성인식 하기 위해서는 대본의 어절을 바탕으로 음성인식 문법을 구성해야 한다. [그림2]와 같이 단어의 반복 문법으로 구성하면 전체 대본을 인식할 수 있다. 한 대본의 평균 단어의 수가 500개 미만이므로 별도의 언어모델은 사용하지 않았으나 문장안의 전이를 모델링하기 어절 사이의 아크를 추가했다. 즉, 대본의 어떤 단어부터 발성이 시작될 수 있고, 한 문장은 어절의 흐름으로 이루어지며, 노드간의 별도의 확률 값은 더해지지 않는다. 덧붙여, Cross-word 모델링을 통해 어절 간의 트라이폰 발음을 정확히 모델링한다.

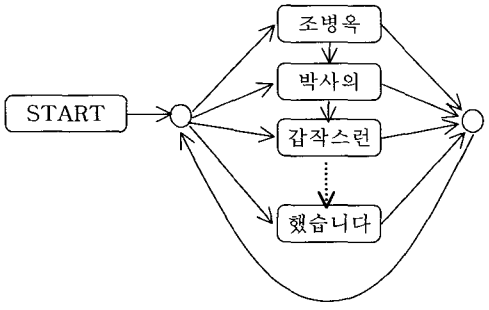

그림 2. 음성인식 문법

2.3 음성인식률의 평가 - Offline

(1) 음성인식률 평가 DB

연속음성인식의 성능을 객관적으로 측정하기 위해 방송용 실제 DB를 바탕으로 평가를 수행하였다. 평가용 음성 DB는 2003년 역사스페셜 녹음 대본과 음성파일을 사용하였다. 프롬프터는 아나운서의 녹화에 사용되는 것이므로, 외부 VCR이나 리포터의 음성은 제외하고 아나운서의 발성 부분만을 그 테스트 대상으로 하였다. 다음 표1은 각 대본의 위치와 첫 문장을 정리한 것이다. 총 3주 방송 대본을 대상으로 하였으며 2003년 4월 19일자 DB1에 대해서는 유인촌 진행자의 발성 부분을 모두 표시하였다.

표1. 연속음성인식엔진 평가용 역사스페셜 DB

DB1 [2003.04.19] 1960년 봄 시민들의 힘으로..		
DB2 [2003.04.26] 저 광활한 만주 벌판을 내준 후부터...		
DB3 [2003.06.07] 이곳은 지난 74년,		
DB1 - <2003. 04. 19.>		
ST1	00:01:20	1960년 봄-시민들의 힘으로...
ST2	00:11:42	지금 제가 서 있는 곳은 당시...
ST3	00:20:42	조병옥 박사의 갑작스런 죽음은...
ST4	00:32:35	3.15부정선거로 유혈사태까지...
ST5	00:42:35	이곳은 당시 효자동의 전차...
ST6	00:54:15	이렇게 해서 1948년8월15일...

(2) 음성인식률 평가 결과

위의 대본을 바탕으로 음성인식률을 계산한 것을 다음의 표에 나타낸다. 대본의 평균 문장 수는 약 16문장이며, 단어수는 약 189개에서 평균 인식률은 약 93.31%를 보였다. 단어수가 그리 많지 않아 90% 이상의 단어 인식률을 얻을 수 있었다. 평가의 방법은 다음과 같이 문장에서의 단어인식률을 사용하였다.

$$WER = \frac{N - D - S}{N} * 100(\%) \qquad (1)$$

표2. 연속음성인식률 평가 (Offline)

	시간(분:초)	문장수	단어수	인식률
ST1	1:57	16	226	91.18
ST2	2:47	20	174	93.49
ST3	2:27	20	231	94.19
ST4	1:18	10	112	96.50
ST5	2:02	16	264	92.55
ST6	1:33	14	124	91.95
평균	1:98	16	189	93.31

(3) 음성인식률 저하의 원인

음성인식률 저하의 원인은 다음과 같이 크게 세가지 카테고리로 구분할 수 있다. 아래의 경우는 대본 편집시에 발음을 한글로 적어주는 것이 인식률 향상을 위해 유리하다.

① 발음 변이- 원문장이 "3.15" 인 경우 이를 역사스페셜에서는 "삼일오 부정선거"와 같이 "삼일오"로

읽게 되는 반면, 생로병사의 비밀에서는 "삼점일오 퍼센트"와 같이 "삼점일오"로 읽게 된다.

② 외래어 - 대본에서 원문장에 "SAY NOTHING"과 같은 외래어가 있는 경우, 이를 "세이나씽"으로 읽을 수도 있고, "세이너씽"으로 읽을 수도 있다.

③ 낭독 중 지시사 '이,그,저'의 첨가 - 대본을 낭독할 때 자연스러움을 위해서 '이,그,저'와 같은 지시사가 첨가되는 경우가 있다.

III. 후처리 및 온라인 테스트

3.1 후처리 알고리즘

후처리 모듈은 음성인식 엔진의 결과를 바탕으로 전체 대본에서 현재 위치를 판단하는 모듈이다. 현재 위치의 판단은 N-Word Window를 기반으로 하여 최근 3~5개의 인식 어휘를 바탕으로 현재 위치를 추적하게 된다. N 의 크기는 실제 프롬프터 S/W의 폰트 크기에 따라 한 라인에 표시가능한 단어의 수에 따라 3에서 5까지의 수로 결정한다. 다음 [표3]은 후처리 알고리즘을 정리한 것이다.

표3. 후처리 알고리즘

1. 최신 음성인식 결과 가져옴
2. 명령어인지 확인하여 명령어 수행
> 명령어: 북마크이전/북마크다음/페이지업/페이지다운
3. 인식결과가 있는지 파악
> 공백을 제거한 TRIMTEXT에서 찾음
4. 현재 인식위치와 프롬프터 위치를 비교
A. 차이가 크면 스크롤 스피드 증가
B. 차이가 작으면 스크롤 스피드 감소
5. 음성 끝점이 검출되면 스크롤 멈춤
6. 다음 읽을 위치를 프롬프터 위치로 이동

음성인식엔진이 약 93% 정도의 정확도를 보이지만 실제 필드에서 사용하기 위해서는 98% 이상의 정확률이 필수적이다. 이에, 후처리 알고리즘에서는 음성인식 오류를 보완하기 위해 다음과 같은 기법을 사용하였다.

① 텍스트에서 특수문자 자동제거: ['`,.?"<>()!-]" 와 같은 특수문자를 제거한다.

② 글자 수를 기준으로 위치 판단: 어절이 아닌 음절수 6글자 이상인 경우 정확한 인식결과로 인정한다.[1)]

③ 공백을 제거하고 위치 검색: 각 어절 사이의 공백을 제거한 Trimmed Text에서 인식결과를 검색한다.

④ 인식결과 위치로의 이동제한: 오인식이 있을 경우 해당 위치로 갑자기 점프하는 것을 방지하기 위해 인식결과가 현재 페이지에 나타나는 경우에만 스크롤을 이동한다.

⑤ 자동발음변환: "십육분/나씽"과 같이 리턴된 결과를 자동으로 "16분/NOTHING" 으로 변환.

⑥ 스크롤 속도의 조절: 급격한 위치 변화가 없도록 4~6 사이의 스크롤 속도로 조절.

⑦ 인식결과가 두 번 이상 등장하는 경우: 현재 프롬프팅 하는 위치보다 아래쪽에 나오는 첫 번째 위치로 이동.

3.2 음성인식 자동 스크롤 정확도 측정 - Online

음성 인식 후 후처리 모듈을 거친 자동 스크롤의 정확도를 측정하기 위해 다수의 인원을 대상으로 다음과 같은 인식 실험을 수행하였다. 낭독 및 인식 실험을 위한 대본으로는 역사스페셜, 생로병사의 비밀, KBS독립영화관 3가지의 프로그램에서 총 9회의 방영분에 대한 MC의 대사를 정리하여 사용하였다. 총 21명(남성 15명, 여성 6명)이 인식 실험에 참여하였으며, 각 참가자들은 준비된 대본들 중 100줄 가량의 대본을 수회에 걸쳐 임의로 선택 후 낭독하여 자동 스크롤의 정확도를 측정하였다.

표4. 프롬프터 S/W Online 실험 결과

No.	100줄 실험 분량	에러 횟수	100줄 평균 에러	No.	100줄 실험 분량	에러 횟수	100줄 평균 에러
1	6회	2	0.3	12	3회	12	4
2	3회	2	0.7	13	2회	2	1
3	3회	1	0.3	14	2회	1	0.5
4	3회	3	1	15	2회	0	0
5	3회	6	2	16	2회	1	0.5
6	3회	2	0.7	17	1회	0	0
7	3회	6	2	18	1회	1	1
8	3회	0	0	19	1회	0	0
9	3회	2	0.7	20	1회	3	3
10	3회	7	2.3	21	1회	11	11
11	3회	4	1.3	100줄 당 전체 평균			1.3회

정확도의 측정 방식은 실험 참가자가 대본을 낭독하여 자동으로 스크롤이 진행되는 과정에서 낭독하고 있는 부분이 화면 안에 들어오지 않는 경우를 에러로 처리하였다. 각 실험 대상자에 대한 에러율을 조사한 결과 100줄 평균 약 1.3회의 에러가 발생하여, 98%이상의

---

1) [a] 3어절: 자 이 의자는 (4자)
[b] 2어절: 운동예방법을 알아봅니다. (11자)
[a]는 발성은 짧은 반면 3어절이어서 오인식의 확률이 높은 반면, [b]는 2어절임에도 발성이 길어 오인식의 확률이 낮다. 따라서 음절기준이 보다 정확하다.

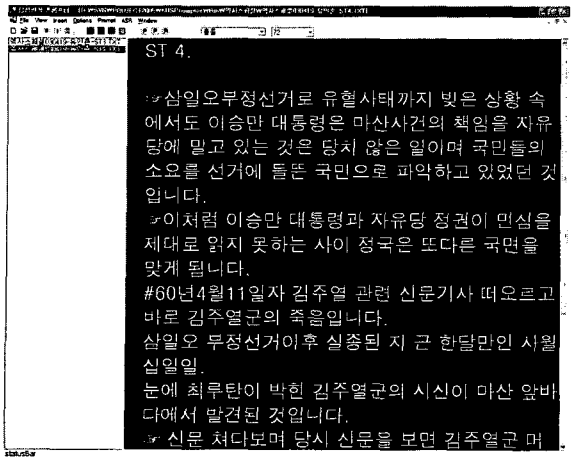

그림 3. 음성인식 프롬프터 S/W 동작화면

필드 정확율을 얻을 수 있었다. [표4]는 실험의 구체적인 내용 및 결과를 나타낸 것이다.

Ⅳ. 음성인식 프롬프터 프로그램

[그림3]이 개발된 음성인식 프롬프터의 편집 화면이다. 좌측에 현재 편집 중인 문서가 리스트 형태로 표현되며 이를 더블클릭하면 해당 문서를 편집할 수 있다. 우측에는 편집 중인 문서가 표시된다. 툴바에서 편집 기능에는 폰트의 크기 조절, 폰트 굵게, 이탤릭, 언더라인 조절, 폰트의 색깔 조절, 폰트의 종류 및 크기 조절 기능을 기본으로 제공한다. 'P' 버튼을 누르면 전체화면 모드로 전환되며, 'A' 버튼을 누르면 음성인식을 시작할 수 있다. 다음 [표5]는 음성인식 프롬프터의 전체 기능을 간략히 정리한 것이다.

Ⅴ. 결론

기존의 프롬프터 S/W는 별도의 오퍼레이터가 아나운서의 발성을 들으면서 수동으로 다음 읽을 위치를 지정해 주는 방식이었다. 그러나 이를 음성인식을 통해 자동화하여 아나운서가 별도의 오퍼레이터 도움 없이도 프롬프터 S/W를 보면서 대본 녹화가 가능하다. 타인의 도움 없이 프롬프터를 보면서 지속적인 낭독을 할 수 있으므로 연습 및 실제 녹화에 소요되는 시간을 단축시킬 수 있으며, 기타 다른 용도로도 충분히 활용할 수 있으리라 판단된다.

표5. 음성인식 프롬프터의 기능

항목	기능
File	New, Open, Save, Close 등
View	Full/Prompter Layout, Unde/Redo
Insert	Page Break, Bookmark, File
Option	Cue Marker, Scroll Control, BG Color, Font, Margins, Alignment
Prompt	Edit Position, Prev Position
ASR	Start ASR, Stop ASR
Window	Cascade, Tile, Close
단축키 및 음성명령	F11/F12: 음성인식 시작/중지 SPACE: 프롬프팅 시작/중지 1-9: 스크롤 속도 조절 1~9단계 </>: 스크롤 속도 느리게/빠르게 ENTER: 스크롤 방향 조절 F7/F8: 이전,다음 북마크 ESC: 프롬프팅 종료

음성인식 기술의 활용 분야는 다양하지만 잡음 환경에 대한 강인성의 부족, 대어휘 인식시 인식률 저하 및 미등록어 발생의 문제 등으로 인해 그 활용분야가 미흡했던 것이 사실이다. 하지만 음성인식 프롬프터의 경우 대본이 명확히 정해져 있고 단어수가 1000단어 미만이며 프로그램의 후처리를 통해 인식 오류를 보완할 수 있다는 점 때문에 상용화가 충분히 가능한 분야이다. 이처럼 새로운 기술을 상용화 가능한 분야에서 성공적으로 적용하여 연구 개발의 효과를 입증하였다.

향후에는 실제 녹화 과정에서 적용하여 제시되는 의견들을 반영하여 S/W를 지속적으로 수정 보완할 예정이다. 또한 영어, 일본어, 중국어 방송 시장에도 적용할 수 있도록 다국어로 도메인을 확장할 예정이다.

참고문헌

[1] TQ-WIN 프롬프터 매뉴얼, TOUCH-Q. http://touchq.new21.net/