

시각장애인의 미디어접근권 향상을 위한 음성자막 저작기술

장인선, 안충현, 서정일, 장윤선*
 한국전자통신연구원 방송통신미디어연구부
 {jinsn, hyun, seoji}@etri.re.kr
 *충남대학교 전자공학과
 jangys@cnu.ac.kr

Audio Subtitles Authoring Technique for Improving the Blind People's Media Accessibility

Inseon Jang, ChungHyun Ahn, Jeongil Seo, Younseon Jang*
 Electronics and Telecommunications Research Institute
 *Chungnam National University

요 약

본 논문에서는 시각장애인이 외국어로 된 미디어 콘텐츠를 이해하는데 필수적인 음성자막 서비스용 콘텐츠 저작방법을 제안하고 그 구현 예를 제시한다. 제안하는 방법은 저작 단계별로 편의 기술을 제공하여 저작자로 하여금 편리하고 경제적으로 음성자막 콘텐츠를 제작할 수 있도록 한다.

1. 서론

최근 들어 미디어 매체가 다양해지면서 장애인들이 더 많은 종류의 미디어 콘텐츠와 관련 서비스에 접근이 용이해졌다. 일 예로, 2014 년에 ETRI 가 한국시각장애인 연합회를 통해 실시한 설문조사 결과에 따르면 인터넷을 통해 외화와 웹툰 서비스를 이용해 본 경험이 있는 시각장애인의 비율이 각각 63.3%, 40%로 인터넷을 통한 미디어 접근이 시각장애인들에게도 많이 확대되었음을 알 수 있다[1]. 하지만, 인터넷은 물론이고 방송에서도 외화의 한국어 자막 서비스가 보편화되고 있는 추세이므로 시각장애인들이 외화를 접하기에는 현실적으로 어려움이 있다.

외국어 미디어에 대한 시각장애인용 보조 서비스로 음성자막(audio subtitles 혹은 spoken subtitles)이 있다. 음성자막 서비스는 미디어 내 외국어로 녹음된 음성을 대상 언어로 변환(번역)한 자막(inter-lingual caption)을 음성으로 제공하는 것이다. 이는 일반적으로 보이스 오버(voice-over)의 형태로써 제공되며 원래의 음성 위에 대상 언어로 번역한 음성을 동시에 들리게끔 한다.

외화에 대한 음성자막 서비스는 2000 년대 중반 네덜란드와 스웨덴에서 시작되었으며 이후 DVB 등의 표준이 마련되었다[2] [3]. 2010 년대 초에 들어서는 장애인의 미디어접근성 향상을 위한 연구 프로젝트 - 예로, ADLAB, Hbb4All 과 ITU-T 의 활동이 활발해지는 등 음성자막에 대한 관심이 높아지고 있다[4] [5] [6]. 우리나라의 경우, 외화 방송에 대해서 기존의 더빙 서비스를 자막 서비스로 대체하고 있는 추세이며 장애인방송 수신기에 보이스 오버를 위한 음성합성이 포함되어 있지 않는 등 외화와 관련한 시각장애인용 방송 서비스가 충분하지 않은 실정이다. 다만, Daum 팟플레이어 등 몇몇 PC 기반의 동영상 재생 플레이어에서 음성자막 서비스를 제공하고 있다[7].

일반적으로 음성자막을 제작하기 위해서는 영상번역가, 편집자, 나레이터(성우) 등이 필요하다. 영상번역가가 외화와 대사 리스트를 기반으로 번역을 수행하고 편집자의 검수를 거쳐 대상 언어의 자막을 완성한다. 번역된 자막은 성우 더빙 및 오디오 믹싱 등을 거쳐 음성자막 콘텐츠로 제작한 후 전송되거나 단말에 직접 전송되어 단말 내 음성합성기를 통해 보이스 오버 서비스를 제공한다.

본 논문에서는 시각장애인의 미디어 접근성 향상을 위한 음성자막 콘텐츠 저작방법을 제안한다. 제안하는 방법은 웹번역 서비스를 활용하여 자동 외국어 번역을 제공하며 번역된 한국어 자막을 TTS(Text-to-Speech)를 통해 음성으로 변환하고 변환된 자막 음성을 마스터 오디오와 믹싱하여 출력함으로써 음성자막 콘텐츠를 효율적으로 저작할 수 있도록 한다. 또한, 각 대사 별로 화자, 감정 등에 대한 정보를 삽입할 수 있도록 하여 방송 등 고품질의 더빙 서비스용 대본 출력을 지원한다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 음성자막 저작기술의 특징을 설명하고 3 절에서는 제안한 방법을 포함하는 음성자막 저작도구의 구현 예를 설명한다. 마지막으로, 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 화면해설 저작기술의 특징

가. TTS 기반 보이스 오버

제안하는 저작 방법에서는 한국어로 입력된 자막파일을 직접 입력 받거나 외국어 자막을 번역하여 그 결과에 TTS 를 적용하고 해당 파형을 디스플레이 하여 저작자에게 음성자막의 오디오 정보를 제공한다. 이는 기존 외국어 음성 대비 한국어 음성 신호의 길이 비교를 용이하게 하므로 음성자막 분량을 선정하는데 활용할 수 있다. 또한, 자연스러운 발음과 억양 표현뿐만 아니라 말의 속도(speech rate), 크기(volume) 및

피치(pitch) 등 세밀한 제어가 가능한 음성합성 SW 를 연동하여 성우의 음성자막 더빙을 대체하는 보이스 오버를 저작할 수 있다.

나. 더빙용 대본 편집

대사에 해당하는 등장인물의 감성 정보를 태깅하여 대본에 지문형태로 삽입함을 제안한다. 이는 고품질의 음성자막 방송 콘텐츠를 제작하기 위한 성우 더빙용 대본으로 활용할 수 있다.

다. 오디오 믹싱

동영상 파일 내 포함되어 있는 마스터 오디오와 음성자막 합성음성의 속성(볼륨, 재생 시간 등)을 조절 후 믹싱하여 최종적인 음성자막 오디오를 생성하도록 한다. 일 예로, 음성자막 합성음성을 외국어 음성보다 약간 지연시켜 재생되도록 함으로써 화자 구분을 용이하게 하고, 합성음성이 더해지는 구간의 외국어 음성 볼륨과 한국어 음성 볼륨을 각각 감소/증가시켜 대사 내용과 화자 구분에 대한 명료성을 증가시킴으로써 음성자막 오디오에 대한 청취자의 이해도를 향상시킬 수 있다.

상기의 특징 외에도, 외화 콘텐츠의 외국어 자막만 존재하는 경우에는 Google 이나 MS Bing 등에서 제공하는 웹 번역 서비스를 활용하여 자동으로 한글 자막을 생성할 수 있다.

3. 구현 예

제안한 화면해설 저작도구의 구현 예는 그림 1 과 같다. 입력된 동영상 파일(MPEG-2 TS, AVI, MP4 및 WMV 지원) 비디오는 Preview 창에, 외국어 음성을 포함하는 오디오 파형은 Audio 창에 각각 보여지며 자막 파일(SMI, SRT)은 Script viewer 에서 확인 가능하다. 외국어 자막의 경우, 웹 번역 서비스를 통하여 한국어 자막을 생성할 수 있으며 생성된 자막은 TTS 엔진을 실행하여 합성음성을 출력한다. 출력된 음성자막 합성음성의 파형은 타임라인에 맞추어 (자막 파일에 기재된) 디스플레이 시작시간에 자동으로 위치하게 되며 자막 텍스트와 시작시간은 사용자 수정이 가능하다. 구현 예에서는 디오텍(Diotek)의 PowerTTS 를 이용하였다. 또한, 각 자막 편집 창에는 감성정보 입력창을 연동하여 대사에 해당하는 인물의 감성정보(성별, 연령, 캐릭터, 감정 등)를 쉽게 기재할 수 있도록 하였다. 타임 라인에 맞추어 배치된 음성자막 합성음성과 마스터 오디오를 믹싱하여 음성자막 보이스 오버 오디오를 wav 파일로 출력할 수 있으며 비디오와 함께 믹싱하여 음성자막 콘텐츠를 생성할 수 있다.

4. 결론

본 논문에서는 음성자막 콘텐츠 저작기술을 제안하고 각 세부 기술에 대해 설명하였다. 제안하는 방식은 음성자막 전반에 대한 편의성을 제공하며 음성자막 서비스를 위한 더빙 대본을 효율적으로 작성하도록 한다.

제안한 음성자막 콘텐츠 저작도구를 기반으로 향후에는 시각장애인용 미디어 콘텐츠 제작에서의 효율성을 검증하고 영상물뿐만 아니라 웹툰 등 다른 서비스 영역으로의 확대를 위해 연구개발을 수행할 예정이다.

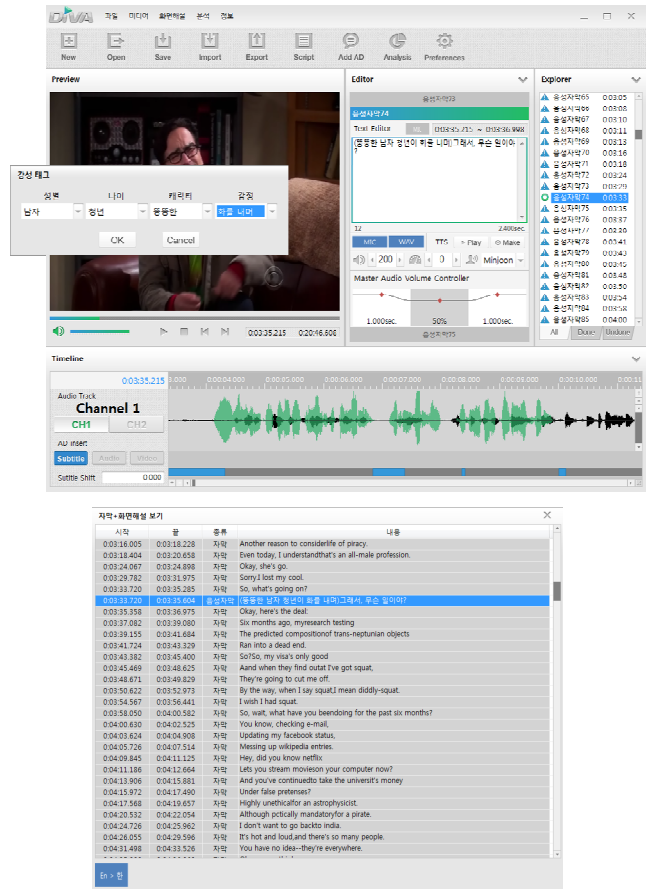


그림 1. 구현의 예: 음성자막 저작도구 및 감성 정보 삽입(위), 스크립트 뷰어(아래)

감사의 글

이 논문은 2015 년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (B0192-15-1001, 시청각장애인 방송접근권 향상을 위한 디지털자막·음성해설 서비스 기술 개발)

참고문헌

- [1] 장인선, 안충현, 서정일, “시각장애인을 위한 미디어 접근성 향상에 대한 연구,” 한국방송공학회 추계학술대회 논문집, 2015.
- [2] ETSI EN 300 468 V1.13.1 (2012-08) Digital Video Broadcasting (DVB): Specification for Service Information (SI) in DVB systems.
- [3] S. Braun, P. Orero, “Audio Description with Audio Subtitling – an emergent modality of audiovisual localization,” Perspectives: Studies in Translatology. vol. 18, issue 3, pp. 173~188, 2010.
- [4] Audio-Description: Life-Long Access for the Blind, <http://www.adlabproject.eu/>.
- [5] Hybrid Broadcast Broadband For ALL, <http://www.hbb4all.eu/>.
- [6] ITU-T Focus Group on Audiovisual Media Accessibility (2013), <http://www.itu.int/pub/T-FG-AVA>.
- [7] <http://daumui.tistory.com/36>