청각장애인을 위한 감성자막 편집기 구현

*김현순 *오주현
*한국방송공사 미디어기술연구소
*{soon71, jhoh}@kbs.co.kr

Implementation of an emotional subtitle editor for deaf and hearing impaired people

*Hyunsoon Kim, *Juhyun Oh
KBS Media Technology Research Institute

요약

디지털화와 기술의 급격한 발전으로 방송 서비스도 고품질 서비스를 보다 편리하게 이용할 수 있도록 진화하고 있다. 이러한 변화하는 방송 환경에서 비장애인 대비 소외계층의 정보 접근성을 높이기 위한 연구에 대한 필요성이 증가하고 있다. 이러한 연구의 일환으로 UHD 자막 방송 서비스를 개선하기 위한 연구인 '감성표현 자막 서비스 기술' 연구를 진행하였다. 감성표현 자막 서비스 기술은 단순한 텍스트의 전달이 아닌 이미지와 폰트 스타일을 포함한 다양한 시각적 표현을 통해 청각장애인의 방송 내용에 대한 이해도를 향상시키기 위한 기술이다.

본 논문에서는 이러한 감성표현 자막 서비스를 소개하고 해당 서비스를 가능하게 하는 관련 기술과 시스템 구현 결과에 대하여 다룬다. 지상파 UHD 방송을 대상으로 개선된 형태의 자막 서비스를 제공하기 위한 핵심 시스템인 감성자막 편집기를 개발하였다. 감성자막 편집기는 화자의 감정 정보 등을 입력, 편집하고 편집된 감성자막을 영상과 싱크를 맞추어 재생하는 기술과 감성자막을 UHD 송출시스템으로 전송하는 시스템이다.

1. 서론

방송이 디지털화되고 UHD 방식이 도입되는 등 방송 서비스가 발전, 확장됨에 캐릭터 수화 방송, 자막 방송 등 장애인을 위한 방송 서비스의 양적, 질적 개선을 위한 노력도 지속되어야 한다[1,2]. 이러한 서비스의 하나인 청각장애인을 위한 지상파 자막 서비스의 경우, 지상파 UHD 본 방송을 시작한 이래 폐쇄 자막 서비스 시스템을 구축하여 서비스를 제공하고 있다[3, 4]. 이러한 폐쇄자막 서비스에 대하여 현재 서비스되고 있는 형태인 단조로운 텍스트 형태의 서비스를 벗어나 다양한 텍스트 스타일 등을 적용하여 풍부하게 내용을 전달하는 것에 대한 요구가 발생하고 있다.

이러한 청각 장애인을 위한 자막 방송 서비스 개선에 관한 연구의 일환으로 진행한 감성표현 자막 서비스 및 기술을 소개하고자 한다. 감성표현 자막 서비스 기술은 단순한 텍스트의 전달이 아닌 다양한 시각적 표현을 통해 청각장애인의 방송 내용에 대한 이해도를 향상시키기 위한 기술이다[5].

본 논문에서는 이러한 기술 중 지상파 UHD 방송을 대상으로 이러한 보다 개선된 형태의 자막 서비스인 '지상파 UHD 기반 감성자막 서비스'를 제공하기 위한 '감성자막 편집기 개발'[5]에 대한 업그레이드 기능과 연관된 소주제로서 감성자막 픽토그램 구현 결과를 위주로 다룬다.

먼저 감성자막 서비스 개념 및 해당 서비스를 제공하기 위한 전체 시스템을 기술한다. 이어서 감성자막 편집기 개발, 감성자막 픽토그램 제작 결과에 대하여 다룬 후, 향후 연구 항목을 제시한다.

2. 서비스 개념 및 시스템 개요

감성자막 서비스는 화자의 감정 정보를 자막 메타데이터에 추가적으로 제공하여, 감정에 따라 다양한 이모티콘이나 다양한 종류의 폰트 스타일로 자막 서비스가 가능하게 하는 자막 방송 서비스이다. 화자의 감정 정보와 콘텐츠 흐름에 대한 이해를 돕기 위한 감정표현 이미지는 애니메이션 이미지인 APNG(Animated Portable Network Graphics) 포맷으로 제작하였다. 전화벨 소리, 발자국 소리 등 장면 이해에 도움이 되는 주요 소리나 화자의 감정 상태 등을 이미지로 표현하고, 이미지를 자막 파일에 추가하여 서비스할 수 있다. 화자구별을 위하여 화면에서 화자가 누구인지 AI 시스템이 인식하고, 화면상 화자의 위치에 자막의 위치를 맞추어 배치하는 서비스도 가능하다. 또한 화자의 감정을 인식하고 이에 대한 정보를 자막 파일에 추가적으로 기입하고 감정 정보에 따라 자막의 폰트 종류, 색상, 크기 등 텍스트 스타일을 다르게 하여 표현할 수도 있다.

감성 표현 자막 서비스를 지상파 UHD 자막 서비스에 추가적인 메타데이터 형태로 추가하여 제공하기 위해서는, 지상파 UHD 자막 표준인 IMSC(TTML Profiles for Internet Media Subtitles and

Captions)[6]에 기반한 자막 메타데이터에 감성 정보(화자의 감정, 화자 위치, 의성어 및 의태어 정보 등)를 추가하여 전송하도록 개발하여야 한다. 이를 위하여 미디어 기반 감정인식 AI 시스템, 감성자막 편집/재생 기술, 영상-자막 동기화 시스템, 감정을 표현하기 위한 감성표현 이미지 제작 등이 요구된다.

이들 기술 중에서 본 논문에서 중점적으로 다루는 것은 감성자막 편집기 개발, 감성표현 이미지 제작이다. 감성자막 편집기는 감정인식 AI 시스템에서 인식한 감정 정보 등을 전달받아 자막 데이터에 추가하고 영상-자막 동기화 시스템과 연동하여 영상과 자막의 싱크를 맞춘다.

APNG 형식의 픽토그램 애니메이션은 다양한 종류의 소리 이벤트를 검토하여 최대한 많은 종류를 표현할 수 있도록 개발하였다. 감성자막 편집기에서는 이러한 감성표현 이미지를 입력, 편집하고 송출 전 재생하여 테스트하는 기능을 제공한다.

그림 1. APNG 이미지(핸드폰) 표출 예

그림 2. 폰트 스타일로 감정 표현 예

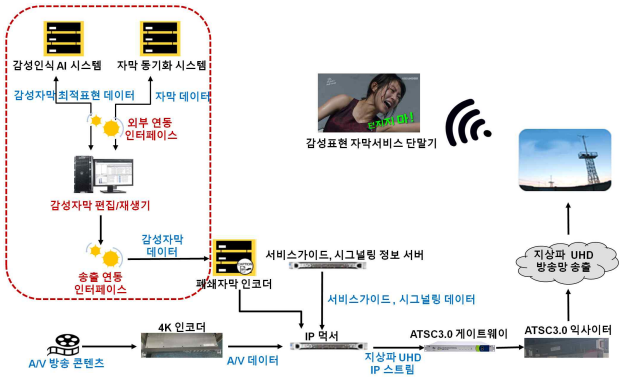

그림 3. 감성자막 서비스 및 시스템 개념도

3. 감성자막 편집기 개발

화자의 감정 정보와 콘텐츠 흐름에 대한 이해를 돕기 위한 이미지를 포함하는 자막인 감성자막을 제공하기 위하여, 감정 정보 등을 입력, 편집하고 편집된 감성자막을 영상과 싱크를 맞추어 재생하는 편집/재생기 기술과 감성자막을 UHD 송출시스템으로 송출하기 위한 기술이 필요하다. 이를 위하여 감성자막 편집/재생기를 개발하였다.

감성자막 편집 모듈은 지상파 UHD 자막 표준인 IMSC 표준을 지원하며, 기존의 자막 텍스트를 포함하여 감성 표현 정보를 편집하는 기능을 수행한다. 즉 감성자막 편집/재생 모듈의 주요 기능은 동영상 및 자막 파일을 입력받아 이를 편집하고, 그 결과를 송출 전에 재생하여 보는 것이다. 이러한 기술 외에도 감성자막 이미지 처리 기능, 감성인식 AI 시스템, 자막 동기화 시스템과의 연동 인터페이스 모듈, 편집/재생기 내 플레이어 기능 등을 포함한 업그레이드 기능을 필요로 한다. 이러한 기능을 수행하기 위하여 동영상 및 자막 파일 입력/파싱/디코딩/재생, 동영상과 자막 간 싱크, 프리뷰 기능, 감정정보 입력/편집/재생 기능, 외부 시스템과의 연동 인터페이스 기능을 지원한다.

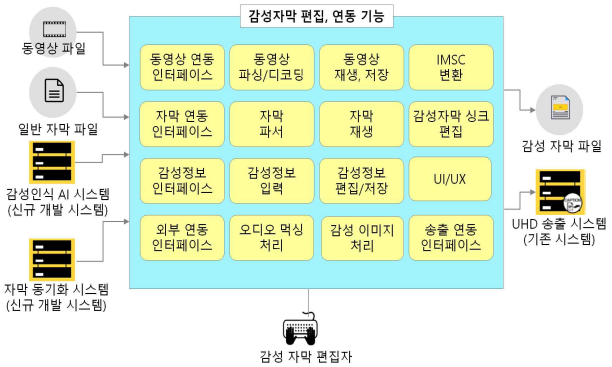

그림 4. 감성자막 편집, 연동 모듈 기능도

감성자막 편집기 화면 구성은 자막 정보 영역, 비디오 영역, 오디오 파형 영역으로 구성된다. 감성자막 편집 모듈에서는 동영상, 자막 파일을 입력받고 시스템 사용자인 감성자막 편집자로부터 감정 정보, 화자 정보, 감성 스타일, 자막 위치, 감성 표현 이미지, 자막 표현 시간 정보 등을 입력 받아 감성자막 파일을 생성, 출력, 저장하는 기능을 제공한다. 자막 정보, 동영상 파일, 오디오 파일을 파싱하고 시스템 사용자가 직관적으로 편집할 수 있도록 UI(User Interface)를 제공한다.

감성자막에서 사용하는 감성 정보는 감정에 따른 폰트 스타일, 화자 정보, 화자의 위치에 맞는 자막 표시 위치, 콘텐츠의 흐름에 대한 이해를 돕는 감성표현 이미지이다. 각각의 감성 정보에 대해 별도의 등록이 가능하며, 자막 목록에서 우클릭 팝업으로 선택 가능하다.

감성에 따른 스타일 설정으로 화자 및 상황에 따른 스타일을 설정할 수 있으며, 스타일의 주요 구성은 글꼴, 글자 크기, 이탤릭 여부, 글자 색상 등이다. 이렇게 감성에 따라 글자 색상 및 글자 크기를 변경함으로써 보다 정확한 감성을 설정할 수 있다.

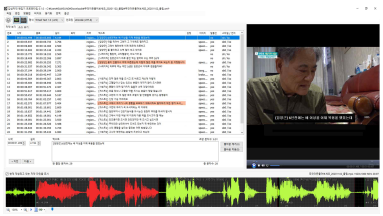

그림 5. 편집기 메인 UI

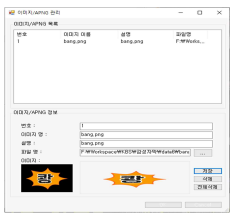

그림 6. 이미지 등록

자막 편집기는 감성인식 AI 시스템, 자막 동기화 시스템과의 인터페이스를 제공한다. 감성자막 편집기에서 동영상(MP4) 파일과 자막(SMI)을 AI Data Agent에 분석을 요청하면, AI Data Agent는 수신한 동영상과 자막 파일을 AI 감정인식 서버를 통해 분석을 수행한다. 감성자막 편집기에서 분석 요청 이후에는 다시 AI Data Agent에 분석한 결과를 요청하여 분석이 완료된 결과를 AI 분석 파일(JSON)으로 수신한다.

자막 싱크를 위해서는 감성자막 편집기에서 동영상 파일과 자막 파일을 자막 싱크 시스템에 전달한다. 자막 싱크 시스템에서 싱크를 처리하며, 감성자막 편집기에서는 동기화가 완료되었는지 확인을 하여 동기

화된 파일을 다운로드 한다.

4. 감성자막 픽토그램 제작

본 논문에서 제안하는 감성자막은 PNG 이미지나 애니메이션을 자막 텍스트와 함께 서비스하는 것을 포함한다. 자막 이미지와 애니메이션은 특히 화면의 시각 정보(visual information)만으로는 알기 어려운 소리 이벤트들(audio events)을 나타내는 데 유용하게 사용될 수 있다. 따라서 영상 콘텐츠에 나타날 수 있는 소리 이벤트들을 나타내는 PNG 애니메이션을 제작하였다. 영상 콘텐츠에 나타날 수 있는 모든 종류의 소리 이벤트를 포함하기 위하여 Google의 AI 오디오 학습 데이터셋인 'AudioSet'[7]을 참고하였다(그림 7).

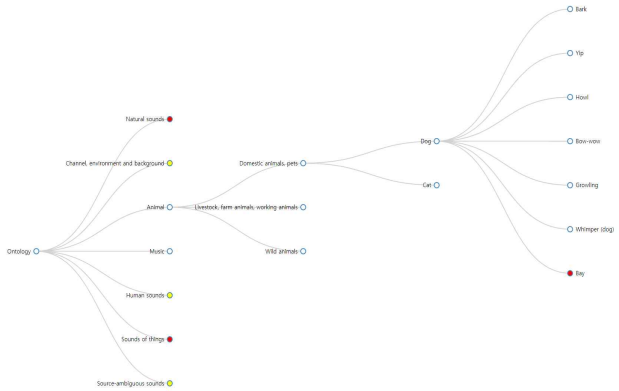

그림 7. Google의 AudioSet 온톨로지 구조

AudioSet은 총 527개의 클래스로 구성되는데, 이들 중 방송 영상 콘텐츠에 등장 빈도가 높다고 판단되는 122개의 클래스를 그림 8과 같이 선정하였다.

이와 같은 소리 이벤트들은 흑백의 '픽토그램(pictogram)' 애니메이션으로 제작하였다. 소리를 이미지로 나타내기 위한 상징성, 흑백 위주인 텍스트 폐쇄자막과의 동질성, 그리고 콘텐츠 영상에 기 삽입된 화려한 오픈 자막(open caption)과의 차별성을 확보하고 일관된 톤(tone)으로 구현하기 위해서이다.

그림 9는 이들 중 일부의 APNG 애니메이션 프레임을 보여준다. APNG 애니메이션은 600×600 픽셀, 32 bit RGBA, 최대 40프레임의 길이로 제작하였다.

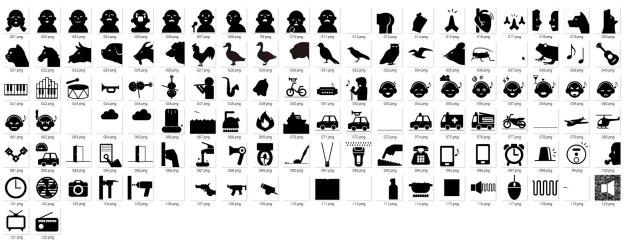

그림 8. 감성자막 픽토그램 리스트

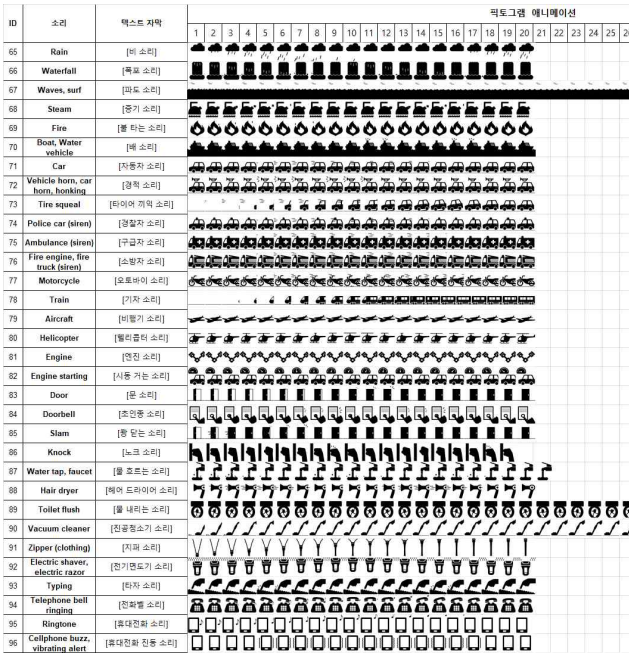

그림 9. 감성자막 픽토그램 애니메이션 일부

5. 결론

화자의 감정 정보와 콘텐츠 흐름에 대한 이해를 돕기 위한 감성자막 이미지를 포함하는 자막인 감성자막을 제공하기 위하여, 감정 정보 등을 입력, 편집하고 편집된 감성자막을 영상과 싱크를 맞추어 재생하는 편집기를 개발하였다. 감성자막 이미지는 화면의 시각 정보만으로는 알기 어려운 소리 이벤트를 표현하기 위하여 감성자막 픽토그램을 제작하여 테스트하였다.

방송 서비스의 발전과 함께 자막 방송 서비스의 질적 개선도 지속되어야 한다. 이를 위하여 본 논문에서 제시한 구현 결과 이외에도 한국어 자막 동기화 등을 포함한 시청각장애인을 위한 방송 서비스 개선을 위한 연구가 계속되기를 바란다.

Acknowledgement

본 연구 논문은 과학기술정보통신부 및 정보통신기획평가원의 정보통신·방송 연구개발 사업의 일환으로 수행중인 한국전자통신연구원 주관 "시청각 장애인의 방송시청을 지원하는 감성표현 서비스 개발" [2019-0-00447] 과제의 지원을 받은 연구결과임.

참고문헌

[1] TTA, "시청각 장애 보조 방송 서비스," TTAK.KO-07.0093/R2, 2018.
[2] 양승준, 안충현, "청각장애인을 위한 동적인 감성 자막에 관한 연구," 한국통신학회 추계종합학술발표회, pp. 85-86, 2014.
[3] TTA, "지상파 UHDTV 방송 송수신 정합," TTAK.KO-07.0127/R4, 2019.
[4] Yunhyoung Kim, "Implementation of Closed Captioning

System for Terrestrial UHD based on ATSC 3.0," SMPTE 2017 Annual Technical Conference & Exhibition.
[5] 김현순, 오주현, "청각장애인을 위한 감성자막 편집기 개발," 한국방송·미디어공학회 하계학술대회, 2020.
[6] "TTML Profiles for Internet Media Subtitles and Captions 1.0 (IMSC1)," W3C, http://www.w3.org/TR/ttml-imsc1.0.1/
[7] J. F. Gemmeke 외, "Audio Set: An Ontology and Man-Labeled Dataset for Audio Events," IEEE ICASSP 2017.