MBC의 미디어AI 서비스

□ **성시훈** / ㈜문화방송

요 약

㈜문화방송(MBC)은 콘텐츠 제작 및 유통 워크플로우에 인공지능(Artificial Intelligence, AI) 기술을 적용한 미디어AI 서비스를 운영하고 있다. 영상아카이브에 보관되어 있는 수십만 개의 아날로그와 SD급 콘텐츠를 대상으로 HD급 수준의 영상화질로 품질을 향상시키기 위해서 AI영상화질개선시스템을 2020년에 개발 구축해서 여러 목적에 활용하고 있으며, HD급 콘텐츠를 대상으로 4K 초고화질급으로 변환하는 기술로 고도화해서 실서비스 적용을 눈앞에 두고 있다. 그리고 2년의 STT(Speech-To-Text, 음성문자변환) 베타서비스를 통해 얻어진 사용성 검증과 운영 경험을 바탕으로 STT HUB 서비스를 개발 구축해서 2022년부터 보도와 시사교양 프로그램의 제작 워크플로우에 적용하고 있다. 이들 서비스의 주요 기능들과 기술적 요소들의 구현, 미디어AI 서비스 운영의 경험을 나누고자 한다.

I. 서 론

수개월 전, OpenAI의 ChatGPT가 미국 로스쿨과 의사 면허 시험에 합격할 정도라는 기사와 영상클립들이 확대 재생산되며 AI기술에 대한 관심이 한층 높아졌다[1-3]. 성능이 향상되었다는 사실을 부정하진 않으나 생성형 AI인 ChatGPT가 완전히 새로운 기술이 아니라 GPT-3와 같은 이전 버전들이 있었고 문제에 대한 답이 모두 정답은 아니라는 사실은 간과하고 있다. 학습되지 않은 상황에 대한 질문에 가끔 어이없는 오답을 만들어 내지만 사람들은 ChatGPT가 AI의 모든 것인 것마냥 열광하고 있다. 결과적으로 OpenAI는 GPT-3.5의 일부 유료화로 인한 재정적 이익과 2023년 3월에 공개한 GPT-4를 위한 더할 나위 없는 프로모션 효과를 얻었다.

1990년대에 당시 이름도 낯설었던 뉴럴네트워크(Neural Network, 신경회로망)의 인식률 1%를 올리기 위해서 수일 밤을 새야만 했고 AI를 위해서는 사람이 혹사돼야 했다. 흔히 이야기하는 'AI기술 2차 암흑기'였지만

<그림 1> AI기술이 적용된 MBC 메이저리그 생중계의 화면 일부(2001년)

〈그림 1〉과 같이 MBC는 2001년부터 미국 메이저리그야구(MLB)와 WBC 등을 생중계했고 마일(mile) 단위가 익숙치 않은 국내 시청자를 위해서 개발한 자동자막변환시스템에 신경회로망 문자인식기술을 적용했다. 상대적으로 검증된 숫자인식이었지만 한 경기당 35만장 이상의 영상을 실시간 처리해야 했고 영상이 숫자인지 여부를 판별하는 등 여러 예외 처리가 필요했기에 긴장하며 방송했었다. 수년간 안정적으로 사용했고 AI기술이 생방송에 고정적으로 사용된 예가 없었기에 여러 차례 인용되기도 했다[4].

성공적이었던 사례임에도 불구하고 'AI는 원천적으로 오류를 내재하고 있지만 오류를 예측할 수 없어서 방송에 활용하기 어렵다'는 판단에 당시는 추가적인 AI기술 적용을 고려하지 않았다. 시간이 지나 수년 전, 회사의 요구로 미디어와 관련된 AI기술들을 다시 검토하게 되었고 다시금 AI기술의 성능적 한계에 적절하게 대처하며 콘텐츠 제작과 유통에 활용하고 있다.

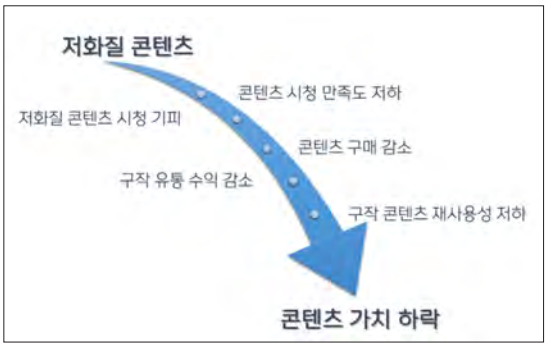
II. AI영상화질개선과 구작 콘텐츠

1. 구작 콘텐츠의 숨겨진 가치

MBC는 K-컬쳐를 대표하는 한국 드라마, 초기 한류를 선도한 대장금, 허준, 전원일기, 무한도전 등 국내외 시청자에게 사랑받은 수많은 명작을 보유하고 있다. 과거의 향수를 느끼고 싶은 40~60대와 잠깐 주류에서 멀어졌던 문화 트렌드에 대한 MZ 세대의 호기심으로 과거 구작 콘텐츠가 강제 소환되어 과거의 명작 드라마와 예능프로그램의 인기가 지속적으로 유지되고 있고 관련 콘텐츠 유통이 늘어나고 있다.

MBC는 HD급 콘텐츠의 수량에 비해 2배에 달하는 SD급 콘텐츠를 영상아카이브로 보유하고 있다. 이들 프로그램들이 감동과 웃음을 주는 스토리를 가진 우수한 콘텐츠임에도 불구하고 당시 방송표준이었던 SD급 영상화질로 제작되고 저장된 탓에, 이제는 대형 디스플레이에서 시청하기에 불편한 콘텐츠가 되어서 방송사의 영상아카이브에 제대로 빛을 보지 못한 채 켜켜이 쌓인 중고 콘텐츠 취급을 받고 있다.

불과 10여 년 전인 2012년에 디지털 HD방송으로 전환되기 전까지 방송된 대부분의 프로그램들이 34만 화소(720x480)의 SD급과 아날로그 방식으로 제작되어서 207만여 화소의 HD화질과 829만여 UHD화질에 비해 선명도가 상대적으로 크게 떨어지는데 반해 최근 출시된 TV는 고급형 뿐 아니라 보급형까지 이미 급격히 대형화되어서 VOD 뿐 아니라 구작을 재방송하는 유료방송에서 시청하는데 불편이 더욱 두드러지게 되었다. SD급 콘텐츠를 이대로 방치한다면 〈그림 2〉와 같이 제대로 가치를 인정받지 못하고 제한된 시청환경에서만 소비되는 콘텐츠로 전락할 수 있다.



<그림 2> 콘텐츠의 가치 소멸

2. AI영상화질개선시스템

제대로 명품 콘텐츠의 원래 가치를 인정받고 보다 우수한 고품질 영상서비스를 제공하기 위해서는 원석을 세공해서 반짝이는 보석으로 가치를 높이듯이 묵혀둔 콘텐츠를 새로 갈고 닦아야 하겠지만 하루하루 편성에 따른 콘텐츠를 생산하기에도 급급한 방송사의 제작 여건상 애초에 불가능한 현실을 직시해야 했다. 더구나 현장에서 작업자가 영상화질을 개선 향상하거나 복원하기 위해서는 NLE 또는 특수영상장비를 이용해서 장면별로 여러 영상처리 필터를 번갈아 적용해 보며 수차례 시도를 한다. 그럼에도 불구하고 영상에서 노이즈를 줄이면서 동시에 선명도를 높이는 작업은 쉽지 않다. 숙련된 영상 전문가의 작업이 요구되며 59.94i 기준 1시간 콘텐츠를 처리하기 위해서는 107,892장의 영상을 처리해야 한다. 수많은 콘텐츠를 일일이 수작업으로 처리하는 것은 제작 시간이나 노동의 한계에 부딪힌다.

영상화질개선기술에 대한 기술타당성을 검토하였고 당시 화질 변환을 연구하던 대다수 국내외 연구기관과 제조사들이 몰두해 있던 HD급 영상에서 UHD급 영상으로의 변환이 아니라 다시 제작할 수 없는 SD급 화질의 구작 명품 콘텐츠를 다시 개선할 수 있는 기술에 대한 필요성, 시장에서의 사업성과 하드웨어 실현가능성을 고려해서 개발목표를 SD급 콘텐츠의 개선으로 설정했다.

수십년 전부터 슈퍼레졸루션(Super-Resolution)이란 이름으로 수많은 영상화질 개선 시도가 있었지만 저화질 원본 영상에 대한 기술적인 한계와 어려움으로 실제 사용가능한 화질개선 기술 개발이 이루어지지 못했다. 기존 연구 방식에서 AI방식으로 전환되면서 학습에 소요되는 영상데이터세트는 무엇보다 중요하다. 대부분의 대학이나 연구기관에서 가상으로 제작된 학습데이터를 사용해서 우수한 결과로 제안된 논문들이 많지만 아쉽게도 실험영상과 조금이라도 다른 특성의 실제 콘텐츠를 적용하면 그저 그런 결과를 내는 사례가 많다.

AI영상화질개선기술은 SD급 영상의 아날로그적 특성 때문에 일일이 나열하기에는 지면이 부족하다 할 만한 많은 기술 개발의 어려움이 있었지만, AI 딥러닝기술과 축적된 미디어처리기술을 바탕으로 MBC 내 영상전문가 검증 등을 통해서 미디어서비스에 활용할 만한 AI영상화질개선시스템이 개발되어 운영되고 있다. 〈그림 3〉은 2003년

<그림 3> MBC 드라마 대장금(2003년작)의 AI영상화질개선 전(좌)과 후(우) 비교

작 드라마 대장금의 인물 장면에서 눈매와 윤곽을 보정하고 영상잡음을 제거한 결과이다[5]. 시스템은 딥러닝 네트워크 최적화, 병렬처리, CUDA 최적화 등을 거쳐서 고속화 개발되었다. 장르, 콘텐츠 내용과 상관없이 단일 학습 파라미터로 동작하므로 작업자의 개입 없이 자동화되어 있다. 2021년에 과거 인제스트 과정에서 발견되는 VCR과 인코더 간의 아날로그 컴포지트 신호 부조화로 인한 여러 영상 노이즈에 대한 대응 알고리듬을 추가 적용해서 콘텐츠 적용범위를 넓혔다. 여전히 수요가 높은 다양한 SD급 콘텐츠를 고화질로 개선해서 IPTV의 VOD 서비스와 케이블 PP에서 방송하고 있다.

3. AI초고화질영상변환 기술 개발

최근 제작되는 HD급 콘텐츠는 UHD 카메라를 일부 사용하고 테이프 인제스트 과정 없이 파일로 직접 편집되므로 화질 열화가 적어서 초기 HD급 콘텐츠에 비해 상대적으로 화질이 우수하다. 그러나 HD급 콘텐츠도 제작시기가 2000년대 초반에 근접할수록 UHD급 콘텐츠에 비해 노이즈가 많고 화질이 낮은 영상 불균일이 발견된다. VCR과 파일 제작 과정이 혼재되어 어쩔 수 없이 인제스트와 테이프 녹화가 반복되었고 제작 비용을 낮추기 위해서 일부 프로그램은 고압축 영상포맷을 사용했기 때문이다. 과거 제작 시점에서 만족할 만한 화질이었던 SD급 콘텐츠가 현재 저화질로 취급받는 현상이 UHD급 콘텐츠가 일반화될 시점에 HD급 콘텐츠에도 반복될 가능성이 예견된다.

HD급 아카이브본의 화질 개선과 UHD급 콘텐츠 화질 변환을 통한 콘텐츠 가치 향상을 목표로 SD급 영상에서 고화질 HD급 영상으로 변환한 기술을 고도화해서 HD급 영상을 초고화질(UHD) 영상으로 변환하는 기술 개발을 진행 중이며 실사용을 눈앞에 두고 있다.

지상파 방송사는 과거와 다른 미디어 환경에서 어려움을 겪고 있는 한편, UHD 전환이라는 중요한 추진과제를 안고 있다. 해당 기술을 적절히 활용한다면 UHD 방송 전환기에 UHD급 콘텐츠 제작에 필요한 고비용으로 인한 콘텐츠의 부족 상황을 보완하며 UHD 생태계를 구축하고 UHD 방송 환경으로 완전 전환하는 데 기여할 수 있을 것으로 기대한다.

III. STT 기술과 콘텐츠 제작 업무

1. STT 기술의 한계

누구, 기가지니, 클로바, 카카오미니 같은 스마트 스피커를 비롯해서 애플 시리, 아마존 알렉사, 구글 어시스턴트, 마이크로소프트 코타나, 삼성 빅스비 등의 음성서비스가 컴퓨터, 스마트폰, 스마트TV를 비롯해서 우리 주변의 많은 생활가전들, 자동차 등에 탑재되어 사용되고 있다[6]. 근래 음성인식기술이 회의록을 작성하고 내용을 요약하는 등 여러 상황에서 도움을 주고 있지만 간단한 호출어도 알아듣지 못해서 답답함을 경험하고 여전히 사용에 제약이 많은 것 또한 사실이다.

STT 서비스도 음성인식기술에 기반을 둔 서비스이지만 10초 이내 단어 수준의 키워드 조합으로 사용되는 스마트 스피커에 비해 음소의 작은 오류가 다른 뜻의 단어로 오류가 확대될 수 있어 문맥 이해에 영향을 미치므로 녹취 상태에 따라 인식성능이 크게 달라진다. 특히 주변 소음 및 배경음악, 전화 통화/현장 녹음, 화자간 음성 겹침, 화자 연령, 발성 습관, 사투리 사용, 마스크 착용 등 모든 음향 요소가 인식성능에 영향을 준다. 최근 폭발적인 학습데이터 증가와 인식 알고리듬 개선으로 주변 잡음이나 다양한 화자, 녹취 환경 등에 대한 영향을 줄여가고 있다.

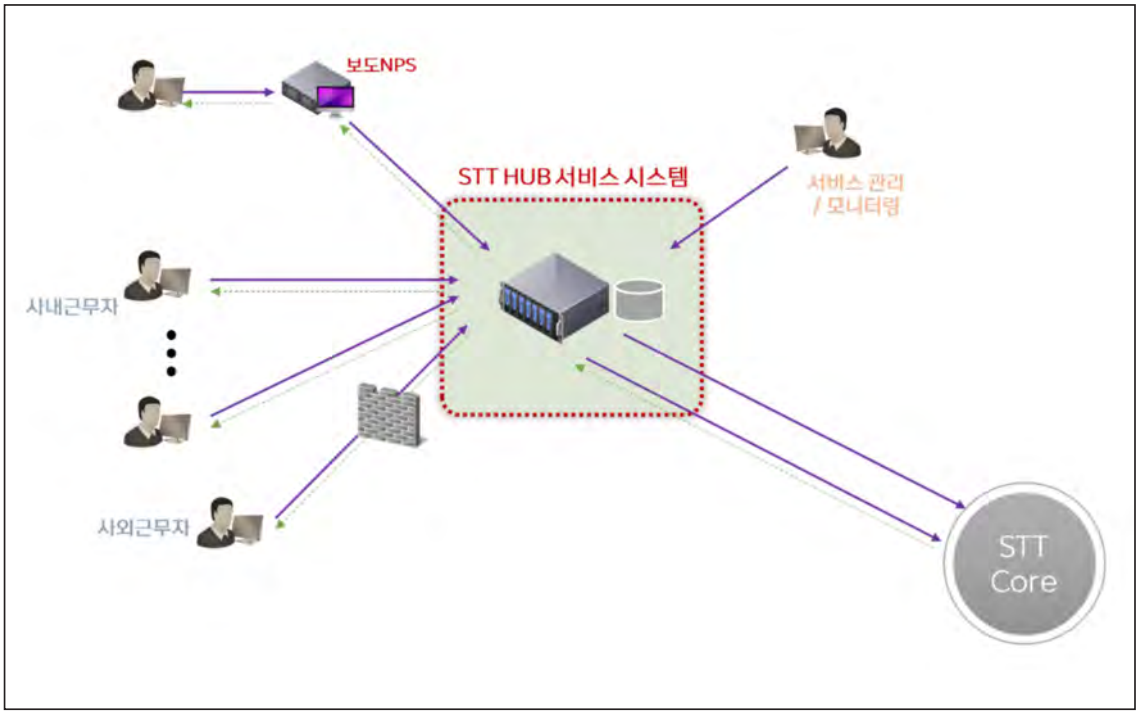
그러나 여전히 음성 콘텐츠에 따라 성능이 달라지므로 사용 용도에 따라 적절히 활용해야 한다. 특히 프로그램 제작진은 AI기술자가 아닌 일반 사용자이므로 기술 이해도와 기대치에 따라 개인별 만족도가 달라지는 주관적인 영향도 무시할 수 없다. 더구나 한글과 영어가 혼재되어 있는 경우가 많고 언어별로 구조적인 차이가 있으므로 인식 성능 뿐 아니라 처리속도 등의 서비스 요소가 STT 서

비스마다 다른 것을 고려해서 상황에 맞는 STT 서비스를 선택해야 하는 사용자의 판단이 필요하다. 따라서 STT 서비스는 사용자의 판단을 대리해서 최적의 STT 결과를 도출할 수 있도록 주어진 콘텐츠의 상황을 판단하고 적절한 서비스를 실행해서 음성-문자변환 결과를 제공해야 한다.

2. STT HUB 서비스

MBC는 AI의 귀로 상징되는 STT 기술을 콘텐츠 제작 과정에 도입하고 있다. 2019년부터 2년간 시험서비스와 기술 검토를 거쳐서 STT 기술이 도움되는 방송 콘텐츠 제작 업무를 확인했으며, 〈그림 4〉와 같이 음성인식기술 기반 STT HUB 서비스와 운영 모니터링 시스템을 2021년 사내 제작시스템과 연계해서 구축하고 2022년부터 본격적으로 제작 환경에 적용하고 있다. 오랜 기간의 검토와 사용자, PD, 작가와의 협업으로 구축되었음에도 사용자에게 인식 성능에 대한 높은 기대를 자제시키고 이해를 구하는 방향으로 운영되고 있다.

STT HUB 서비스 시스템은 구축과정에서 사용자의 의견을 충분히 반영하고 꾸준한 성능 업데이트로 현행 업무를 그대로 대체할 수 있도록 개발되어 기존 업무와 STT 업무를 무리없이 연결할 수 있었다. 서비스를 적용한 지 이제 1년이 되었지만 꾸준히 월평균 1,400여 시간분, 3,500여 건의 요청에 서비스를 제공하고 있으며 높은 사용자 만족도를 보이고 있다. STT HUB 서비스는 상황을 판단해서 다수의 언어를 빠르게 인식할 수 있고 여러 요청을 동시에 처리할 수 있다. 1시간분 콘텐츠 기준으로 1~2분 내 문자변환결과를 제공한다. 영상을 보거나 음향을 들어서 제작 소재를 확인해야만 하는 현재의 방송 콘텐츠 제작업무에 시간정보와 연동된 문자변환 결과를 제공함으로써 비용 절감과 업무효율성 향상에 도움을 주고 있으며 STT를 활용한 제작 범위를 넓혀가고 있다.



<그림 4> STT HUB 서비스

3. STT 기술의 활용

콘텐츠 내용을 이해하는 방법과 수단은 영상, 음성, 문자 등 여러 가지를 고려할 수 있다. 문자로 된 정보와 기록은 다양한 미디어 중 인간이 가장 쉽게 접근이 가능한 수단으로써 내용 검색 등 다양한 응용분야에 활용하고 있다. 그러나 영상에 대한 AI의 분석 능력은 특정 영역에 국한되어 있다. 음성에 대한 이해 능력도 참고할 수준은 되지만 결과에 대한 검수가 반드시 필요하다. 특히 화자 분리와 인식 등은 상대적으로 성능의 신뢰도가 낮아서 집중적인 연구개발을 통한 성능 향상이 필요하다. 향후 STT 기술에 대한 신뢰도를 높일 수 있다면 다양한 콘텐츠 메타데이터를 추출하는 업무를 비롯한 콘텐츠의 제작, 분석, 검색을 위한 주요 기술로 자리할 수 있을 것으로 기대한다.

IV. 결론

여러 전문가들은 폭발적인 AI기술의 성장 요인에 대해 하드웨어 기술의 발전과 빠른 기술 공유를 통한 AI 원천기술 성능 향상, 사회 곳곳에서 디지털 기술로 생산되고 축적된 방대한 학습데이터 구축을 꼽는다. 무엇보다 AI에 대한 사용자들의 인식 전환을 간과할 수 없다. AI기술이 집안 곳곳 생활가전에 일상적으로 적용되고 '하이엔드급 AI가전'으로 홍보되면서 사람들 사이에 나도 한번 써보고 싶은 '고급 기술'로 인식되었다. 이런 인식은 방송 미디어 산업에도 스며들고 있다.

MBC는 우수한 콘텐츠 자산을 풍부하게 보유하고 있다. 구작 콘텐츠를 사랑해 주는 시청자들을 위해서 많은 사랑을 받았던 명작 드라마와 스테디셀러를 순차적으로 고화질로 변환해서 명작 콘텐츠에 숨을 불어넣어 되살리고 있고 초고화질 영상변환으로 UHD 서비스까지 적용범위를 확장하고 있다. 한편으로는 방송제작 워크플로우에 현시점에서 적용 가능한 AI기술을 판단하고 객관적인 시선으로 적절히 활용함으로써 효율적으로 제작 업무를 보조하고 있다.

글로벌 사업자까지 가세해서 경쟁이 불가피한 미디어 환경 속에서 MBC를 비롯한 방송사들은 지금까지 축적한 방송 기술을 한층 발전시키는 한편, 각자가 스스로 쳐놓은 편견의 벽을 깨기 위한 새로운 시도와 노력 외에 또 무엇이 필요한지 더욱 치밀하게 고민해야 한다. 미디어AI 기술을 통해 콘텐츠 제작과 유통 환경을 혁신함으로써 미디어 경쟁의 바람을 헤쳐 나가길 기대한다.

참 고 문 헌

[1] ChatGPT, https://chat.openai.com

[2] David, "'ChatGPT' 와튼스쿨 MBA, 美 의사면허 시험 통과 입증," 사이언스모니터, 2023. 1. 25, http://scimonitors.com/chatgpt-와튼스쿨-mba-의사-면허-시험-통과

[3] Tiffany H. Kung, Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, and Victor Tseng, "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models," *medRxiv*, https://www.medrxiv.org/content/10.1101/2022.12.19.22283643v2

[4] Si-Hun Sung and Woo-Sung Chun, "Knowledge-based numeric open caption recognition for live sportscast," *Proc. IEEE 16th ICPR, Aug.* 2002, pp. 822~825

[5] *Youtube*, https://www.youtube.com/watch?v=mEGp8Zzsr80

[6] 스마트 스피커, *wikipedia*, https://ko.wikipedia.org/wiki/스마트_스피커

저 자 소 개

성 시 훈

- 1995년 : 경북대학교 전자공학과 학사
- 1997년 : 경북대학교 전자공학과 석사
- 2002년 : 경북대학교 전자공학과 박사
- 2000년 ~ 현재 : ㈜문화방송 콘텐츠메타데이터파트장
- 주관심분야 : 미디어AI 서비스기술, 차세대방송 기반기술, 파일기반 제작시스템 개발