

## SER 기술을 이용한 대화형 시뮬레이션 게임 제안

이강희<sup>o</sup>, 전서현<sup>\*</sup>

숭실대학교, 글로벌미디어학부<sup>o</sup>

숭실대학교, 글로벌미디어학부<sup>\*</sup>

e-mail: kanghee.lee@ssu.ac.kr, sylph196@naver.com

## A Proposal of an Interactive Simulation Game using SER (Speech Emotion Recognition) Technology

Kang-Hee Lee<sup>o</sup>, Seo-Hyun Jeon<sup>\*</sup>

Global School of Media, Soongsil University<sup>o</sup>

Global School of Media, Soongsil University<sup>\*</sup>

### ● 요약 ●

본 논문에서는 단순히 필요한 정보를 얻기 위한 수준에 그쳤던 현대의 인공지능을 SER (Speech Emotion Recognition) 기술을 이용하여 사용자와 직접적으로 대화하는 형식으로 발전시키고자 한다. 사용자의 음성 언어에서 감정을 추출하여 인공지능 분야 및 챗봇과 대화함에 있어 좀더 효과적으로 해석할 수 있도록 도움을 준다. 이것을 대화형 시뮬레이션 게임에 접목시켜 단순한 선택형 대화 방식이 아닌 구어체로 대화하며 사용자에게 높은 몰입도를 줄 수 있다.

**키워드:** SER (Speech Emotion Recognition), 대화(communication), 감정(emotion)

### I. Introduction

모바일 양산형 게임 등 새롭지 못한 게임이 끊임없이 쏟아져, 새로운 게임을 시작해도 중반을 넘어서기 전에 흥미를 잃어서 중도하차 혹은 봉인해버리는 현상, 즉 게임불감증이 증가하고 있는 지금, 인공지능 기술을 대화형 시뮬레이션 게임에 접목시켜 단순한 선택형 대화 방식이 아닌 구어체로 대화하며 사용자에게 높은 몰입도를 줄 수 있도록 한다.

데이터를 새로 추가한다[3]. 각 언어별 특성을 데이터 러닝을 통해 인식하고, 인식률을 최대치로 높이는 것에 중점을 둔다.

### II. The Proposed Scheme

#### 1. 관련 연구

관련 연구로 hcmllab의 emovoice은 SVM으로 구현되어있는 감정인식 프로그램과 Koshort가 있다. 전자는 윈도우 플랫폼에서만 작동하며 따로 부가적인 서비스는 없다[1]. 후자는 한국어 오픈소스 자연어처리로 라이브러리 모음으로 형태소분석기의 경우 여러 분석기를 같은 인터페이스로 사용하게 해주는 KoNLPy를 많이 사용한다. 다만 Mecab은 윈도우 파이썬에서 지원을 안한다는 단점이 있다[2].

#### 2. 시스템 구성

먼저 기존 SVM(Support Vector Machine, ‘기계학습알고리즘’중 하나)로 구현되었던 감정인식 오픈소스 모델을 ANN(Artificial Neural Network, 인공신경망)으로 수정하고 한국어에 대한 인식

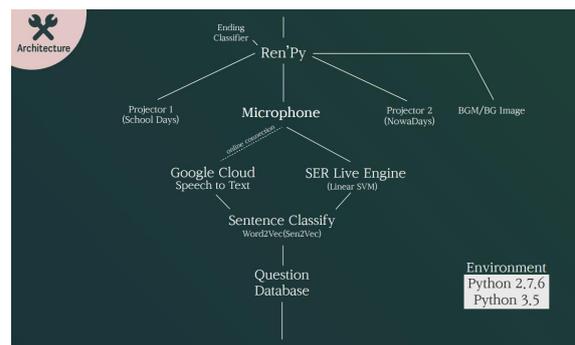


Fig. 1. System Architecture

원활한 프로그래밍을 위해, 파이썬을 기반으로 제작된 Ren'Py를 사용하여 게임의 전체적인 사나리오 및 스크립트를 제작하고 음성을 인식할 수 있는 헤드폰, 혹은 마이크로 사용자의 음성을 인식한다. 구글 클라우드 오픈 소스를 통해 사용자의 음성을 텍스트로 변환하고, 그와 동시에 SER Live Engine을 이용하여 음성의 높낮이 변화를 이용해 감정을 인식한다. 인식된 감정은 총 보통, 기쁨, 슬픔, 화남, 지루함, 놀람 등 총 6개의 감정으로 분리된다. 두 가지 결과치를

추합하여 좀 더 정확한 결과값을 낼 수 있도록 사용자가 말한 문장을 Word2Vec을 이용해 형태소 별로 쪼개어 긍정과 부정, 평이문과 의문문 등 정확한 해석을 할 수 있도록 돕는다[4]. 그 결과에 따라서 입력해둔 사나리오가 어느 방향으로 진행될 것인지 결정한다. 즉, 사용자의 대화 내용 뿐만아니라 감정 상태까지 고려하여 각 사용자마다 각기 다른 스토리 라인으로 게임을 진행해 무수한 타임라인을 가질 수 있다.

### III. Conclusions

단순히 필요한 정보를 얻기 위한 수준에 그쳤던 현대의 인공지능을 SER (Speech Emotion Recognition) 기술을 이용하여 사용자와 직접적으로 대화하는 형식으로 발전시키고자 한다. 사용자의 음성 언어에서 감정을 추출하여 인공지능 분야 및 챗봇과 대화함에 있어 좀더 효과적으로 해석할 수 있도록 도움을 준다. 이것을 대화형 시뮬레이션 게임에 접목시켜 단순한 선택형 대화 방식이 아닌 구어체로 대화하며 사용자에게 높은 몰입도를 줄 수 있다. 이 외에도 인공지능의 감정 인식은 다양한 활용 방안을 줄 수 있다.

- VR 시뮬레이션 게임 : 사용자가 게임 내부 캐릭터(NPC 등)와 대화할 경우, 실제 음성에서 느껴지는 플레이어의 감정이 게임의 진행도, 루트, 엔딩에 실시간으로 영향을 미친다. (ex) 심리/추리, 연애 등 다양한 게임 장르에 적용 가능)
- 면대면식 커뮤니케이션을 중요시하는 직종 대상으로 한 교육용 시뮬레이션을 제작한다. (ex)서비스 직종의 톤 교정, 클레임에 대처하는 롤플레이팅 등)
- 일대다 가상 교육플랫폼에 응용 가능하다.

SER (Speech Emotion Recognition) 기술은 미래의 XR (Extended Reality)에도 꼭 필요한 기술이며, 대화형 선택 및 Chatbot 기술 UX의 마지막 과제일 것이다. 더 완벽해지고 더 빨라진다면 아무런 괴리감없이 컴퓨터와 대화할 수 있는 세상이 올 것이다.

### ACKNOWLEDGEMENT

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2017R1D1A1B05036028). 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5A2A03927522)

### REFERENCES

[1] Build your own Real-time Speech Emotion Recognizer, available: <https://github.com/hcmlab/emovoice>

[2] Koshort, available: <https://koshort.github.io/>  
 [3] Björn Schuller, Gerhard Rigoll, and Manfred Lang, "Hidden Markov Model-based Speech Emotion Recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.  
 [4] Wrod2Vec using Tensorflow, available: <https://tensorflow-korea.gitbooks.io/tensorflow-kr/content/g3doc/tutorials/wrod2vec/>