

<https://doi.org/10.7236/JIIBC.2024.24.4.9>  
JIIBC 2024-4-2

# 고령화 사회를 위한 음성 인식 챗봇 시스템 : 기술 개발과 맞춤형 UI/UX 설계

## Voice Recognition Chatbot System for an Aging Society: Technology Development and Customized UI/UX Design

정윤지\*, 유민성\*, 오주영\*, 황현석\*, 허원희\*\*

Yun-Ji Jeong\*, Min-Seong Yu\*, Joo-Young Oh\*,  
Hyeon-Seok Hwang\*, Won-Whoi Hun\*\*

**요약** 본 연구는 고령화 사회의 노년층 우울증과 고독감 문제를 해결하기 위해 음성 인식 챗봇 시스템을 개발하였다. 이 시스템은 Whisper 모델, GPT 2.5, XTTS2를 활용하여 고성능 음성 인식과 자연어 처리, 텍스트-음성 변환 기능을 제공한다. 사용자는 이를 통해 감정과 상태를 표현하고 적절한 반응을 얻을 수 있으며, 지인의 목소리를 이용한 음성 인식 기능으로 친숙함과 안정감을 느낄 수 있다. UX/UI는 스마트 시니어 세대의 인지 반응과 시력 저하, 운동 능력 제약 등을 고려하여 설계되었다. 명도와 선명도가 높은 색상, 가독성이 좋은 서체등을 활용하여 고령자의 사용 편의성을 높였다. 이 연구는 음성 기반 인터페이스를 통해 노년층의 삶의 질 향상에 기여할 것으로 기대된다.

**Abstract** This study developed a voice recognition chatbot system to address depression and loneliness among the elderly in an aging society. The system utilizes the Whisper model, GPT 2.5, and XTTS2 to provide high-performance voice recognition, natural language processing, and text-to-speech conversion. Users can express their emotions and states and receive appropriate responses, with voice recognition functionality using familiar voices for comfort and reassurance. The UX/UI design considers the cognitive responses, visual impairments, and physical limitations of the smart senior generation, using high contrast colors and readable fonts for enhanced usability. This research is expected to improve the quality of life for the elderly through voice-based interfaces.

**Key Words** : Aging Society, Chatbot System, Emotional Support, Social Connections, Voice-Based Interface, Voice Recognition

### 1. 서론

질병 예방의 강화와 진보된 치료 기술, 향상된 건강관리 시스템 등으로 인해 의료 기술이 빠르게 발전하면서

인류의 수명이 연장되고 있다. 그러나 이는 아이러니하게도 노인 인구의 급격한 증가를 초래하고 있다. 이러한 현상은 우리나라뿐만 아니라 전 세계적으로도 나타나고 있다. 통계청 자료에 따르면, 65세 이상 고령인구는

\*학생회원, 성결대학교 미디어소프트웨어학과

\*\*정회원, 성결대학교 미디어소프트웨어학과(교신저자)

접수일자 2024년 6월 9일, 수정완료 2024년 7월 9일

게재확정일자 2024년 8월 9일

Received: 9 June, 2024 / Revised: 9 July, 2024 /

Accepted: 9 August, 2024

\*Corresponding Author: wonwhoi@daum.net

Dept. Media Software, Sungkyul University, Korea



간으로 텍스트로 변환하며, 변환된 텍스트는 GPT 2.5를 통해 이해되고 적절한 응답이 생성된다. 이후 XTTS2 모델이 이 텍스트를 다시 자연스러운 음성으로 변환하여 사용자에게 전달한다. 이 시스템은 HTTP 통신을 통해 실시간으로 사용자의 음성 데이터를 처리하며, 토큰(Token)을 사용하여 모델 간의 효율적인 데이터 전송을 실현한다. 이를 통해 사용자는 지인의 목소리로 대화하는 듯한 경험을 할 수 있다.

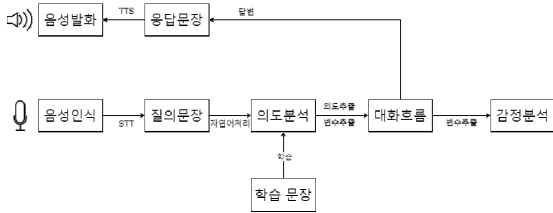


그림 2. 음성 챗봇 순서도  
 Fig. 2. Voice chat-bot flowchart

#### 가. 음성-텍스트 변환(Speech-to-Text, STT)을 통한 음성 인식 (Speech-to-Text, STT)

본 시스템의 음성 인식은 OpenAI에서 개발한 Whisper 모델을 사용한다. 이 모델은 680,000시간 이상의 다국어, 다방언 오디오 데이터를 학습하였으며, 이로 인해 다양한 음성 변형 및 억양에 대한 적응력이 높다. 또한, 잡음 제거 기능을 포함하고 있어, 배경 소음이 있는 환경에서도 정확한 음성 인식이 가능하다. 본 시스템은 Whisper 모델을 통해 음성 입력을 실시간으로 텍스트로 변환하고, 변환된 텍스트는 이후의 자연어 처리 단계로 전달되어 사용자와의 원활한 상호작용이 가능하다.

#### 나. 대화 시스템(Chatbot)을 통한 질의 문장의 의도 분석

대화 생성은 OpenAI의 GPT 아키텍처를 기반으로 한 GPT 2.5 모델을 사용한다. GPT 2.5는 속도가 빠르면서도 다양한 상황에 대한 대응이 가능한 모델로, 본 연구에서는 웰니스 데이터셋을 기반으로 학습하였다. 웰니스 데이터셋은 강남 세브란스에서 전달받은 상담 데이터를 기준으로 이를 통해 문장별 분리 및 대화 의도를 분류하여 분류별로 유저-챗봇 대화 스크립트를 생성했다. 이를 통해 사용자의 질문이나 요청에 대해 의미 있는 답변을 생성할 수 있다. 모델은 모든 말을 숫자로 변경하여 학습하며, 이 과정에서 코버트(Token)를 사용하여 텍스트를 효과적으로 처리할 수 있어 사용자의 입력에 대한 실시간 응답이 가능하다.

#### 다. 대화 흐름 분석을 통한 감정 분석

의도를 분석하는 과정에서 텍스트 변환 결과를 바탕으로 사용자의 의도를 파악한다. 이 과정에서 자연어 처리(NLP) 기술을 활용하여 사용자의 질문이나 명령을 이해한다. 그리고 변수 추출의 단계를 통해 대화의 텍스트 내용과 음성 특성에서 감정 분석을 위한 변수를 추출한다. 이는 단어 선택, 문장 구성, 음성의 톤 등을 포함할 수 있다. 마지막으로 감정 분석의 단계에서 추출된 변수를 감정 분석 모델에 입력하여 사용자의 감정 상태를 분석한다. 이 모델은 머신 러닝이나 딥러닝 기술을 활용하여 구축할 수 있으며, 훈련 데이터에 따라 다양한 감정(기쁨, 슬픔, 분노 등)을 식별할 수 있다.

#### 라. 텍스트-음성 변환(Text-to-Speech, TTS)을 통한 응답 문장 음성 발화

텍스트-음성 변환은 XTTS2 모델을 사용한다. XTTS2는 PyTorch 프레임워크를 사용하여 구현되었으며, Hugging Face와 Coqui 라이브러리를 통해 개발된 최신 TTS 모델이다. 이 모델은 VITS(Variational Inference Text-to-Speech)의 인코더-디코더 동시 학습 방식을 기반으로 텍스트 입력을 음성으로 변환하는 과정을 최적화한다. 인코더는 입력 텍스트의 특징을 추출하고, 디코더는 이를 바탕으로 자연스럽게 유창한 음성을 생성한다. 특히, XTTS2는 레퍼런스 음성(reference speech)에서 음의 정보(prosodic information)를 추출하여 이를 음성 생성 과정에 반영하는 기능을 추가하였다. 이를 통해 생성된 음성은 단순히 텍스트를 읽는 것을 넘어, 발화자의 감정이나 의도 등을 반영하여 더욱 생동감 있는 음성을 제공한다.

#### 2. 음성 인식 챗봇 시스템 애플리케이션 UX/UI

본 연구의 음성 인식 챗봇 시스템을 기능으로 탑재한 애플리케이션의 UX 및 UI를 제작하기 위해 고려대학교에서 주관하고 을지대학교, 온더아이티, 이니션, 테크빌 교육에서 연구한 ‘스마트 시니어 세대의 문화 향유를 위한 인지 반응 맞춤형 UI/UX 기술 개발’을 참고하였다.<sup>[5]</sup> 이 연구는 증가하는 스마트 시니어 세대에게 공평하고 원활한 문화 향유 서비스를 제공하기 위해 진행되었다. 이를 위해 스마트 시니어 세대의 인지 반응 요소를 다각도로 측정하고 평가하는 기술을 개발하고, 이들의 성향과 행동을 분석하여 맞춤형 콘텐츠를 제공하는 것을 목표로 설정하였다.<sup>[5]</sup>

색상은 단순히 정보를 전달하는 도구로 사용되는 것이 아니라, 사람들이 색을 인지하는 능력과 제한 사항을 고려하여 신중하게 선택되어야 한다. 고령화가 진행됨에 따라 시각적 변화로 인해 스펙트럼의 하위 범주보다는 상위 범주에 있는 색상을 더 잘 구별하게 되며, 명도가 높고 선명한 노란색이나 빨간색이 더 잘 인식된다. 따라서 주황(#FF8C00)을 애플리케이션의 주요 색상으로 선택하였고 주황의 보색인 초록(#4BC8AA)을 보조 색상으로 선택하여 주요 색상과 대비되어 정보를 구별하기 편할 수 있게 하였다.

고령자들은 노화로 인해 시력이 감퇴하므로 이를 고려하여 글자의 크기를 적용해야 한다. 일반인보다 1~2pt 정도 커야 쉽게 읽히는 것으로 나타나 제목의 경우 18~24pt를 권장하고, 본문의 경우 12~14pt를 사용하도록 권장된다.<sup>[6]</sup> 따라서 해당 애플리케이션의 글자의 크기는 해당 권장 사항을 준수하여 개발하였다. 또한, 글자의 서체는 가독성이 좋은 산세리프형 서체를 사용하였다. 이처럼 선행된 연구 결과를 참고하여 애플리케이션의 UX/UI를 제작하였다.

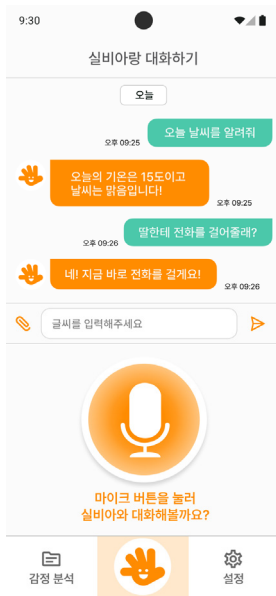


그림 3. 사용자 음성 수집 및 응답 문장 음성 발화 홈 화면  
Fig. 3. User voice collection and response sentence voice utterance home screen

그림 3은 해당 애플리케이션의 홈 화면으로, 사용자와 챗봇 간의 대화를 시각적으로 확인할 수 있다. 사용자가 가운데에 있는 마이크 버튼을 누르고 말하면, 음성이 텍

스트로 변환(Speech-to-Text, STT)되어 화면의 오른쪽 말풍선에 표시된다. 이뿐만 아니라, 사용자는 직접 텍스트를 입력하여 챗봇을 이용할 수도 있다. 이후 챗봇이 사용자의 음성을 분석하여 응답 문장을 화면에 표시하고, 사용자가 설정한 목소리로 음성 응답을 들을 수 있다. 또한, 사용자가 필요한 기능을 홈 화면에서 모두 찾을 수 있도록 내비게이션 바를 통해 감정 분석 기능과 설정 버튼을 배치하였다.

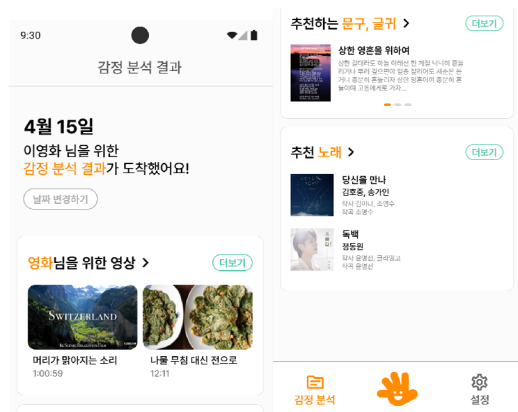


그림 4. 감정 분석 결과에 의한 피드백  
Fig. 4. Feedback based on sentiment analysis results

감정 분석 화면에서는 사용자가 챗봇을 이용하는 과정에서 인공지능이 분석한 감정을 통계적으로 시각화하여 제공한다. 사용자는 이러한 통계 자료를 통해 자신의 감정 상태를 객관적으로 이해하고, 그 변화를 쉽게 파악할 수 있다.

또한, 본 애플리케이션은 분석된 감정을 기반으로 사용자 맞춤형 콘텐츠를 추천하는 기능을 갖추고 있다. 사용자의 현재 감정 상태에 맞춰 적절한 영상, 글귀, 노래 등을 추천하여 사용자가 자신의 감정을 보다 긍정적으로 관리할 수 있도록 돕는다. 예를 들어, 스트레스를 많이 받은 날에는 마음을 진정시키는 음악을, 기분이 우울한 날에는 격려의 메시지를 담은 영상을 제공하는 방식이다. 이를 통해 사용자는 개인의 감정 상태에 맞춘 맞춤형 콘텐츠를 활용하여 감정 조절 및 개선에 도움을 받을 수 있다.

따라서, 이 시스템은 사용자에게 그날의 감정을 객관적으로 파악할 수 있는 도구를 제공할 뿐만 아니라, 감정 상태에 따른 맞춤형 콘텐츠를 통해 정서적 안정을 도모하는 데 기여한다.

### 3. 음성 인식 챗봇 시스템의 성능 및 유용성 평가를 위한 추가 연구 방향

해당 연구는 고령화 사회에서 노년층의 정서적 지원과 소통을 위한 음성 인식 챗봇 시스템의 개발을 목표로 하였다. 이를 위해 OpenAI의 Whisper 모델, GPT 2.5, XTTS2를 활용하여 고성능의 음성 인식, 자연어 처리, 텍스트-음성 변환 기능을 제공하는 시스템을 구축하였다. 본 연구에서 제안하는 챗봇 시스템은 노년층 사용자의 정서적 안정과 사회적 연결감 강화에 기여할 가능성을 제시한다. 이 연구를 통해 개발된 시스템의 성능과 유용성을 평가하기 위해선 후속 연구가 필요하며, 이를 위해 다음과 같은 내용을 연구에 포함할 것을 제안한다.

#### 가. 사용자 경험 및 만족도 조사

노년층 사용자가 실제로 본 챗봇 시스템을 사용했을 때의 경험과 만족도를 조사하여 시스템의 유용성과 사용자 친화성을 평가한다. 사용자의 피드백을 바탕으로 시스템의 인터페이스와 기능을 개선할 수 있는 방향을 제시한다.

#### 나. 다양한 언어 및 방언에 대한 적응성 평가

Whisper 모델은 다양한 언어와 방언에서도 높은 정확도로 음성을 인식할 수 있는 능력을 갖추고 있다. 이에 대한 평가를 통해 시스템이 다양한 문화적 배경을 가진 사용자들에게도 효과적으로 서비스를 제공할 수 있는지 검증한다.

#### 다. 정서 분석 기능의 추가 및 평가

사용자의 음성이나 대화 내용에서 정서적 상태를 분석할 수 있는 기능을 추가하여, 사용자의 감정 변화를 파악하고 이에 적절히 대응하는 챗봇의 능력을 평가한다. 정서 분석을 통해 사용자가 우울감이나 고독감을 느낄 때 적극적으로 지원할 방안을 모색한다.

#### 라. 장기적 사용에 따른 효과 분석

시스템을 일정 기간 지속해서 사용한 사용자들을 대상으로 장기적인 효과를 분석한다. 이를 통해 챗봇 시스템이 노년층의 정서 안정, 사회적 연결감 강화, 우울증 완화 등에 실질적으로 기여할 수 있는지 평가한다.

#### 마. 보안 및 개인정보 보호 문제 해결 방안 연구

음성 인식 및 처리 과정에서 사용자의 개인정보와 데

이터 보안을 확보하는 것은 매우 중요하다. 사용자의 정보를 안전하게 보호하면서도 효과적인 서비스를 제공할 수 있는 기술적, 정책적 해결 방안을 연구한다.

#### 바. 다양한 응용 분야에서의 활용 가능성 탐색

건강관리, 사회적 소통 활성화 외에도 본 시스템이 적용될 수 있는 다양한 분야를 탐색하고, 해당 분야에서의 구체적인 활용 사례와 효과를 분석한다.

이러한 후속 연구를 통해 본 연구의 챗봇 시스템이 노년층의 삶의 질 향상에 더욱 기여할 방안을 모색하고, 음성 기반 인터페이스 기술의 발전과 사회적 적용 가능성을 더욱 확장할 수 있을 것이다.

## III. 결 론

노령화 사회에서는 노년층의 정서적 지원과 소통이 중요한 이슈로 주목받고 있다. 우울증이나 고독감 등을 겪는 노인 인구가 증가함에 따라, 이들에게 사회적 연결감을 제공하고 정서적 안정을 도모할 수 있는 기술의 필요성이 더욱 커지고 있다. 본 연구에서 개발된 음성 인식 챗봇 시스템은 이러한 문제에 대응하기 위해 설계되었다. 이 시스템은 OpenAI의 Whisper 모델, GPT 2.5, XTTS2를 활용하여 고성능의 음성 인식, 자연어 처리, 텍스트-음성 변환 기능을 제공한다. 이를 통해 사용자는 마치 사람과 대화하듯이 음성으로 챗봇과 소통할 수 있으며, 이 과정에서 자신의 감정이나 상태를 표현하고 적절한 반응을 얻을 수 있다. 특히, 지인의 목소리를 이용한 음성 인식 기능은 사용자에게 친숙함과 안정감을 제공할 수 있으며, 이는 정서적 소통에 긍정적인 영향을 미칠 것으로 기대된다. 또한, 기술적인 측면에서 볼 때, 본 연구에서 사용된 Whisper 모델, GPT 2.5, XTTS2는 각각의 분야에서 최신 기술을 대표한다. Whisper 모델은 다양한 언어와 방언, 소음이 섞인 환경에서도 높은 정확도로 음성을 인식할 수 있으며, GPT 2.5는 사용자의 의도를 정확히 이해하고 적절한 응답을 생성하는 능력을 갖추고 있다. XTTS2는 생성된 텍스트를 자연스러운 음성으로 변환하여 사용자에게 전달하는 기능을 제공한다. 이러한 기술의 통합은 사용자에게 더욱 향상된 경험을 제공하며, 음성 챗봇 기술의 발전에 기여하고 있다.

또한, 음성 인식 챗봇 시스템이 탑재된 애플리케이션의 UX/UI를 개선하기 위해 스마트 시니어 세대의 인지 반응

을 고려한 UI/UX 기술을 참고하였다. 애플리케이션의 주요 색상으로 명도와 선명도가 높은 주황(#FF8C00)과 보조 색상으로 초록(#4BC8AA)을 선택하여 가시성을 높였다. 또한, 고령자의 시력을 고려하여 글자 크기를 18~24pt(제목)와 12~14pt(본문)로 설정하고, 가독성이 좋은 산세리프 서체를 사용하였다. 음성 수집 및 응답 기능을 포함한 홈 화면과 감정 분석 결과를 시각화하여 제공하는 기능을 통해 사용자는 자신의 감정 상태를 쉽게 파악하고, 맞춤형 콘텐츠를 통해 정서적 안정을 도모할 수 있도록 설계하였다. 이러한 접근은 스마트 시니어 세대에게 공평하고 원활한 문화 향유 서비스를 제공하는 데 기여한다.

결론적으로, 본 연구는 노년층의 정서적 지원과 소통을 위한 음성 인식 챗봇 시스템의 개발을 통해 음성 기반 인터페이스의 가능성을 탐색하였다. 이는 기술이 인간의 삶에 긍정적인 영향을 미칠 수 있음을 보여주며, 향후 더욱 진보된 자연어 이해 및 음성 처리 기술의 개발을 위한 기반을 마련한다. 본 연구의 결과는 고령화 사회에서 노년층의 삶의 질을 향상하는 데 기여할 것으로 기대된다.

## References

- [1] Statistics Korea. (2022). Population Projections 2020.
- [2] Kee, B., "Depression in the elderly", Psychiatry, Geriatrics, Chung-Ang Medical University Hospital, Vol. 3, No. 3, pp. 61-71, 1999.
- [3] Kim, K.-O. & Lim, J.-H., "Construction of a Structural Model about the Perceived Health Status, Pain, Sleep Pattern and Quality of Life in the Elderly", Journal of the Korea Academia-Industrial Cooperation Society, Vol. 15, No. 7, pp. 4336-4345, 2014. DOI: <https://doi.org/10.5762/KAIS.2014.15.7.4336>.
- [4] Yun-Ji Jeong, Min-Seong Yu, Joo-Young Oh, Hyeon-Seok Hwang, Won-Whoi Hun, "Case Analysis for the Emotional Communication Contents of New Silver Generation", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 24, No. 2, pp. 23-28, April 2024.
- [5] Heesuk Im, Deokyoung Jung, Soonwoo Choi, Yeonkyung Ahn, Dongseok Ryu, Hyeoseong Ji, "Development of Customized UI/UX Technology for Cognitive Responses for the Smart Senior Generation's Cultural Enjoyment", Ministry of Culture, Sports and Tourism, Korea University, Mar. 2016. TRKO201700000054. DOI: <https://doi.org/10.23000/TRKO201700000054>.
- [6] Hyeonji Geum, "A Study on Improving GUI Design of Mobile Healthcare Applications for the Elderly: Focusing on Android-based Esencare Application", Master's thesis, Korea University of Technology and Education, 2021.

## 저 자 소 개

### 정 윤 지(학생회원)



- 2020년 3월 ~ 현재 : 성결대학교 미디어소프트웨어학과 학생
- 관심분야 : App/Web Mobile, IT, 디자인

### 유 민 성(학생회원)



- 2020년 3월 ~ 현재 : 성결대학교 미디어소프트웨어학과 학생
- 관심분야 : 인공지능, IT

### 오 주 영(학생회원)



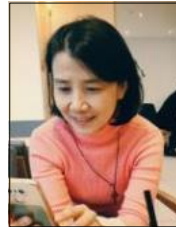
- 2019년 3월 ~ 현재 : 성결대학교 미디어소프트웨어학과 학생
- 관심분야 : 모바일, IT, 디자인

### 황 현 석(학생회원)



- 2019년 3월 ~ 현재 : 성결대학교 미디어소프트웨어학과 학생
- 관심분야 : 인공지능, 통신, 서버

### 허 원 희(정회원)



- 1993년 2월 : 국민대학교 전자공학과
- 1997년 5월 : Pratt Institute Computer Graphics(MFA)
- 2012년 8월 : 서울과학기술대학교 디지털콘텐츠디자인전공 (디자인학박사)
- 2004년 3월 ~ 현재 : 성결대학교 미디어소프트웨어학과 교수
- 관심분야 : 3D, 모바일, IT, 콘텐츠디자인