디지털산업정보학회 논문지 제20권 제2호 - 2024년 6월
http://dx.doi.org/10.17662/ksdim.2024.20.2.027

눈동자 추적 기반 입력 및 딥러닝 기반 음성 합성을 적용한 루게릭 환자 의사소통 지원 시스템*

박 현 주** · 정 승 도***

Communication Support System for ALS Patient Based on Text Input Interface Using Eye Tracking and Deep Learning Based Sound Synthesis

Park Hyunjoo · Jeong Seungdo

〈Abstract〉

Accidents or disease can lead to acquired voice dysphonia. In this case, we propose a new input interface based on eye movements to facilitate communication for patients. Unlike the existing method that presents the English alphabet as it is, we reorganized the layout of the alphabet to support the Korean alphabet and designed it so that patients can enter words by themselves using only eye movements, gaze, and blinking. The proposed interface not only reduces fatigue by minimizing eye movements, but also allows for easy and quick input through an intuitive arrangement. For natural communication, we also implemented a system that allows patients who are unable to speak to communicate with their own voice. The system works by tracking eye movements to record what the patient is trying to say, then using Glow-TTS and Multi-band MelGAN to reconstruct their own voice using the learned voice to output sound.

Key Words : MelGAN, Multiband-MelGAN, ALS, Voice Synthesis, Eye Tracking, TTS, Deep Learning, Communication System

Ⅰ. 서론

말은 본인의 의사를 전달하는 가장 기본적인 소통의 수단으로 누구나 자연스럽게 할 수 있는 것으로 여겨진다. 그러나 사고나 질병으로 인하여 말을 자유롭게 하지 못하는 사람들 또한 존재한다. 대표적으로 루게릭병이라고도 불리는 근위축성 측색경화증(ALS)은 퇴행성 신경질환으로 초기증상은 매우 미미하나 점차적으로 근육이 약화되고 목소리를 내기 위한 근육에도 마비가 진행되어 의사소통이 어려워지는 현상까지 발생한다[1]. 루게릭병뿐만 아니라 다른 이유에 의해 의사소통이 어려운 환자

* 이 논문은 2023학년도 상명대학교 교내연구비를 지원받아 수행 하였음.
** 상명대학교 스마트정보통신공학과 부교수(주저자)
*** 상명대학교 스마트정보통신공학과 부교수(교신저자)

들을 지원하기 위한 방법이 다양하게 연구되고 있다. 상용화가 이루어진 의사소통 시스템 중 의사소통 글자판은 핵심 단어를 먼저 종이에 적어 가리키면 환자가 간단한 동작을 취해 대화가 이어져 나가는 방법으로 되어있다. 시스템에 사용된 한글 입력 키보드는 컴퓨터의 자판 배열을 따르므로 왼쪽의 자음, 오른쪽의 모음 배치로 글자를 완성한다. 하지만 이 시스템 이용 시 눈동자의 많은 움직임으로 인해 정확도가 떨어지거나 눈의 피로도가 증가하여 장시간, 많은 양의 의사 소통에는 문제가 있다[2-4]. 이는 움직이기 힘든 환자가 이용하고 있는 시스템으로서는 실효성 측면에서도 한계점이 있는 것으로 보인다.

말과 거동이 자유롭지 못한 환자들의 의사소통 지원 체계는 글자판, 대화 그림, 터치패드, 안구 마우스 등을 기반으로 하는 다양한 연구가 존재한다. 기본적인 소통 지원 체계로는 글자판과 대화 그림 등을 예로 할 수 있는데, 이는 원하는 글자에서 눈을 깜박여 소통하는 것을 기본으로 한다.

〈그림 1〉 의사소통 글자판

삼성전자가 개발한 EYECAN+ 안구 마우스는 원하는 글씨로 눈을 움직이면 컴퓨터 화면에 글이 작성되어 의사소통을 가능하게 한다. 다른 도구에 비해 환자의 생각을 글로 더 잘 표현할 수 있으나 가격이 고가라는 단점이 있다. 또한 안구 마우스 훈련 소프트웨어를 이용하여 사용법을 익히는 과정도 복잡할 뿐만 아니라 추가적인 비용이 소요된다.

본 연구에서는 이를 개선하기 위해 웹캠 기반의 시선 추적을 이용한 방법을 제안하여 직접적인 발성에 의한 의사소통이 어려운 환자를 지원하기 위한 시스템을 제안한다. 환자 스스로 눈으로 원하는 단어를 입력할 수 있는 한글 입력 인터페이스를 구현하였다[5]. 이는 언어장애나 루게릭병과 같이 근력 약화 증상을 겪고 있는 환자들의 의사소통을 지원하는 시스템에서 편리한 입력을 위한 효과적인 인터페이스로 활용될 수 있다[6].

위축성 측색경화증(ALS) 환자인 경우 성대 근육이 가장 늦게 마비 되어 환자가 목소리를 완전히 잃기 전에 정해진 문장을 녹음한 음성을 보관하여 재생하는 것이 아닌, 환자가 표현하고자 하는 문장을 복원된 개인의 목소리를 통해 의사소통 되도록 하고자 한다[7].

이를 실현시키기 위해서 Python환경에서 OpenCV와 Deep Learning을 사용하였다. 또한 움직이기 힘든 환자들도 눈을 이용해 쉽게 사용할 수 있도록 단어나 문장을 입력할 수 있는 입력 UI(User Interface)를 구현하였으며, 이 UI를 통해 본 연구의 최종 목적인 개인화 TTS기반 위축성 측색경화증(ALS) 환자를 위한 의사소통 시스템을 개발하였다[8].

Ⅱ. 관련 연구

2.1 의사소통 지원 하드웨어

의사소통 지원체계에는 글자판, 대화그림, 터치패드, 안구 마우스가 있다.

〈그림 2〉 안구 마우스 사용사진

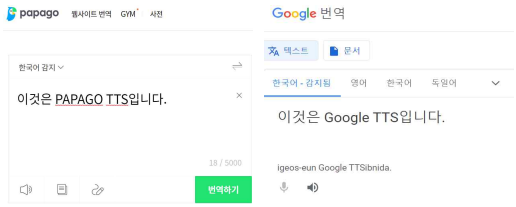
이 중에서 안구 마우스는 원하는 글씨로 눈을 움직이면 컴퓨터 화면에 글이 작성되어 의사소통을 가능하게 한다[3]. <그림 2>는 삼성전자가 개발한 EYECAN+ 안구 마우스를 실제로 사용하고 있는 사진이다. 다른 도구에 비해 환자의 생각을 글로 더 잘 표현할 수 있으나 가장 큰 문제점은 안구 마우스의 가격이 고가라는 점이다. 안구 마우스의 가격은 환자용이 아닌 Gaming 용 안구 마우스를 포함하여 수 십만 원에서 많게는 천만 원대까지의 가격을 호가하기 때문에 환자들에게 비용 부담을 가중 시킨다.

2.2 의사 소통 지원 소프트웨어

다음 그림 <그림 2>는 안구 마우스 훈련용 소프트웨어이다. 안구 마우스는 비싼 가격에 더불어 기기 사용법을 익히는 데에 긴 시간이 걸리고 추가 비용이 든다.

〈그림 3〉 안구 마우스 훈련용 소프트웨어

그러나 본 연구에서는 laptop의 내장된 카메라를 사용하기 때문에 낮은 비용이 사용되므로 가격에서 많은 이득을 얻을 수 있다. 또한 아이트래킹으로 글씨를 선택하면 출력할 수 있지만 개인화 음성 합성 기술이 결합되어 있지 않아 사용자의 음성 특성을 나타내지 못한다.



〈그림 4〉 Naver TTS, Google TTS 사진

<그림 4>는 Naver의 PAPAGO TTS와 Google TTS를 나타낸다[8]. 기존의 TTS 기능은 Google 번역, Naver의 PAPAGO 등과 같이 정해진 음성으로 텍스트를 재생시키는 데에 그친다. 하지만 본 연구에서는 개인화 TTS를 통해 환자 본인의 목소리를 복원하여 입력한 텍스트를 재생시키기 때문에 본인의 목소리로 대화하는 듯한 시스템을 제안하였다.

Ⅲ. 본론

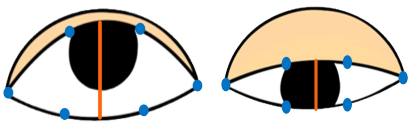
3.1 시선 추적

본 연구에서 제안하는 인터페이스는 안구의 움직임 추적에 기반하고 있다[2-4]. 이를 위해 얼굴을 포함하는 웹캠 입력에 대해서 Dlib 68 안면 랜드마크 예측 알고리즘을 적용하여 얼굴상의 특징점 68개를 추출한다. 이 중에서 눈에 해당하는 특징점만을 선별하여 추출하고 흰자 비율, 눈의 수직, 수평 길이의 비율 등을 계산하여 상하좌우, 깜빡임을 판별하여 글자 입력을 위한 동작으로 활용하였다.

〈그림 5〉 눈의 흰자 비율로 좌 / 우 응시 판별

<그림 5>는 좌우 응시를 판별하는 방식을 보여주고 있다. 흰자와 검은자를 구분하고 중앙을 기준으로 치우침 정도를 정의하여 응시 방향을 결정한다.

<그림 6>은 상하 응시를 구분하는 방법을 보인 것으로 상하 응시의 경우 좌우 응시와 같은 방식으로 적용하면 판별 오류가 많아진다. 여기에서는 하단을 응시하는 경우 눈꺼풀이 내려가는 점을 이용하기 위하여 중앙점을 기준으로 상하 간의 거리를 추가로 사용하였다.

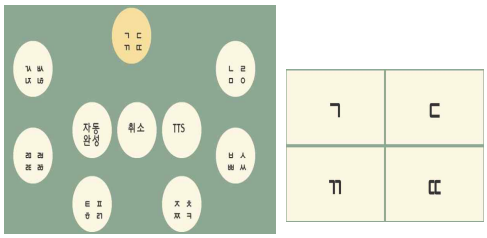


〈그림 6〉 눈꺼풀의 길이 비율로 상 / 하 응시 판별

글자를 선택할 때 사용하는 깜빡임은 눈의 수직, 수평 길이를 구하고 두 길이의 비율을 계산하여 판별한다. 일상적인 눈 깜빡임과 글자를 선택할 때의 깜빡임을 구분하기 위해 일정 시간(약 1초) 눈을 감는 것을 기준으로 하였다.

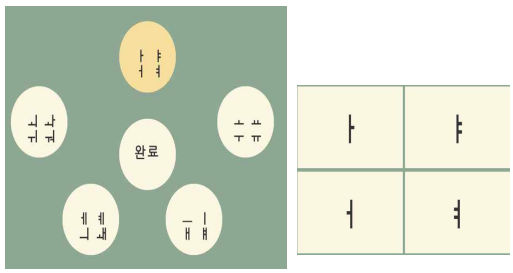
3.2 한글 입력을 위한 자판 배치 개선

기존의 자판 배치를 그대로 사용하는 경우 사용자 눈의 움직임이 커서 글자 선택에 어려움을 겪거나 눈의 피로도를 증가시키는 단점이 있다. 새로운 입력 키 보드는 기존 키보드보다 직관성을 높여 한눈에 알아보기 쉽도록 구성하였다. 한글 지원을 위하여 자음과 모음을 분석하여 그룹별로 순환되는 형식의 자판 배치를 한다. 제안하는 자판의 자음 배치는 <그림 7>과 같다.



〈그림 7〉 자음의 배치

<그림 7>에서 보이는 바와 같이 외부에 원형으로 배치된 자음 그룹이 순차적으로 이동되며 원하는 자음이 있는 경우 눈을 깜빡여서 선택한다. 선택되고 난 이후에는 <그림 7>의 오른쪽에 보이는 바와 같이 4개의 자음 배치가 나타나며 여기에서 상하좌우 눈동자 움직임을 이용하여 원하는 자음을 선택할 수 있도록 구성하였다. 다음으로 모음을 선택할 수 있는 자판 배치가 보이게 된다.



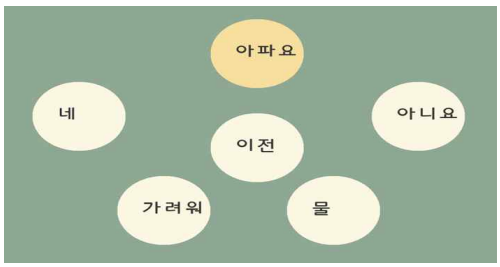
〈그림 8〉 모음의 배치

이때의 모음 자판 배치는 <그림 8>와 같다. 그림에서와 같이 그룹별로 순환하는 과정에서 깜빡임으로 선택하고 그룹 내 선택은 안구 상하좌우의 움직임을 이용하도록 하였다.

한글의 경우 초성, 중성, 종성으로 이루어져 눈으로 글자를 입력할 시 모든 글자가 각각 입력되는 문제가 있다. 이러한 문제를 해결하기 위해 초성, 중성, 종성에 따라 한 글자를 선택한 후에 리스트에 순차적으로 넣고 글자가 완성될 때 자모 결합 알고리즘을 적용하여 결합하였다.

3.3 입력 편의를 위한 단어 구성

본 시스템은 루게릭병 환자들을 위해 고안한 장치이므로 의학적 용어 또는 몸 상태를 표현하기 위해 쓰이거나 병원 치료 관련 단어를 조사하였다. 자주 사용하는 용어의 표현을 쉽게 하려고 조사된 단어에 사용된 자음, 모음의 빈도와 기본적인 일상적 표현에 쓰이는 단어의 자음, 모음의 빈도를 적절히 혼합하여 배치한 입력 인터페이스를 새롭게 제안한다.



〈그림 9〉 입력 인터페이스

3.4 개인화를 위한 음성 합성

환자와 대화하는 보호자 또는 청취자 입장은 환자의 목소리를 그대로 들을 수 있는 것을 더 좋아할 수 있다. 제안하는 TTS 기술은 환자의 음성을 보존하여 환자가 시선 추적 입력 장치를 통해 입력한 문장을 본인 음성 그대로 들려 줄 수 있는 기술이다.

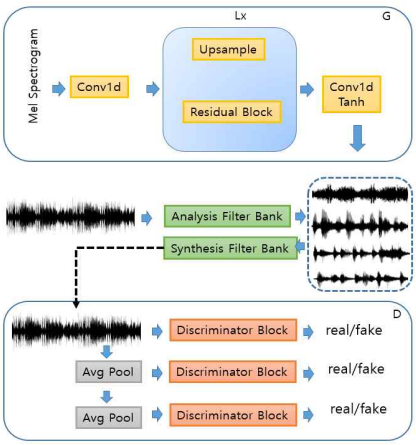
3.4.1. 음성 합성 기술과 TTS 시스템 구현

(1) Glow-TTS를 이용한 TTS 시스템

흐름 기반 생성 모델 및 동적 프로그래밍의 속성을 활용하여 텍스트와 음성을 빠르고 안정적으로 정렬하는 TTS 모델이다 [8].

(2) Multi-band MelGAN을 이용한 음성 합성

기존에 제안한 MelGAN[9]을 개선한 방식으로 다중 대역 구조를 통해 더 나은 음질과 안정성을 제공한다 [10]. 주파수 대역별 별도의 생성기 사용으로 합성 시간을 줄이고 변환손실 함수를 사용하여 음질을 향상한다.



〈그림 10〉 multiband MelGAN 구조

3.4.2. 제안 시스템 구현

다음의 과정을 통해 본인의 목소리로 원하는 문장을 출력할 수 있다.

(1) 시선을 추적하여 원하는 단어를 선택

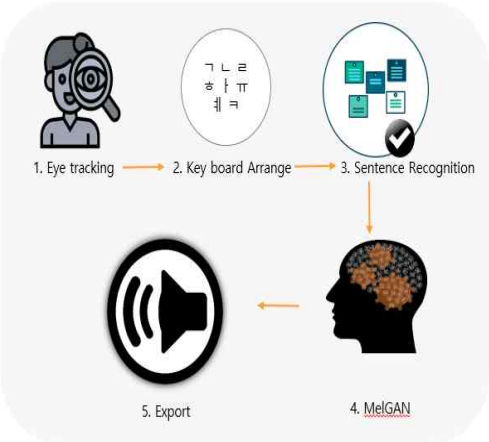
한글 입력 편의를 위한 인터페이스를 이용하여 표현하고자 하는 글자를 쉽게 선택할 수 있다.

(2) 원하는 문장 인식

시선 추척에서 선택한 문장으로 본인이 말하고자 하는 내용의 문장을 완성한다.

(3) 본인의 목소리로 출력

시선 추적 및 빠른 변환을 위해 Glow-TTS를 이용하여 TTS 모델을 완성한다. 화자의 목소리 이용이 가능하도록 성대 근육을 잃기 전 녹음된 목소리 파일로 학습된 음성을 multiband MelGAN을 이용하여 합성한다[11]. 표현하고자 하는 문장이 본인의 목소리로 매끄럽게 출력된다. 다음 <그림 11>은 전체적인 제안 시스템의 흐름을 보여준다.



〈그림 11〉 제안 시스템 흐름도

시선추적을 이용하여 입력된 문장이 본인의 목소리로 출력됨을 알 수 있다.

Ⅳ. 성능 평가

4.1. 실험환경

제안하는 시스템의 구현은 모두 파이썬 3.9 버전을 기준으로 하였다. 눈동자 인식 및 추적을 위해서 OpenCV를 사용하였다. 또한 자신의 목소리 음성 녹음을 하기 위해 Mimic Recording Studio를 사용하였다. 녹음 데이터를 Glow와 Multi-band MelGAN에서 학습을 진행하기 위해서 GPU를 이용할 수 있는 Google Colab을 사용하였다[12].

4.2. 실험 데이터 구성

학습을 시키기 위해서 3시간 분량의 녹음 데이터, 약 3000개 이상의 문장을 녹음하였다. 녹음 데이터는 Deep Leaning에서 사용할 수 있는 파일 형식으로 변환시켜 사용 한다[13-14].

4.3. 한글 입력 실험

Qwerty 자판을 사용하고 눈의 움직임으로 마우스를 제어하여 입력하는 것을 기준으로 하여 제안하는 시스템의 입력 속도에 대한 개선 정도를 비교 평가하였다. 실험 결과는 <표 1>에 요약하여 제시하였다.

〈표 1〉 입력 방식에 따른 개선정도

시스템 / 입력길이	Qwerty	의사소통 글자판	제안 시스템	개선율
단어	1	0.82	0.61	63.9%
단문	1	0.71	0.55	81.8%
장문	1	0.85	0.73	36.9%

먼저 간단한 단어 단위의 입력 실험을 진행 하였고, 100개 단어에 대한 실험을 진행한 결과 평균을 비교하였다. 일반 단어 입력인 경우 전체 입력 시간이 짧기 때문에 3개 방식에서 모두 큰 차이는 없었으나 제안하는 인터페이스를 이용하여 입력한 경우의 성능이 가장 좋게

나타났다. 실험 결과 평균적으로 Qwerty 자판의 속도를 기준으로 63.9%가 개선됨을 확인하였다.

단문 입력 실험에서는 “물 좀 갖다주세요”와 같은 10개 어절 이하의 문장을 완성하는 실험 결과를 비교하였다. 실험 결과 Qwerty 자판 입력 방식 대비 제안하는 방법이 81.8% 개선됨을 확인하였다. 짧은 단어 입력에 대한 비교 실험에서는 실험한 입력 방식이 모두 비슷한 성능을 보였지만 단문 입력 실험에서는 제안하는 입력 방식의 성능 개선의 정도가 매우 크게 나타남을 확인하였다. 장문 입력 실험은 Qwerty 자판 보다 빠르고 의사소통 글자판 방식의 입력과 비슷한 성능을 보였는데 이는 눈의 움직임이 많아져 피로도 증가로 인해 자판 선택의 오류가 증가한 것에 기인한 것으로 분석된다. 그러나 실제 환경에서 장문 입력을 하는 경우가 드물기 때문에 환자 지원 시스템으로 적용하는 것에는 무리가 없다.

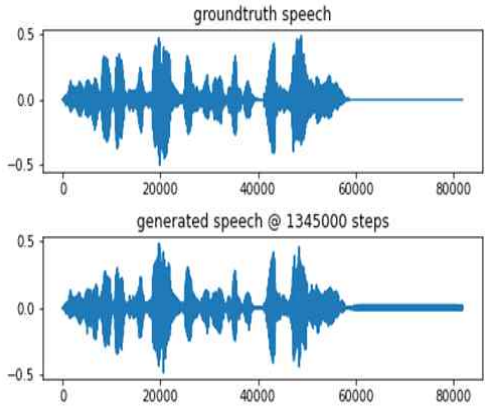
4.4. 파형 비교 실험

음성 데이터 생성의 성능을 평가하기 위하여 첫 번째 실험으로 원본 음성의 파형과 생성한 음성의 파형을 비교한다. 이때 “당신에게 자주 안부 전하지 못해 미안해요”의 문장을 실험에 사용하였고, 원본 음성의 파형과 제안하는 시스템의 음성 생성 모듈의 출력 결과의 파형을 비교 평가하였다.

두 번째 실험에서는 사용자에 따라 그 사용자의 억양 등의 요소가 반영되어 음성이 생성되는지를 비교 평가하였고, 본 실험에서는 “그것이 그 사람을 더 추하게 만들어요.”라는 문장을 이용하여 진행하였다.

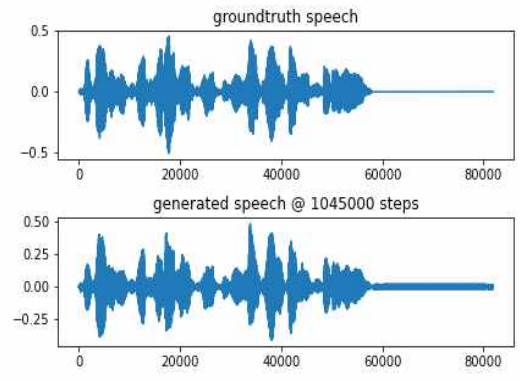
4.4.1. 실험 결과

첫 번째 실험의 결과를 <그림 12>에 제시하였다. 실험 결과의 그림에서 보여지는 바와 같이 A 화자의 원본의 음성과 TTS를 거친 파형이 비슷한 파형을 생성해내는 것을 알 수 있다. TTS 파형을 Deep learning 을 이용하여 목소리를 복원하였을 때 사람의 목소리와 똑같이 들리는 것을 알 수 있다. 실제 실험에서 복원된 소리와 원본 목소리의 구분이 어려웠다.

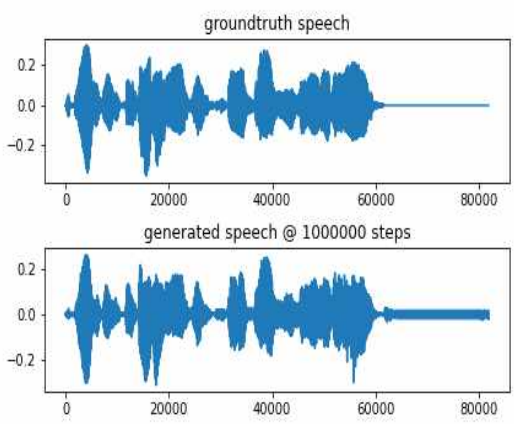


〈그림 12〉 A화자의 원본 음성과 생성 음성의 음성파형

두 번째 실험은 화자에 따른 특성이 반영되어 동일한 문장이나 화자에 따라 다른 음성 데이터가 생성됨을 확인하는 실험이다. 본 실험에서는 A화자와 B화자가 똑같은 문장을 말하게 하였고, 각각의 원본 음성 파형과 제안하는 시스템의 음성 생성 모듈에서 생성된 파형을 서로 비교하였다.



〈그림 13〉 A화자의 원본 음성과 생성 음성의 음성파형



〈그림 14〉 B화자의 원본 음성과 생성 음성의 음성파형

<그림 13>은 A화자가 "그것이 그 사람을 더 추하게 만들어요"라는 문장을 말한 것이고 원본 음성 및 생성 음성의 파형을 보여준다. <그림14>는 B화자가 "그것이 그 사람을 더 추하게 만들어요"라는 동일한 문장을 말한 것이고 원본 음성 및 생성 음성의 파형을 보여준다. 결과에서 보여지는 바와 같이 화자에 따라서 동일한 문장이지만 각기 다른 음성 파형을 생성하는 것을 알 수 있다.

이로써, TTS를 이용해 원본의 음성과 생성된 음성이 비슷한 음성임을 알 수 있으며 화자에 따라서 각자 다른 특성이 반영되어 본인의 목소리에 근접한 다른 음성 파형이 생성됨을 확인하였다.

Ⅴ. 결론

기존의 시선추적 기반 입력은 키보드 자판에 맞추어 동작하는 것이 일반적이다. 이를 개선하고자 본 연구에서는 상하좌우 움직임과 깜빡임 및 응시만을 사용하여 제어하고 빠르고 정확한 입력을 위한 한글 특징에 맞는 자판을 새롭게 제안하였다. 본 논문에서 제안한 시스템은 기존의 의사소통에서 사용되는 키보드와는 달리, 직관적이며 눈에 피로도를 줄인 키보드 패턴을 사용하였다. 입력 방식도 안구의 움직임을 최소화하도록 구성하였다. 따라서 기존의 인터페이스와 비교하여 제안하는 시선 추적에 기반한 한글 입력 인터페이스는 안구 움직임이 많지 않아 눈의 피로는 최소화하면서 단어를 빠르고 간편하게 입력할 수 있는 직관적 인터페이스라 할 수 있다.

시선을 추적하여 루게릭병 환자가 입력한 단어를 Glow-TTS와 MelGAN을 이용하여 자신의 목소리로 복원하여 재생할 수 있는 의사소통 시스템을 제안하고 구현하였다. 실험을 통해 실제 환자 목소리와 유사한 음성을 생성할 수 있음을 확인할 수 있었다. 음성 합성에 대해 Deep Learning을 이용하여 지속적인 훈련이 진행된다면, 목소리의 톤과 억양도 실제 목소리와 거의 비슷하게 음성 합성이 나오게 될 것이라고 예상된다. 또한 본 연구의 확대 분야로 환자가 이미 목소리를 잃은 경우 환자의 동성 가족의 음성과 환자의 구강구조 특징을 사용하여 재생이 가능한 음성 파일을 생성할 수 있다. 실험 결과로 알 수 있듯이 두 사람의 발성이 서로 다른 사람임을 알 수 있는 다른 파형을 출력하고, 화자의 발성톤이 거의 비슷하게 복원되어 톤, 발성을 되살려 냈으므로 본 연구의 의의를 찾을 수 있다.

향후 본 연구는 계속 발전되고 있는 TTS모델을 사용하여 음성 합성의 정확성을 더 높일 수 있으며 본인의 목소리로 네비게이션 목소리 바꾸기, 어린이 교육 분야에서 책 읽어주기, 구현동화 들려주기 등에 활용할 수 있다. 또한 아이트래킹의 기술을 마케팅 분야, 스포츠 학습 분야, 장애인 복지 분야, e-sports, 게임 등 다방면에서 활용될 수 있다.

참고문헌

[1] HANYANG University Medical Center, "What is Lou Gehrig's disease?," https://seoul.hyumc.com

/hyctc/als/info.do?action=view&bbsId=ALSdiseases&nttSeq=10861, 2018.7.10.

[2] Sung-hyeon Jo, “Introduction to Eye Tracking Technology,” The Magazine of the IEEE, Vol 45, No. 8, 2018, pp.23-32.

[3] Kim-JongHa. “Eye-tracking and Perception,” Review of Architecture and Building Science, Vol.58, No. 9, 2014, 21-26.

[4] Byoung-jin Kim, Suk-ju Kang, “Reliability Measurement Technique of The Eye Tracking System Using Gaze Point Information,” Journal of Digital Contents Society, Vol. 17, No. 5, http://dx.doi.org/10.9728/dcs.2016.17.5.367, Oct. 2016, pp.367-373.

[5] Eunseo Jeon, Chaeyeon Kim, Aeri Shin, Yeongeun Jeon, Dong-Ok Won. “Mobile Platform Communication System based on Eye Movement Detection for Amyotrophic Lateral Sclerosis Patients,” Journal of the Korean Society of Information Science, 2022, pp.1701-1703.

[6] Jong-Hyun Kim, “Efficient Mobile Writing System with Korean Input Interface Based on Face Recognition,” Journal of The Korea Society of Computer and Information, Vol.25, No.6, June 2020, pp.49-56, https://doi.org/10.9708/jksci.2020.25.0

[7] Hyungseop Son, “A Study on Efficient Korean Voice Synthesis Model for The Visually handicapped,” Master’s thesis, Hanbat University, Daejeon, Korea, 2022. Amazing Talker. https://www.amazingtalker.co.kr/blog/ko/kr-en/62043/

[8] Lee Hee-man, Kim Ji-young, “Voice synthesis engine for TTS application Speech Synthesis Engine for TTS,” The Journal of Korean Institute of Communications and Information Sciences, Vol 23, No. 6, 1998, pp.1443-1453.

[9] Geng Yang, Shan Yang, Kai Liu, Peng Fang, Wei Chen, Lei Xie, “Multi-band MelGAN: Faster Waveform Generation for High-Quality Text-to - Speech,” arXiv:2005.05106, https://doi.org/10.48550/arXiv.2005.05106

[10] Jaehyeon Kim, Sungwon Kim, Jungil Kong, Sungroh Yoon, “Glow-TTS: A Generative Flow for Text-to-Speech via Monotonic Alignment Search,” arXiv:2005.11129, https://doi.org/10.48550/arXiv.2005.11129

[11] Kundan Kumar, Rithesh Kumar, Thibault de Boissiere, Lucas Gestin, Wei Zhen Teoh, Jose Sotelo, Alexandre de Brebisson, Yoshua Bengio, Aaron Courville, “MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis,” https://doi.org/10.48550/arXiv.1910.06711

[12] Su-hyeon Oh, Jin-Seob Kim, Byungwook Lee, Minseong Choi, Eunwoo Song, “Effective Learning Rate Scheduling for Speaker-Adaptive TTS,” Korean Society of Electronics Engineers Conference, 2022, pp.1173-1175.

[13] T. Okamoto, T. Toda, Y. Shiga and H. Kawai, “Tacotron-Based Acoustic Model Using Phoneme Alignment for Practical Neural Text-to-Speech Systems,” IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) Singapore, 2019, pp. 214-221, doi: 10.1109/ASRU46091.2019.9003956.

[14] Ahn, H.; Yim, C. “Convolutional Neural Networks Using Skip Connections with Layer Groups for Super-Resolution Image Reconstruction Based on Deep Learning,” Applied Sciences, Vol.10, No.6, 2020, pp.1959-1968, https://doi.org/10.3390/app10061959

■ 저자소개 ■

박 현 주
(Park Hyunjoo)

2017년 9월~현재
상명대학교 스마트 정보통신공학과
부교수
2011년 2월 상명대학교 컴퓨터과학과
(이학박사)
2001년 2월 홍익대학교 전산학과(이학석사)
1998년 2월 상명대학교 전산학과(이학사)

관심분야 : 정보통신, 엔터테인먼트 콘텐츠, 인공지능 인터페이스
E-mail : cathy2369@smu.ac.kr

정 승 도
(Jeong Seungdo)

2015년 3월~현재
상명대학교 스마트 정보통신공학과
부교수
2007년 8월 한양대학교 전자통신전파공학과
(공학박사)
2001년 2월 한양대학교 전자통신전파공학과
(공학석사)
1999년 2월 한양대학교 전자통신전파공학과
(공학사)

관심분야 : 딥러닝, Tensor 응용, 멀티미디어
E-mail : sdjeong@smu.ac.kr

논문접수일 : 2024년 5월 31일
수정접수일 : 2024년 6월 13일
게재확정일 : 2024년 6월 16일