

최근 정보통신 기술 음성 변조 어플과 그에 대응하는 딥 보이스 페이크 대응 프로그램에 관한 연구

박이현, 박민수, 김균하, 백시우, 이재형
울산과학기술대학교, 국민대학교, University of California San Diego
gya07055@naver.com

A Study on the Voice modulation system and resulting prevention of deepfake voices systems

Ihyun Leo Park, Park minsu, Baek siwoo, Lee Jaehyeong, Kim gyunha
Chosun Univ., Ulsan College. University of California San Diego
gya07055@naver.com

요 약

본 연구는 실시간 음성 변조와 딥 보이스 페이크 탐지 시스템을 개발하여 사용자 경험을 극대화하고 보이스 데이터의 안전성과 활용도를 높이는 것을 목표로 합니다. 보이스 변조 시스템은 영화, 애니메이션, 음악 제작 등에서 특수 음성 효과와 유명 인물의 목소리를 모방하는 데 활용될 수 있습니다. 또한 딥 보이스 페이크 (Deep Voice fake) 변조의 위험을 사람들에게 알리며 경각심을 가지게 할 수 있습니다. 딥 보이스 페이크 탐지 시스템은 보이스 데이터를 안전하게 상용화하며, 생체 인식 기술에도 활용될 수 있는 기초 작업입니다. 이로 인해 보이스 데이터의 범죄나 악용 사례를 줄이고, 보이스 데이터의 신뢰성과 다양한 산업 분야에서의 활용도를 극대화할 수 있으며 보이스 데이터의 새로운 창작 기회와 안전한 사용을 촉진합니다.

I. 서 론

최근 몇 년간 인공지능과 딥러닝 기술의 발전으로 음성 변조와 딥 보이스 페이크 기술이 급속도로 발전하고 있습니다. 이러한 기술은 엔터테인먼트, 보안, 의료 등 다양한 분야에서 혁신적인 변화를 가져올 잠재력을 가지고 있지만, 동시에 개인 정보 보호 및 사회적 신뢰성에 심각한 위협이 될 수 있습니다. 본 논문에서는 실시간 목소리 변조 어플리케이션과 딥 보이스 페이크 판단 프로그램을 개발하여 이러한 기술의 장단점을 분석하고, 더 나아가 사회적 경각심을 고취시키고자 합니다.

먼저, 목소리 변조 어플리케이션 개발을 통해 사용자의 목소리(음성 데이터 인풋)을 받으면 STT(Speech to Text)를 사용해 텍스트로 변환 후, 변조 프로그램과 TTS(Text-to-Speech) 기술을 활용하여 음성 시그널을

아웃풋하여 유명인의 목소리 데이터베이스를 기반으로 변조된 음성을 생성할 것입니다. 이를 통해 사용자들은 자신들의 목소리를 유명인의 목소리로 변조하는 체험을 하게 되며, 딥 보이스 페이크 기술에 대한 이해와 경각심을 높일 수 있을 것입니다.

다음으로, 딥 보이스 페이크 판단 프로그램을 개발하여 딥 보이스 페이크 음성의 판별 정확성을 향상시키고자 합니다. 이를 위해 깃 허브의 오픈 소스를 활용하여 사전 데이터로 훈련된 모델을 사용하고, Whisper 인코더와 LFCC, LCNN, MGCC 등의 시스템을 적용할 것입니다. 이러한 시스템들은 특정 충격음 및 주파수를 감지하고 걸러내는 기능을 갖추고 있어, EER(Equal Error Rate) 값을 비교하여 딥 보이스 페이크 판단의 오차율을 줄일 수 있습니다. 변조된 음성은 사용자 목소리 데이터에 없는 주파수를 다수 포함하고 있어 높은 EER 30% 이상

을 나타내는 반면, 일반 음성은 10% 미만의 EER 을 나타내는 특성을 활용할 것입니다.

본 연구를 통해 딥 보이스 페이크 기술의 사회적 인식을 제고하고, 이를 판별하는 기술의 정확성을 향상시키는 데 기여하고자 합니다. 이러한 노력은 딥 보이스 페이크 기술의 악용을 방지하고, 보다 안전하고 신뢰할 수 있는 디지털 환경을 구축하는 데 중요한 역할을 할 것입니다.

II. 본론

- 온 디바이스 실시간 음성 변조 어플 개발

본 연구에서는 사용자의 목소리를 실시간으로 변조하여 즉시 저장 및 사용이 가능하게 하는 어플리케이션을 개발하고자 합니다. 이를 위해 음성 변조, 변조 데이터, 통신, 어플 제작의 네 가지 주요 요소를 사용하였습니다.

1. **음성 변조**:

- 실시간 음성 변조를 위해 시그널 소스를 이용한 목소리 변조 오픈소스 프로그램을 사용합니다.ⁱ

2. **변조 데이터**:

- 다양한 음성 변조 데이터를 AI hub 에서 수집하고, 디스코드 서버를 통해 활용합니다.ⁱⁱ

3. **통신**:

- 음성 변조 앱과 DF 판단 프로그램 간의 통신을 위해 FastAPI 를 사용합니다.ⁱⁱⁱ

4. **어플 제작**:

- 어플리케이션 개발에는 안드로이드 코틀린(Kotlin)을 사용하여 직관적이고 간단한 UI 를 구현하면서도 다양한 음성 변조 옵션을 제공하여 사용자 경험을 극대화합니다.^{iv}

5. **어플 구상도**



본 연구의 시제품 목표는 다음과 같습니다.

- ◆ 유명 연예인의 목소리 변조 : 최소 10~15 명 이상의 유명 연예인의 목소리로 변조 가능하도록 구현하고 이를 통해 사용자가 더 많은 관심을 가질 수 있도록 유도합니다.
- ◆ 편리한 UI 구성 : 직관적이고 간단한 UI 를 제공하여 사용자가 쉽게 접근하고 사용할 수 있게 합니다.
- ◆ 실시간 변조 기능 : 짧은 인사나 대화를 실시간으로 변조할 수 있도록 제작하여 사용자에게 즉각적인 피드백을 제공할 것 입니다.

- 실시간 딥 보이스 페이크 판단 프로그램 개발

딥 보이스 페이크 판단 프로그램을 개발하기 위해 Whisper 라는 딥러닝 모델을 활용합니다. 이 모델은 LFCC, LCNN, MGCC 등 주파수 탐지법을 기반으로 동작하는 프로그램이며 딥 보이스 페이크 페이크와 일반 음성 데이터의 주파수 특징을 학습하여 딥 보이스 페이크와 일반 음성 데이터의 차이를 구별 EER 값으로 나타냅니다.

1. **딥 보이스 페이크 탐지 프로그램**:

- Whisper 오픈 소스 프로그램을 활용하여 음성 인풋의 딥 보이스 페이크 여부를 추출합니다

2. **통신**:

- 음성 변조 앱과 DF 판단 프로그램 간의 통신을 위해 FastAPI 를 사용합니다.^v

3. **UI 제작**:

- 사용자 인터페이스는 파이썬 (python) 을 사용하여 개발

4.**예상 구상도**



본 연구의 시제품 목표는 다음과 같습니다.

- ◆ 변조된 음성 데이터와 변조되지 않은 음성 데이터 구분: 변조된 음성 데이터와 변조되지 않은 음성 데이터를 구분하여 변조 진위 여부를 파악합니다.
- ◆ EER 값 제공: 분류 시스템의 성능을 측정하는 지표로 EER(Equal Error Rate) 값을 제공하여 모델의 성능을 평가합니다.
- ◆ 개발 어플과의 연동: 개발된 어플과 연동하여 변조된 음성을 실시간으로 판단하는 기능을 목표로 합니다.

III. 결론

본 연구에서 개발된 보이스 변조 시스템과 딥 보이스 페이크 탐지 시스템은 다양한 방면에서 큰 기여를 할 수 있다고 생각합니다.

- 보이스 변조 시스템

보이스 변조 시스템은 단순히 재미와 유머를 제공하는 것 외에도 엔터테인먼트 및 창작 활동에 많은 도움을 줄 수 있습니다. 예를 들어, 영화나 애니메이션 제작 시 특수 음성 효과를 만들거나, 유명 인물의 목소리를 모방하는 데 사용될 수 있고 나아가, 이 기술이 더욱 발전하면 음악 제작 과정에서도 활용될 가능성이 크다고

생각합니다. 이러한 다양한 활용 가능성은 보이스 변조 시스템의 가치를 더욱 높이며, 사용자들에게 새로운 창작의 기회를 제공할 수 있을 것입니다.

- 딥 보이스 페이크 탐지 시스템

딥 보이스 페이크 탐지 시스템의 가장 큰 기대효과는 보이스 데이터를 더욱 안전하게 상용화하는 것입니다. 현재 딥 보이스 페이크 기술의 발전으로 인해 보이스 데이터의 사용이 제한되고 있으나, 딥 보이스 페이크 탐지 시스템이 발달하고 가볍게 사용될 수 있다면, 보이스 데이터는 생체 인식 기술에도 활용될 수 있습니다. 또한, 딥 보이스 페이크 탐지 시스템의 발전은 범죄나 악용 사례를 줄이는 데 기여할 것으로 예상됩니다. 이를 통해 보이스 데이터의 신뢰성을 높이고, 다양한 산업 분야에서의 활용도를 극대화할 수 있을 것이라고 예상합니다.

- 종합

본 연구는 실시간 음성 변조와 딥 보이스 페이크 탐지 기술을 통해 사용자 경험을 극대화하고, 보이스 데이터의 안전성과 활용도를 높이는 데 기여하고자 합니다. 보이스 변조 시스템은 엔터테인먼트 및 창작 활동에 새로운 가능성을 제공하며, 딥 보이스 페이크 탐지 시스템은 보이스 데이터의 상용화와 안전성을 보장합니다. 이러한 기술적 발전은 보이스 데이터의 활용 범위를 넓히고, 사용자들에게 보다 신뢰할 수 있는 서비스를 제공할 것으로 예상합니다.

참 고 자 료

ⁱ Github

RVC-Project/Retrieval-based-Voice-Conversion-WebUI

<https://github.com/auspicious3000/contentvec/>

<https://github.com/jaywalnut310/vits>

<https://github.com/jik876/hifi-gan>

<https://github.com/gradio-app/gradio>

<https://github.com/FFmpeg/FFmpeg>

ⁱⁱ Hugging face

<https://huggingface.co/SeoulStreamingStation/IU-Voice-MultiLanguage-V1/tree/main>

<https://huggingface.co/lj1995/VoiceConversionWebUI/tree/main>

ⁱⁱⁱ <https://fastapi.tiangolo.com/ko/>

^{iv} <https://developer.android.com/courses/kotlin-android-fundamentals/overview?hl=ko>

^v <https://fastapi.tiangolo.com/ko/>