검색 기반 음성 변환 기술의 발전과 응용 사례 연구

허태성O, 오주헌*
O인하공업전문대학 컴퓨터정보공학과,
*인하공업전문대학 컴퓨터정보공학과
e-mail: tshur@inhatc.ac.krO, 202447003@itc.ac.kr*

Advances and Applications of Retrieval-based Voice Conversion Technology

Hur Tai-sungO, Oh Ju Heon*
ODept. of Computer Science Engineering, Inha Technical College,
*Dept. of Computer Science Engineering, Inha Technical College

요 약

검색 기반 음성 변환(Retrieval-based Voice Conversion)은 주어진 음성을 다른 화자의 음성으로 변환하는 혁신적인 기술로, 다양한 응용 분야에서 그 중요성이 증가하고 있다. 본 논문에서는 RVC의 기본 원리와 주요 구성 요소를 설명하고, 최신 연구 동향과 기술적 발전을 탐구한다.

RVC의 핵심은 음성 모델을 구축하고, 입력 음성의 특징을 추출하여 가장 유사한 음성을 검색한 후, 이를 바탕으로 자연스러운 음성 변환을 수행하는 것이다. 이러한 과정에서 음성의 주파수 성분, 음색, 억양 등의 특성을 효과적으로 처리하는 기술이 사용된다. RVC는 고품질의 음성 변환을 제공하며, 처리 시간이 빠르기 때문에 실시간 음성 변환이 가능하다는 장점을 지닌다. 그러나 모델의 크기와 다양성에 대한 의존성, 복잡한 음성 특성의 처리 문제 등 몇 가지 도전 과제도 존재한다. 본 논문은 이러한 장점과 단점을 균형 있게 분석하며, RVC 기술의 현재와 미래를 조망한다. 이를 통해 음성 합성, 언어 번역, 가상 비서 등 다양한 분야에서 RVC의 잠재적 응용 가능성을 제시하고자 한다.

▸ Keyword : AI, VITS, RVC, 검색 기반 음성 변환, 음성 변조, 음성 합성, 실시간 음성 변환

I. Introduction

음성 변환 기술은 최근 몇 년간 급격한 발전을 이루며, 음성 합성, 언어 번역, 가상 비서 등 다양한 응용 분야에서 중요한 역할을 하고 있다. 이러한 기술 중에서도 검색 기반 음성 변환(Retrieval-based Voice Conversion, RVC)은 주목할 만한 접근 방식으로, 입력 음성을 다른 화자의 음성으로 변환하는 데 있어 높은 품질과 실시간 처리를 가능하게 한다.

II. Preliminaries

1. 음성 변환

1.1 검색 기반 음성 변환

검색 기반 음성 변환이란 입력 음성의 음색, 억양 등을 분석하여 특징을 추출한다. 추출된 음성 특징을 바탕으로 모델에서 유사한 음성 샘플을 검색하여 입력 음성을 목표 화자의 음성으로 변환하는 기술이다. 모델에 저장된 실제 음성 샘플을 활용하기 때문에 더 자연스럽고 고품질의 음성을 생성할 수 있는 장점이 있다.

III. The Proposed Scheme

1. 음성 모델 구축

본 연구에서는 RVC를 구현하기 위해 본인의 음성 데이터를 기반으로 한 음성 모델을 구축하였다. 음성 모델 구축 단계는 음성 데이터 수집, 전처리, 음성 특징 추출, 그리고 모델 학습의 네 가지 주요 단계로 구성된다.

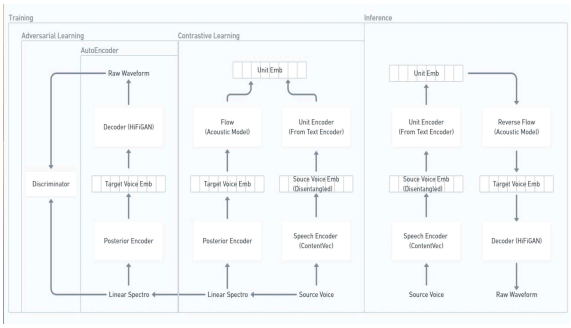


Fig. 1. Retrieval-based Voice Conversion Architecture

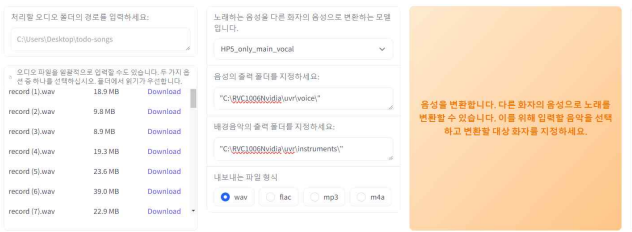


Fig. 2. 목소리 분리 및 잔향 제거 UI

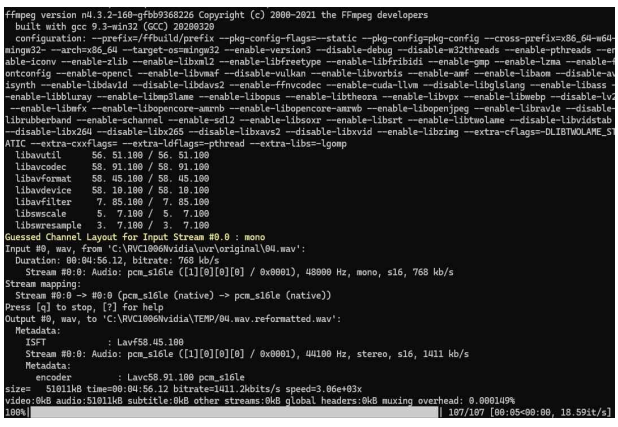
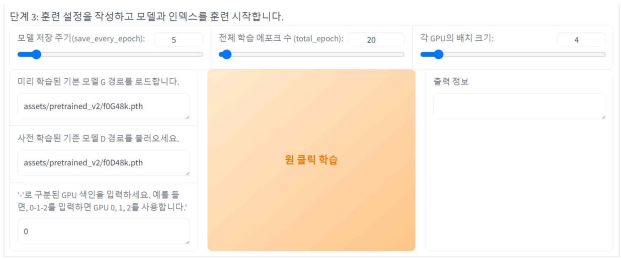


Fig. 3. 목소리 분리 및 잔향 제거 결과



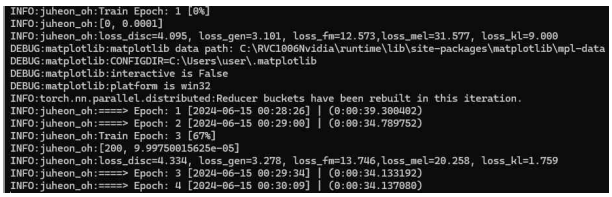


Fig. 4. 모델 학습

Fig. 5. 음성 변환

IV. Conclusions

본 연구는 검색 기반 음성 변환 시스템의 구현을 다루었다. 단일 화자의 음성 데이터를 활용하여 고품질의 음성 모델을 구축하고, 이를 통해 입력 음성을 사용자의 음성으로 변환하는 과정을 상세히 설명하였다. RVC 시스템은 음성 특징 추출, 검색, 음성 변환의 세 가지 주요 단계로 구성되며, 각 단계의 효율적인 수행을 통해 자연스럽고 일관된 음성 변환을 가능하게 한다.

References

[1] https://en.wikipedia.org/wiki/Retrieval-based_Voice_Conversion

[2] https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI

[3] https://music-audio-ai.tistory.com/22

[4] https://alltommysworks.com/?s=rvc

[5] https://github.com/w-okada/voice-changer