

2022 AI보안연구센터 인턴 미팅

Research REPORT

22년 9월 26일

20170622

이종헌



Audio Deepfake

특정인의 목소리를 딥 러닝 기술로 학습시켜,
문자 음성 자동변환 기술(TTS-Text to Speech)로
해당 특정인이 하지 않은 말을 마치 한 것처럼 만들어 내는 기술이다

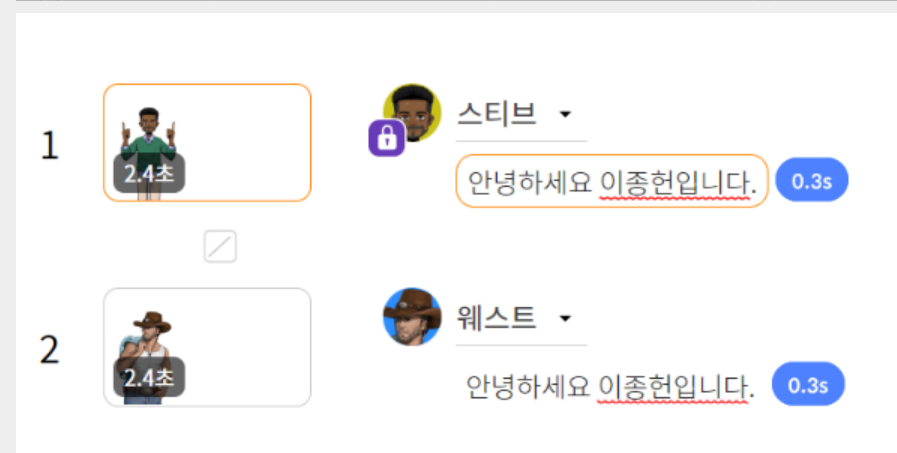
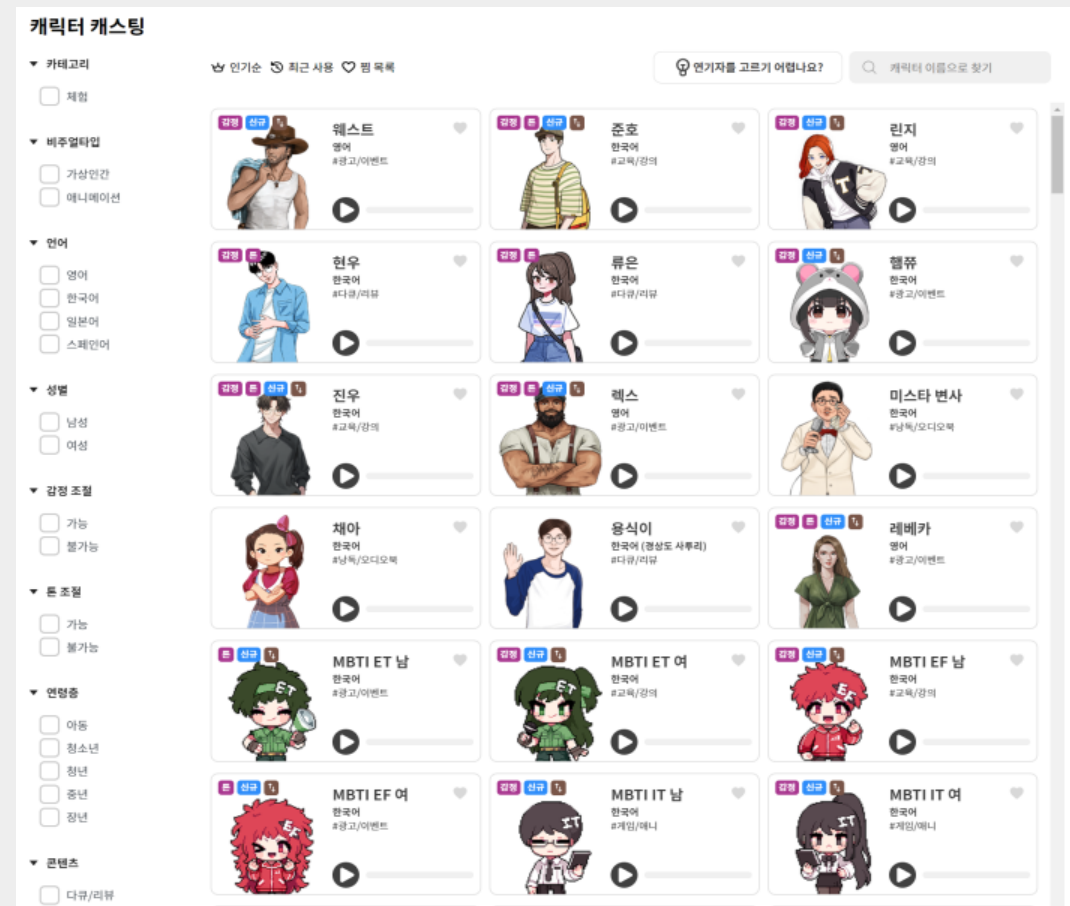


TTS(Text to Speech)

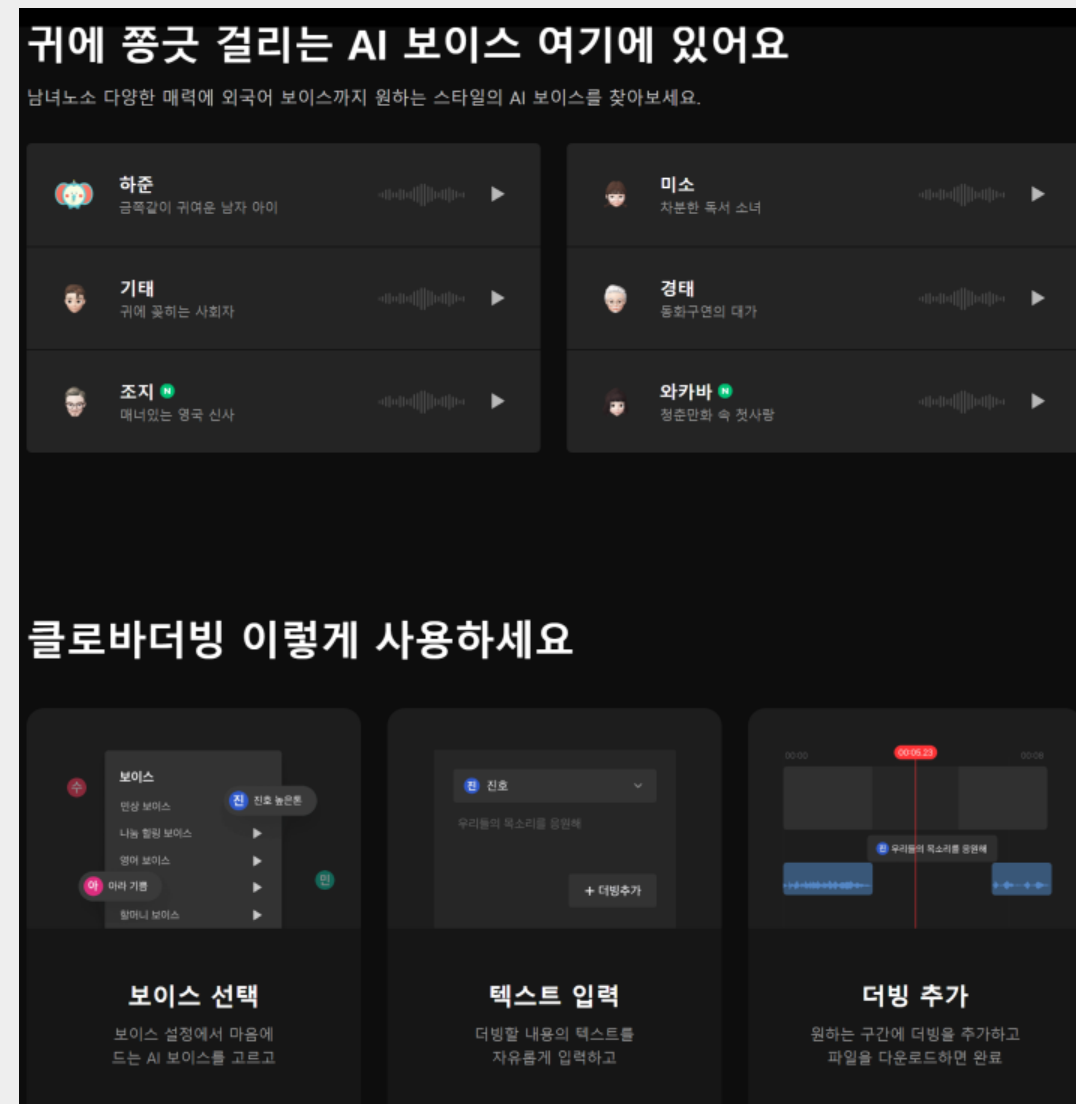
TTS(음성합성)는 Text to Speech로 딥러닝을 통해 인위적으로 사람의 소리를 합성하는 시스템이며, 텍스트를 음성으로 변환한다.
STT vs TTS -> STT : 음성인식, TTS : 음성합성



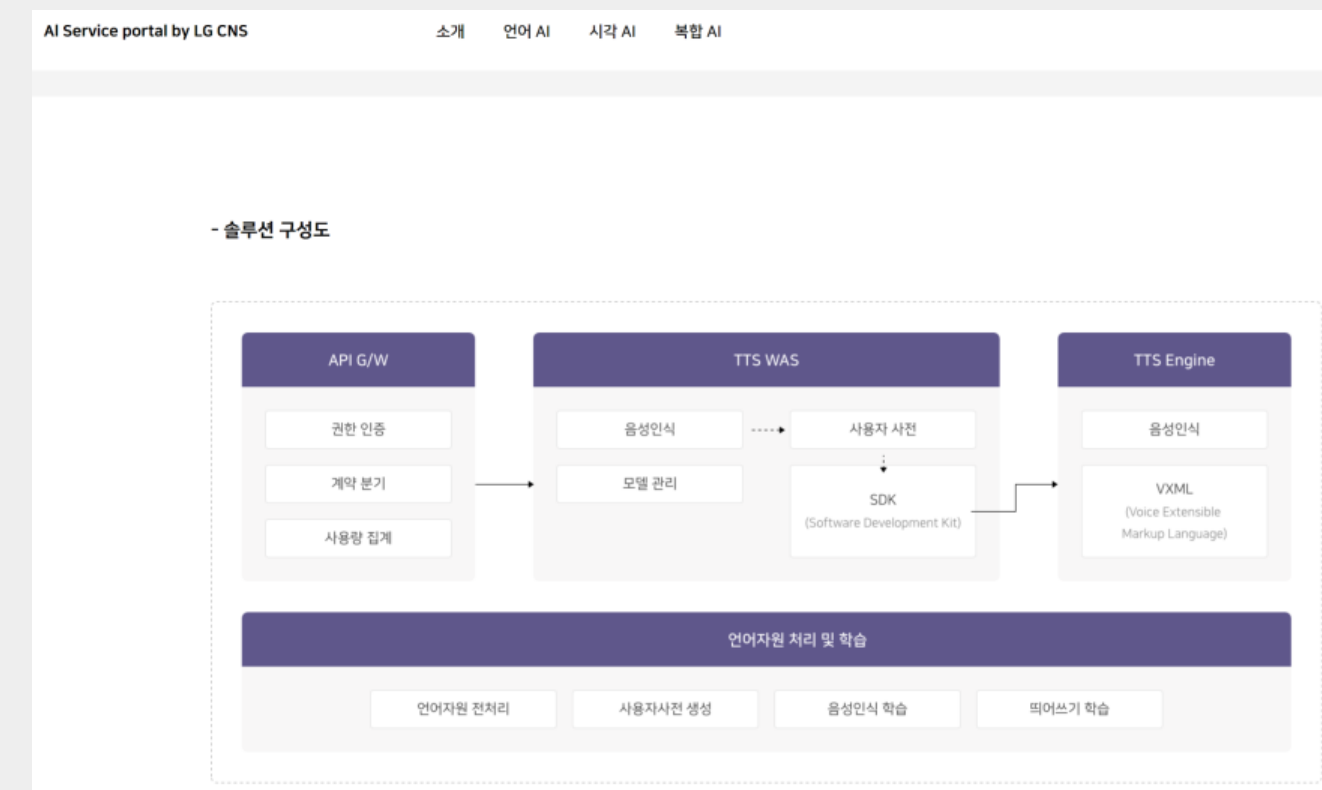
TTS(Text to Speech) 서비스 사례



<Typecast의 TTS 서비스>



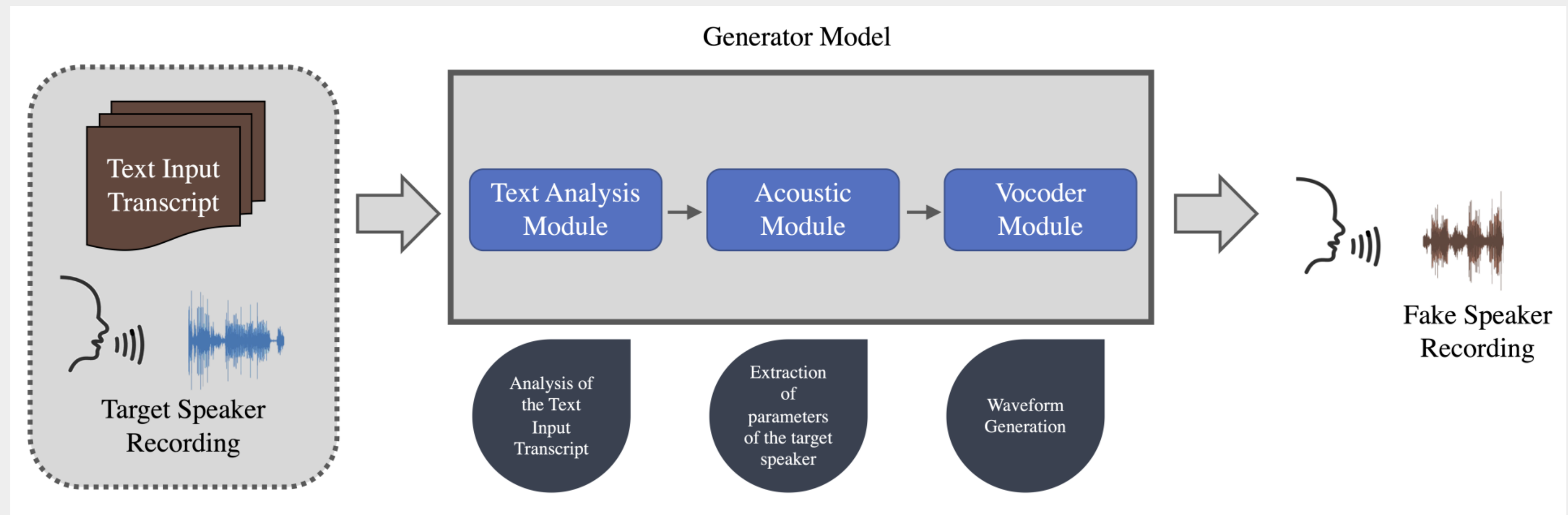
<Naver의 Clova Dubbing>



<LG CNS의 TTS API>

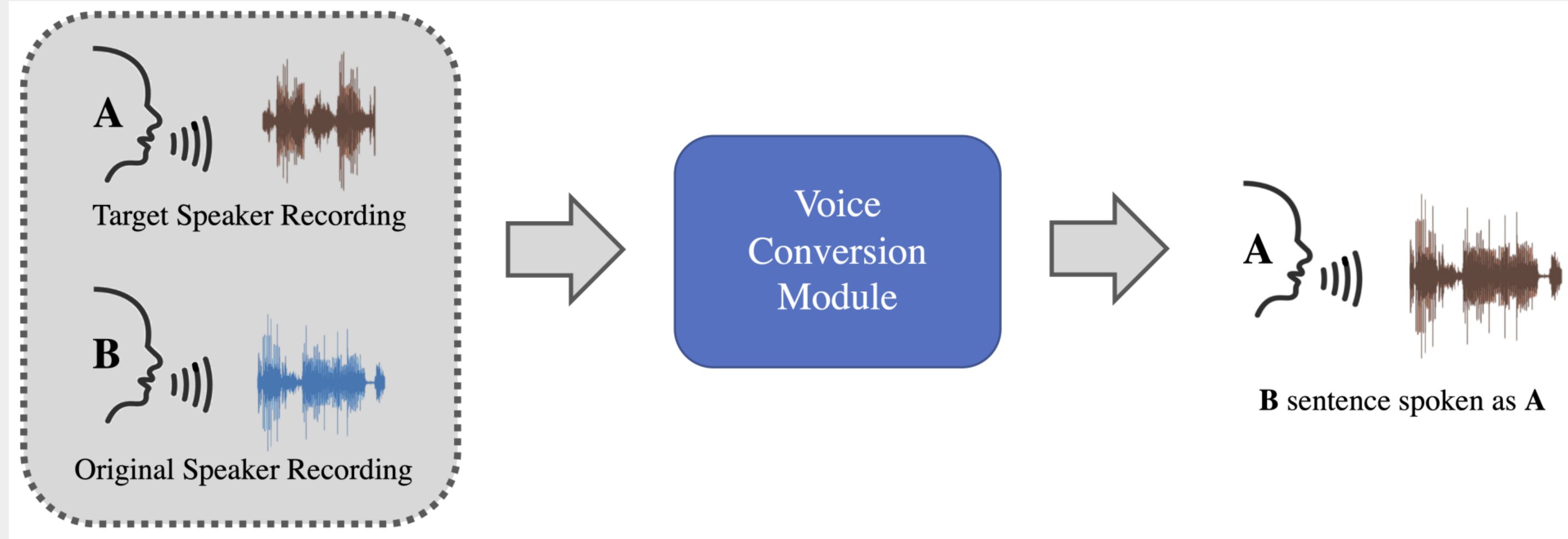
Audio Deepfake(합성 기반)

합성기반의 Audio Deepfake는 TTS(음성합성)를 활용한 딥페이크 기법
텍스트 분석 모델, 음향 모델, 보코더 모델 세가지로 구성



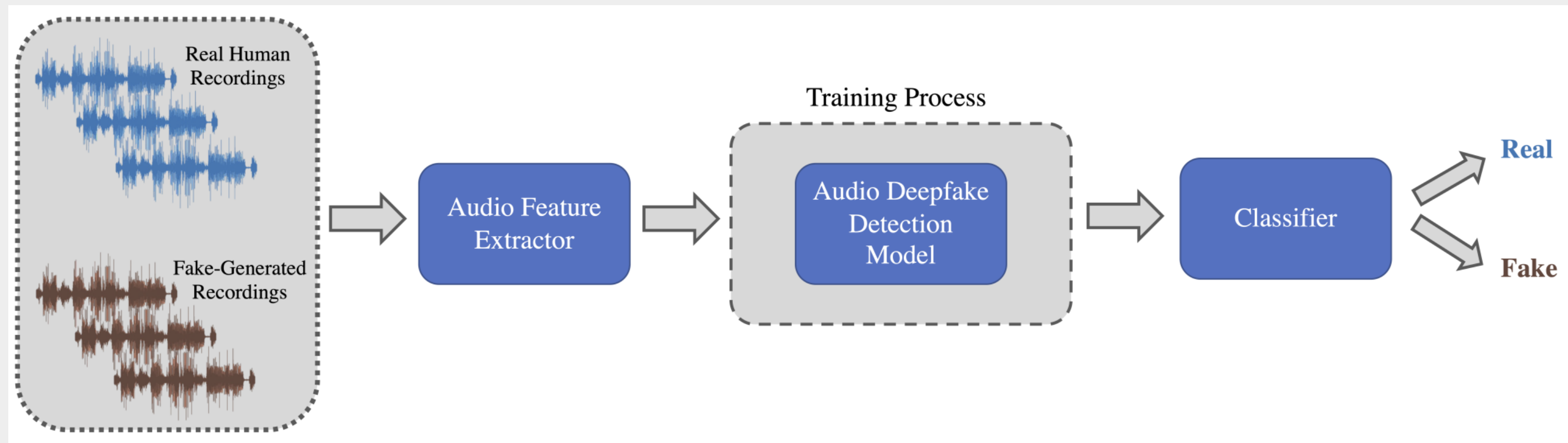
Audio Deepfake(모방 기반)

모방기반의 Audio Deepfake는
Original Speaker Recording을 Target Speaker Recording로 변환하는 기법
생성적 적대 신경망 (GAN)을 통하여 음성을 생성



Audio Deepfake 감지


오디오 딥페이크 감지 모델을 통해 감지



05 Voice Conversion

VC(Voice Conversion)

음성에 포함되어 있는 발화자의 특징을 변환하여 타인의 음성을 생성하는 기법

 Audio

Voice Conversion





100 papers with code • 1 benchmarks • 2 datasets

Voice Conversion is a technology that modifies the speech of a source speaker and makes their speech sound like that of another target speaker without changing the linguistic information.

Source: [Joint training framework for text-to-speech and voice conversion using multi-source Tacotron and WaveNet](#)




Benchmarks

These leaderboards are used to track progress in Voice Conversion



Trend	Dataset	Best Model	Paper	Code	Compare
	ZeroSpeech 2019 English	 VQ-CPC			See all

Libraries ①

Use these libraries to find Voice Conversion models and implementations

 s3prl/s3prl	3 papers	1,455 ★
 andi611/Self-Supervised-Speech-Pret...	3 papers	1,455 ★
 espnet/espnet	2 papers	5,497 ★


Datasets

 ESD  VESUS



Most implemented papers

Most implemented Social Latest No code


Search for a paper, author or keyword





StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks

 liusongxiang/StarGAN-Voice-Conversion •  PyTorch • 6 Jun 2018


This paper proposes a method that allows non-parallel many-to-many voice conversion (VC) by using a variant of a generative adversarial network (GAN) called StarGAN.

 Paper



 Code

 11


StarGAN-VC은 GAN 아키텍처를 적용하여 실시간 음성합성이 가능한 방법론을 제안하고 비교적 적은 Non-parallel 데이터를 이용하여 실제 음성과 비슷한 음성을 생성할 수 있다는 장점을 갖고 있는 논문





AUTOVC: Zero-Shot Voice Style Transfer with Only Autoencoder Loss

 liusongxiang/StarGAN-Voice-Conversion •  PyTorch • 14 May 2019

On the other hand, CVAE training is simple but does not come with the distribution-matching property of a GAN.

 Paper

 Code

 10

AutoVC는 AutoEncoder의 BottleNeck 구조를 활용하여 음성으로부터 화자의 특징정보와 내용정보를 분리하는 방법을 제안하고 Reconstruction Loss를 활용하여 안전하게 모델을 학습한 후 학습에 사용한 화자뿐만아니라 학습데이터에 없는 화자(Zero shot Conversion)에 대한 고품질의 음성변환 결과를 도출하였습니다.