

딥러닝을 통해 목소리를 만든다고?

FASTSPEECH2를 활용한 TTS 구현

#경영경제학부 #21600569 이해찬

#전산전자공학부 #21800815 이요한

#글로벌리더십학부 #22100678 정용빈



Table of Contents

- #1 주제 선정 배경
- #2 What to do?
- #3 TTS에 대하여
- #4 FASTSPEECH2 란?
- #5 모델링 개요
- #6 모델링 시연 및 설명
- #7 한계 및 보완점
- #8 프로젝트 의의

주제 선정 배경



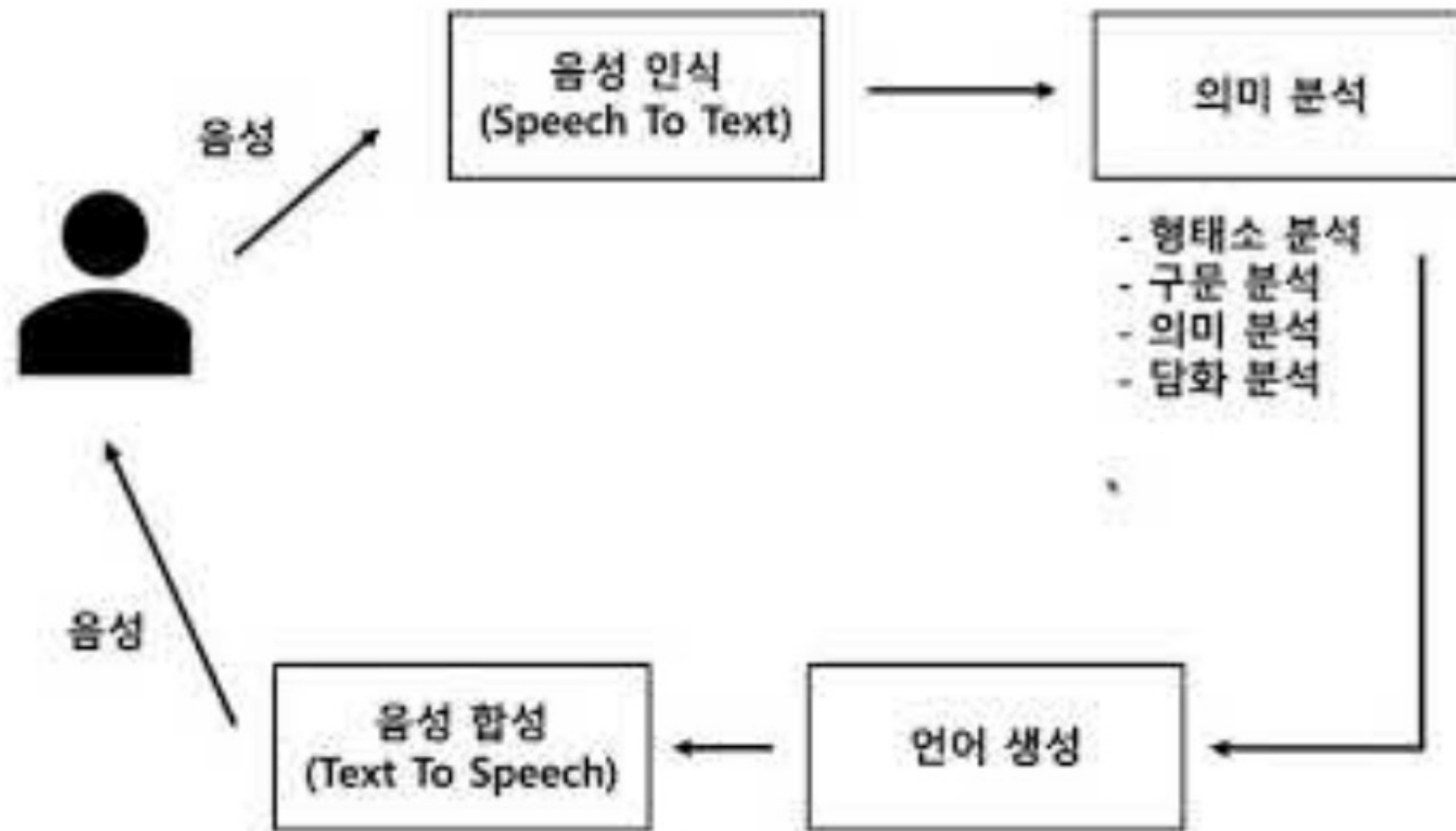
다시 한번,
그들의 목소리를 들을 수 있다면...



딥러닝을 통해 인간의 회복을 돕는 두 가지 영상을 보며
불의의 사고로 목소리를 낼 수 없는 사람의 회복을 돕거나
다시는 만나지 못 하는 누군가를 그리워하는 사람을 위로할 수 있는
방법에 대하여 고민해보다 이 프로젝트를 진행하게 되었습니다.

#01 주제 선정 배경

What to do?



#02

WHAT TO DO?

FASTSPEECH2 를 활용해 특정 인물의 음성데이터를 학습시켜
내가 원하는 텍스트를 학습시킨 인물의 목소리로 TTS를 구현하는 것

TTS란?

Text To Speech

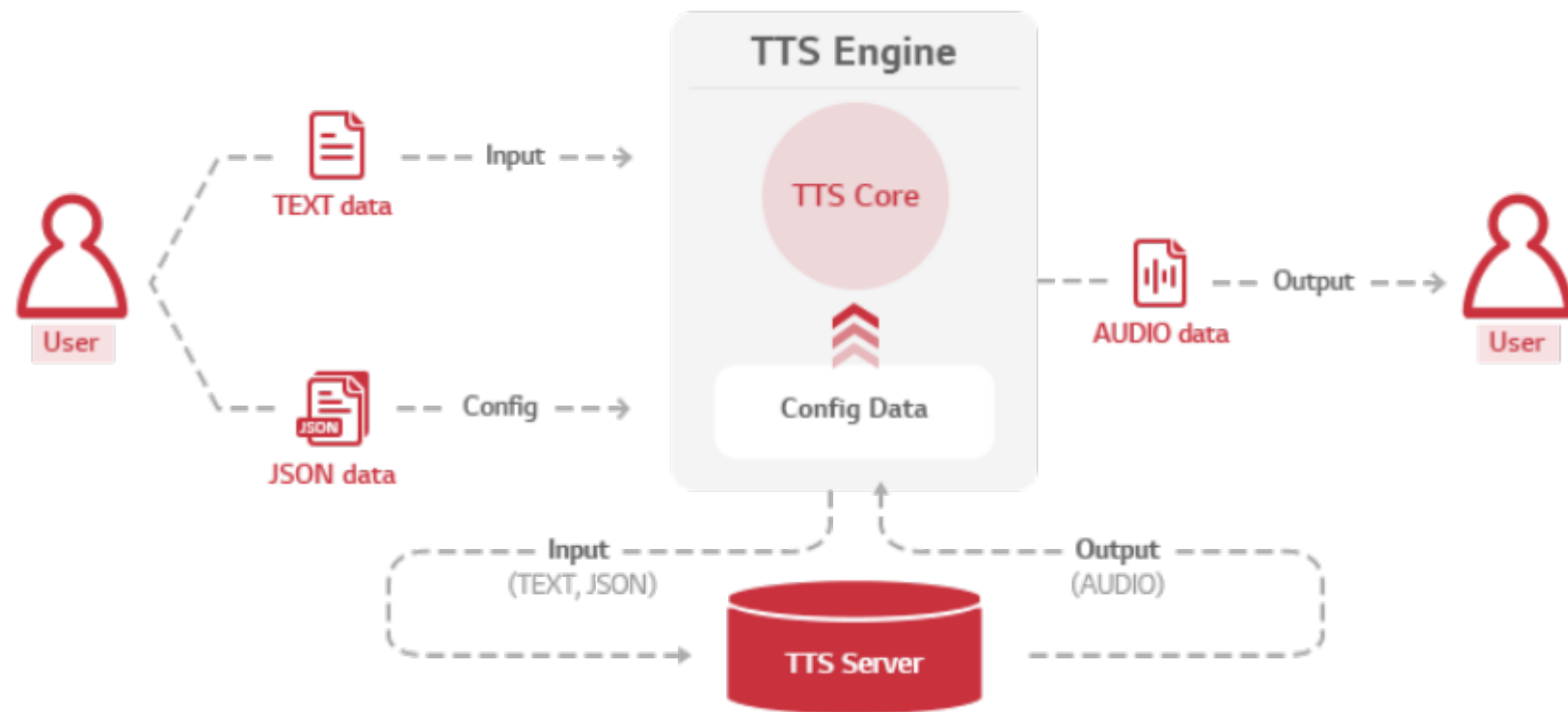


#03

TTS란?

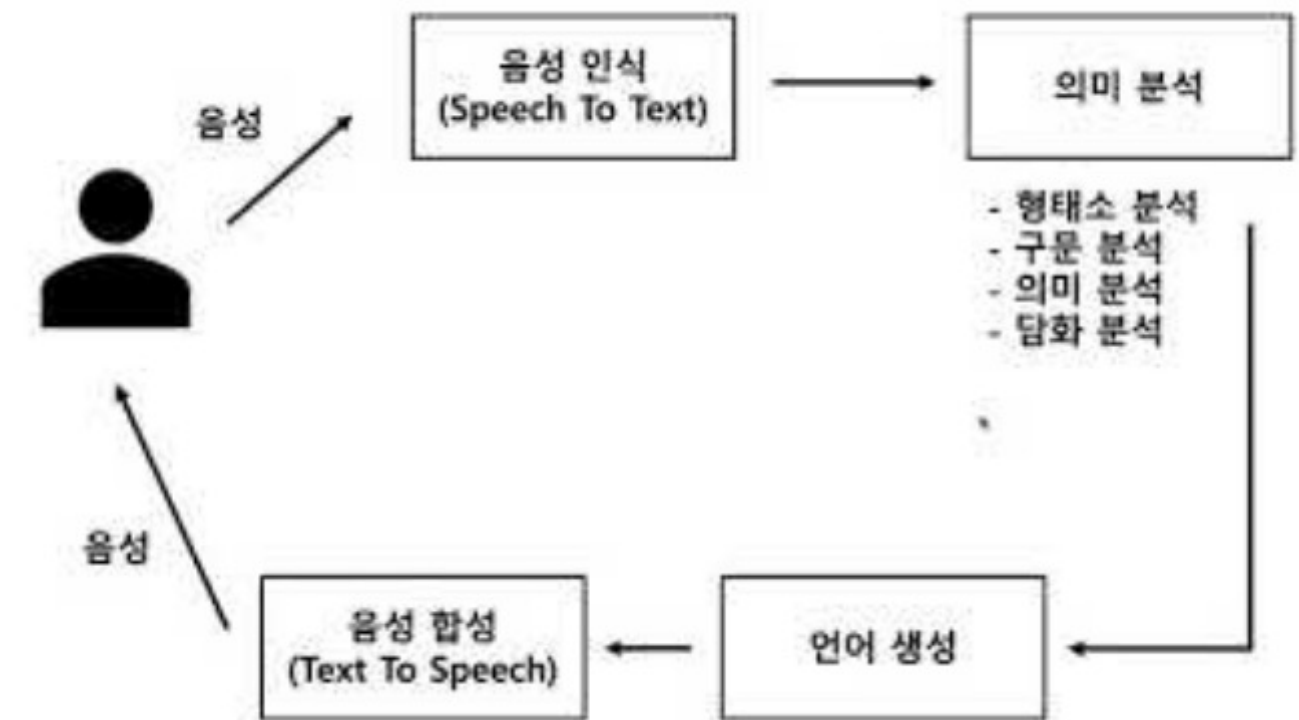
TEXT TO SPEECH 의 약자로 음성을 합성하여
원하는 텍스트로 읽어 주는 기능을 의미 함
eg. 스마트폰에서 설정 시 문자가 오면 문자를 읽어주는 기능

A. 기존의 TTS

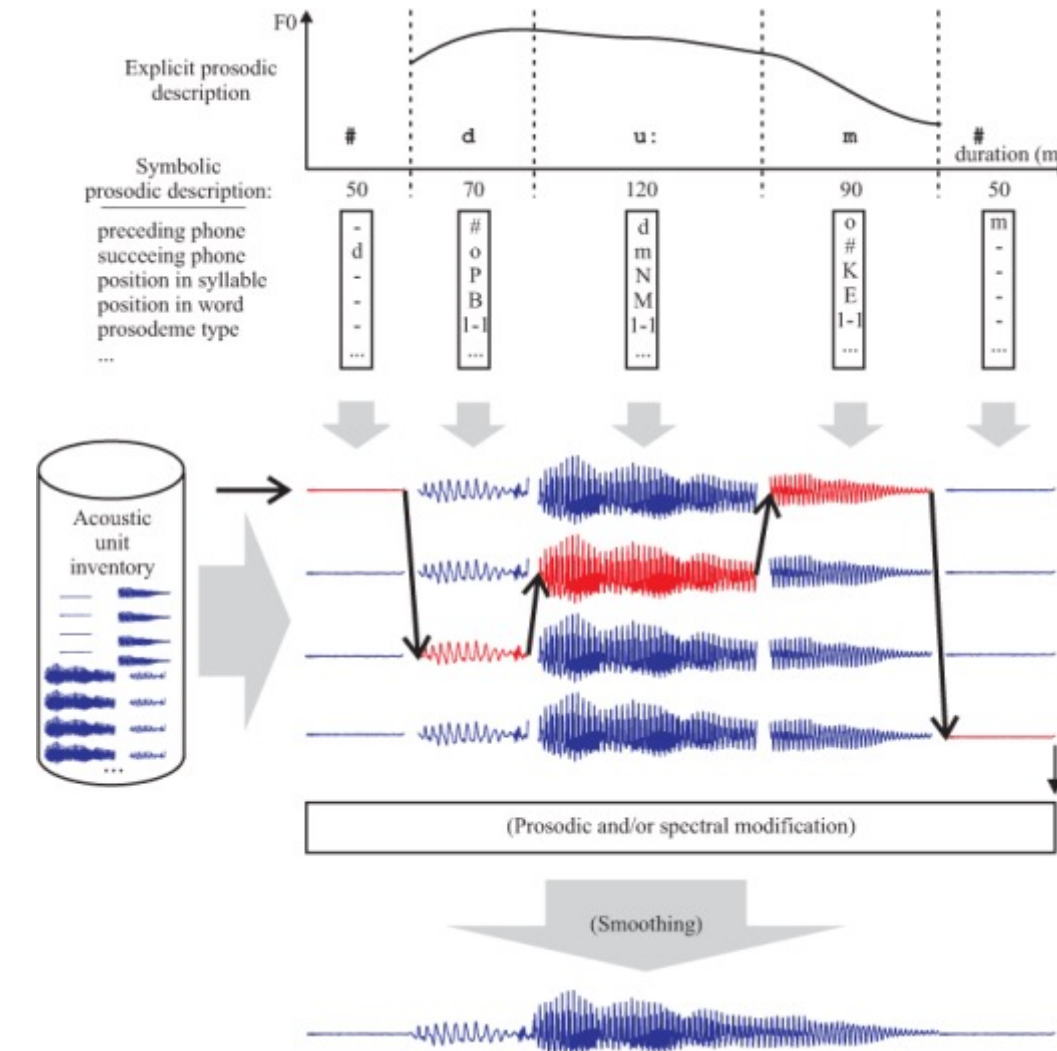
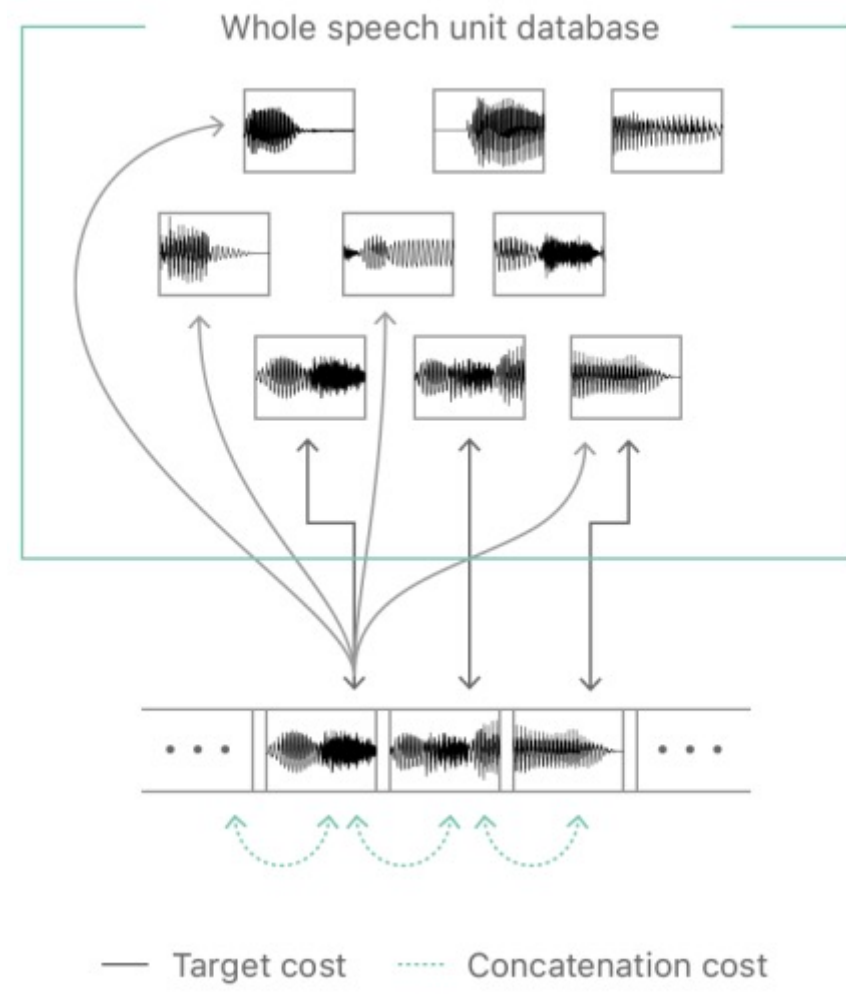


Concatenation 방식

B. 딥러닝 기반의 TTS

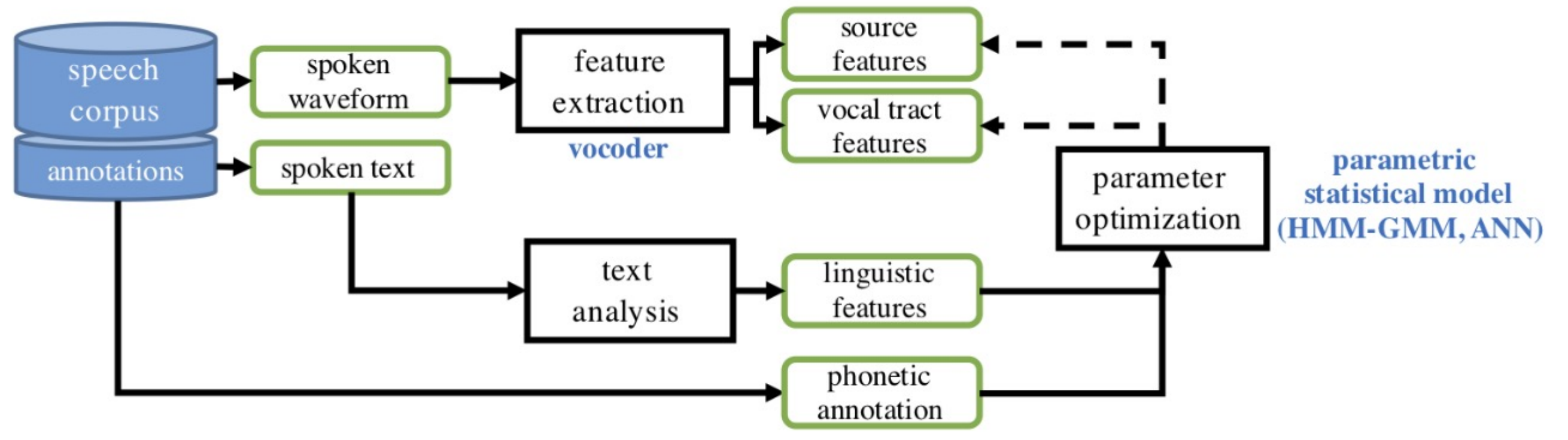


Fast Speech2 (딥러닝) 방식



치명적인 단점

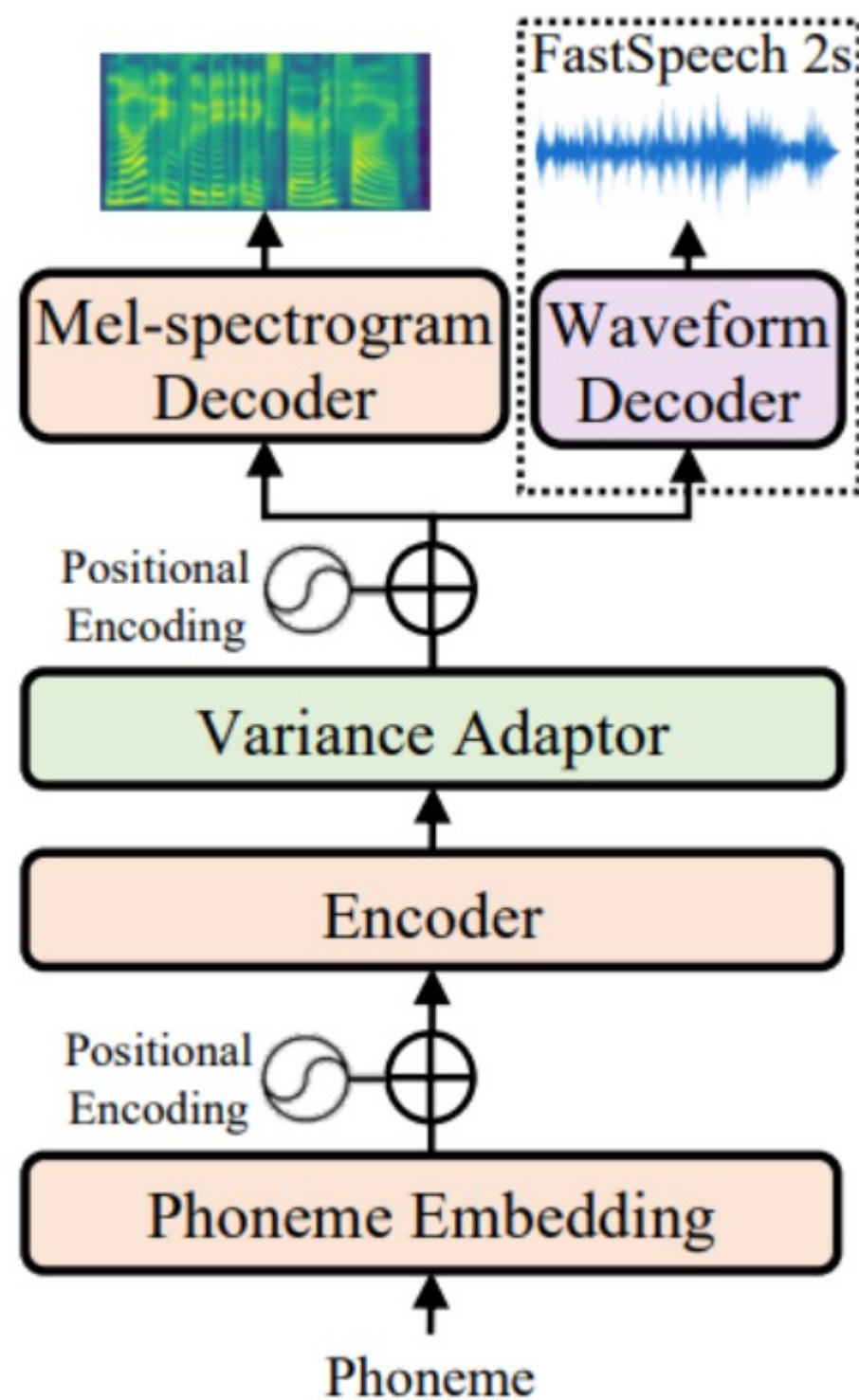
1. 데이터베이스에 들어있는 목소리로만 구현 가능
2. 부자연스러운 억양



장점

1. 언제든지 원하는 사람의 목소리로 구현 가능
2. 실제와 유사한 억양

FAST SPEECH2



(a) FastSpeech 2

FASTSPEECH 2: FAST AND HIGH-QUALITY
END-TO-END TEXT TO SPEECH

System Overview

Input

Text

"김영섭 교수님 사랑합니다"

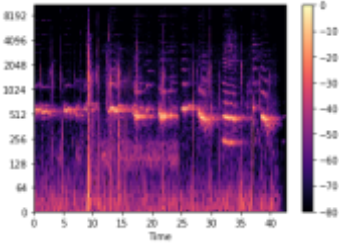
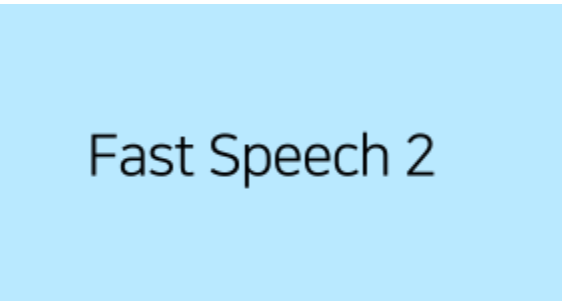


Phoneme

"ㄱ | ㅁ | ㅈ | ㅇㅅㅈ | ㅂ | ㄱㅍㅅㅈ..."



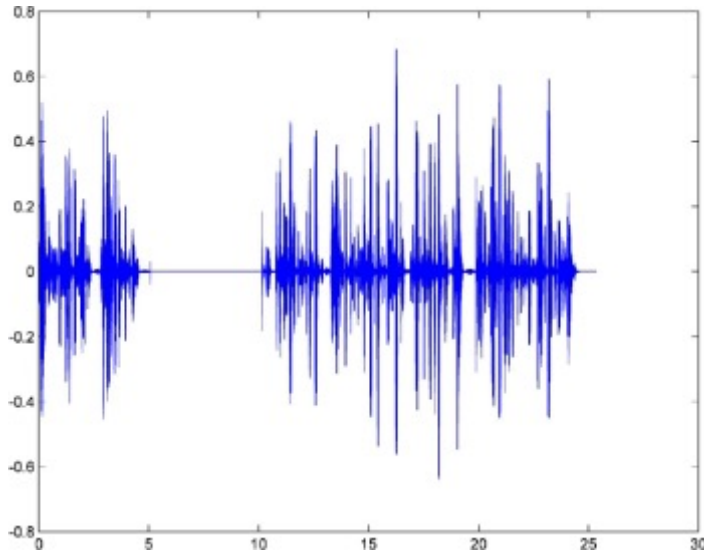
TTS System



Mel-spectrogram



Output

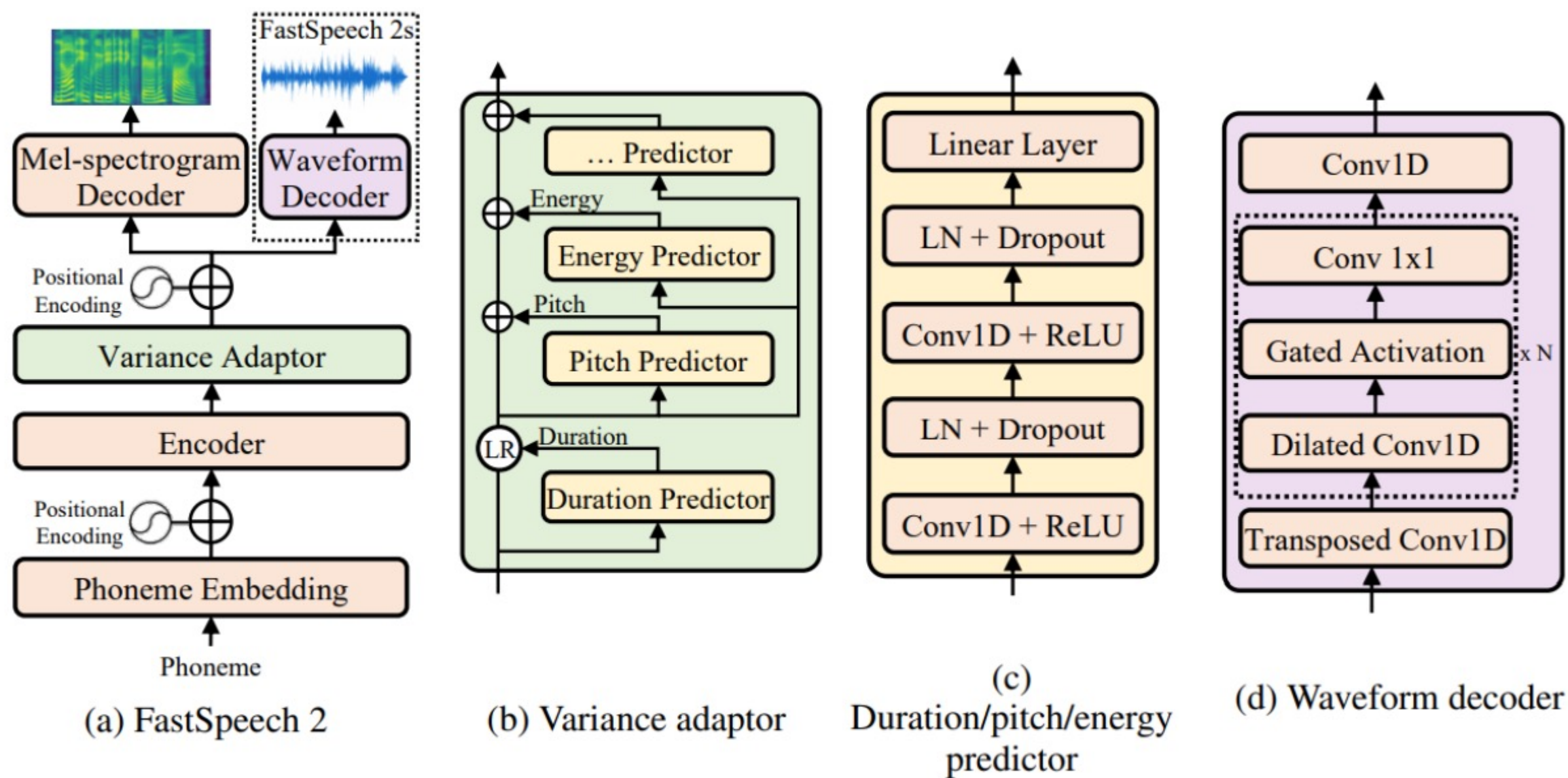


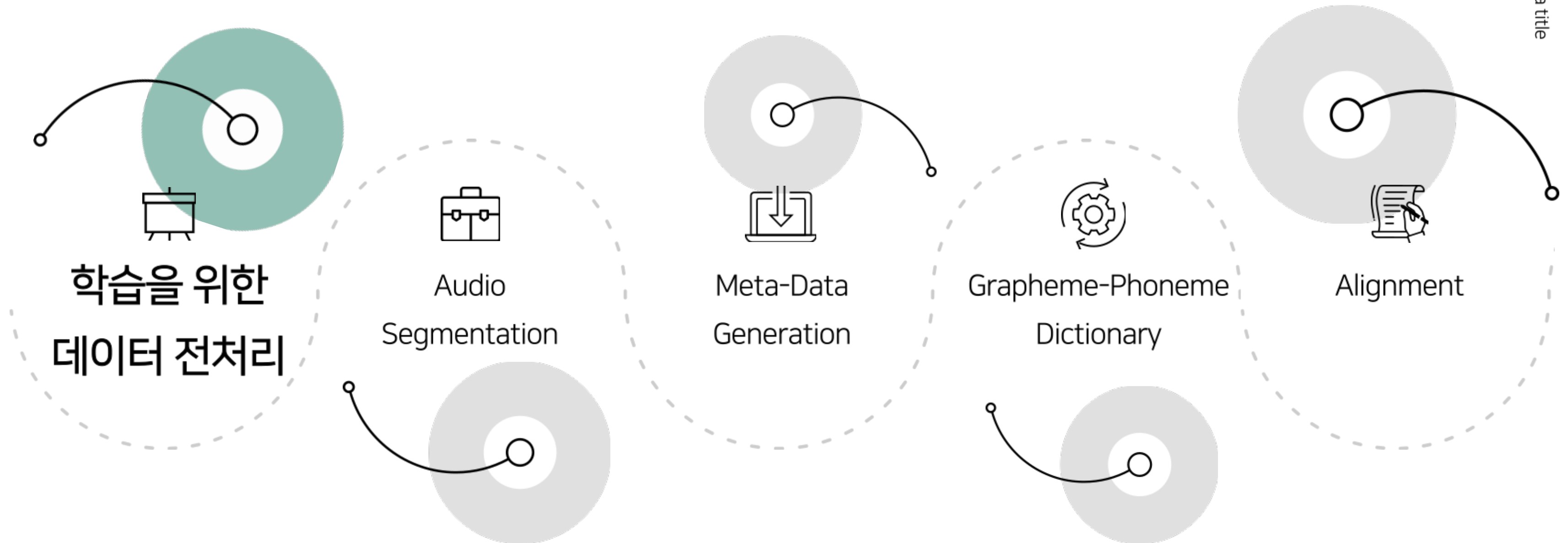
Audio

Mel-spectrogram Generator

Vocoder

Model Architecture





#04 한계점 및 보완점

한계점

#1 컴퓨터 사양의 부족

새로 학습시킬 데이터는 5~6시간의 분량으로
약 3.75GB의 용량이며 낮은 사양의 pc로
학습시키기에는 오랜 시간이 걸림

#2 전처리를 위한 오픈 소스의 부재

데이터를 학습시키기 위한 오픈 소스는 존재하지만
손쉽게 음성을 자르고 자른 음성에 맞게 LABELING 하는
오픈소스가 없어 비전문가가 다루기에는 어려움이 있음

보완점

#1 적절한 오픈소스 개발

음성데이터 만으로 음성 SEGMENTATION,
META DATA, GRAPHEME AND PHONEME
DICTIONARY 등을 만들고 ALIGNMENT 까지
한번에 쉽고 간단하게 되는 것이 존재한다면
누구나 쉽게 접근 할 수 있을 것이다.

#2 원하는 데이터 학습

김영섭 교수님의 목소리로 구현해 보기 위하여 텍스트와 음
성을 모두 추출하였지만 2와 같은 문제로 전에 실패하였기
에 PRE-TRAINED 가 아닌
USER-TRAINED 데이터의 전처리 문제 해결의 필요성

#05 프로젝트 의의

#1 개인화 음성학습 가능

기존과 달리 전문가가 이미 구축한 제한적인 목소리 뿐만 아니라
필요한 사람에 따라 원하는 목소리로 구현 가능

#2 기술을 통한 위로

돌아가신 부모님의 목소리를 복원하기, 목소리를 낼 수 없는 사람의 사전 기록된
목소리를 활용한 간접적 치유 등과 같이 인공지능을 활용하여 인간을 도움

#3 딥러닝을 통한 기술의 급진적 발전

딥러닝을 통한 TTS의 급진적인 발전과 같이 다양한 분야에서 기술 적 문제
로 인간과 유사하게 구현하지 못했던 부분을 손쉽게 해결 가능