Clova

사용자 적응형 WaveNet Vocoder

Clova Voice

Clova

10분 음성으로 TTS 만들기

Clova Voice

Clova

발표내용

Text-to-speech (TTS) 란 기계가 사람처럼 **텍스트를 읽어주는** 기술입니다. 최근 각광받고 있는 **딥러닝** 기반의 TTS 시스템은 **WaveNet** 라는 neural network 보코더를 사용합니다. 일반적으로 고품질 WaveNet 모델을 학습시키기 위해서 **수시간의 음성** 데이터가 필요합니다. 과연 **적은양의 음성** 데이터로 (ex. **10분**) 고품질 WaveNet을 만드는게 가능할까요?

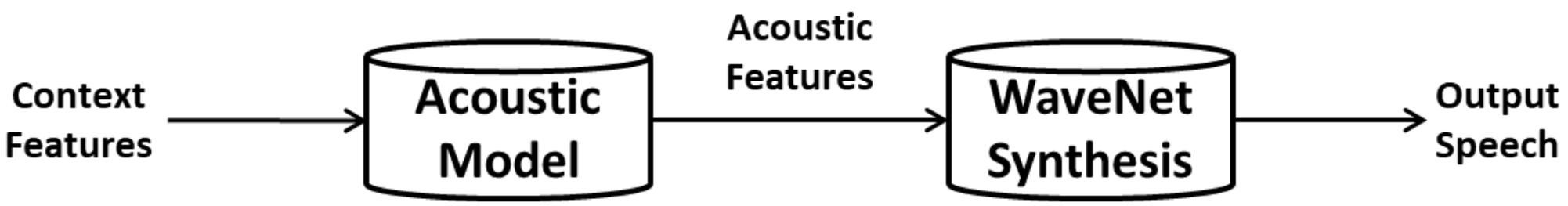
본 발표에서는 **"저비용 but 고품질" WaveNet TTS**를 구축하기 위한 아래의 두가지 핵심 기술에 대해 설명드리고자 합니다.

ExcitNet : WaveNet 합성기 성능 끌어올리기

Speaker Adaptation : 저비용 합성기 만들기

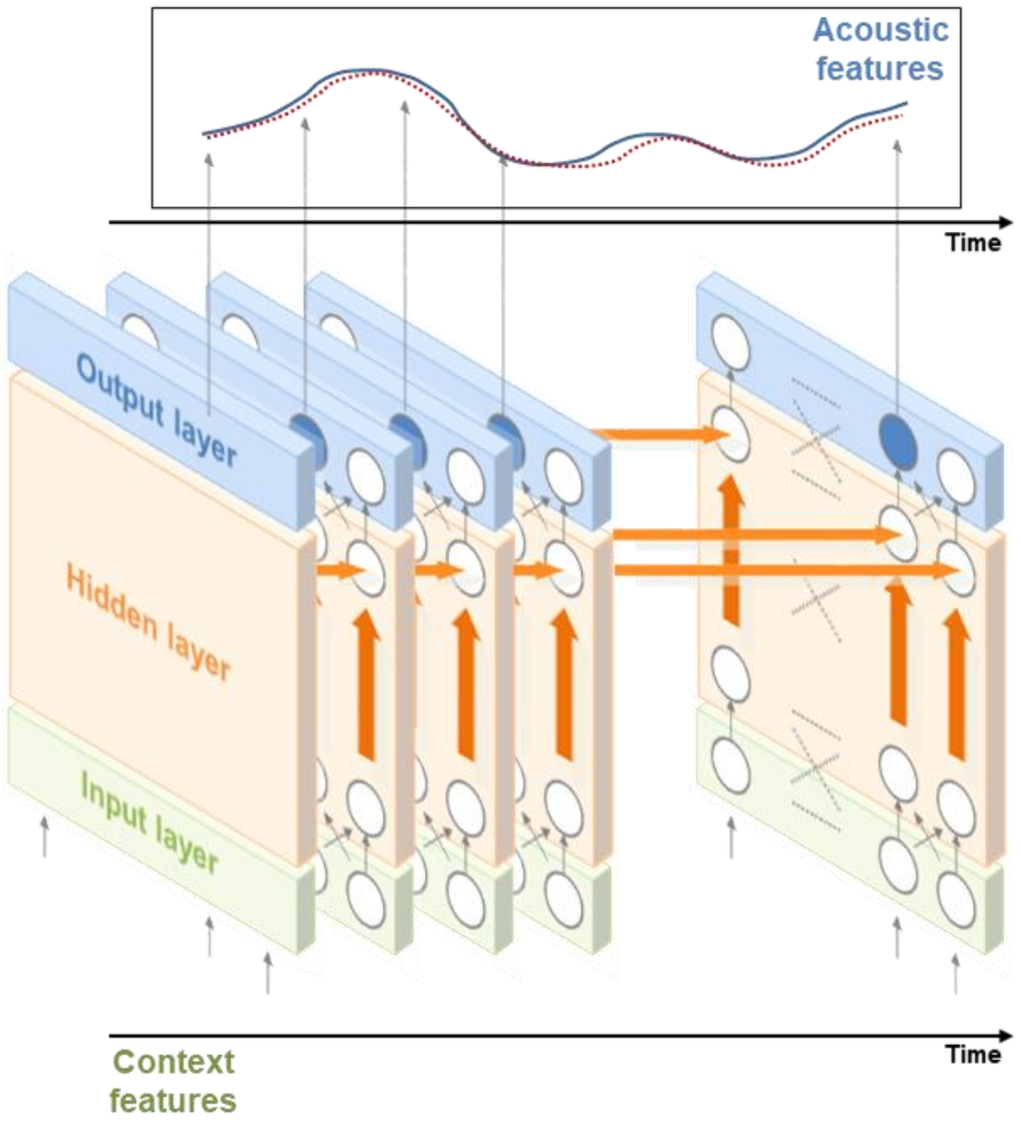
Overview

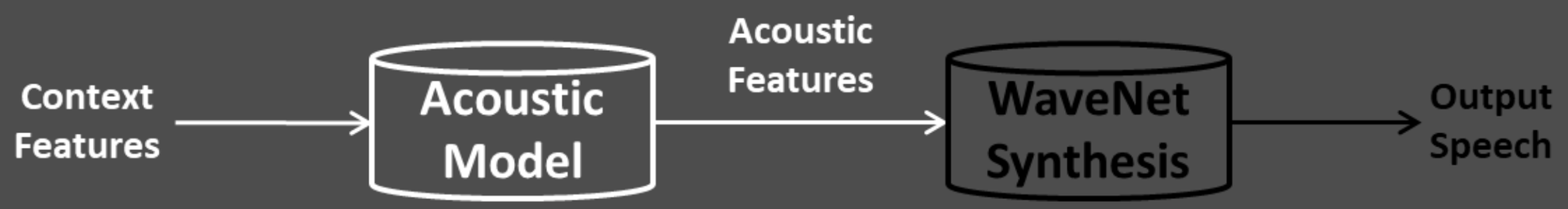
딥러닝 기반 음성합성 시스템
통계 모델을 활용하여 음성 신호 생성



Overview

Clova

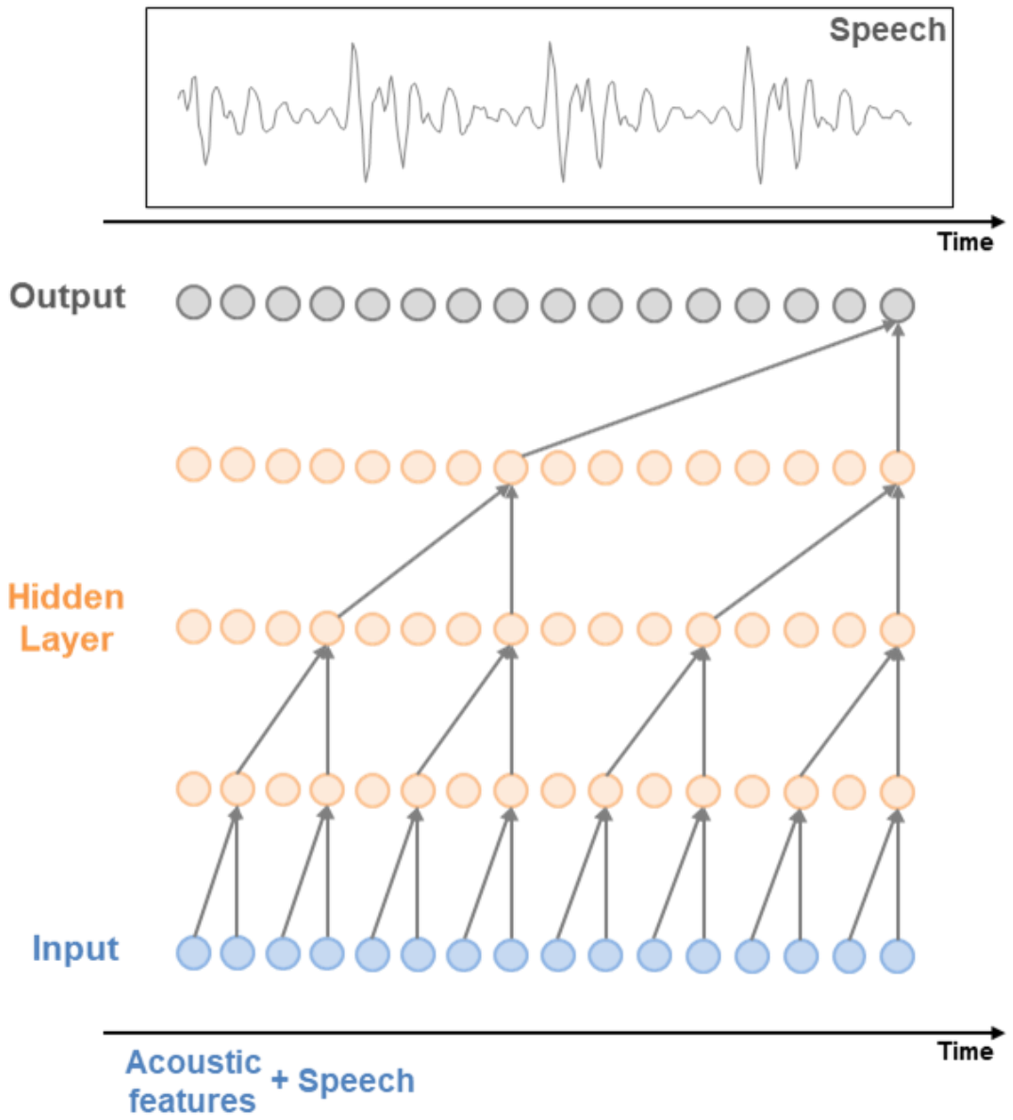


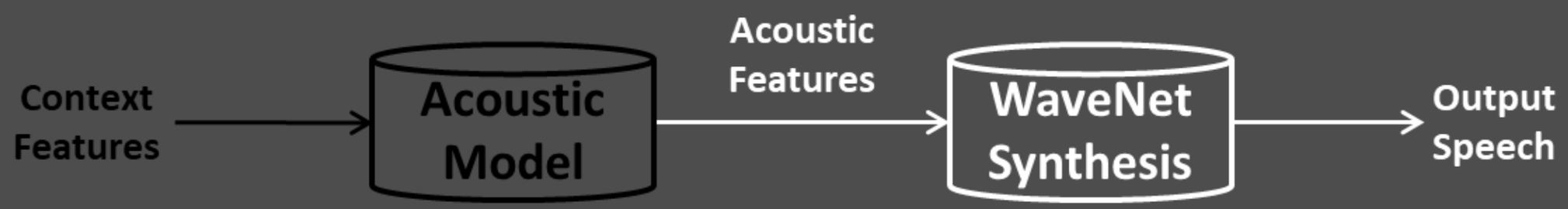


음성 파라미터 추정

Overview

Clova





음성 생성

Clova

문제점

위에서 설명드린 **" Acoustic Model + WaveNet Synthesis"** 방법은 일반적인 DNN-TTS 시스템에서 사용되는 대표적인 알고리즘 들입니다.
다만, 두가지 모델을 학습시키기 위한 **"충분한 양의 음성데이터"** 가 있어야 고품질의 합성음을 생성할 수 있다는 제약이 있습니다.

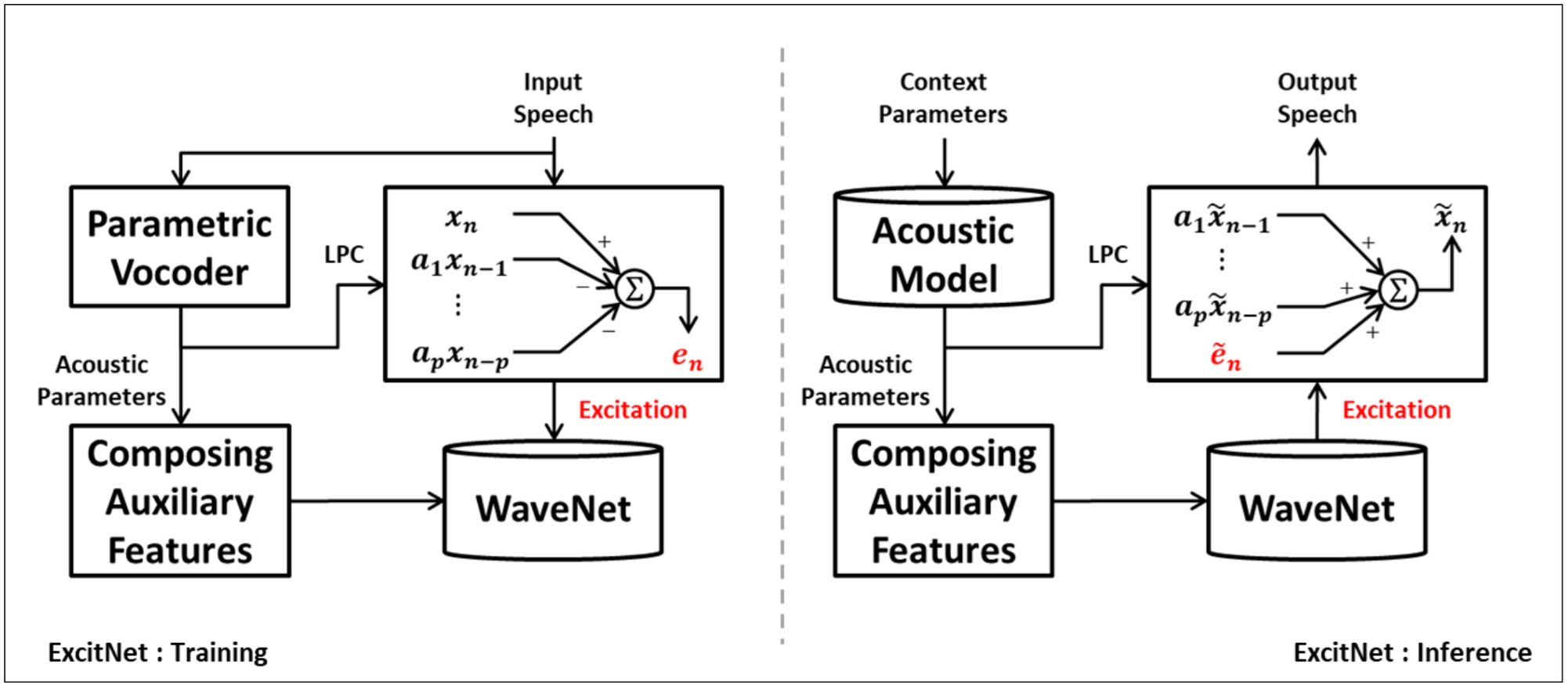
본 발표에서는 **"저비용 but 고품질" WaveNet TTS**를 구축하기 위한 아래의 두가지 핵심 기술에 대해 설명드리고자 합니다.

ExcitNet : WaveNet 합성기 성능 끌어올리기

Speaker Adaptation : 저비용 합성기 만들기

Clova

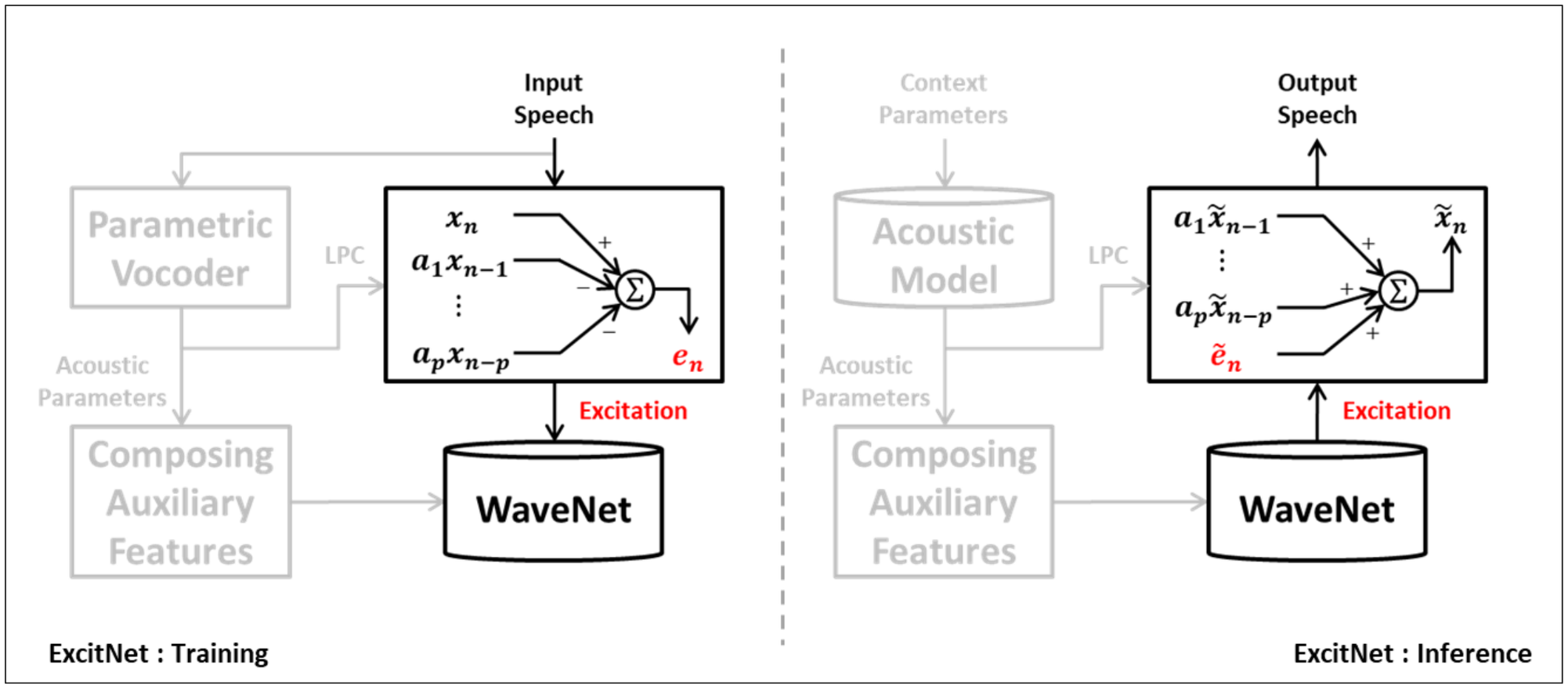
해결방법



ExcitNet : WaveNet 합성기 성능 끌어올리기

Speaker Adaptation : 저비용 합성기 만들기

해결방법

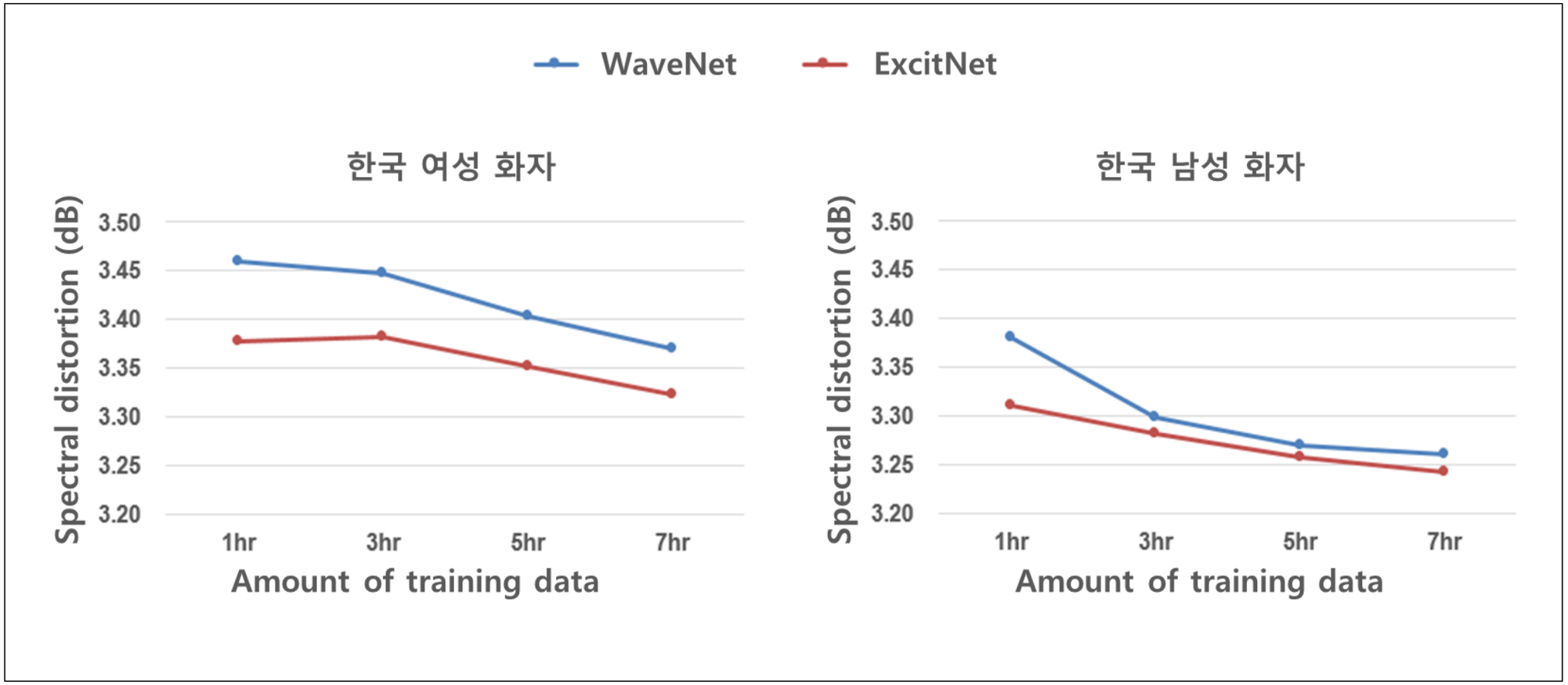


Linear Prediction 적용

WaveNet Vocoder 추정 성능 향상

Clova

해결방법



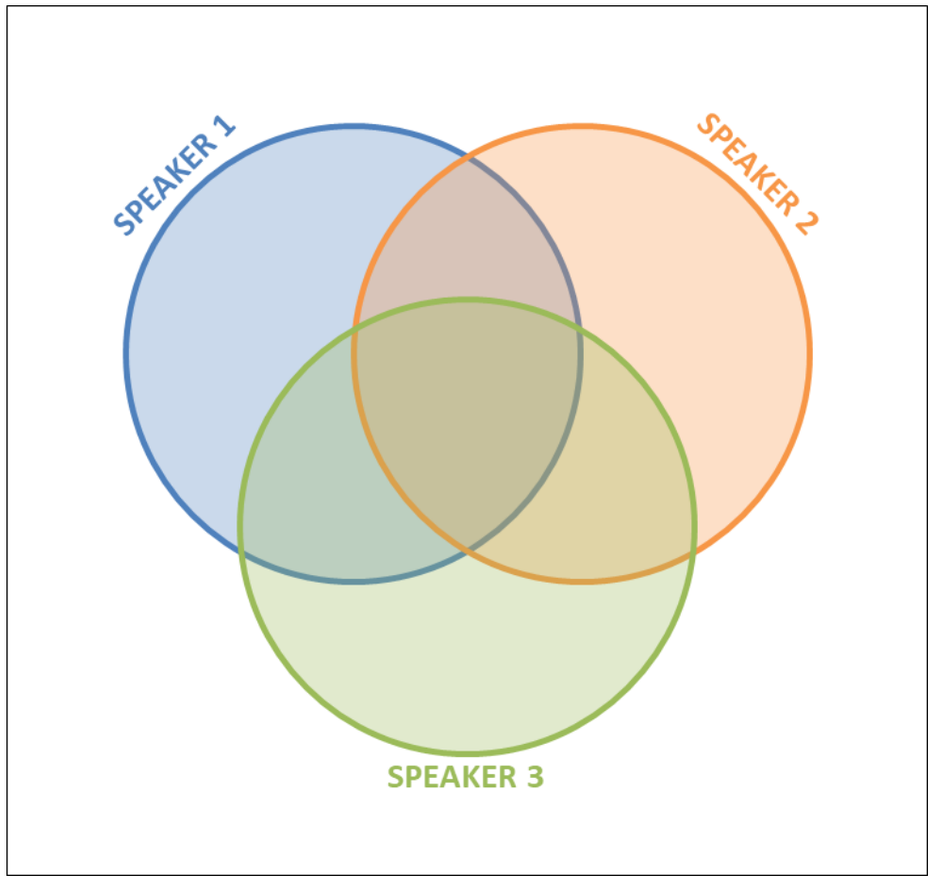
http://arxiv.org/abs/1811.04769

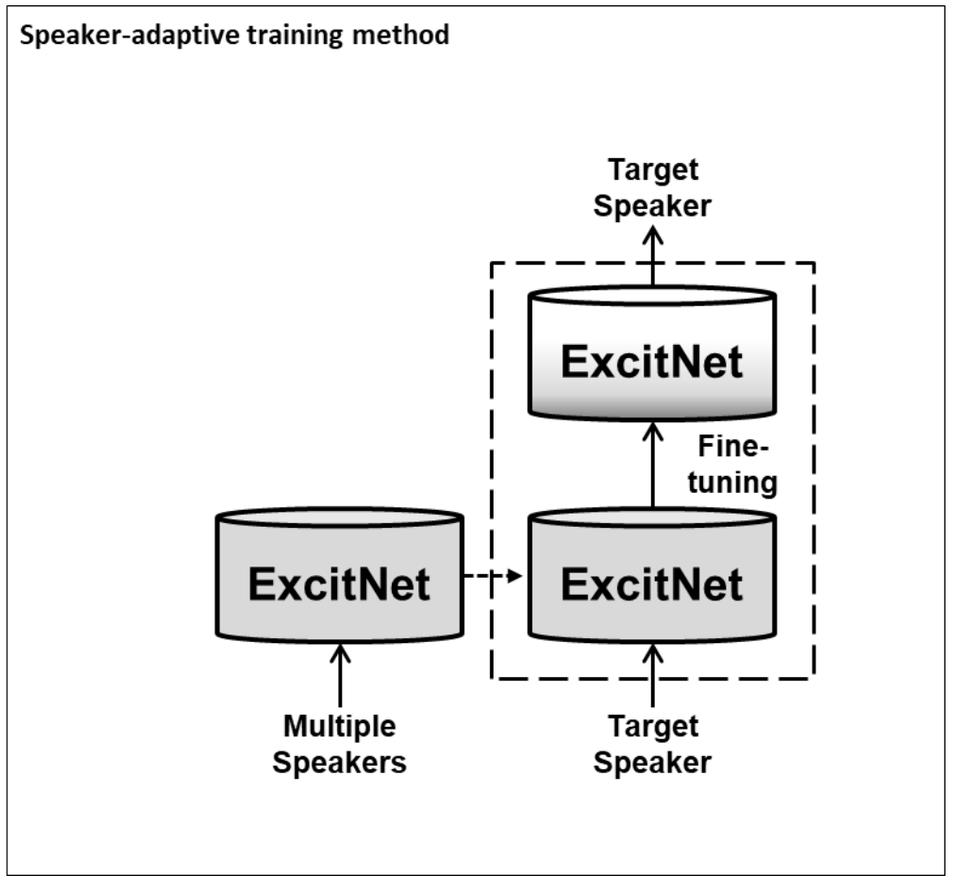
Linear Prediction 적용

WaveNet Vocoder 추정 성능 향상

Clova

해결방법



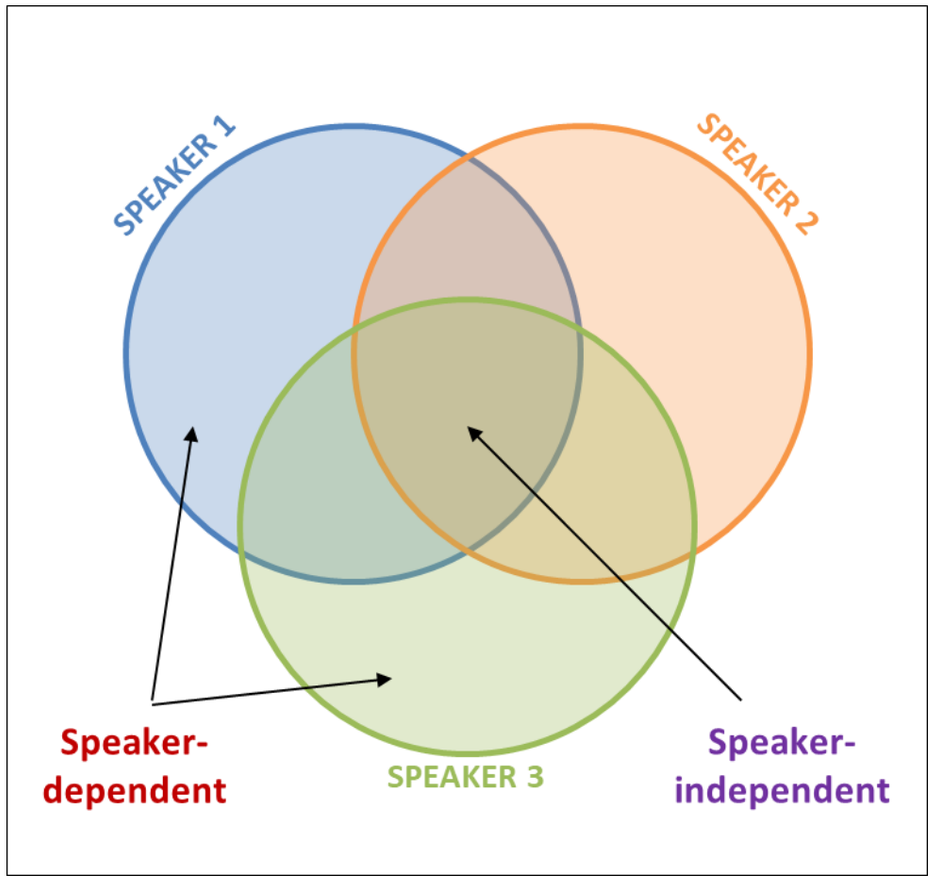


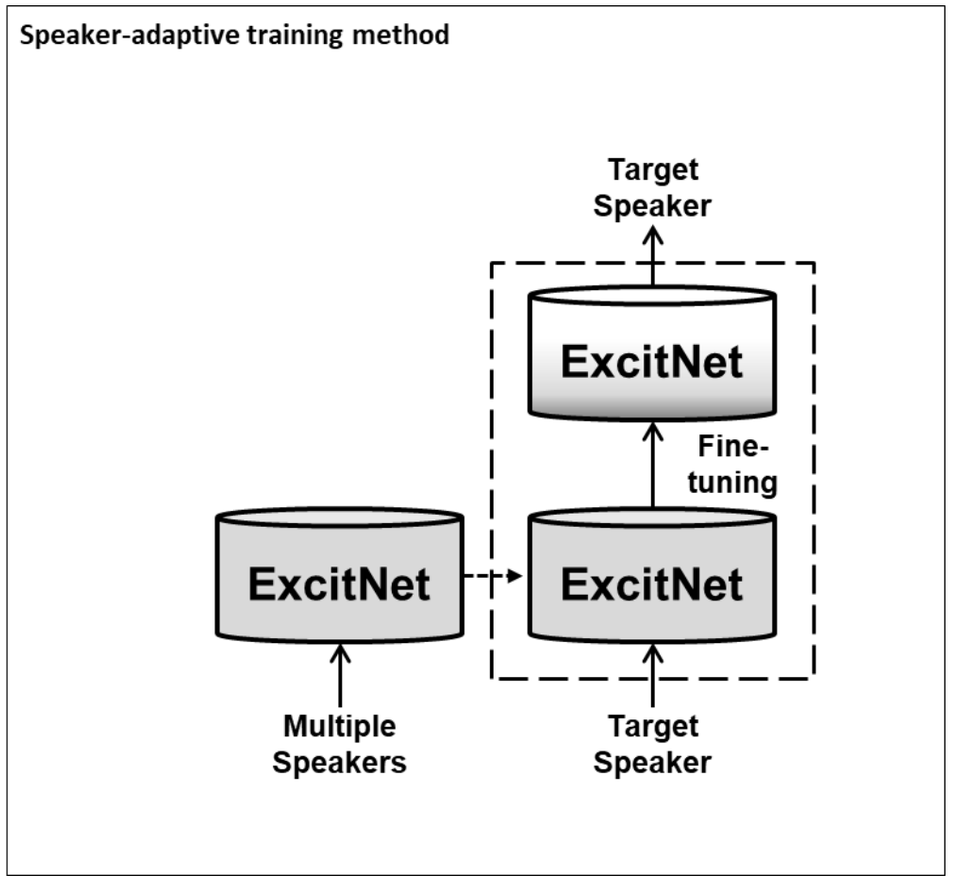
ExcitNet : WaveNet 합성기 성능 끌어올리기

Speaker Adaptation : 저비용 합성기 만들기

Clova

해결방법





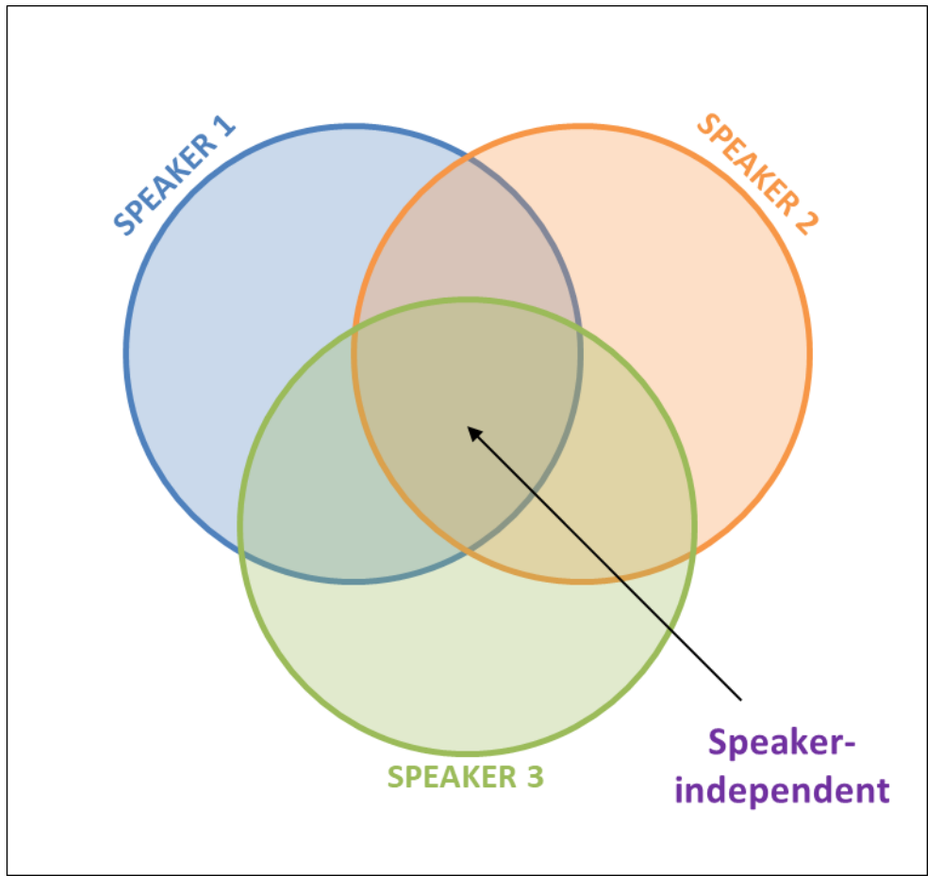
Speaker Independent

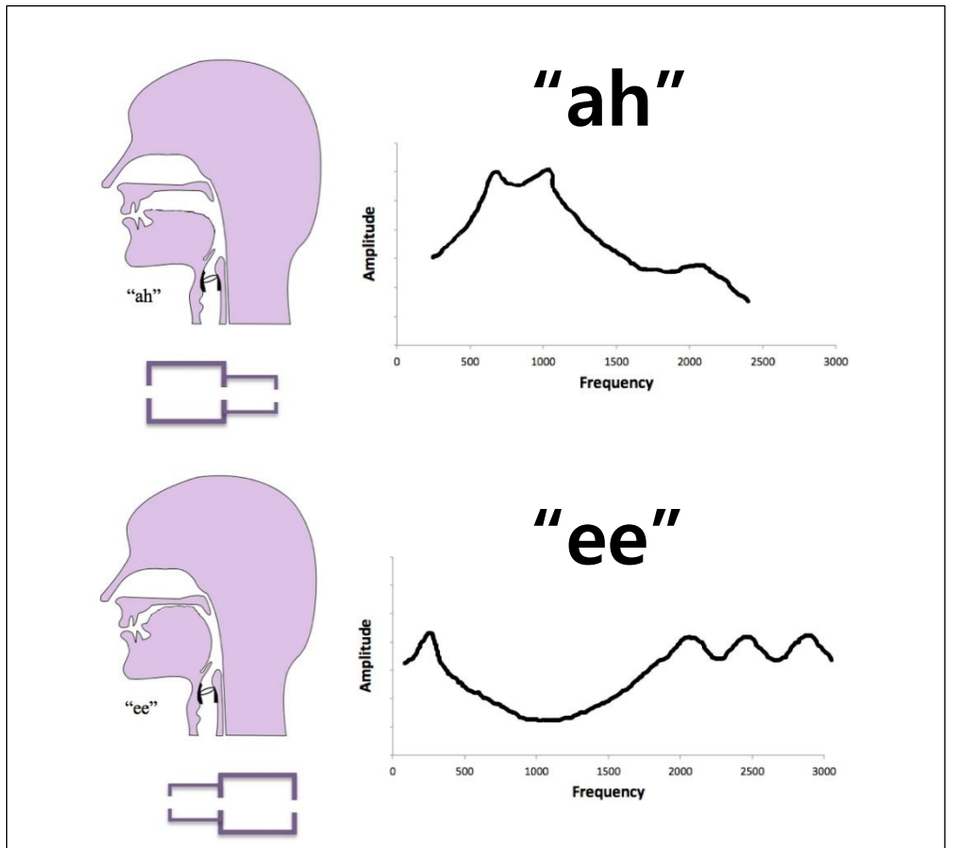
vs

Speaker dependent

Clova

해결방법



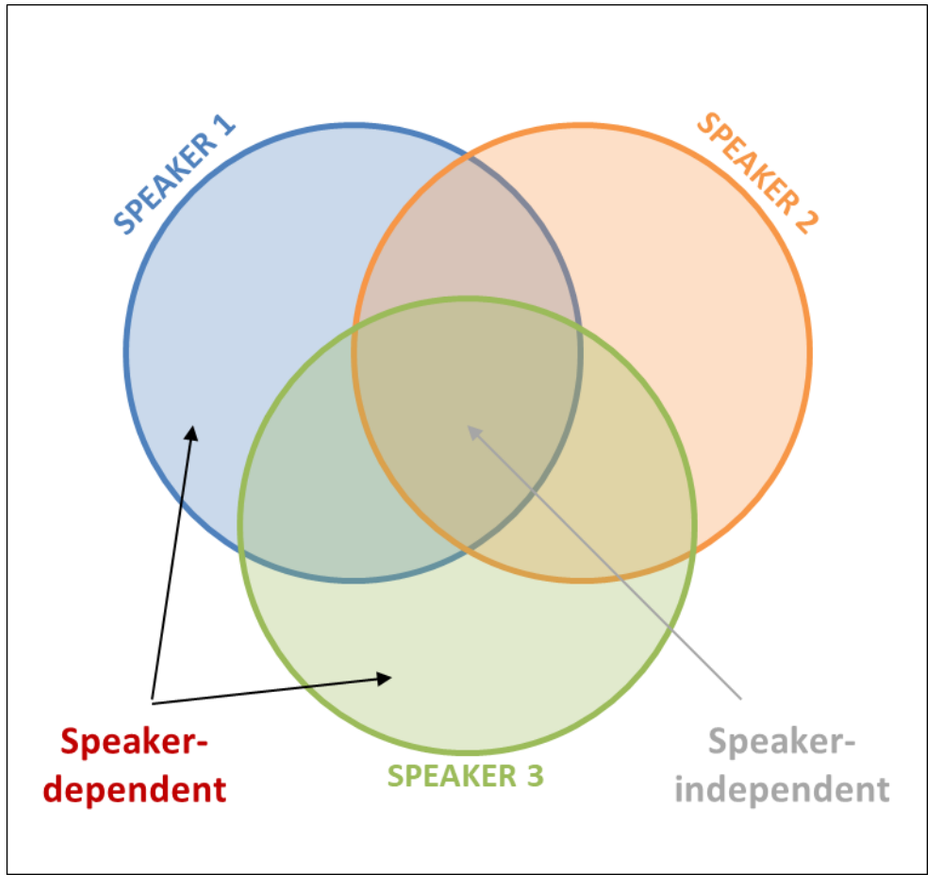


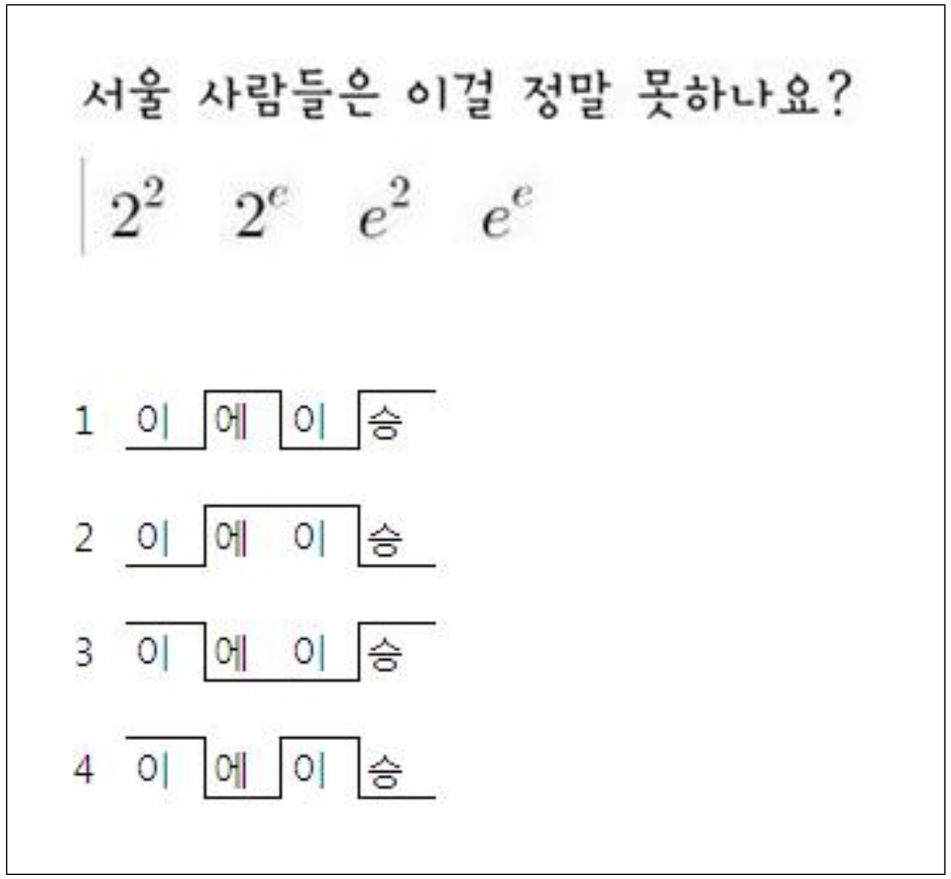
http://kyleforinash.altervista.org/S/14Voice.html

Speaker Independent

여러 화자들이 갖는 공통적인 특성

해결방법





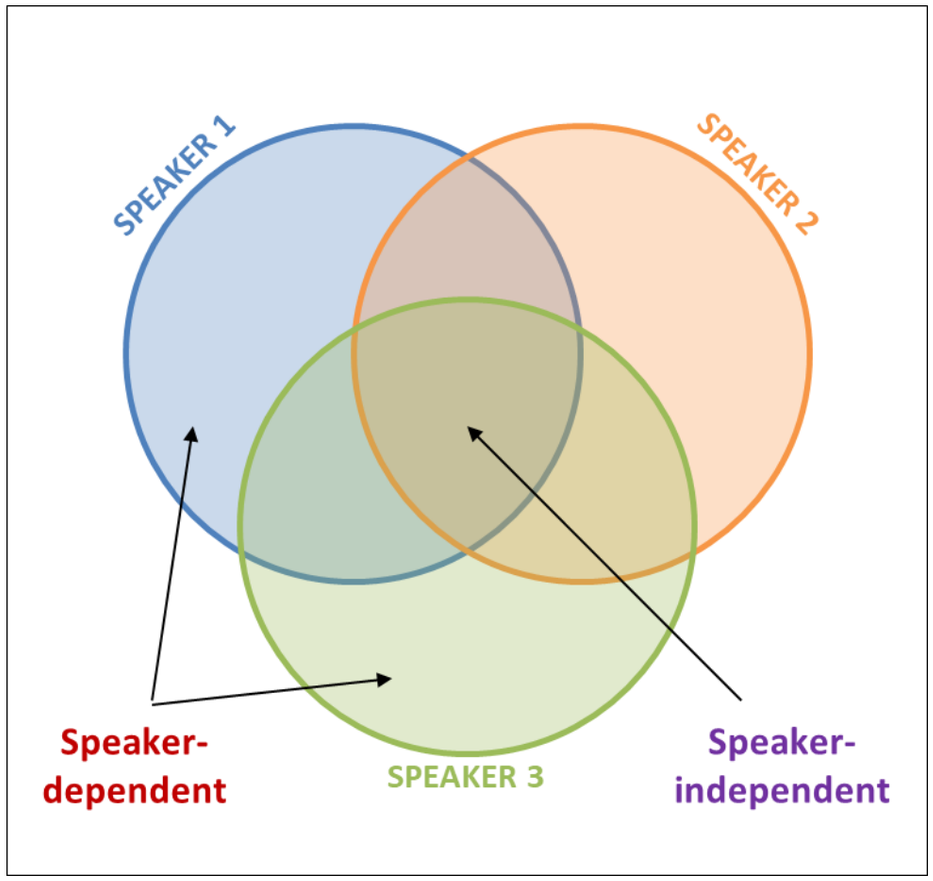
http://cywsc32.egloos.com/m/3062323

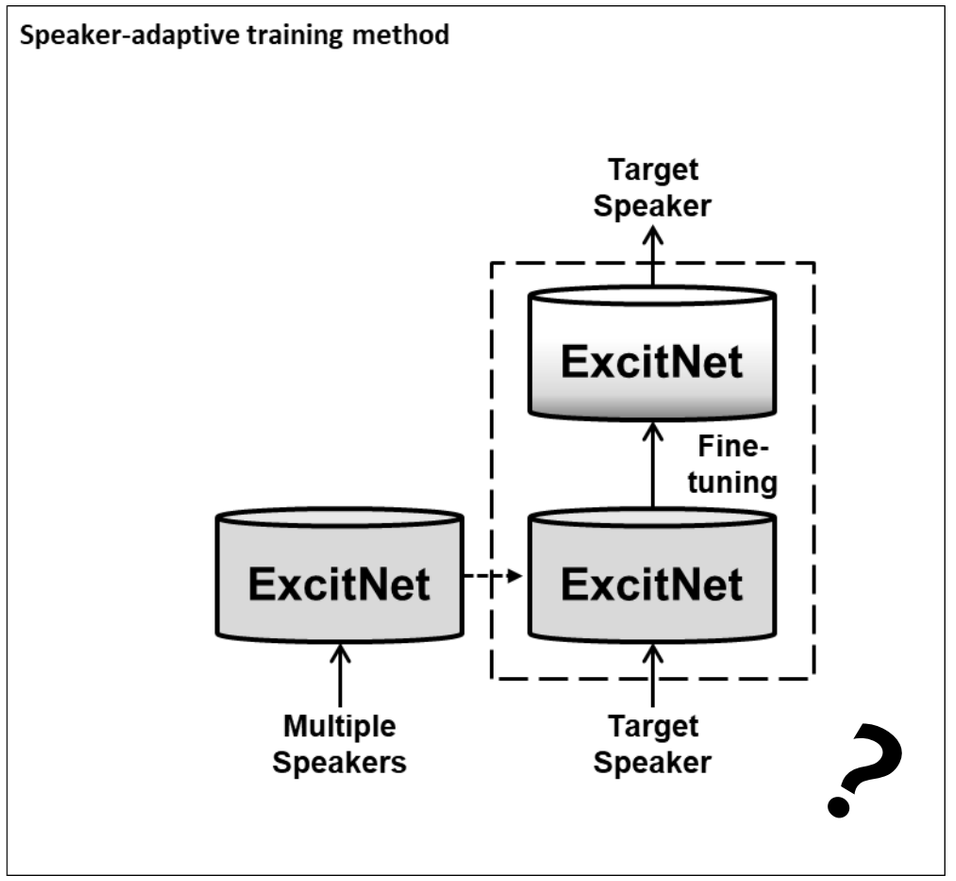
Speaker Dependent

화자 고유의 특성

Clova

해결방법

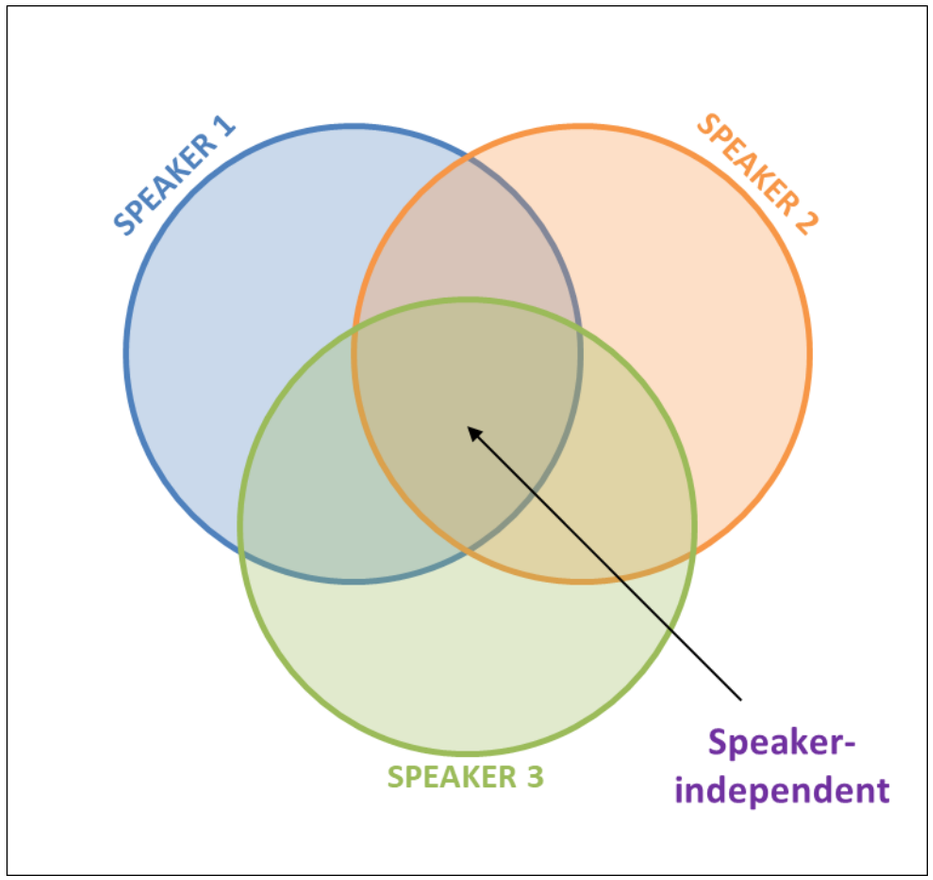


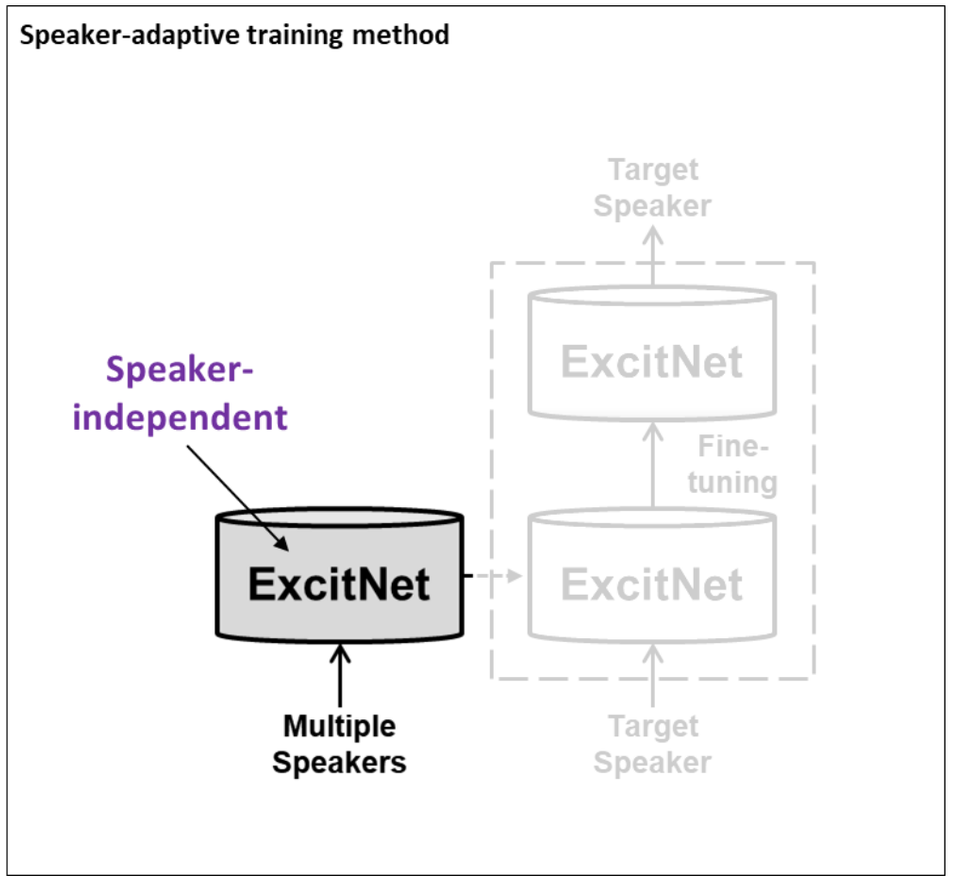


Independent vs Dependent

모델이 학습할 수 있을까 ?

해결방법



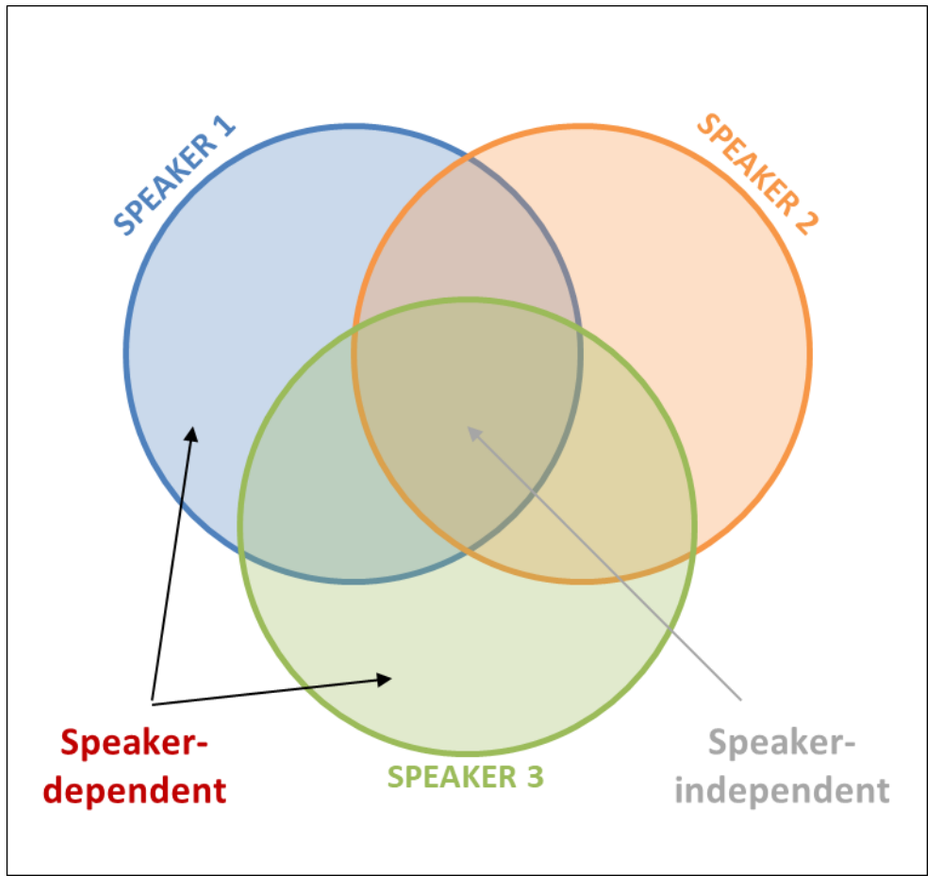


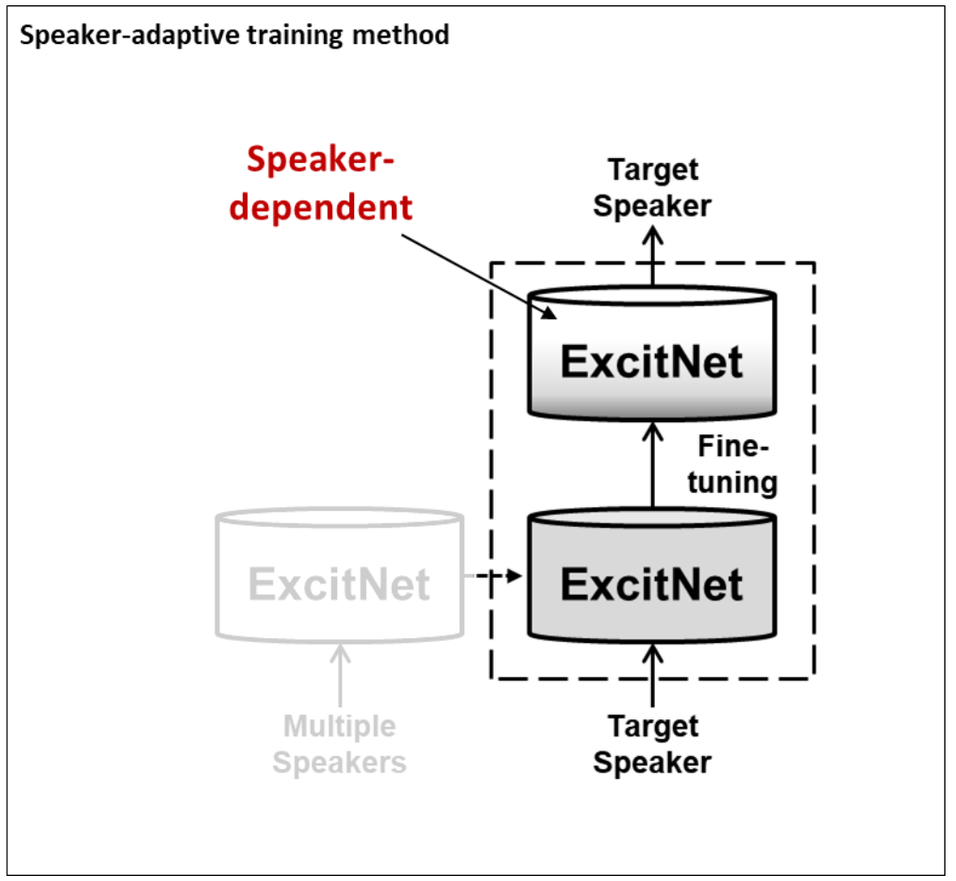
Speaker Independent Model

여러 화자들을 섞어서 모델 훈련

Clova

해결방법

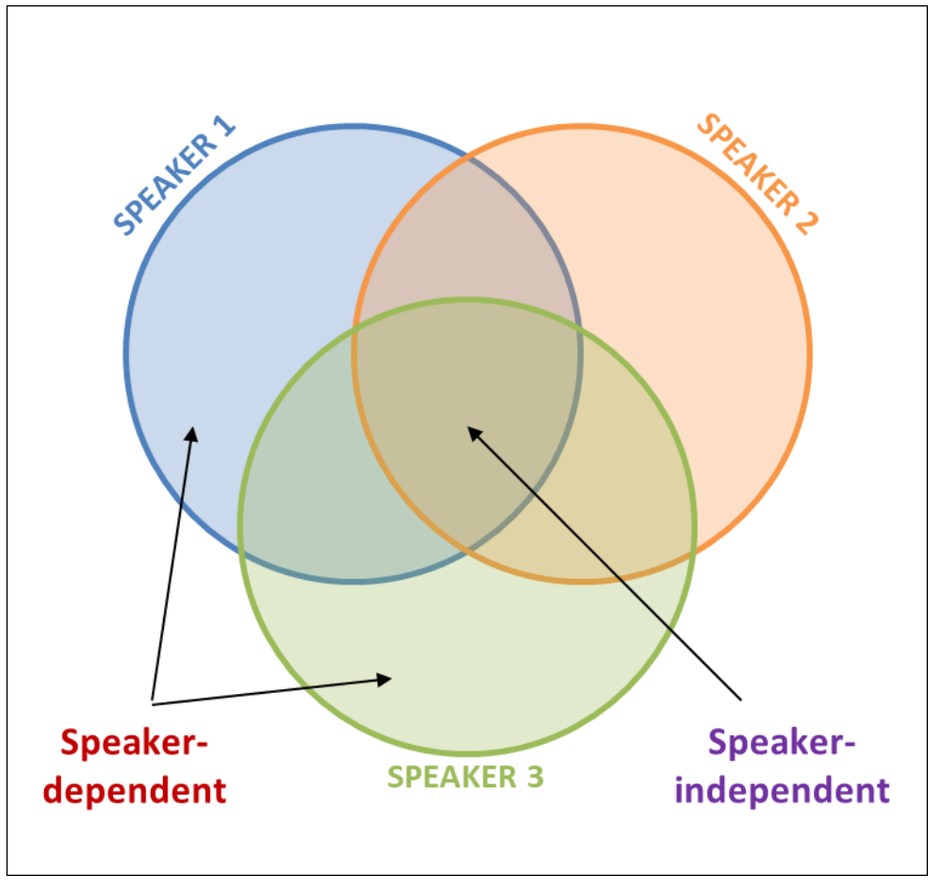


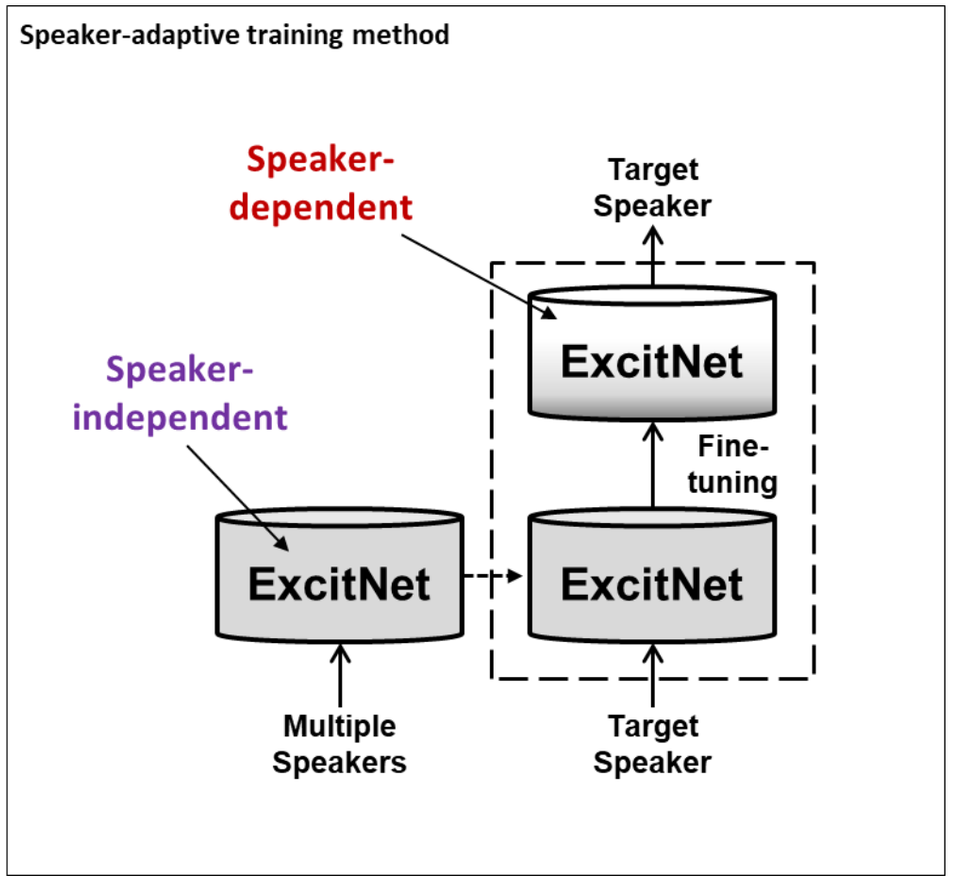


Speaker Dependent Model

특정 화자로 Adaptation

해결방법



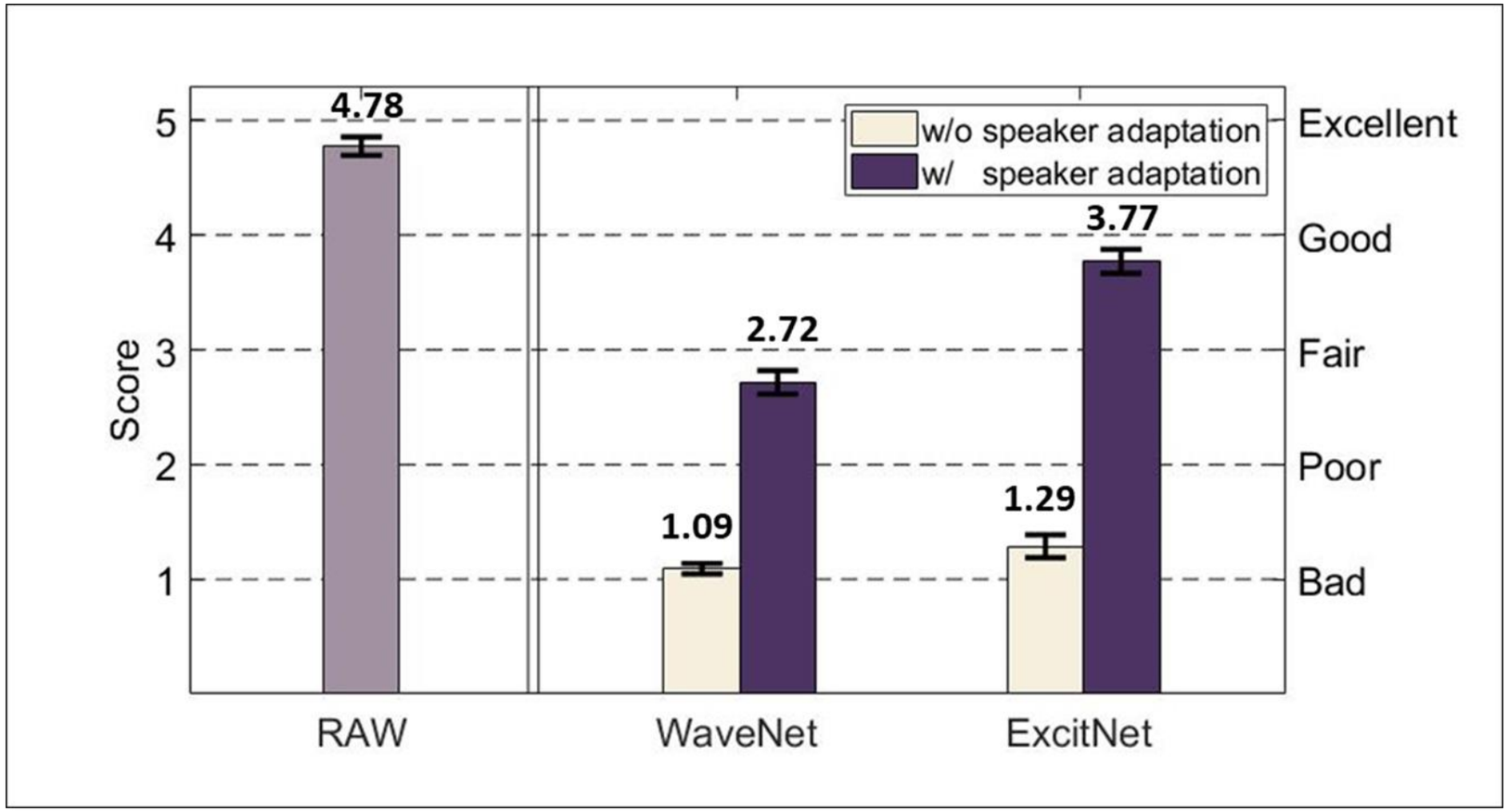


Independent vs Dependent

모델이 학습할 수 있었다.

Clova

개선효과

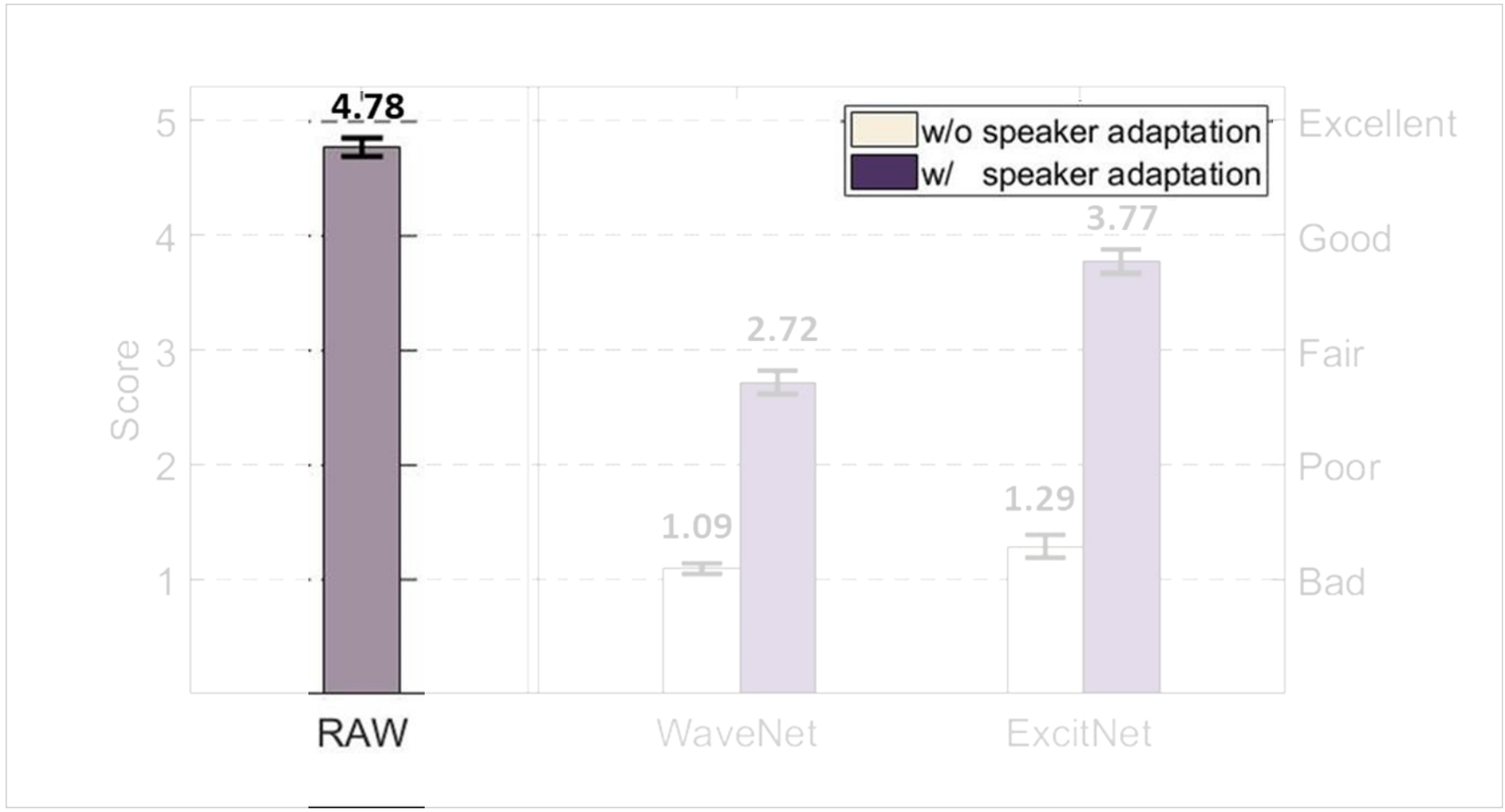


http://arxiv.org/abs/1811.03311

10분 음성으로 Adaptation

WaveNet 대비 3.4배 성능 우위 (1.09 → 3.77)

Demo

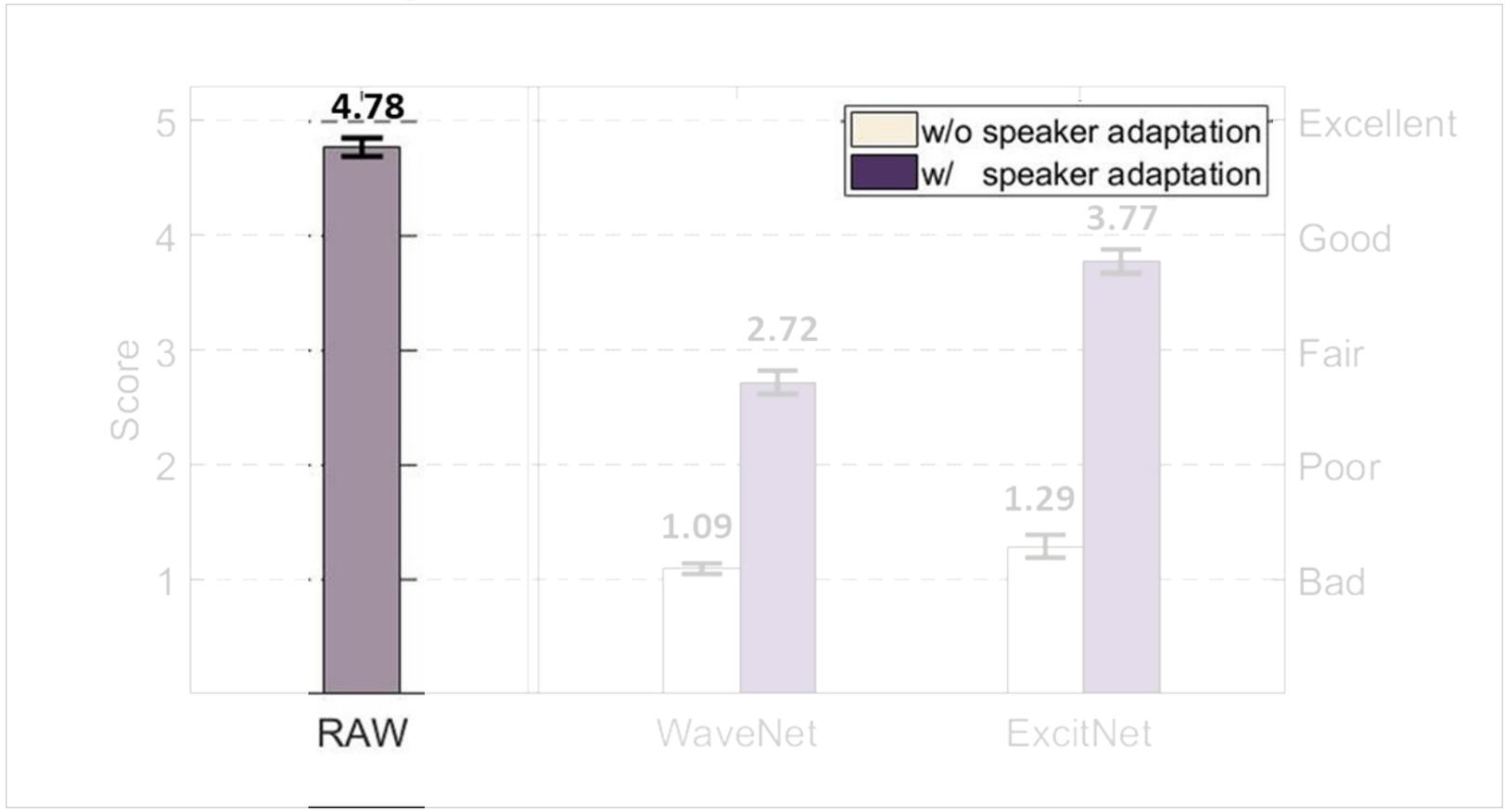


http://arxiv.org/abs/1811.03311

"DB 암호화 시장의 성장 전망은
밝은 편이다."

Clova

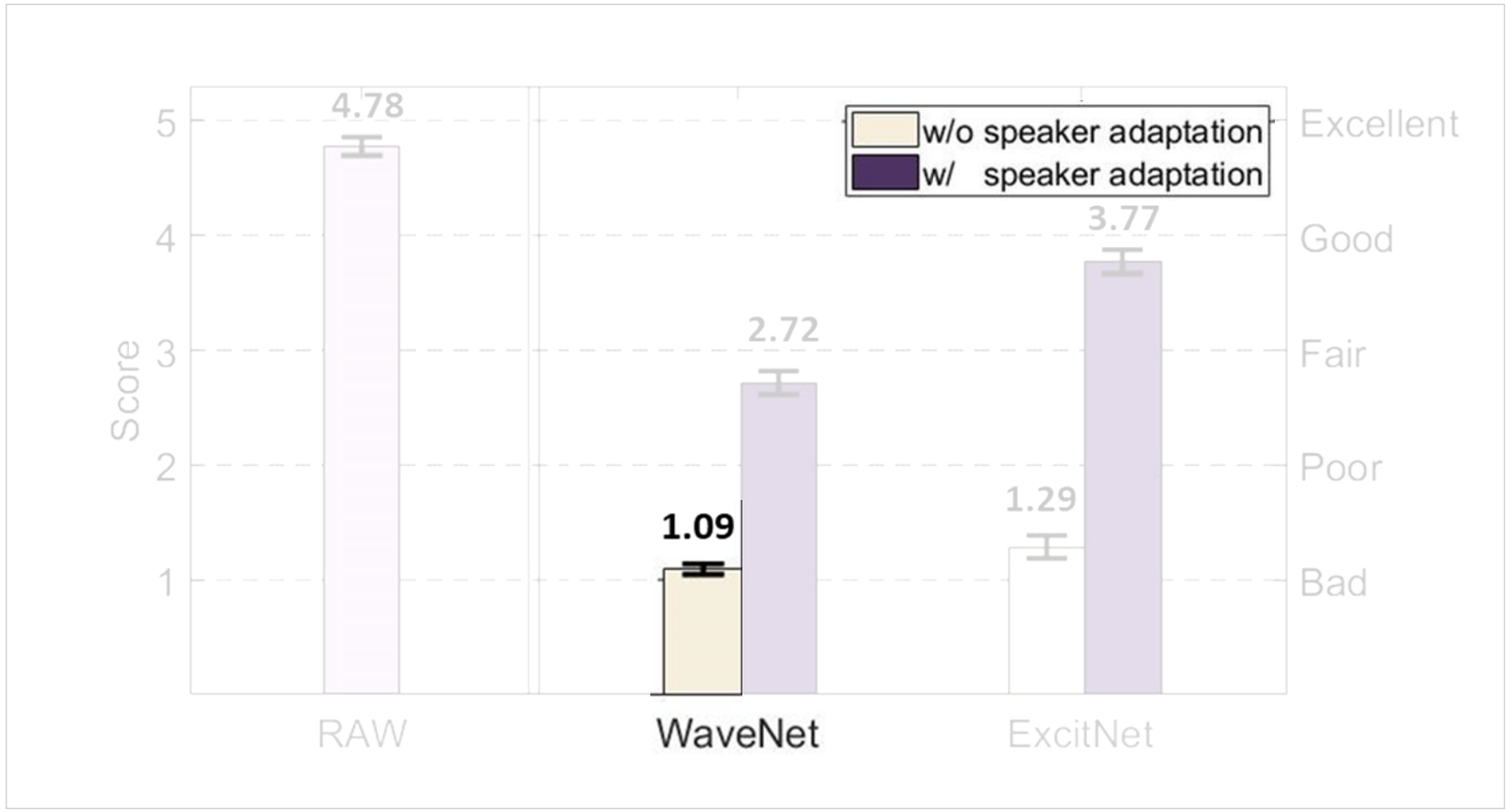
Demo



http://arxiv.org/abs/1811.03311

"DB 암호화 시장의 성장 전망은 밝은 편이다."

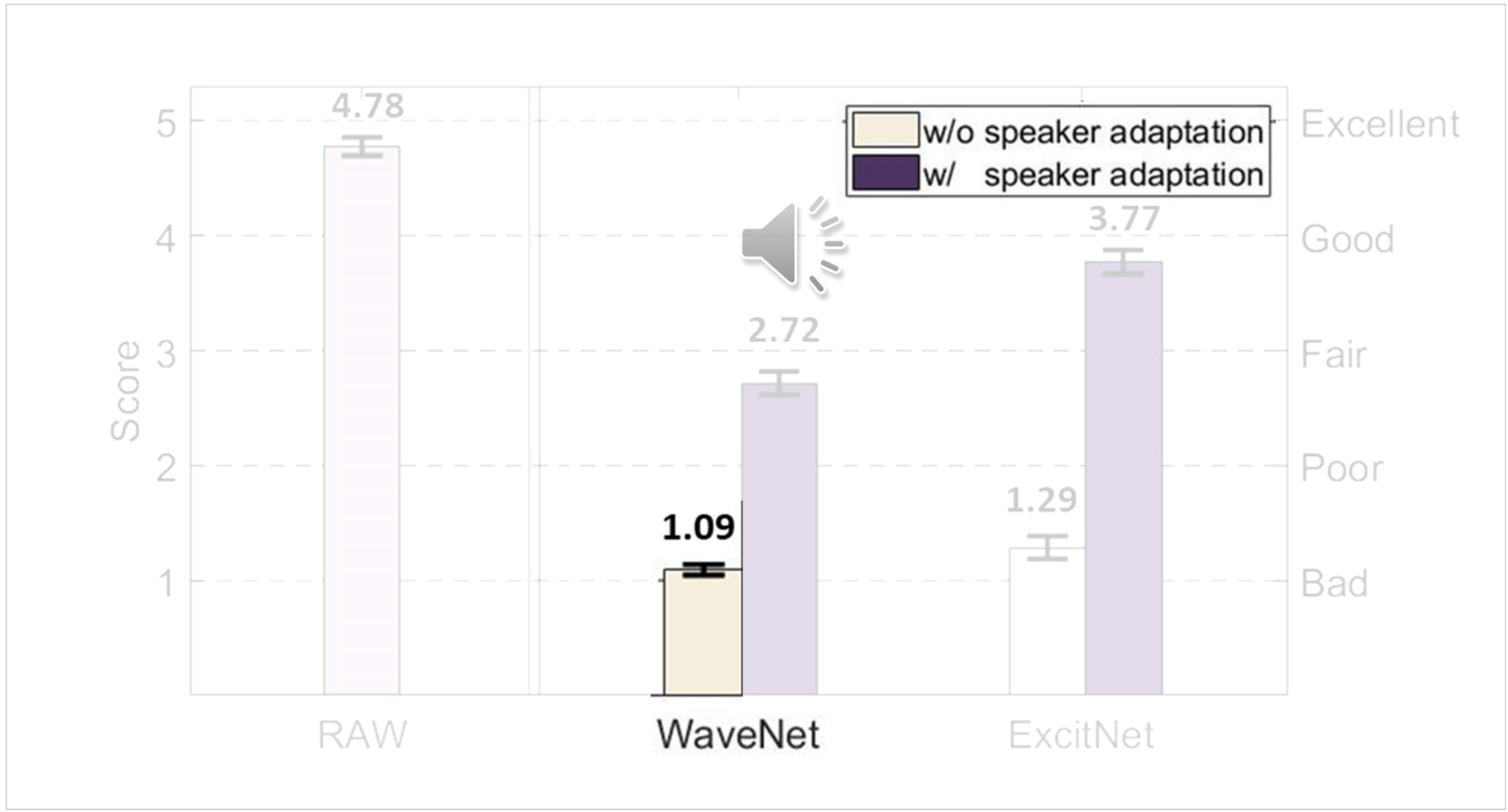
Demo



"DB 암호화 시장의 성장 전망은 밝은 편이다."

Clova

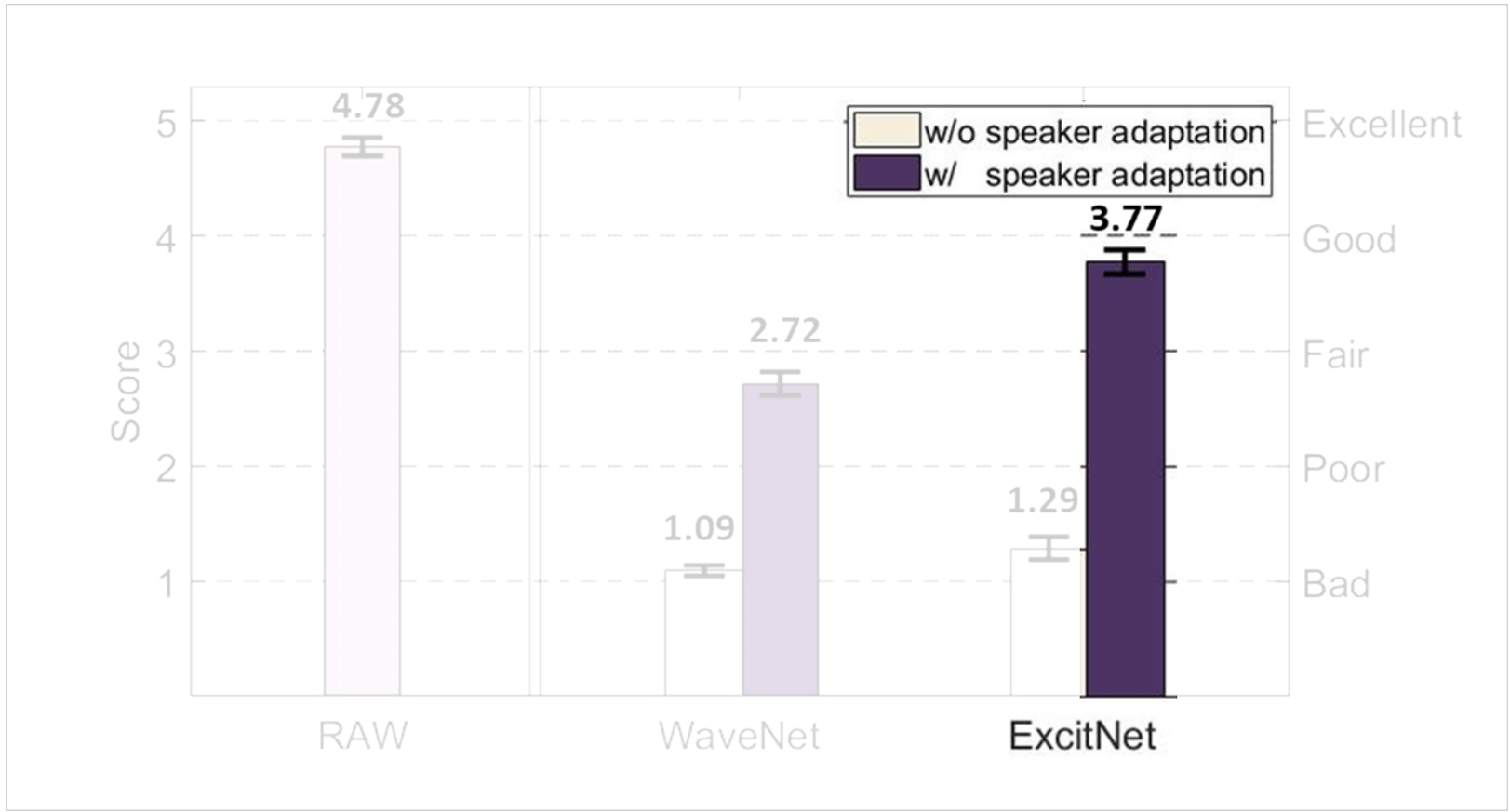
Demo



http://arxiv.org/abs/1811.03311

"DB 암호화 시장의 성장 전망은 밝은 편이다."

Demo

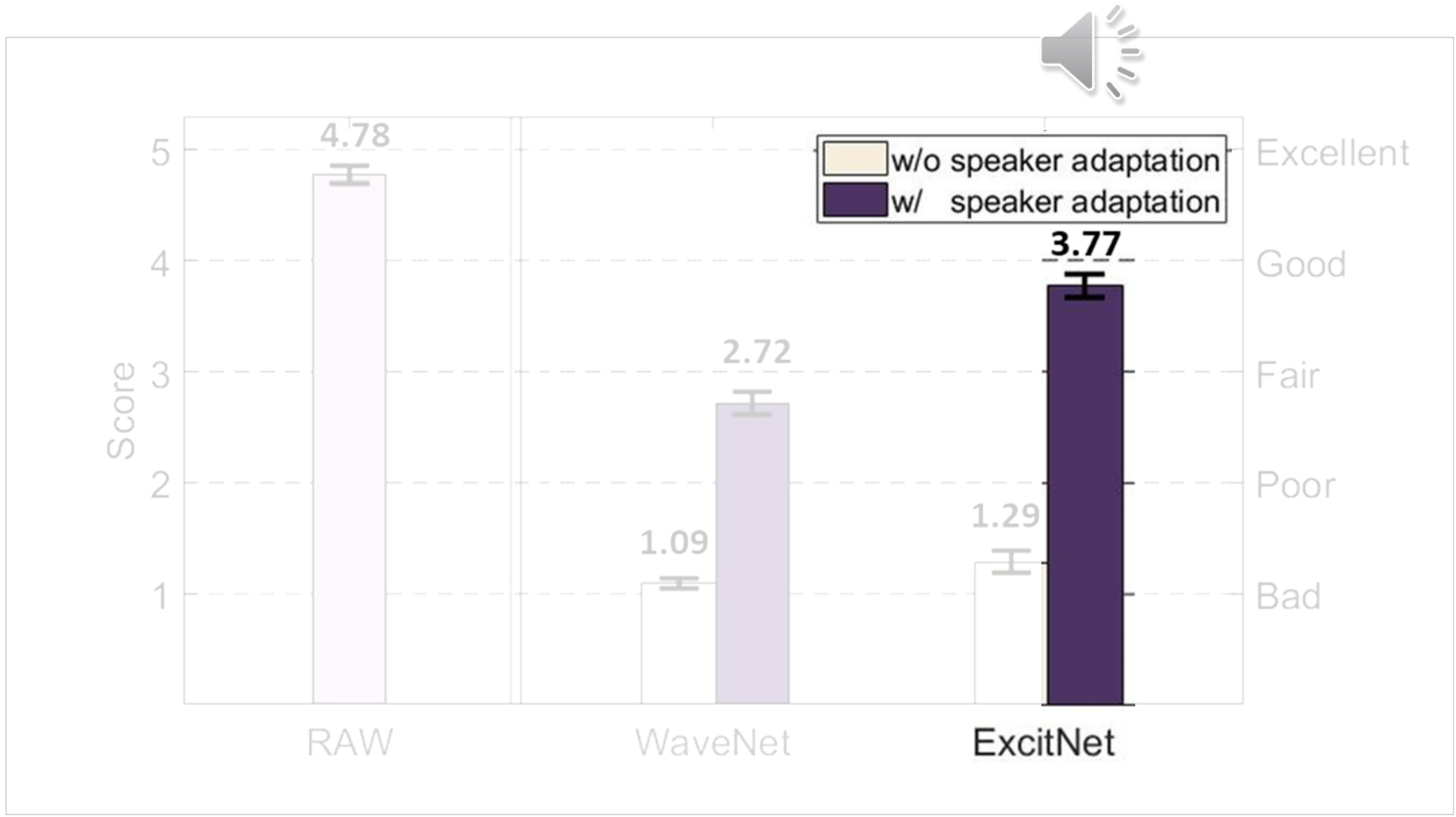


http://arxiv.org/abs/1811.03311

"DB 암호화 시장의 성장 전망은
밝은 편이다."

Clova

Demo



http://arxiv.org/abs/1811.03311

"DB 암호화 시장의 성장 전망은
밝은 편이다."

Thank You