

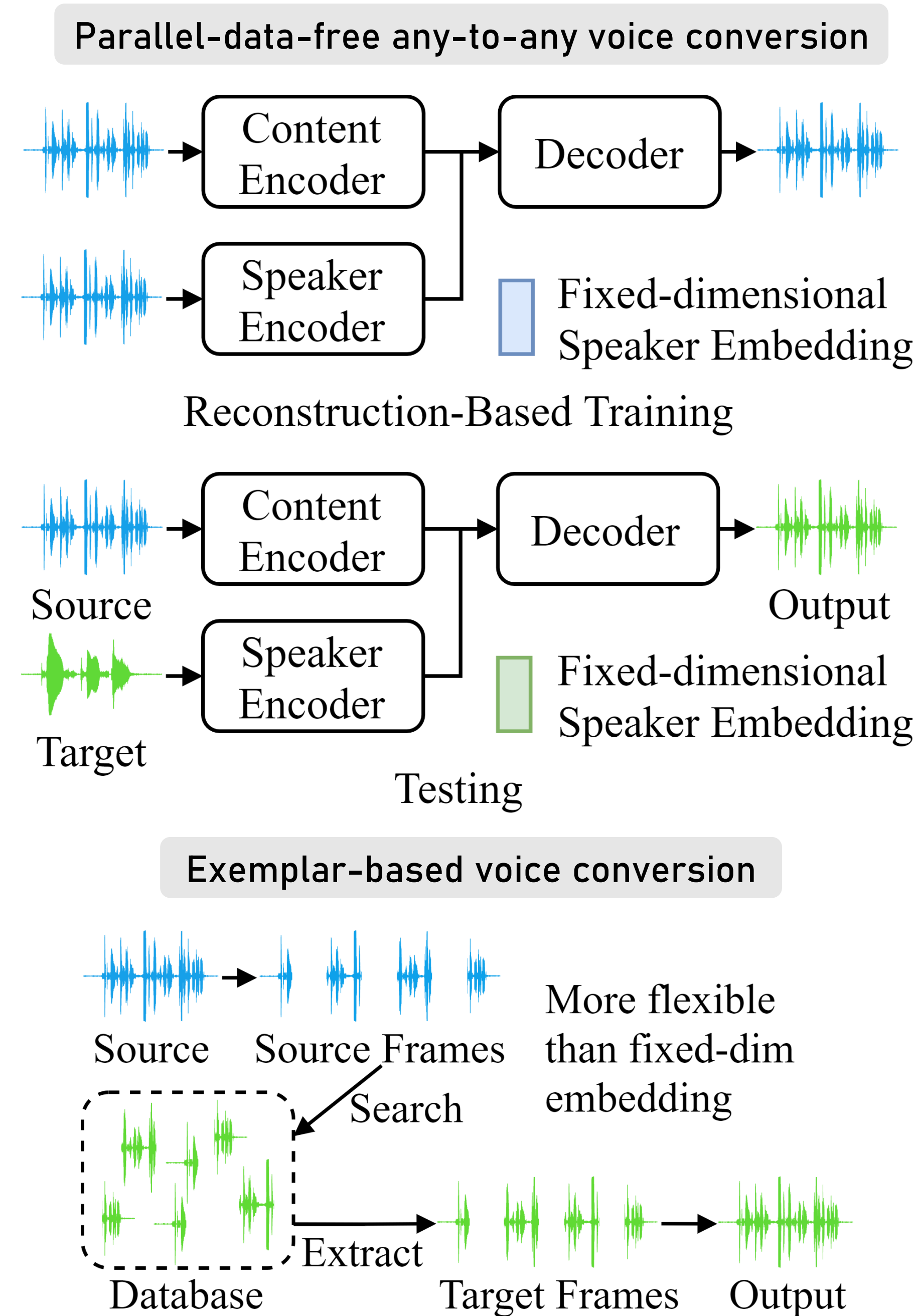


* These authors contribute equally

FragmentVC: Any-to-Any Voice Conversion by End-to-End Extracting and Fusing Fine-Grained Voice Fragments With Attention

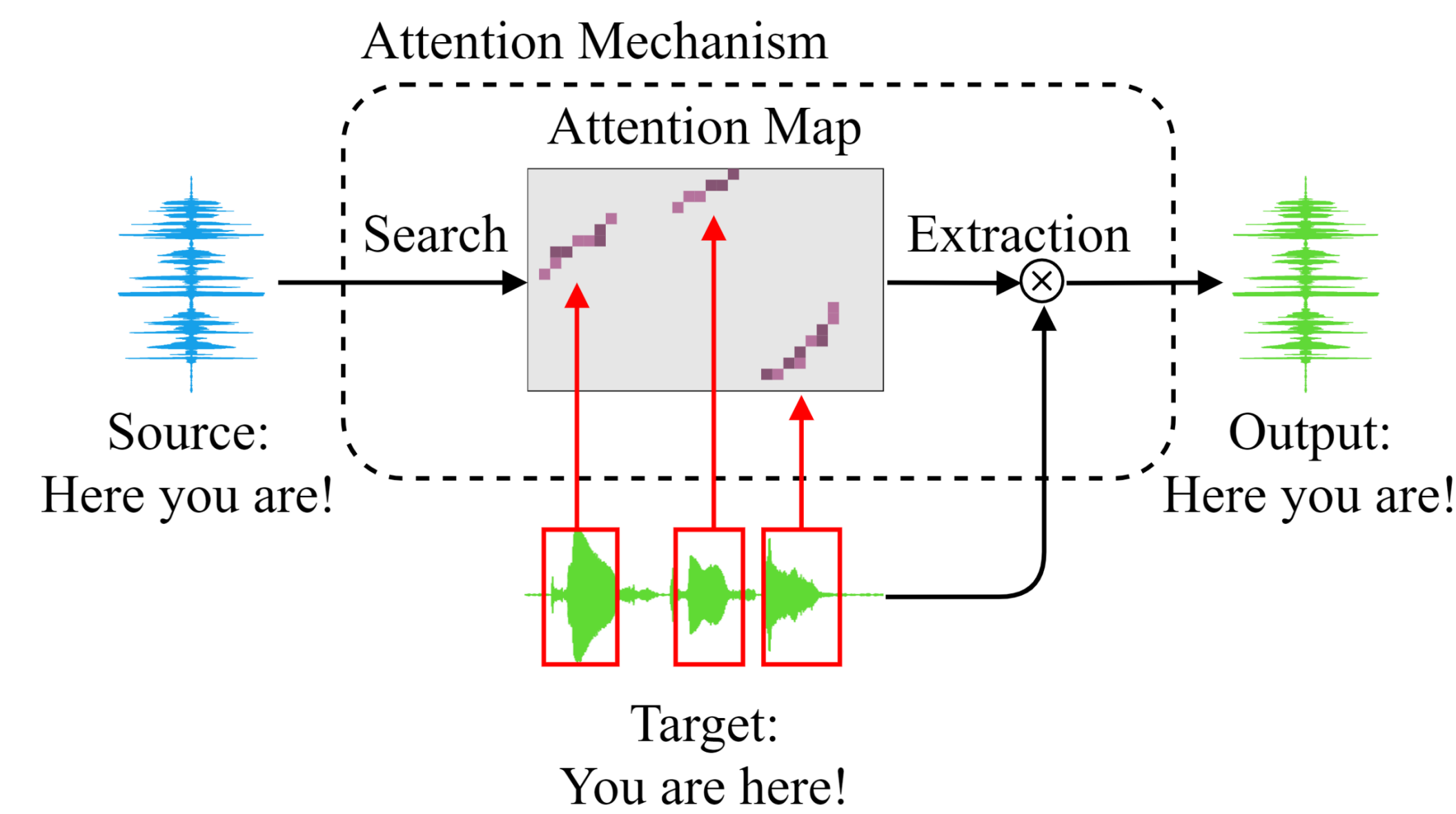
Yist Y. Lin*, Chung-Ming Chien*, Jheng-Hao Lin, Hung-yi Lee and Lin-shan Lee
College of Electrical Engineering and Computer Science, National Taiwan University

I. Prior Arts



II. Motivation

- Exemplar-based voice conversion with end-to-end neural network?
- Attention mechanism!**

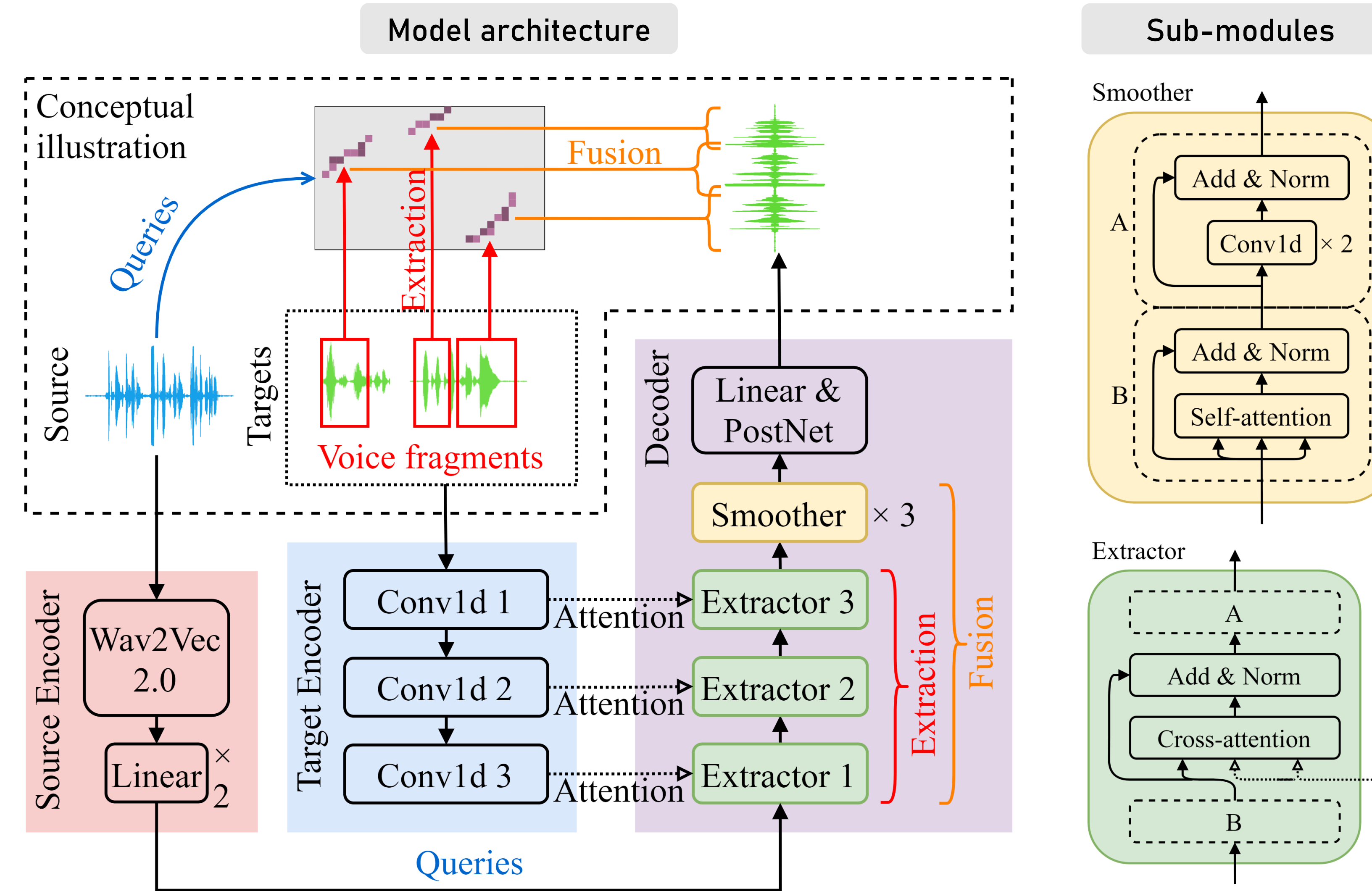


Code



Audio samples

III. Proposed Methods



- Pretrained Wav2Vec**: extracting phonetic structure from the source utterance, fixed during training
- Extractor**: transformer blocks with cross-attention to extract voice fragments from target speaker utterances based on the source Wav2Vec features
- Smoother**: smoothing the extracted voice fragments and generating the output

IV. Experiments

Automatic speaker similarity evaluation

Ratio of utterances passing a speaker verification system

	Proposed	Proposed w/o finetune	AdaIN-VC [1]	AutoVC[2]
Seen-to-seen	94.8	94.7	97.8	39.3
unseen-to-unseen	92.5	99.8	87.1	19.0

Subjective evaluation

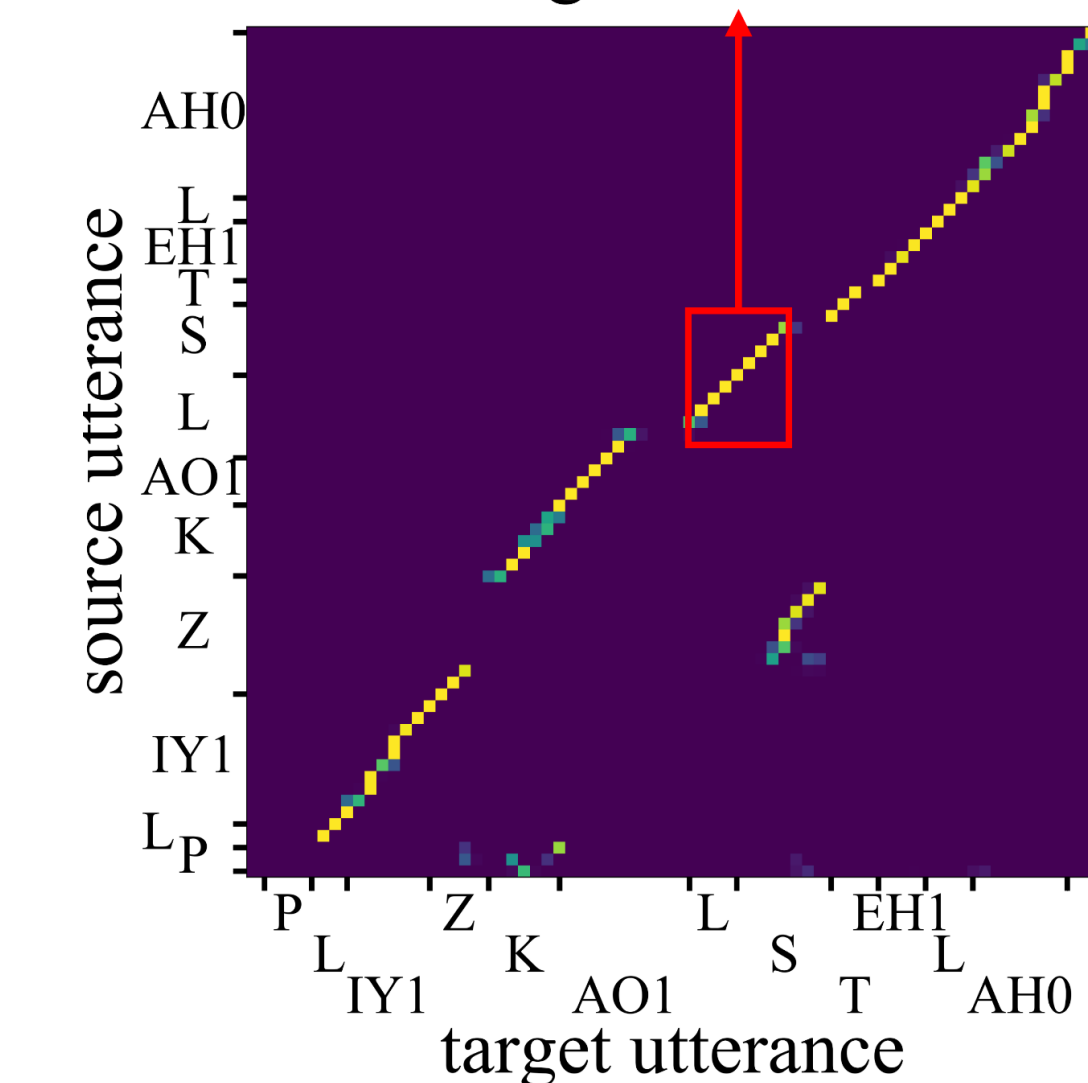
MOS on speaker similarity and naturalness

	Proposed	Proposed w/o finetune	AdaIN-VC	AutoVC	Real
Similarity	3.32±.15	3.81±.15	2.75±.15	2.12±.14	-
Naturalness	3.26±.12	2.73±.11	2.52±.12	2.31±.12	4.09±.12

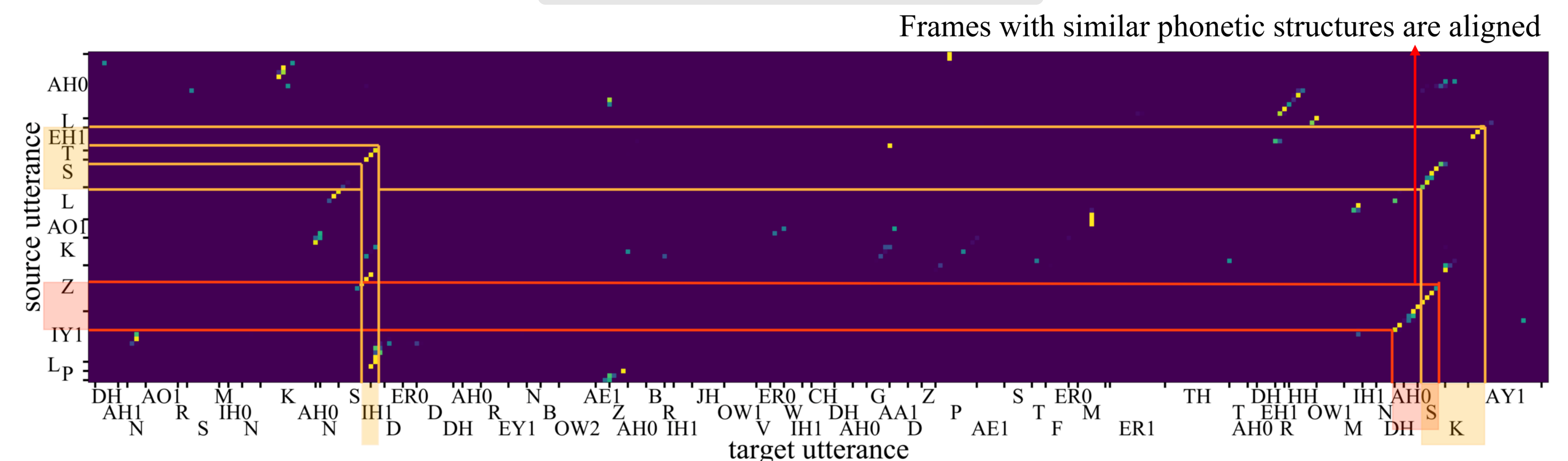
*Seen-to-seen: VCTK corpus, Unseen-to-unseen: CMU Arctic Database
 [1] Chou et al., One-Shot Voice Conversion by Separating Speaker and Content Representations with Instance Normalization
 [2] Qian et al., AUTOVC: Zero-Shot Voice Style Transfer with Only Autoencoder Loss

V. Attention Analysis

Same content different speakers
Voice fragments are extracted



Different content different speakers



VI. Conclusion

- A SOTA approach to any-to-any voice conversion
- Utilize attention mechanism to end-to-end
- Extract** target speaker fragments phonetically similar to the source fragments
- Fuse** the extracted fragments to achieve voice conversion