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

College of Electrical Engineering and Computer Science, National Taiwan University

I. Prior Arts

Parallel data-free any-to-any voice conversion

II. Motivation

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

III. Proposed Methods

Model architecture

Sub-modules

Exemplar-based voice conversion

Training

IV. Experiments

Automatic speaker similarity evaluation

Ratio of utterances passing a speaker verification system

Subjective evaluation

MOS on speaker similarity and naturalness

V. Attention Analysis

Same content different speakers

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 phonemically similar to the source fragments
Fuse the extracted fragments to achieve voice conversion