

# PPG-BASED SINGING VOICE CONVERSION WITH ADVERSARIAL REPRESENTATION LEARNING

*Zhonghao Li, Benlai Tang, Xiang Yin, Yuan Wan, Ling Xu, Chen Shen, Zejun Ma*

ByteDance AI Lab

{lizhonghao.01, tangbenlai, yinxiang.stephen}@bytedance.com

## ABSTRACT

Singing voice conversion (SVC) aims to convert the voice of one singer to that of other singers while keeping the singing content and melody. On top of recent voice conversion works, we propose a novel model to steadily convert songs while keeping their naturalness and intonation. We build an end-to-end architecture, taking phonetic posteriorgrams (PPGs) as inputs and generating mel spectrograms. Specifically, we im-

source and the target speakers to sing the same songs during the training phase.

As parallel singing corpus is rare, several works have been conducted to solve this problem. Referring to advanced achievements from voice conversion [4, 5], [6] builds an autoencoder framework to train the conversion model. The autoencoder model, consisting of a WaveNet [7] encoder to compress acoustic information and a WaveNet decoder to recover waveform with a speaker embedding table, maps the