MDX-GAN: Enhancing Perceptual Quality in Multi-Class Source Separation Via Adversarial Training

Ke Chen[1,2], Jiaqi Su[2], Zeyu Jin[2]

[1]University of California San Diego [2]Adobe Research

ICASSP 2024 KOREA

Introduction

MDX-GAN is an efficient and high-fidelity source separator for enhancing the perceptual quality.

The paper contains:

- A conditional separator based on MDX-Net and post-processing network for

  separating a wide range of audio events

- A training process with

  data simulation of real-world speech-centered scenarios

  perceptually-motivated loss functions and adversarial training paradigm

- A comprehensive evaluation in a real-world setting on

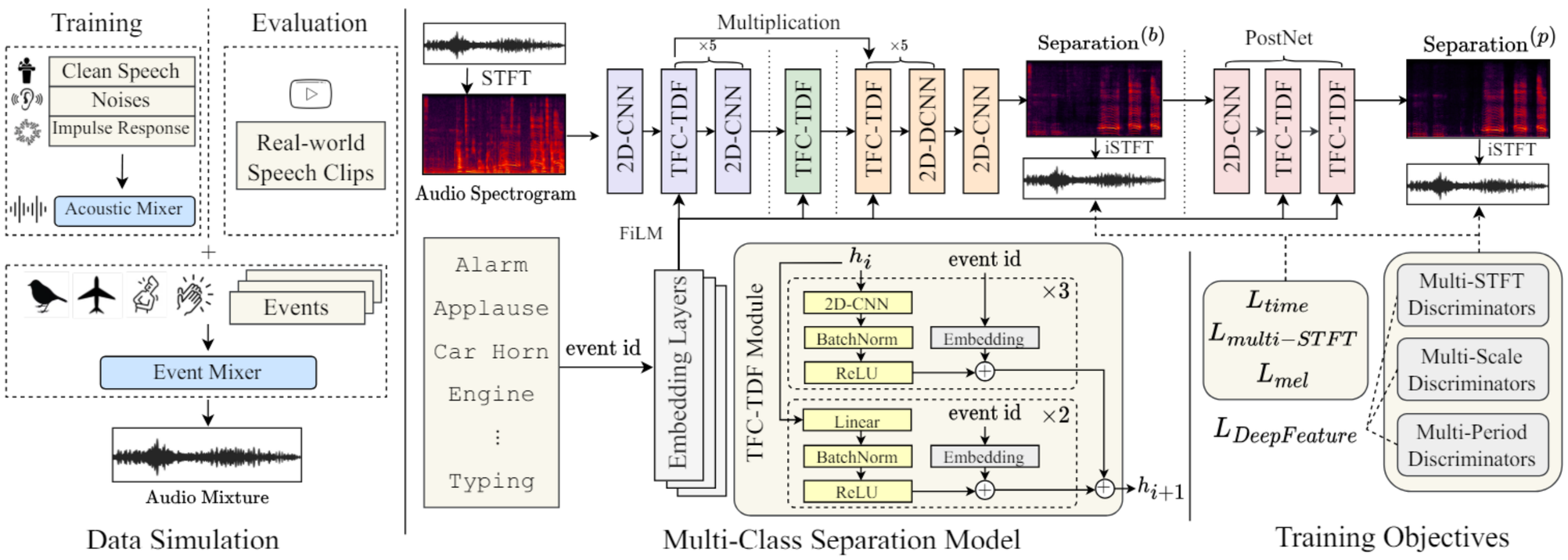
  objective (SDRi) and subjective (MOS score) experiments

  discussion and visualization within the ablation study

Demo

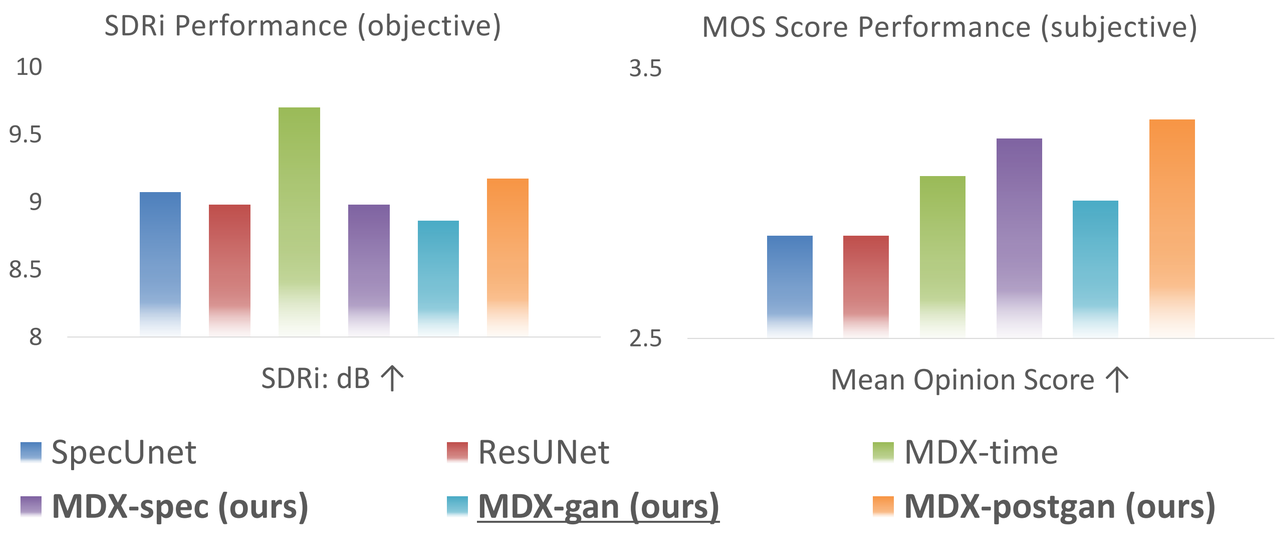
Paper

Model Architecture



Experimental Results

We observe a discrepancy between the MOS score and the SDRi metric across separation results of ten-class sound events.



Visualization (more details in the demo page)