

# Supplementary Material

In this supplementary material, we provide additional audio samples that were not included in the main paper, titled "Do Not Mimic My Voice: Teacher-Guided Unlearning for Zero-Shot Text-to-Speech". These samples illustrate the performance of the proposed unlearning system in generating synthetic speech. The audio files showcase various speaker and text combinations, offering a more in-depth understanding of how the system handles both remain and forget speaker samples.

This supplementary material provides audio samples from both the forget and remain datasets.

- For the remain speakers, we include the inference results from various methods: the re-implemented VoiceBox as the original model, and baseline methods such as Negative Gradient (NG), Kullback-Leibler Divergence (KL) and Sample-Guided Unlearning (SGU). Additionally, we provide results from our proposed method, Target-Guided Unlearning (TGU), as well as the prompt audio samples used for synthesis.
- For the forget speakers, we present inference results for the same methods, along with their respective prompt audio files. To illustrate the ability of the unlearned model to generate speech in a random voice when given a forget speaker, we provide three different audio samples synthesized from the same prompt audio. These samples demonstrate the model’s capability to produce varying voices based on the forget speaker input.

## File Structure

The supplementary material is organized as follows:

```
AudioSamples/
├── remain_speakers/
│   ├── 1/
│   │   ├── prompt.wav
│   │   ├── ground_truth.wav
│   │   ├── original.wav
│   │   ├── TGU.wav
│   │   ├── SGU.wav
│   │   ├── NG.wav
│   │   └── KL.wav
│   └── ...
├── forget_speakers/
│   ├── 1/
│   │   ├── prompt.wav
│   │   ├── original_0.wav
│   │   ├── original_1.wav
│   │   ├── original_2.wav
│   │   ├── TGU_0.wav
│   │   ├── ...
│   │   └── SGU_0.wav
│   └── ...
```

