Introduction

Many audio processing/synthesis applications target 16kHz.
- Computationally easier target (relative to 48kHz).
- Less immersive listening experience.

We propose:
- BWE as postprocessing for bridging from 8k/16kHz to 48kHz
- A waveform BWE method using GAN that achieves audio quality typically indistinguishable from real 48kHz audio.

Existing BWE methods: limited extension up to 16kHz
- Spectral methods compensate energy, but over-smooth spectrogram and introduce artifacts by phase approximation.
- Waveform methods still not close to real high-quality.

Evaluations show that our BWE method:
- Achieve close to real 48kHz audio quality for 16k-to-48k BWE; greatly improve over previous methods for 8k-to-48k BWE.
- Bring consistent quality boost to denoisers and vocoders.

Method

Adapt from HiFi-GAN [5]: Feed-forward WaveNet with deep feature matching in adversarial training

- Discriminator on 128-coefficient log mel spectrogram
- Multi-scale discriminators on downsampled waveforms
- L1 loss, spectrogram losses, and feature matching loss with deep features of the discriminators
- Weight normalization to speed up convergence

Experiments

- Experiment 1: BWE baseline comparison
  - Numerical values of objective measures do not correlate well with perceptual quality.
- Experiment 2: BWE for denoising
  - Consistent audio quality boost for various speech denoisers and vocoders
- Experiment 3: BWE for vocoding
  - 16k-to-48k BWE by HiFi-GAN+ is typically indistinguishable from real 48kHz.

Conclusions

- Our waveform-to-waveform bandwidth extension method based on HiFi-GAN can achieve audio quality typically indistinguishable from real 48kHz audio.
- Applying BWE as post-processing can consistently boost quality in a wide variety of audio applications (denoisers, vocoders etc.).
- Existing objective measures do not correlate well with perceptual quality (for the 48kHz BWE task).