Audio Super Resolution with Respeecher

respeecher

Summary

Audio Super Resolution with Respeecher

Audio Super Resolution with Respeecher

Probably most of you have heard about image super resolution. This technology enhances the resolution of an image from low-resolution to high. It is usually used for the following cases:

- **Medical purposes:** Generating high-resolution MRI from otherwise low-resolution MRI images.
- **Media:** Reducing server costs, since media can be sent at a lower resolution and quickly upscaled.
- **Surveillance**: detecting, identifying, and performing facial recognition on low-resolution images from security cameras.

While neural super resolution for images has seen a lot of commercial success (e.g. NVIDIA's DLSS library for fast high resolution rendering), similar techniques in the audio domain remain less explored.

Initially driven by internal needs, Respeecher built an audio version of the super resolution algorithm to help our team deliver the highest resolution audio across the board, even in cases when the client doesn't have the high-res sources available. Being sound professionals ourselves, we quickly realized that other sound designers and editors out there might also be in need of a similar tool. That's why we decided to provide it as a service to everyone and start gradually transforming it into a standalone product for speech enhancement.

If you've got old audio tapes for restoration, or some files from the web in a compressed format, your project will benefit from recovering the ultra-high frequencies that initially were not present in the recording.

What is Audio Super Resolution

Audio super resolution is similar to image super resolution, but in the audio domain.

- **Neural image super resolution:** an artificial neural network acts like a restoration artist, “imagining” the missing details and increasing the spatial resolution of an image in a natural way.

- **Neural audio super resolution:** an artificial neural network adds missing details in the time domain by increasing the effective sampling rate of an audio signal.

For example, if you have a recording that was captured using a video conferencing software, skype or zoom, the audio quality will have a noticeable high-end roll-off. Or in case you are working with some older quality mp3s or some old recordings and you want to recover that high-end, that’s when audio super resolution is especially useful.

Low resolution audio recordings (like the ones recorded using very old hardware) lack the ‘air’ and ‘brightness’ of high-res recordings. They lack energy in the high frequencies, which is why they are often referred to as band-limited. Audio super resolution helps to recover such recordings.

Audio Super Resolution Use Cases

For sound professionals

Band-limited audio usually comes from old audio tapes for restoration or files from the web in a compressed format. These are pretty unusable in any modern production-quality dialogue, because the lack of the high frequencies is immediately noticeable when heard alongside other high-res recordings in the mix.

For speech synthesis apps

To make TTS synthesized speech sound natural, the painstaking process of honing its timbre, smoothness, placement of accents and pauses, intonation, and other areas is a long and unavoidable burden.

Many state-of-the-art text-to-speech systems generate audio in 22 or 24 kHz. Increasing the native output sampling rate of these TTS systems could prove challenging and may result in sacrificing performance and robustness.

Using audio super resolution instead, could yield a significant increase in the sound quality without needing to change anything in the company's existing speech synthesis pipeline.

At Respeecher

Internally, we use super resolution as a post processing step, similar to the TTS use case above, but applied to voice conversion. It adds the missing high frequencies when our models generate low-resolution audio (often due to the lack of high-resolution recordings from the client).

Audio Samples

click to listen

How Audio Super Resolution Works

In the nutshell, our super resolution network is a GAN-based neural audio enhancer that adds extra resolution to recordings with limited bandwidth.

The enhancement is performed by an artificial neural network that analyzes the frequency range of the input low resolution audio and completes its spectrum by generating a high frequency signal that blends smoothly with the original audio.

From a different perspective, the operation of super resolution can be viewed as imputation of new time-domain samples in an audio signal. From this angle, it is similar to the image super-resolution problem, where individual audio samples are analogous to pixels.The technology has to "fill-in" the missing samples.

During the training process, we show the network examples of high-resolution audio tracks together with their artificially downsampled versions. The network is tasked to predict the high-resolution signals given only the downsampled versions on the input.

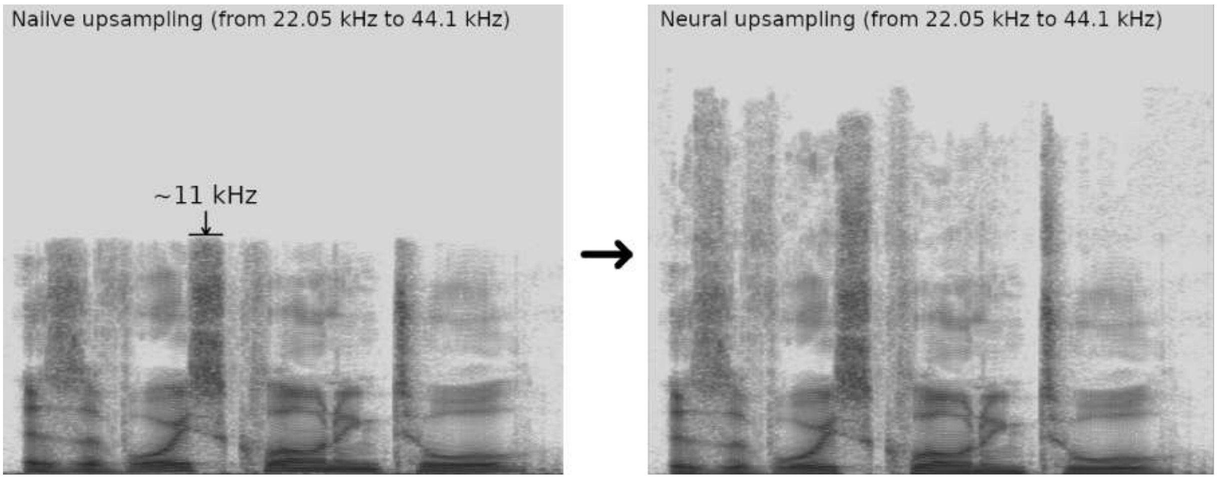

Image: Demonstration of the effect of super resolution in the spectral domain.

The spectral enhancement with Super48 could add ‘air’ or ‘brightness’ and improve the overall listeners’ experience.

- Respeecher current version supports upsampling from **22.05kHz to 44.1kHz**.
- The upcoming version will support a minimum rate of **16kHz and will upsample it to 48kHz**, which is more than enough to cover the human hearing range.

Future Development of Respeecher Audio Super Resolution

Currently with Respeecher audio super resolution is available as a service. We’re working to package it as a standalone app.

Future versions of Respeecher audio super resolution also include:

- **Bit-depth super resolution:** changing from 8-bit to 24-bit audio
- **Decompression:** increasing dynamic range by undoing compression
- Noise reduction and dereverberation