# Convolutional Neural Networks for Speech Audio Enhancement

Kyle Fisher, Adam Scherlis
Department of Computer Science, Stanford University

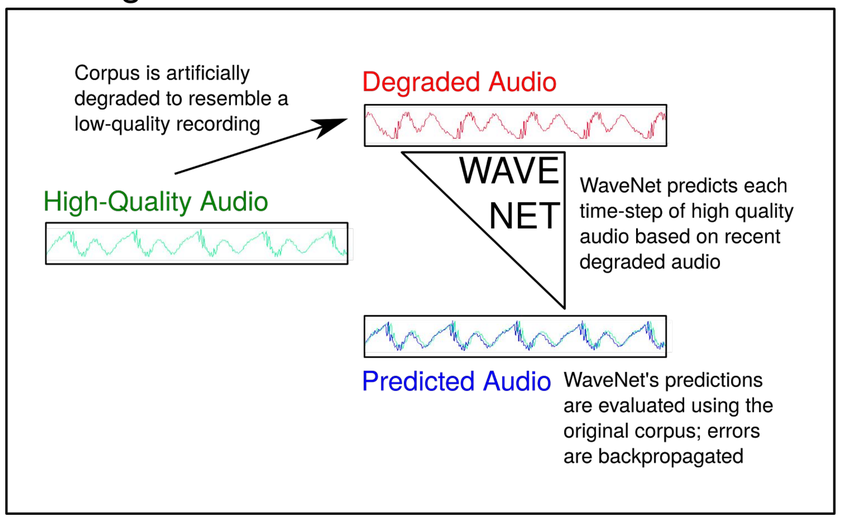
## Overview

Although voice telephony bandwidth has been gradually increasing with time, many of today's speech transmission systems still use a highly compressed representation of voice audio data which noticeably degrades the perceived quality of the received voice signal to human listeners. The dominant technique used to post-process decoded narrowband speech audio signals is *artificial bandwidth extension* (ABE). However, there are unavoidable drawbacks to using ABE, such as the introduction of certain types of unnatural artifacts including frequency distortion and whistling. Therefore, a novel approach toward enhancing narrowband voice codings is to be desired.

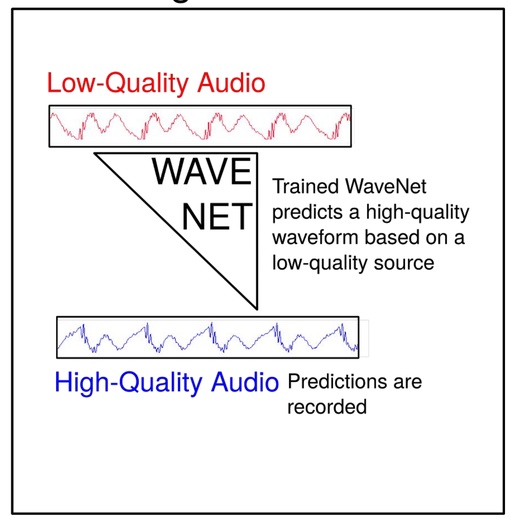
In our approach to speech audio enhancement, we deemphasize the frequency domain and instead attempt to correct the raw, time-domain signals of the degraded audio. We build upon Google DeepMind's recently-released *WaveNet*, a convolutional neural network model originally designed as a framework for text-to-speech synthesis. In our project, we adapt WaveNets toward the purpose of reconstructing high-quality voice audio from degraded signals by training the WaveNets on synchronized clean and degraded audio waveforms. We then repurpose the WaveNets TTS model to perform enhancement rather than generation, and experiment with enhancing audio which has suffered from various degradations.

## Our Approach

### Training WaveNet

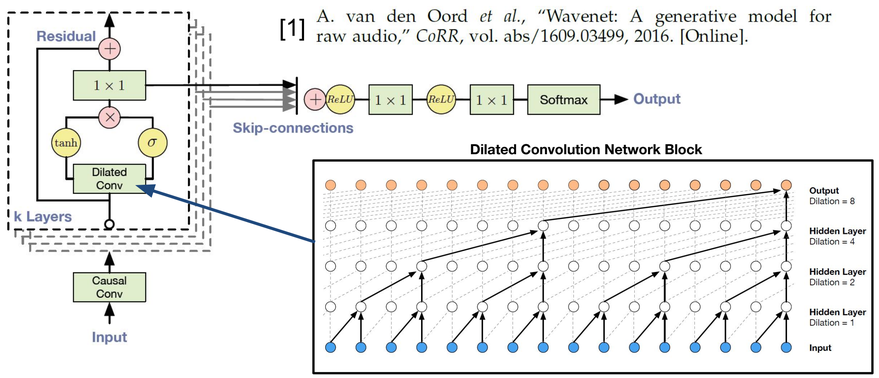



### Enhancing Audio




## The *WaveNet* Model

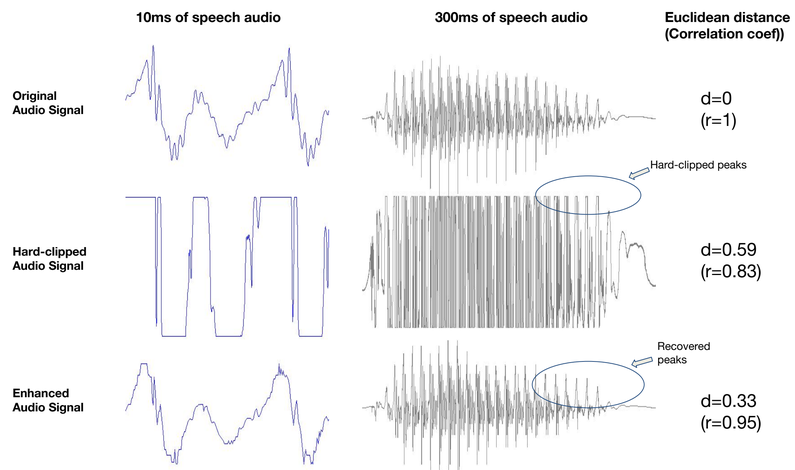
*WaveNet* is a deep, convolutional neural network model designed for text-to-speech synthesis (TTS). It uses blocks of dilated causal convolutions with residual and skip connections to make learning from raw audio and generating waveforms tractable. For TTS, WaveNets are conditioned upon textual information (not shown). Figures have been drawn from [1].

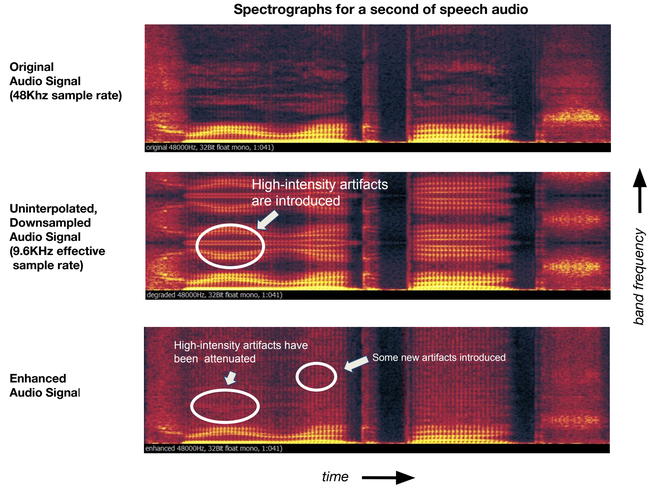



[1] A. van den Oord *et al.*, "Wavenet: A generative model for raw audio," *CoRR*, vol. abs/1609.03499, 2016. [Online].

## Experimental Results

### Learning to recover hard-clipped audio




### Learning to recover downsampled audio