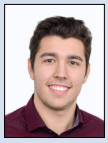
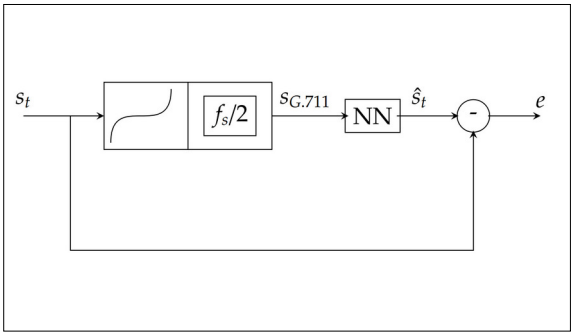


Gian-Reto Wiher

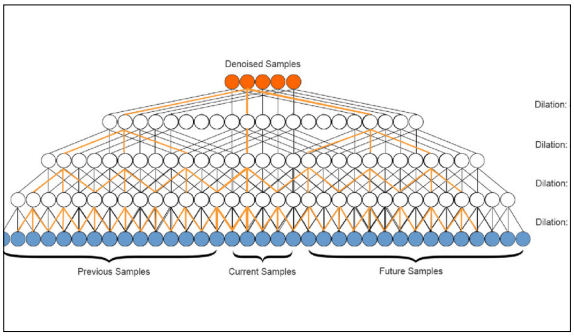
Student	Gian-Reto Wiher
Examiner	Prof. Dr. Guido Schuster
Subject Area	Sensor, Actuator and Communication Systems

Speech Denoising with Neural Networks

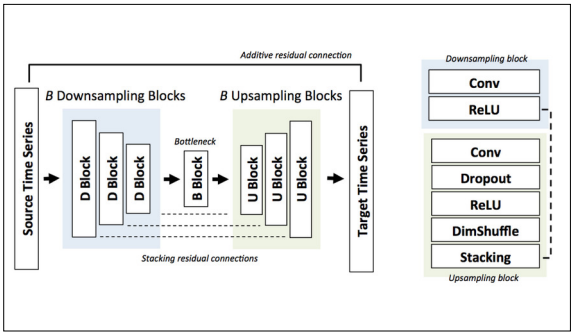
Removing quantization noise and increasing the sampling frequency of G.711 coded speech



The task of the neural network is to minimize the error e, based on the signal received by a phone call.
Ozhiker, Graph of -law a-law algorithms,



The dilated convolutions in the modified Wavenet. The Non-causal structure predicts a target field from input samples.
D. Rethage, A wavenet for speech denoising



Deep Residual Neural Network of the Super Resolution network. On the right, sublayers of the down and upsampling layers.
V. Kuleshov, Audio super resolution using neural networks

Introduction: While a lot of communication channels emerge and improve their quality and reliability, traditional telecommunication and even VoIP still use the G.711 standard which was released nearly 50 years ago. The motivation for this thesis was to address the quality disadvantage that such communication channels bring with them. The goal was to come up with a deep neural network that can increase the sampling frequency and remove the quantization noise that is present in a G.711 audio codec.
The most commonly used audio processing techniques modify the magnitude of the spectrogram to solve their tasks. While a lot of problems can be solved with this approach, it neglects the phase. To overcome this limitation, processing in the time domain should be used as the approach of Wavenet.

Approach: The task was tried to solve with supervised training of a model. A python script was written to generate the high-quality speech and the lower quality G.711 coded speech. For this thesis the A-law quantization was used.
First, the Wavenet was tested without any sampling rate difference between the input and the output. The structure of the model was tested in an empiric way to check if the quality estimated speech improves.
The first approach put emphasis mostly on removing the quantization noise but not on increasing the sampling rate. A second try with the Super-Resolution network tries to do exactly this.
For a short trial, a network was build based on the assignment. This was not successful because there was no data to train on and the structure of the model did not fit. The other two models were trained on the prepared data and a lot of model structures were tested.

Conclusion: This thesis shows, denoising and upsampling audio files is not an easy task to implement. The Wavenet was first tested without any sampling rate difference between the input and the output data. No one of the tested network structures could remove the quantization noise successfully. On a second approach, the Super-Resolution network was tested to increase the sampling rate. This approach also did not yield a positive result in the given time.
It is possible that either one of the models or a combination of the two can solve the task. The Super-Resolution network can be used to increase the sampling rate at the beginning and then as a second stage a Wavenet solution similar to the one used in this thesis to remove the quantization.
Nevertheless, I learned a lot of new things and will use this knowledge in future projects.