

# Removing Noise from Speech with Deep Learning

Andor Kiss  
Balázs Glávits  
Márk Konrád

# Introduction

Our task was to reduce noise from speech using deep learning.  
Preserve sound quality.

# Motivation

Cool noise reducing hardware.



Figure 1: Sennheiser GSP-500

But this is hardware, and we are computer scientists, not electrical engineers.

# Motivation

Noise cancelling software.

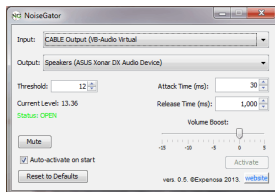


Figure 2: NoiseGator Software

If sound is above the threshold, it goes through.  
Else it is cancelled.

Not flexible enough.

Deep learning could do a better job.

# Existing implementations

- Autoencoder based
- GAN based
- WaveNet based

# Autoencoder based

- Denoising Autoencoder

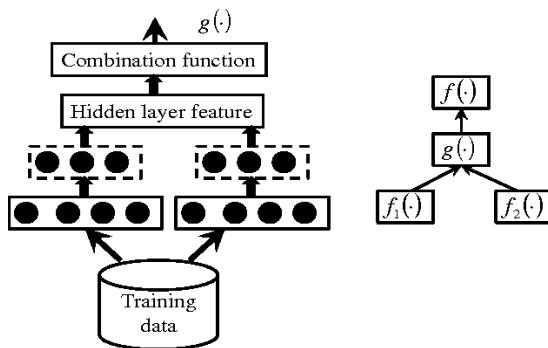


Figure 3: DAE

# Autoencoder based

- Denoising Autoencoder with Multi-branched Encoders

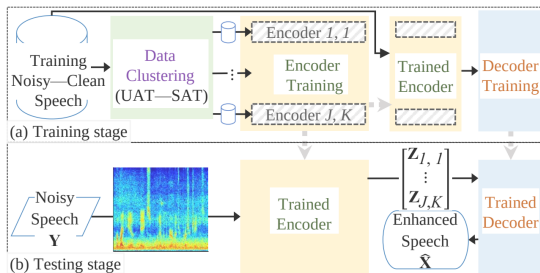


Figure 4: DAEME

# GAN based

- Speech Enhancement Generative Adversarial Network

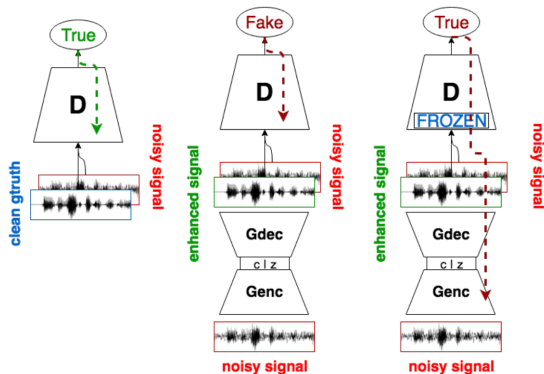


Figure 5: SEGAN



- Wavenet for Speech Denoising

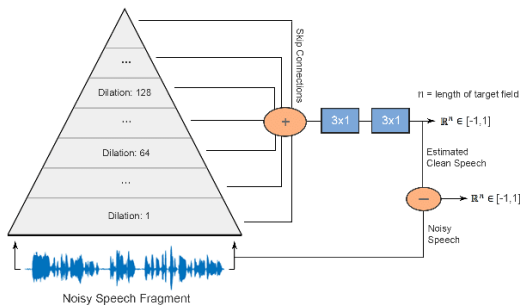


Figure 6: Speech denoising WaveNet

# Training and testing data

- ~23000 samples
- 56 different voices and noise conditions

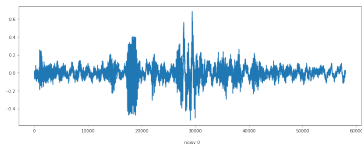


Figure 7: Noisy data

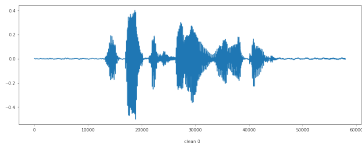


Figure 8: Clean data

# Data pipeline

Training phase.

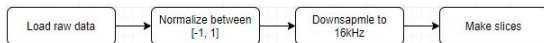


Figure 9: Training preprocessing

We do this on the noisy and clean data as well.

Input: Noisy slices

Output: Clean slices

Data augmentation: Overlapping slices

# Full data pipeline

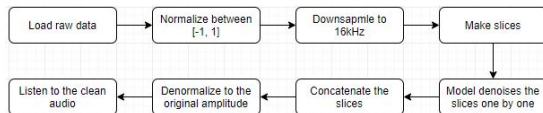


Figure 10: Full pipeline

Model is a black box now, it will be elaborated later.

# Original WaveNet

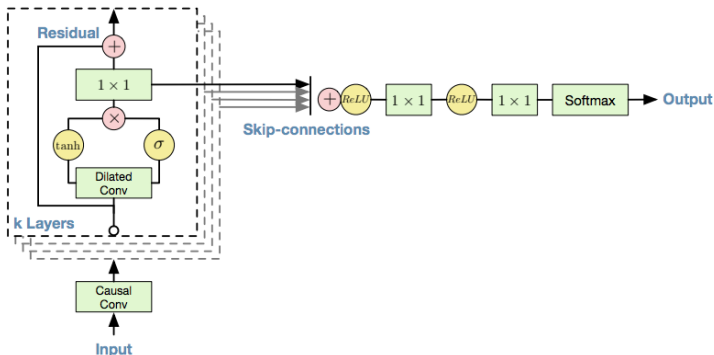


Figure 11: WaveNet

Causal convolutions, mu-law transform and softmax distribution.

# Modified wavenet

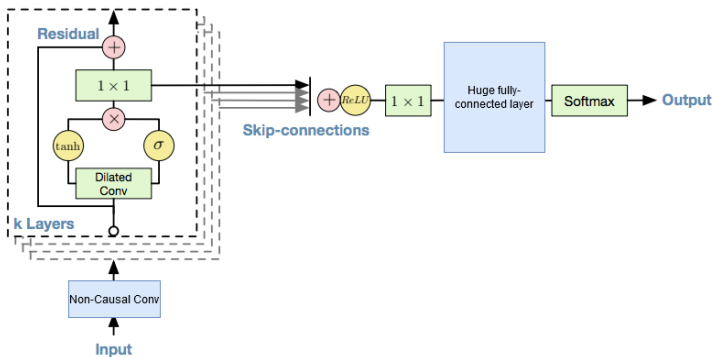


Figure 12: Modified WaveNet

Non-causal convolutions, and dense output layer.



# WaveNet based autoencoder

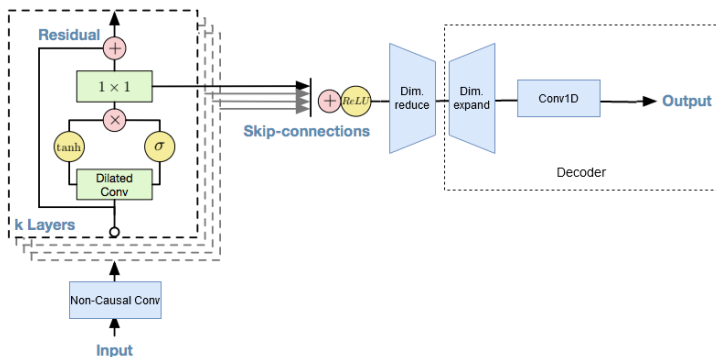


Figure 14: Wavenet based autoencoder



# Regression with convolutional layers

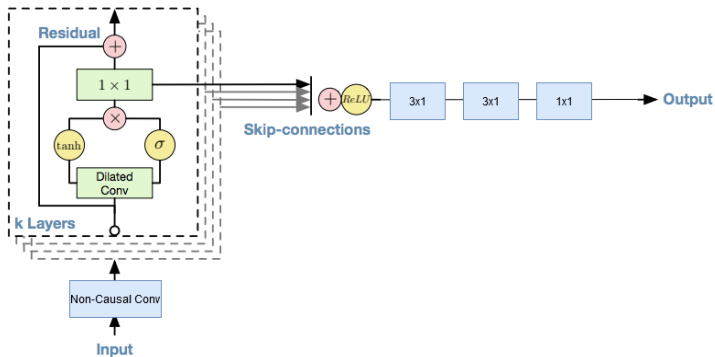


Figure 15: Wavenet based autoencoder

WaveNet with non-causal convolutions, regression, and extra one dimensional convolutional layers on the output.

# Training

- Google Cloud Platform
- Clean & Noisy slice generator
- MAE loss
- SGD optimizer
- ReduceLROnPlateau

# Demo

Separately, in an .ipynb

# Summary

Success, but. . .

# Thank you for your attention

## Sources:

- Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. “WaveNet: A Generative Model for Raw Audio”. In: (2016) arXiv:1609.03499
- Dario Rethage, Jordi Pons, and Xavier Serra. “A Wavenet for Speech Denoising”. In: (2018) arXiv:1706.07162