

## Audio denoising papers

### Noise Reduction Techniques and Algorithms For Speech Signal Processing (Algo\_Speech.pdf)

Different types of noise :

- Background noise
- Echo
- Acoustic / audio feedback (Mic capture loudspeaker sound and send it back)
- Amplifier noise
- Quantization noise when transformning analog to digital (round values), neglectable at sampling higher than 8kHz/16bit
- Loss of quality due to compression

Linear filtering (Time domain) : Simple convolution

Spectral filtering (Frequency domain) : DFT and back

ANC needs a recording of the noise to compare it to the audio

Adaptive Line Enhancer (ALE) doesn't need it.

Smoothing : noise is often random and fast change, so smoothing can help again white and blue (high freq) noise.

### A Review of Adaptive Line Enhancers for Noise Cancellation (ALE.pdf)

Doesn't need recording of noise. Adaptive self-tuning filter that can spearate periodic and stochastic component. Detect low-level sin-waves in noise

...

### A review: Audio noise reduction and various techniques (Techniques.pdf)

Some filters : Butterworth filter, Chebyshev filter, Elliptical filter

### Employing phase information for audio denoising (Phase.pdf)

### Audio Denoising by Time-Frequency Block Thresholding (Block\_Threshold.pdf)

### Speech Denoising with Deep Feature Losses (Speech\_DL.pdf)

Fully convolutional network, work on the raw waveform. For the loss, use the internal activation of another network trained for domestic audio tagging, and environnement detection (classification network). It's a little bit like a GAN.

Most approaches today are done on the spectrogram domain, this one not. Prevents some artefacts due do IFT. Methods that are in the time domain often use regression loss between input and output wave. Here, the loss is the dissimilarity between hidden activations of input and ouput waves. Inspired from computer vision (-> Perceptual\_Losses.pdf)

Details of The main network are given in papers, section II-A-a. Different layers in the classification/feature/loss network correspond to different time scales. The classification network is inspired by VGG architecture from CV, details in paper II-B-a. II-B-b explain how to transoorm activations / weights to a loss.

Train the feature loss network using multiple classification tasks (scene classification, audio tagging). Train the speech denoising using the [1] database. They used the clean speeches and some noise samples and created the training data by combining them together, then they are downsampled.

Experimental setup : compared with Wiener filtering pipeline, SEGAN, and the WaveNet based one used as a baseline. Used different score metrics (overall (OVL), the signal (SIG), and the background (BAK) scores)). It was better than all the baselines. Also evaluated with human testers, also better than the others.

Now this is for speech, and it might not work as well for general sound/music

## **Recurrent Neural Networks for Noise Reduction in Robust ASR (RNN.pdf)**

SPLICE algorithm is a model that can reduce noise by finding a joint distribution between clean and noisy data, ref to article in the paper's reference, but could not find it online for free.

We could simply engineer a filter, but it's hard, and not perfect .

Basic idea : We can use L1 norm as the loss function. This type of network is known as denoising autoencoder (DAE). Since input has variable length, we train on a small moving window

More advanced : Deep recurrent denoising autoencoder, we add connection "between the windows"  $\implies$  Input is [0 1 2] [1 2 3] [2 3 0], we give each one one to a NN with e.g. 3 hidden layer, with layer 2 recursively connected, and it gives [1] [2] [3] as the output. Uses Aurora2 corpus, with noisy variants synthetically generated

## **Investigating RNN-based speech enhancement methods for noise-robust Text-to-Speech (RNN\_Speech\_Enhancement.pdf)**

### **Audio Denoising with Deep Network Priors (DN\_Priors.pdf)**

### **Spectral and Cepstral Audio Noise Reduction Techniques in Speech Emotion Recognition (Spectral\_Cepstral.pdf)**

### **Raw Waveform-based Speech Enhancement by Fully Convolutional Networks (RawWave\_CNN.pdf)**

### **SEGAN: Speech Enhancement Generative Adversarial Network (Speech\_GAN.pdf)**

### **A Wavenet for Speech Denoising (WaveNet.pdf)**

## **Audio super-resolution papers**

### **Audio Super-Resolution using Neural Nets (SuperRes\_NN.pdf)**

Paper + webpage + github on super resolution with deep networks <https://kuleshov.github.io/audio-super-res/#>

**Adversarial Audio Super-resolution with Unsupervised Feature Losses (Adversarial.pdf)**

**Time Series Super Resolution with Temporal Adaptive Batch Normalization (TimeSerie\_Batch.pdf)**

## **Ideas from images**

**Perceptual Losses for Real-Time Style Transfer and Super-Resolution (Perceptual\_Losses.pdf)**

**The Unreasonable Effectiveness of Deep Features as a Perceptual Metric (Perceptual\_Metric.pdf)**

## **Datasets**

### **1: Voice database with noisy and clean version**

<https://datashare.is.ed.ac.uk/handle/10283/1942>

### **2: New version of [1], also voice**

<https://datashare.is.ed.ac.uk/handle/10283/2791>

### **3: Speech database with clean and noisy**

<https://github.com/dingzeyuli/SpEAR-speech-database>

### **4: Aurora2**

<http://aurora.hsnr.de/aurora-2.html> Some script that can generate noisy data