Removing Noise from Speech with Deep Learning

Glávits Balázs
Department of Measurement and Information Systems
Budapest University of Technology and Economics
email@email.email

Kiss Andor
Department of Automation and Applied Informatics
Budapest University of Technology and Economics
email@email.email

Konrád Márk
Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics
konrad0816@gmail.com

Abstract

TODO

1 Data acquisition and exploration

For our training and testing data, we used a dataset called "Noisy speech database for training speech enhancement algorithms and TTS models" by the University of Edinburgh. It consists of $\sim$23000 clean-noisy pairs from 56 different speakers. The samples are stored in separate .wav files of varying length.

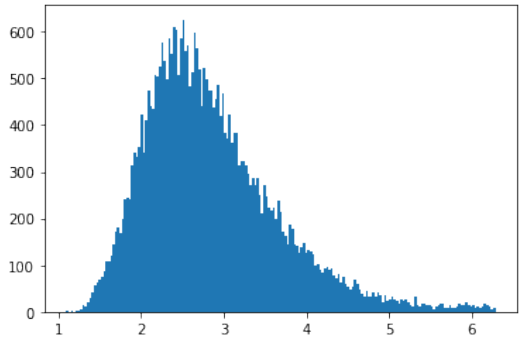


Figure 1: Duration distribution histogram of a subset of speeches

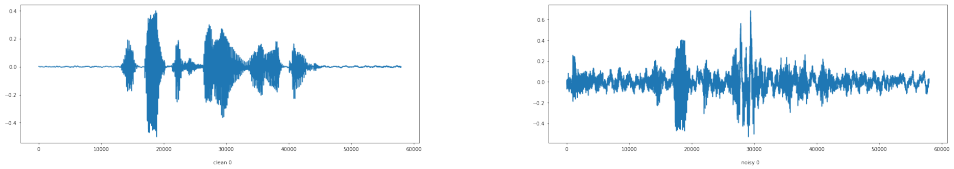
2 Data preprocessing

2.1 Data selection by duration

We intend to zero pad the data in order to make the samples the same length. To reduce the number of zeroes we only use the n samples that are closest to each other in duration. This way we reduce the number of training samples, but hopefully, the efficiency of our network will be much better.

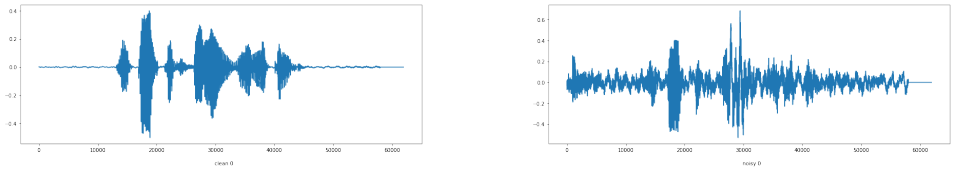
2.2 Zero padding

Our network architecture only supports samples of the same length. To fulfill this constraint, we add zeroes for the end of each recording. The number of zeroes was calculated from the difference between the longest and the current sample.



(a) clean (b) noisy

Figure 2: An example clean-noisy pair

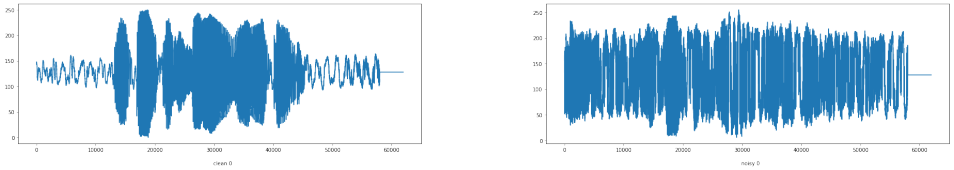


(a) clean (b) noisy

Figure 3: An example clean-noisy pair after zero padding

2.3 μ-law transformation

To make future computations faster we, reduce the samples to 8 bit with μ-law transformation.



(a) clean (b) noisy

Figure 4: An example clean-noisy pair after padding and μ-law transformation