Deep Learning, Fall 2021

Final Project: Enhancing Voices for Better Speech Intelligibility

8 november 2021, Wojciech Rubino, s00128

Synopsis

Motivation: Many people struggle to understand speech in challenging acoustics, e.g. inside a noisy bar. Therefore, enhancing the intelligibility of noisy speech signals is one of the key challenges for any producer of modern communication devices.

The problem is often tackled by dividing a noisy speech signal into a number of frequency bands and attenuating the ones where the signal-to-noise ratio is insufficient. This approach, while effective in some situations, often leads to poor results, and sometimes even exacerbates the problem it is trying to solve.

The hearing aids, aside from attempting to remove the noise, also amplify the frequency bands in which the wearer has a hearing loss. The gain applied at these frequencies depends on the user's audiogram, which means that broadly speaking, all users with the same audiogram are given the same gain setting. However, the sound perception among the individuals sharing the same audiogram can vary greatly, and thus finding a better way of adapting the frequency-gain curve to the individual needs has the potential for increasing the improvement of speech intelligibility by the device.

In this project, we will aim to train a model for obtaining the frequency-gain curve that will result in the highest speech intelligibility. The model is implemented using the PyTorch differentiation package and evaluated on the TIMIT speech dataset. We will use an objective function correlated with HASPI speech intelligibility auditory processing mode.

Plan:

8 November - 14 November:	Get test audio, build a minimal working model
15 November - 21 November:	Enhance the model
22 November - 28 November:	Enhance the model
29 November - 7 December:	Create a poster
8 December:	Exam (Poster Session)
10 December - 1 Januar:	Write the report
2 Januar:	Deadline report