Neuroprothetics Exercise 7
CI Vocoder

Dominik Scherer

27. January 2024

1 Cochlear Implant Vocoder

The cochlear implant (CI) is the most successful neural prosthesis developed to date [1]. While making it possible for many people to understand speech again, its overall sound information transduction is limited by the amount of available frequency channels. In this exercise, a recorded word is processed like in a CI to get a feeling of how the sound encoding works and what quality of hearing can be expected from a patient who received a CI.

2 Audio and noise spectograms

For a demonstration of the sound processing in our virtual CI, the word "Neuroprosthetics" was recorded using Python's sounddevice library with a sampling rate of $sr = 21\,\text{kHz}$ and with 32 bit resolution. The word was chosen since it contains plosives as well as fricative letters. The recorded signal was visualized using a spectrogram, which displays the power spectral density (PSD), which is the relative power of a certain frequency, in time windows of 15 ms with an overlap of 5 ms. This ensures a relatively smooth representation of the frequency components of the signal at a given point (rather window) in time. The spectrogram of the recorded word is shown in Figure 1.

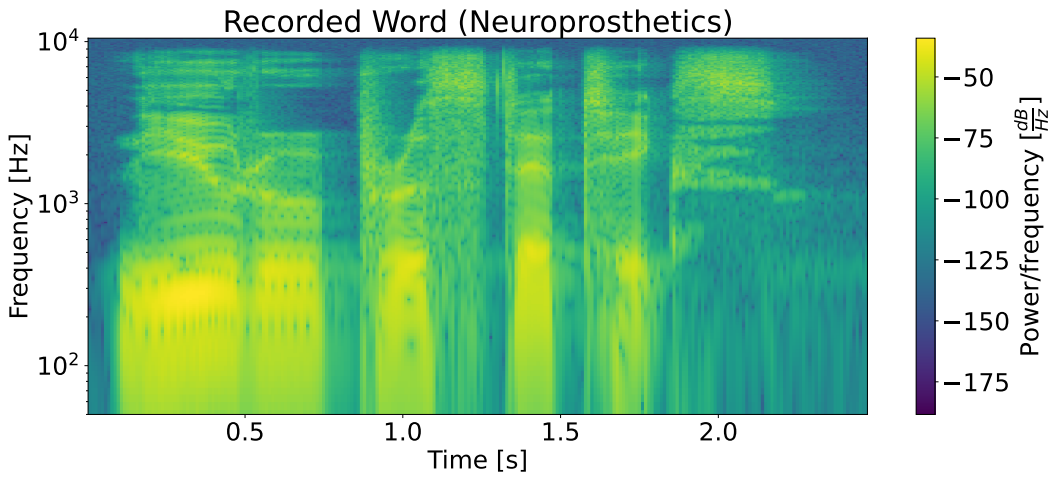

Figure 1: Spectrogram of the word "Neuroprosthetics".

The reason why a spectrogram and not a spectrum is necessary to obtain all the important information is simply because a spectrum only shows the overall contribution of each frequency to the whole signal. Therefore, the information about the time at which the frequency is present is completely lost. Figure 2 shows what a spectrum of the recorded word looks like. It is obvious that it's impossible to reconstruct a word from a pure spectrum. Still, it is fascinating to see what frequencies are most dominant over the whole time that was recorded and in what frequency range speech is happening.

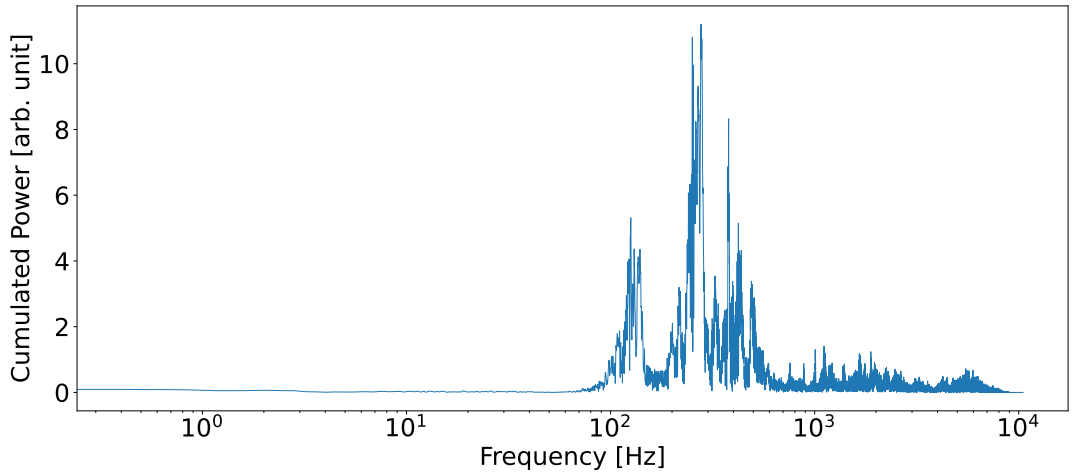

Figure 2: Frequency spectrum of the word "Neuroprosthetics".

To generate the signal that will be sent to the electrodes of the CI, we also need some white noise. For that, a random number for each timestep of the recorded word was generated with an average value of 0 and a standard deviation of 1 (Gaussian white noise). The respective spectrogram is shown in Figure 3.

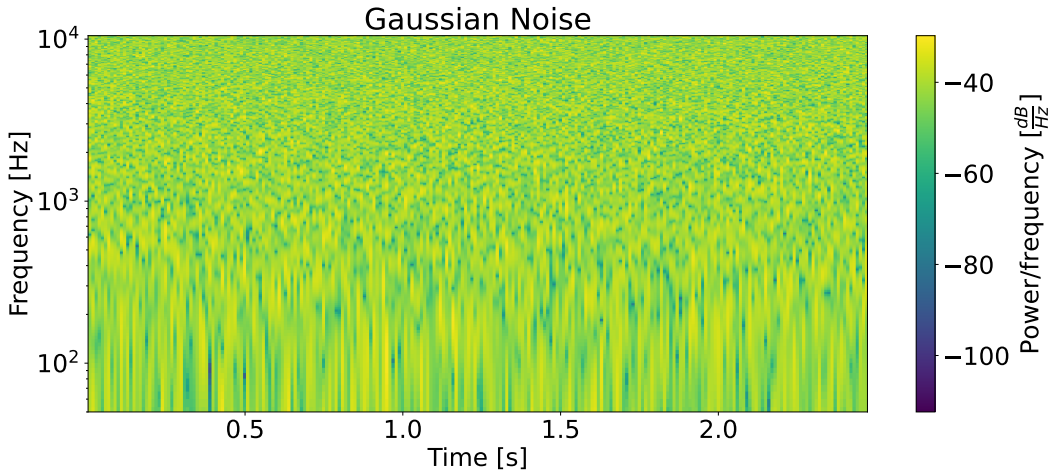

Figure 3: Spectrogram of Gaussian white noise.

3 Filters

Since the CI is limited by the amount of stimulating electrodes, thus frequency channels, the speech signal will be filtered into 10 frequency bands. For that, a filter bank with 10 bandpass filters is implemented for a logarithmically spaced range of 100 Hz − 8 kHz. The filter bank with its border frequencies at a gain of −3 dB is shown in Figure 4.

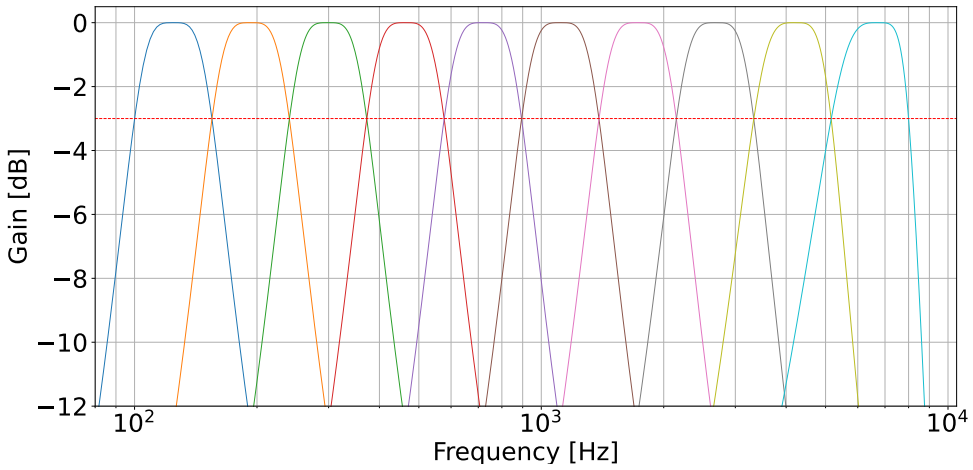

Figure 4: 10-channel filterbank with border frequencies at a gain of −3 dB of each channel coinciding with the adjacent ones.

The filter bank was applied to the speech and the noise signal to obtain each 10 frequency-specific signals. The voice and the noise signals for the 2nd and the 9th filter are shown in Figures 5 to 8.

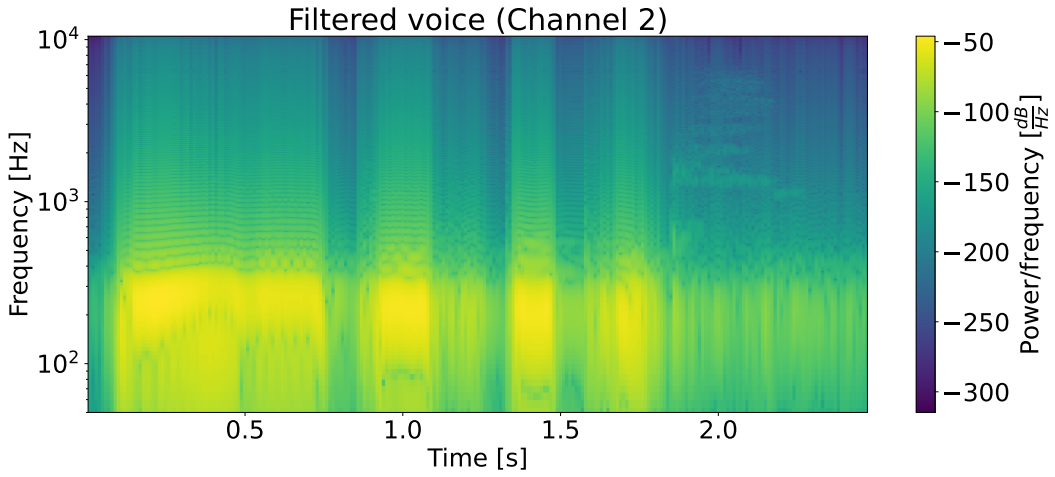

Figure 5: Spectrogram of voice signal filtered by the second frequency channel.

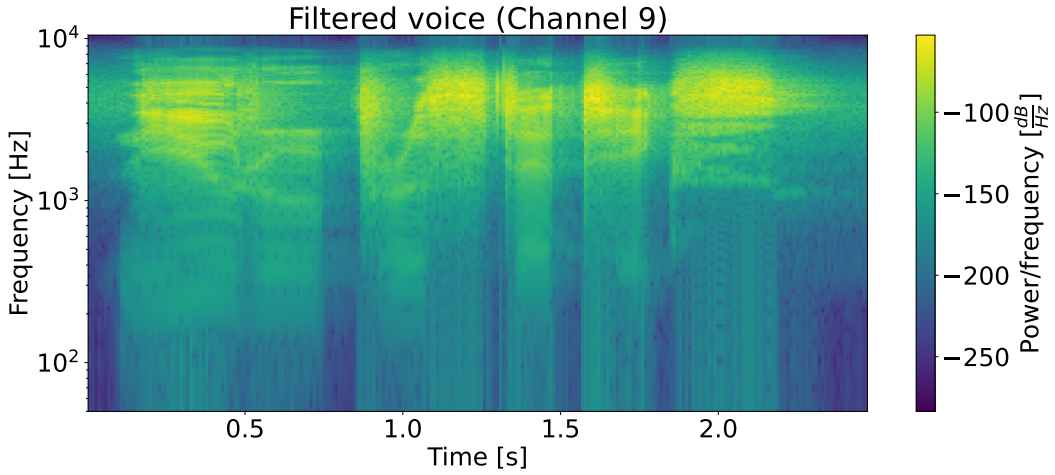

Figure 6: Spectrogram of voice signal filtered by the ninth frequency channel.

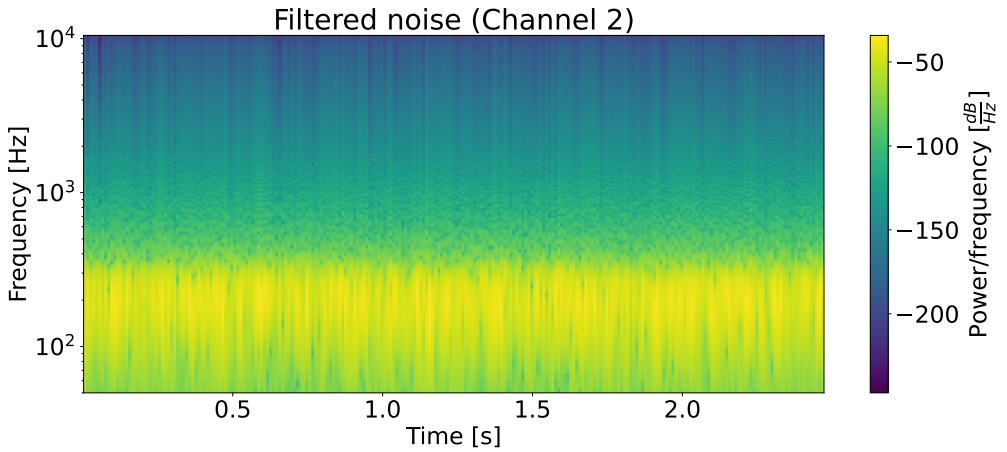

Figure 7: Spectrogram of noise signal filtered by the second frequency channel.

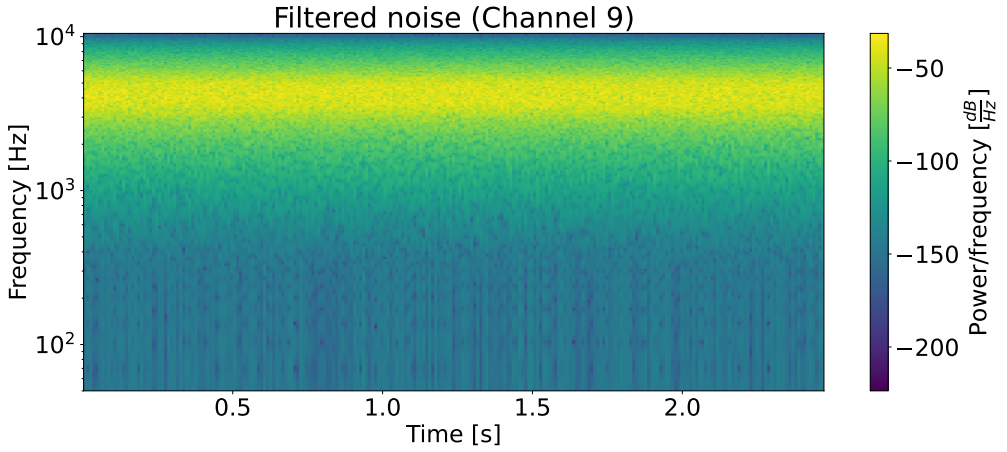

Figure 8: Spectrogram of noise signal filtered by the ninth frequency channel.

When played, the voice signal filtered by the 2nd frequency channel sounds muffled, and it's hard to understand single letters. The 9th channel sounds more sharp or bright. Especially the letter "s" can be understood very well. This is visible in the spectrogram in Figure 6 where shortly after $t = 1\,\mathrm{s}$ and at $t = 2\,\mathrm{s}$ the PSD is the largest as at these points in time the two "s" of the word "neuroprostetics" lay. The explanation for this is simply that the letter "s" consists of higher frequencies than letters like "o" or "n". The "o" is visible only in the voice signal filtered by the 2nd channel at around $t = 0.7\,\mathrm{s}$, but in the voice signal filtered by the 9th channel, it is not really visible. Since the letters are only distinguishable with the frequency information, it is important to cover as many frequency channels as possible to improve the quality of hearing with a CI.

4 Vocoder

To generate the signal in the CI, the amplitude information over time of the different frequency bands is needed. To get this information, the absolute value of the envelope of the filtered signal is obtained by a Hilbert transform for each frequency channel.

The electrodes used in a CI can usually only employ a much smaller dynamic range than human hearing. To simulate this restriction, the so far processed signal of each channel is compressed by applying this equation to the signal S:

$$S_{comp} = \frac{\log_{10}(1 + 300 \cdot S)}{\log_{10}(1 + 300)} \,. \tag{1}$$

Finally, the compressed magnitude information of each filtered frequency band can be translated back to a sound signal by multiplying it with its respective filtered noise signal. This is called modulation and is done timestep by timestep to get a signal vector of the same length as the original signal vector. Adding each channel together results in a complete vocoded version of the originally recorded sound like it would be the output of a CI. The spectrogram of the total vocoded speech signal is shown in Figure 9

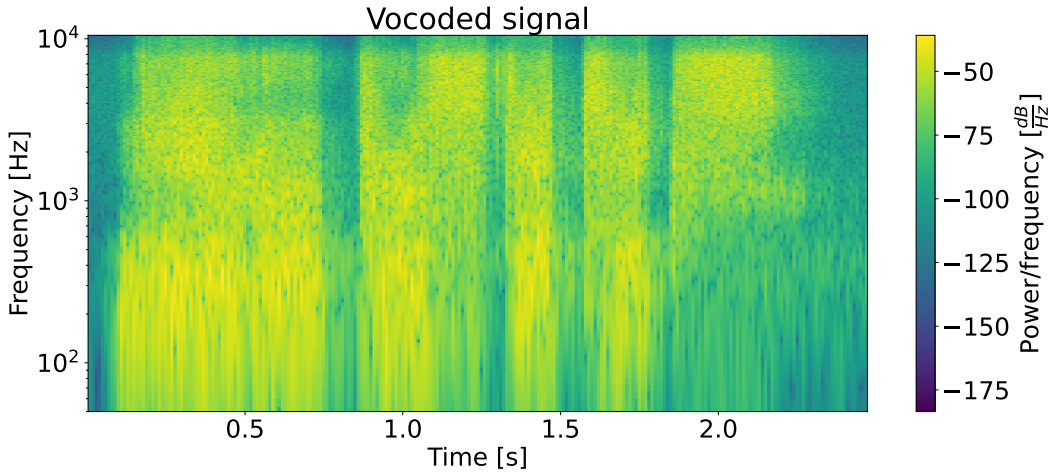

Figure 9: Spectrogram of vocoded signal.

When listening to the vocoded signal, it is definitely harder to understand everything as clearly as the original record, but it is possible. It sounds like a bad robotic voice from an 80's movie and, in general, more rough or tinny than the original record.

I tested several other words and phrases and also asked friends whether they could understand purely the vocoded signal without information about the initial record. While some words were hard to grasp, most of the time, it was possible for them to understand the vocoded signal when I spoke clearly and slowly in the original record. Some even said

after listening to different vocoded sounds that they felt like they were adapting to it and were getting better at understanding such sounds. While my friends may just have imagined this effect, a real adaptation is found in patients, as long as the implantation happens not too long after hearing loss [2].

References

[1] Blake S Wilson, Michael F Dorman, Marty G Woldorff, and Debara L Tucci. Cochlear implants: matching the prosthesis to the brain and facilitating desired plastic changes in brain function. *Progress in brain research*, 194:117–129, 2011.

[2] Anne-Lise Hiel, Jean-Marc Gerard, Monique Decat, and Naïma Deggouj. Is age a limiting factor for adaptation to cochlear implant? *European Archives of Oto-Rhino-Laryngology*, 273:2495–2502, 2016.