Bachelor thesis

Artificial Intelligence

Radboud University

Electrical sounds: preprocessed music for cochlear implant users

Author:
Lizzy Grootjen
s1001148

First supervisor:
dr. Y. Güçlütürk
Artificial Intelligence
y.gucluturk@donders.ru.nl

Second supervisor:
dr. U. Güçlü
Artificial Intelligence
u.guclu@donders.ru.nl

January 27, 2021

Abstract

Music perception for cochlear implant users is limited due to inaccurate pitch perception. As music is embedded in the cognitive system of humans, it is important to improve music perception for cochlear implant users. This study investigates how a neural network can improve pitch perception by amplifying the resonant frequencies in music. In the end, all frequencies were amplified instead of only the resonant frequencies, due to time constraints. The neural network started with a loss of 120,000 and ended with a loss of 1000, which indicates it learned features from the spectrogram of the music snippets. The online questionnaire did not show a significant result for the comparison between the original vocoded music and optimised vocoded music. There was a high variability between participants. The model can be improved by only amplifying the resonant frequencies, either implicitly or explicitly. The model can also be combined with other preprocessing strategies, such as F0mod and semitone mapping.

Contents

Chapter 1

Introduction

In a column written by Hennie Epping, he explains how deafness and the cochlear implant influences his life, in particular with music [4]. After he got his implant, he had to learn again how to enjoy music, as it sounds different than before he became deaf. He describes music with a cochlear implant as "one mess of sounds". It took him several years to be able to enjoy music again. He's not the only one; about half of all CI-users have trouble enjoying music after surgery, but the quality of life for all CI-users is expected to increase if the quality of music perception increases [10].

In any case, it is possible for deaf people with an intact auditory nerve to retrieve some hearing ability with a cochlear implant. A cochlear implant stimulates the auditory nerve directly, and therefore bypasses the cochlea and hair cells. However, the sound processing within the cochlear implant is based on speech perception, rather than music perception. CI-users experience music differently after their surgery, especially songs without verbal cues where CI-users cannot rely on words [11]. When listening to only one instrument, recognising the song is still doable, but once there are two or more instruments, it is too complex for a CI-user to decipher the song. Most songs make use of more than one instrument, especially in classical music. As music is very complex, in melody, rhythm and the number of instruments used, it is not enjoyable for CI-users to listen to [14]. Consequently, research has shown that people with a CI do not listen to music after their surgery [11].

This is problematic, because music is embedded in our cognitive system. Even though not everyone can play an instrument, everyone is able to sing or hum a tune [17]. Therefore it is important for cochlear implant users to be able to enjoy music again. However, listening to music with a cochlear implant is something that a CI-user needs to learn. It also depends on the capability of a patient's brain to adapt to the new device. Not everyone might be able to experience music again to the same extent as they did before the surgery. Therefore, listening to music through the cochlear implant must be simplified.

In this study, one solution will be examined which tries to speed up and simplify the process of learning to perceive music again after a cochlear implant surgery. The focus will be on classical music, as this is highly complex music without verbal cues that cochlear implant users can rely on. Therefore, the research question is as follows: "How can classical music perception be improved without verbal cues for cochlear implant users using a neural network for preprocessing the music pieces?".

As explained before, the current technology is not optimised for music perception, because music is more complex than speech [11] [13] [14]. Since it is inconvenient for CI-users to adjust the implant -they will need to have another surgery- this study will be focused on the preprocessing of music, such that it becomes more understandable for CI-users.

In order to improve music perception, a temporal convolutional neural network (TCN) will be trained on classical music. The TCN learns how to optimise the songs such that it becomes more enjoyable for CI-users. This neural network will be programmed to optimise the pitch, as CI-users have difficulty hearing pitches and melodies [12]. In order to test the network, the original and optimised songs will be put through a noise-vocoder -a simulation of what listening through a CI sounds like- and participants of an online survey will be asked which music piece they find more pleasant to listen to: either the vocoded optimised song or the vocoded original song. This way, it can be determined if CI-users benefit from preprocessing the music by optimising pitch.

In related work, other preprocessing strategies for music will be discussed. However, there have only been few attempts trying to improve music perception for CI-users. The strategies with promising results are still ongoing and not yet implemented in all cochlear implants. These strategies have a fundamentally different approach compared to the strategy proposed by this study, which will explained in more detail later. Therefore, it might be beneficial to combine them for future research.

The structure of this study is as follows: first, a short explanation will be provided with some preliminary knowledge. Then, related work will be discussed, as the models in this study are based upon related research. After that, the research itself will be discussed in detail, including the models. This will be followed by the conclusions drawn from the study.

Chapter 2

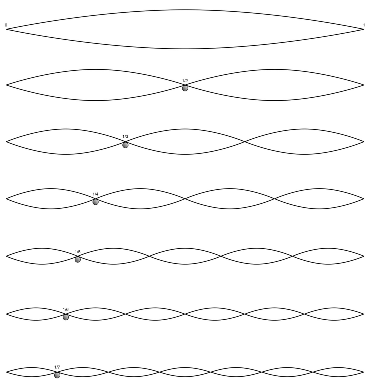
Preliminaries

In this chapter, the cochlear implant and relevant limitations are explained. To this end, some prerequisite knowledge is required. As a cochlear implant simulates normal hearing, the components of a sound wave and how this wave is processed in a healthy ear are explained. This is compared to an impaired hearing patient, to highlight where the sound processing is disturbed. Using a cochlear implant, the disturbance is bypassed in the impaired hearing patient. The implant is a new pathway to the auditory nerve, therefore skipping impaired components in the ear. In the sections below, each step is explained in further detail.

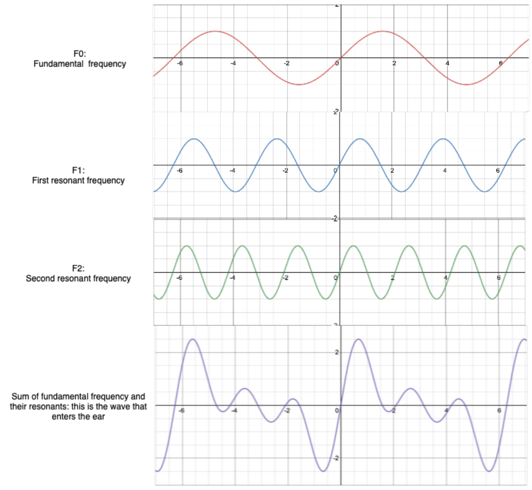
2.1 Sounds and the ear

Sound waves are generated by the vibration of an object, which causes the air surrounding the object to oscillate [5] [8]. These oscillations are changes of pressure in the air. Sound waves have three important domains: temporal domain, the amplitude and the frequency. The temporal domain determines when the change of pressure takes place. The amplitude denotes how much air started to vibrate, also known as the volume. The frequency determines how often the sound wave repeats itself per second, known as the pitch. The object causing vibrations may be a person talking or an instrument being played.

To demonstrate how sound waves work on an instrument, the A-string on a guitar is used as an example. Plucking the A-string on a guitar with standard tuning will produce the most prominent frequency in 110.0 Hz. However, this is not the only frequency in the sound wave. Every tone consists of the fundamental tone, which in this case is the A on 110.0 Hz, and several resonant frequencies [14]. Resonant frequencies are sound waves which fit exactly in the fundamental tone, as can be seen in figure 2.1a. Therefore, the first resonant frequency on this A will have a frequency of 220.0 Hz, the second resonant frequency 330.0 Hz, etc. The sound the A-string produced is obtained by summing the frequencies, which is visualised in figure 2.1b. The timbre of an instrument is determined by which resonant frequencies are more prominent [16]. The pitch is determined by the fundamental frequency, but in absence of the fundamental frequency, pitch can still be recognised due to the resonant frequencies [13] [19].



(a) Visualization of the resonant frequencies shapes compared to the fundamental frequency shape. Source: Wikipedia [23]



(b) Visualization of the fundamental frequency and resonant frequencies in one sound wave.

Figure 2.1: Fundamental frequency and resonant frequencies visualised

The oscillations caused by the guitar will change the pressure in the air. People in range will recognise this change as the sound made by the guitar. For normal hearing people, the sound wave will cause the eardrum to vibrate on the same frequencies. The ossicles are connected to the eardrum, and thus also start to vibrate [7]. The stirrup, the ossicle connected to the cochlea, will then stimulate the oval window. Due to the moving oval window, the pressure within the cochlea changes. As the cochlea is filled with a fluid, this fluid starts to move. The vibrations in the fluid will stimulate the outer hair cells, which then stimulate the basilar membrane. This membrane is connected to the inner hair cells, which are connected to the auditory nerve.

As the pitch perception is established in the cochlea, it will be examined in more detail. The cochlea has a tonotopical mapping, which means each place in the cochlea maps to a certain frequency in a specific ordering. The fluid inside the cochlea moves due to the sound wave, which stimulates the basilar membrane. Due to the varying thickness and width of the basilar membrane, the basis of the cochlea is sensitive for high frequencies. And vice versa, the apex of the cochlea is sensitive for low frequencies. This means when a sound has a lower frequency, it first passes through the higher frequency. This is probably also the reason why people tend to lose high pitch

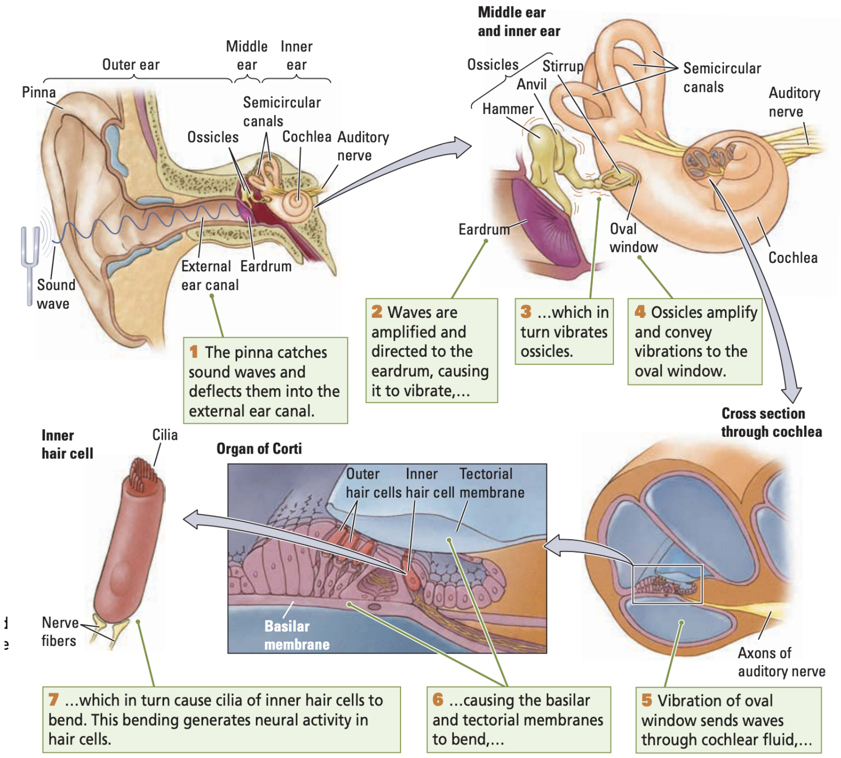


Figure 2.2: The anatomy of the human ear. Source: An introduction to brain and behaviour - Bryan Kolb [7]

hearing ability first, but this has not yet been confirmed. When a sound wave enters the cochlea, the fundamental frequency is mostly stimulated. However, the resonant frequencies are stimulated too, which is important for pitch recognition [13] [19].

When the outer hair cells are damaged, frequencies entering the cochlea will not be processed by the basilar membrane, and thus cannot reach the brain. As the implant bypasses all components up until the inner hair cells in the cochlea and stimulates the auditory nerve directly, some hearing ability can be retrieved.

2.2 Structure of a Cochlear implant

A cochlear implant consists of a microphone, a sound processor, a transmitting coil, a receiving coil and an array of electrodes [8]. The sound which normally is caught by the pinna and eardrum, is now caught by the microphone. The microphone transfers the sound to the sound processor, which processes and converts the sound to an electrical signal. Then the electrical signal is transferred to the transmitting coil, which passes the signal to the receiving coil. The receiving coil sends the electrical signals to the array of electrodes. Each electrode is matched to a position in the cochlea and thus every electrode stimulates a different frequency. The electrode array contains around twenty electrodes. As inserting the electrode array in the cochlea damages the tissue, it is unusual to use more than twenty electrodes. Therefore, a trade-off must be made between the amount of damage and the number of electrodes inserted.

As a cochlear implant consists of an electrode array with a certain number of electrodes (usually twenty electrodes [2] [13] or twenty-two [8]), a sound wave with a continuous range of frequencies needs to be mapped to these twenty electrodes. For this, a filterbank is used. A schematic image of a filterbank is shown in figure 2.3.

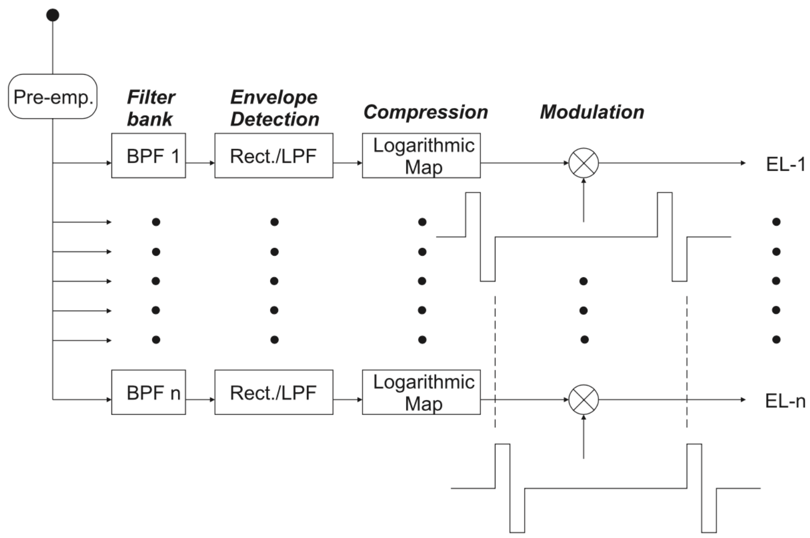


Figure 2.3: Schematic filterbank. By: J. Laneau [8]

A way to simplify the sound wave to facilitate mapping to the electrode array, is to use the envelope of a sound signal. This is done in the preprocessing step. The envelope of a sound wave is an imaginary smooth line that is drawn on top of the time domain signal [5]. Then, the envelope is passed through the filterbank. The filterbank uses bandpass filters to filter out the frequencies which cannot be mapped to the electrodes, as there is no electrode at the position where the frequency needs to be stimulated. Bandpass filters are filters which cut off frequencies in a sound wave [5]. A visualisation of this can be seen in figure 2.4.

As explained earlier, one tone does not consist of only one frequency. The different frequencies needs to be mapped to the electrode array. Compared to the cochlea, this electrode array is a discrete scale. When an instrument plays a tone of which the prominent frequency is exactly between two points

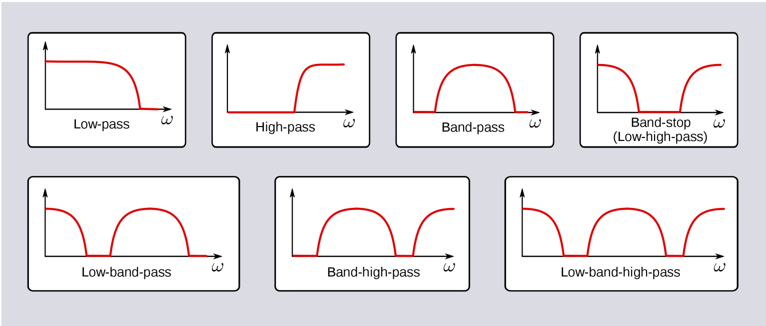


Figure 2.4: Visualisation of different filterbanks. Source: Wikipedia [22]

on the array, an algorithm has to choose which position is the best, or filter out the frequency with a bandpass filter. The second limitation is that once a tone is processed, the harmonic structure is not kept, thus the resonant frequencies are absent [14]. Even when the fundamental frequency is absent, resonant frequencies facilitate in perceiving and recognising the pitch [19]. Therefore, the absence of resonant frequencies makes it even harder to distinguish pitch.

2.3 Issues with the Cochlear Implant

As cochlear implants are designed with speech perception in mind [11] [14] [13], the perception of music for CI-users is not as good as their speech perception. In order to improve music perception, it is important to know which aspects are insufficient in the current technology of cochlear implants. Several studies found two major issues in a cochlear implant which could be the cause of disturbed music perception.

Firstly, participants experience a pitch shift, which results in the perception of sound which is between one and two octaves lower than the actual sound [2] [21]. Why this happens, is not entirely clear. It could be due to the frequency allocation for a specific electrode in the sound processor [21].

Secondly, some musical aspects are hard to recognise for cochlear implant users, as the harmonic structure of music is complex [14]. This mostly involves pitch, melody and timbre [11] [12]. This makes sense, as these three things are closely related. A melody consists of a sequence of pitches together with rhythms. Rhythms are not significantly different for CI-users. However, when one component in a melody changes it becomes harder to to recognise the melody. The timbre depends on the instrument and their resonant frequencies spectrum. Another study has shown that CI-users on average cannot distinguish two tones that are less than seven semitones apart, whereas healthy hearing users can distinguish two tones less than a tenth of a semitone [8].

Chapter 3

Related Work

This chapter discusses related studies and compares the differences between this study and the related studies.

As explained earlier, the disturbances in music perception is caused by the discrete scale of the cochlear implant combined with the processing strategy used in the implants, advanced combination encoders (ACE) [8] [14]. There are various existing solutions to amplify the fundamental frequency in musical tones, for example by using the F0mod sound processing strategy or by using the semitone mapping(s) Smt-LF and Smt-MF [8] [15]. These three methods are discussed in further detail.

The first sound processing strategy is F0mod [8]. This strategy is developed by J. Laneau and is meant as a replacement for the advanced combination encoder. This method consists of two steps: the computational model and the pattern generation. The computational model extracts the pitch a normal hearing user would perceive. Then, the stimulation pattern is generated. This stimulation should result in an equally high pitch as the normal hearing users experience. This stimulation is then used in the cochlear implant. The stimulation pattern is generated as follows: firstly, the value of the fundamental frequency is estimated based on autocorrelation of the input signal. Then the method selects maxima in the sound wave, compresses the signal and maps the waveform to a stimulation pattern. The main issue with this approach is that the complex harmonic structure is not kept, as the waveform is compressed.

The second and third optimisation strategies are the semitone mappings Smt-LF and Smt-MF, developed by Omran et al [13] [14] [15]. These methods focus on improving pitch perception by making sure the fundamental frequencies of adjacent tones are assigned to different electrodes. The Smt-LF method covers the lower and middle frequencies, whereas the Smt-MF covers the middle and higher frequencies. The idea is to investigate the fundamental frequencies and sort those into the channels on the electrode array. If two adjacent tones are played (e.g an A on 440 Hz and a A# on

466.2 Hz), the mapping makes sure those two tones are allocated to different electrodes. This way, those two tones have a different stimulation and thus the cochlear implant user is able to distinguish between the two semitones. The results show a significant difference in pitch perception and melody recognition between cochlear implant users with semitone mapping compared to users without semitone mapping. During the mapping, the resonant frequencies are also mapped in such a way that they are consistent with the fundamental frequency. An issue that may arise with this approach is the context-sensitive stimulation. Depending on the adjacent notes, it is possible that a certain note is first stimulated on electrode 1 but later in the song, it is stimulated on electrode 2. This could change the melody.

The main difference between the other optimisation strategies and the strategy proposed in this study, is the focus on which frequencies to amplify or adjust. These strategies focus on improving the fundamental frequency. However, there is no solution yet which increases the perception of resonant frequencies, which also facilitates in recognising pitch. Especially when there is no ideal matching electrode for a certain fundamental frequency, and thus filtered out by the filterbank, it is even possible to decide to drop the fundamental frequency and focus on amplifying the resonant frequencies. This prevents the two problems which arise in the above methods: the compression of the sound wave, which leads to disruption in the harmonic structure, and the context-sensitive stimulation, which leads to a different melody compared to normal hearing users.

Chapter 4

Research

This chapter discusses the details of this study, which includes the exact research question, hypothesis, and the components needed to perform the study and reproduce the results. Two implementations are made supporting the study, namely the noise-bandpass vocoder and the temporal convolutional neural network. The results of the neural network is presented to the participants in an online survey, to determine if the neural network was able to reach its goal.

4.1 Definitions and Research Question

As little is known about amplifying the resonant frequencies in music for cochlear implant users, this study focuses on improving pitch perception by making the resonant frequencies more prominent in a tone. Firstly, the exact research question is formulated as follows:

> How can classical music perception be improved without verbal cues for cochlear implant users using a neural network for preprocessing the music pieces?

The hypothesis of this study is when the resonant frequencies are amplified in music, it becomes easier for a cochlear implant user to perceive pitch. Therefore, this network is going to optimise the pitch. As this is still a vague concept, pitch is defined as follows:

> The word 'pitch' is used to describe the perceptual differences that arise when the fundamental frequency of a musical sound is changed, while other physical characteristics are held constant. [2]

From this, the study defines optimising the pitch as follows:

> To optimise the pitch is to maximize the amount of resonant frequencies in a music piece. This is done to preserve the harmonic structure as well as possible when the fundamental frequency is filtered out by the filterbank in the cochlear implant.

The goal of this study is to make the optimisation happen automatically.

4.2 General Research setup

In order to answer the research question, a study has been performed. This study consists of three parts: two implementations and a online survey. The two implementations are a noise-bandpass vocoder (in short, vocoder) and a neural network. The vocoder simulates the cochlear implant and the neural network optimises the songs. The idea is as follows: a certain music piece is optimised by the neural network (called TCN). Then the optimised song is put through the vocoder. The vocoder outputs two songs: the optimised song without additional noise and the optimised song with additional noise. The noise is included to simulate different situations while wearing a cochlear implant, as sometimes the implant itself also has some noise or the user is in a noisy environment. The original song is also put through the vocoder, which results again in two songs: the original song without additional noise and the original song with additional noise. Then, the online survey is conducted. This survey consists of two parts: a part with songs including noise and a part with songs not including noise. For each trial, the participant was presented with two versions of a song: the vocoded original and the vocoded optimised. The participant had to indicate which song they found more pleasurable to listen to. The flow of the research is displayed in figure 4.1. A statistical analysis indicates if there is a significant difference between the original songs and the optimised songs. The details on the implementations and the online survey are discussed in further detail in the next sections.

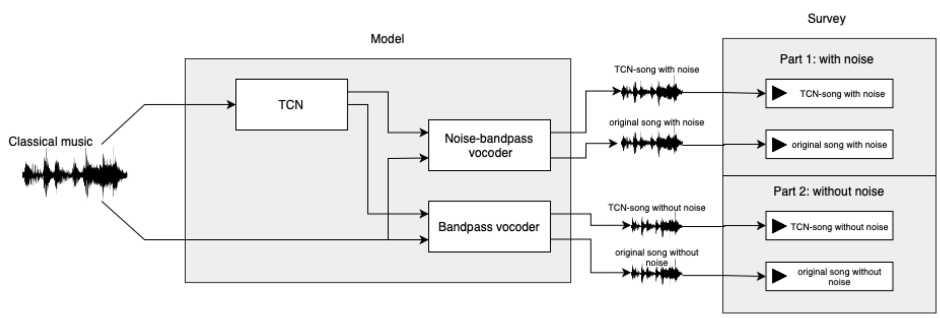


Figure 4.1: Research setup diagram

4.3 Implementations

In order to create the vocoder and train the network, data needs to be provided. The data used is the MusicNet dataset [18]. This dataset contains 330 songs with labels of classical music, which is separated in a train set, validation set and a test set. Note that the labels are not used in this study, as regression is performed rather than classification.

4.3.1 Noise-bandpass vocoder

For the simulation, another study investigated which acoustic models were suitable for a CI-simulation [8]. Multiple models were investigated, but the simple, yet effective, noise vocoder was considered the best option. Therefore this study also uses a noise-bandpass vocoder. This noise-bandpass vocoder uses the Butterworth bandpass filter. This filter matches the filterbank of the cochlear implant: there are a few electrodes matched to frequencies which can be stimulated, and the rest is cut out. As described earlier, cochlear implant users can only distinguish sounds within seven semitones, and in general the sound is perceived around one to two octaves lower. These features are incorporated in the vocoder. The center frequencies are based on a cochlear implant with deep-insertion, as deep-insertion facilitates in recognising pitch [6]. To reduce complexity, this study simulated a cochlear implant with six electrodes instead of twenty.

The implementation used in this study is inspired by an implementation made by Alexandre Chabot-Leclerc [3]. This implementation consists of three building blocks: the bandpass filter, the envelope extraction and adding noise. This vocoder simulates a cochlear implant by filtering the signal with a Butterworth filter, based on the center frequencies provided. Then, it extracts the envelope of the filtered sound signal and adds noise. This is displayed in figure 4.2

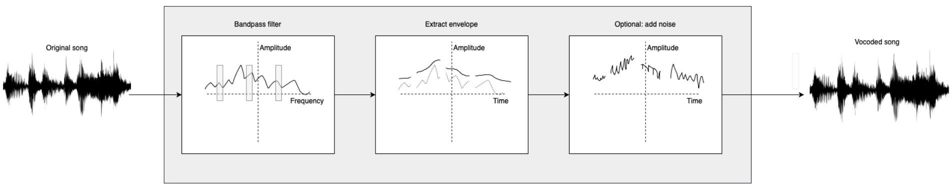


Figure 4.2: Noise-bandpass vocoder diagram

In order to make this code work with PyTorch and tensors, some modifications were made. Firstly, the vocoder is adjusted such that it can work with batches. Secondly, two options are added: an option to disable adding noise, and an option to amplify frequencies after filtering. Thirdly, the functions from NumPy were replaced with functions for PyTorch, except the filter function. The filter used for this implementation is created by Floris

Laporte, and is a faster version than the one provided by the torchaudio library [9]. By replacing NumPy functions with PyTorch functions, the vocoder is able to deal with tensors and the network is able to backpropagate.

4.3.2 Temporal Convolutional Network

Next, this vocoder needs to be incorporated with the neural network. As music is time series data, the choice on what kind of network to use is limited. The straightforward choice for time series data is a LSTM, but the temporal convolutional network (TCN) has outperformed it due to the attention mechanism [20]. A PyTorch implementation made by Shaojie Bai is used for this study [1]. The network consists of the following layers: an input layer of one node, four hidden TCN layers which contain 8, 16, 16 and 8 nodes respectively, a linear layer with 8 nodes and an output layer of one node. The mean squared error is used as a loss function. The optimizer for this neural network is Adam, with a learning rate of 0.001 and a weight decay of 0.99. The learning rate is adjusted every 10 epochs with a factor of 0.1. The TCN layer has a drop-out rate of 0.1. The model is trained for 250 epochs and a batch size of ten for the training and validation set. The test set has a batch size of one.

For training the network, a random sample of eight seconds is taken from a song. This song is put through the network, after which the network will output an optimised song. As described in the previous paragraph, the network consists of several temporal convolutional layers. These layers capture the time-based features of music which are considered important. The optimised song is then passed through the vocoder. While the network is learning, the vocoder does not add noise, to prevent the network to learn denoising audio instead of optimising pitch. The original song is also passed through the vocoder. After vocoding the original song, the frequencies which are not filtered out by the filterbank are amplified. Next, the spectrogram using Short Time Fourier Transform is calculated from the vocoded original song and the vocoded optimised song. This yields an array containing frequencies against time. These two arrays are put into the loss function, which evaluates whether the network captured all important features. The idea is that the neural network tries to bring the spectrogram of the optimised song as close as possible to the original amplified song. Initially, the project was designed to only amplify the resonant frequencies. However, this turned out too complex to be finished in time. Pitch tracking of a musical instrument including all the resonant frequencies is already known as a difficult problem. On top of that, for this study, the pitch tracking function should also be differentiable in order for the neural network to backpropagate. The learning process is visualised in figure 4.3. The full code of this implementation can be found on GitLab [1].

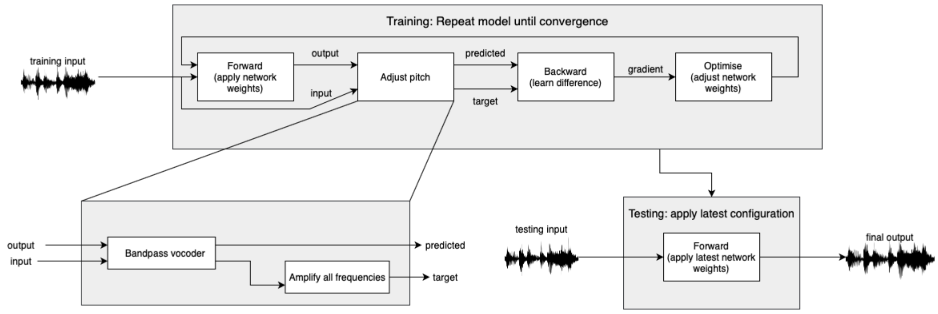


Figure 4.3: TCN network diagram

After training the neural network, the test set is put through the network. The network outputs the optimised version of the song which was input. Both songs -the optimised and the original- will be put through the vocoder again, both with and without added noise. These sound pieces are used for the online survey.

4.4 Online survey

The online survey consists of two parts: one part where the music contained noise, and one part where the music did not contain noise. Each part consisted of ten trials. As described earlier, each trial the participant is presented with two sound fragments: the vocoded original song and the vocoded optimised song. The participants do not know which song is the original and which one is the optimised. Participants are asked to choose which sound fragment they find more enjoyable to listen to. For each part, it is registered how often the participant chose the optimised song. This way, the model is evaluated and determined if it succeeded in making music more enjoyable to listen to by optimising the pitch. Also, it can be determined if noise influences the result.

4.5 Results

The plot containing the losses of the network is displayed in figure 4.4. The figure indicates the neural network was able to learn some of the features. A notable observation is the high loss value at convergence. At a first glance, it looks like the neural network is done learning. However, the loss is the difference between the original vocoded spectrogram and the optimised vocoded spectrogram. As the frequencies are amplified in the

[1] https://gitlab.socsci.ru.nl/L.Grootjen/neural-network-cochlear-implant-and-music

original vocoded spectrogram, this song is the target. As the loss is high even after convergence, the target is still not reached.

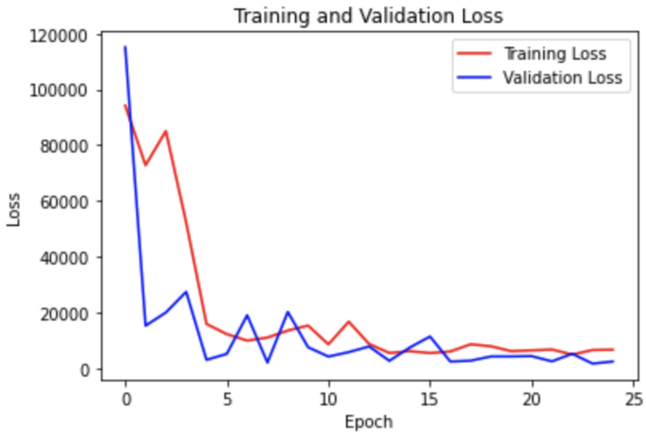


Figure 4.4: The loss of the network during training

The online questionnaire had 27 responses. To determine if there is a significant difference between using the neural network to optimise music and not optimising at all, a binomial significance test was used. For the songs including noise, there was no significant result found ($p = 0.1094$). For the songs without noise, there was also no significant result found ($p = 0.7539$). Pooling the data together, there was no significant result ($p = 0.2632$). A remarkable observation is the variability between participants. The results are shown in figure 4.5

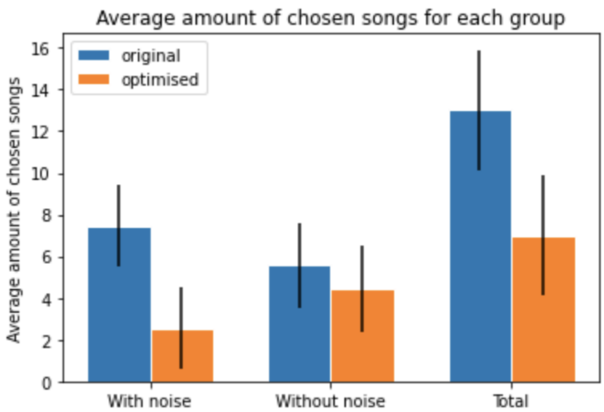


Figure 4.5: Bar plot showing the results of the online survey

Chapter 5

Discussion

This chapter draws a conclusion from the results of the survey and what this means for the field. Furthermore, limitations of this study are discussed and suggestions for future research are given.

In order to answer the research question, this study investigated how a neural network can be used to improve music perception by amplifying the resonant frequencies. From the survey it can be concluded the current model is not yet a good solution to improve music perception, as there was no significant difference found in all three groups. However, it is interesting to note that there was a high variability between participants. This indicates that some people did prefer the optimised song.

Although no significant difference was found, this model can be used as a start for the automatisation of optimising music for cochlear implant users. The study by Omran et al. found that isolating the fundamental frequencies from the resonant frequencies in the electrode array had a positive effect on pitch perception, it is likely that an improved model where resonant frequencies are amplified when the fundamental frequency is filtered out by the filterbank will also have a positive effect, as resonant frequencies facilitate in recognising the pitch [13] [14] [15] [19].

Moreover, there are several limitations which possibly contributed to not finding a significant result. Firstly, the response rate to the survey was low (27). This makes it harder to investigate the performance of the model. Secondly, there were no cochlear implant users involved in this study. The noise-bandpass vocoder was used as a simulation for perceiving music through an implant. However, future research would benefit by asking cochlear implant users to participate in research. Thirdly, due to computational limits, it was not possible to add more hidden layers and hidden units. As music is highly complex, the model might not be able to represent all features of music in the current number of hidden layers and units. Therefore, this model might benefit with a more complex model, especially as the loss is still quite high after convergence. Fourthly, the way how the resonant frequencies are amplified could have played a role. In this study, all frequencies which were not filtered out, are amplified. The model might perform better if only the resonant frequencies were amplified. This was the initial idea of the study, however, during the process it was proven too complex to also track the pitch of music including their resonant frequencies and be able to backpropagate this through the network.

For future research on this topic, a few suggestions can be made. As this neural network optimises the spectrogram, it might be beneficial to make it specifically optimise the resonant frequencies. As described earlier, this can be done by tracking the pitch of the music. This can be implemented in two ways: in an explicit approach and an implicit approach. In the explicit approach, the program amplifies the resonant frequencies manually by adjusting the original sound file. With the implicit approach, the amount of resonant frequencies in a song can be measured. This way, the neural network tries to optimise the amount of resonant frequencies, and thus amplifying them. These options are visualised in figure 5.1.

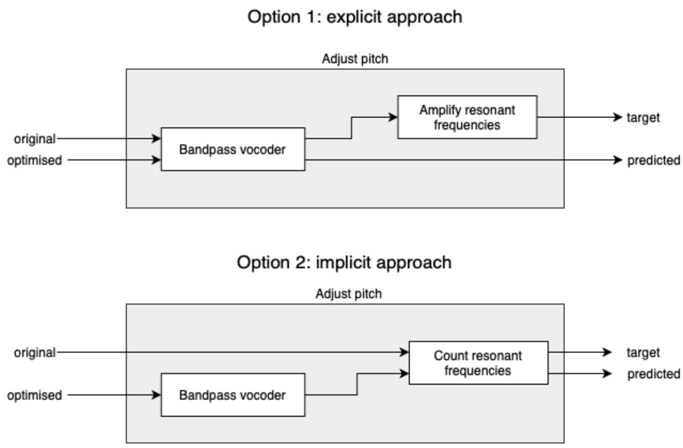


Figure 5.1: Future research suggestions

In order to implement this suggestion, a major leap has to be solved: an efficient and differentiable approach needs to be developed to track pitch of an music audio file. This approach needs to be able to detect the fundamental frequency and the present resonant frequencies and their amplitudes at a certain time point.

Another suggestion on a more higher level for future research is to combine an improved version of this model with other pitch optimisation methods such as Smt-LF and Smt-MF. As the model used in this study focuses on resonant frequencies and the models from Omran et al. on the fundamental frequency, they will complement each other.

In short, the proposed model with suggested improvements can be a good preprocessing strategy for cochlear implant users. Future research is required to investigate the full abilities of using a neural network to optimise pitch perception. Other literature has shown that resonant frequencies do facilitate in recognising the pitch, and thus it would be beneficial to improve this model.

Bibliography

[1] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*, page 14, 2018.

[2] Peter J. Blamey, Gary J. Dooley, Elvira S. Parisi, and Graeme M. Clark. Pitch comparisons of acoustically and electrically evoked auditory sensations. *Hearing Research*, 99(1-2):139–150, 1996.

[3] Alexandre Chabot-Leclerc. Vocoder, 2015.

[4] Hennie Epping. Mijn ervaring met muziek, 2018.

[5] James M Hillenbrand. The Physics of Sound Sound and Vibration. Technical report, Western Michigan University, 2016.

[6] Ingeborg Hochmair, Erwin Hochmair, Peter Nopp, Melissa Waller, and Claude Jolly. Deep electrode insertion and sound coding in cochlear implants. *Hearing Research*, 322:14–23, 2015.

[7] Ian Q. Kolb, Bryan ; Whishaw. *An Introduction To Brain And Behavior - Bryan Kolb fourth edition.* Kevin Feyen, fourth edi edition, 2014.

[8] Johan Laneau. *When the deaf listen to music.* PhD thesis, Katholieke universiteit Leuven, 2005.

[9] Floris Laporte. PyTorch lfilter, 2020.

[10] Luis Lassaletta, Alejandro Castro, Marta Bastarrica, Rosa Pérez-Mora, Rosario Madero, Josefa De Sarriá, and Javier Gavilán. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica*, 127(7):682–686, 2007.

[11] Mariana C Leal, Young Je Shin, Marie Laurence Laborde, Marie Noëlle Calmels, Sebastien Verges, Stéphanie Lugardon, Sandrine Andrieu, Olivier Deguine, and Bernard Fraysse. Music perception in adult cochlear implant recipients. *Acta Oto-Laryngologica*, 123(7):826–835, 2003.

[12] Valerie Looi, Hugh McDermott, Colette McKay, and Louise Hickson. Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing*, 29(3):421–434, 2008.

[13] SA Omran, Michael Büchler, Waikong Lai, and Norbert Dillier. Pitch Perception Using a Semitone Mapping to Improve Music Representation in Nucleus Cochlear Implants. *Uzh.Ch*, (Mci):2–5, 2007.

[14] Sherif A Omran, Waikong Lai, and Norbert Dillier. Melody contour identification and instrument recognition using semitone mapping in Nucleus Cochlear Implant recipients. *Jahrestagung der Deutschen Gesellschaft für Audiologie*, (Helmholtz 1954):1–6, 2007.

[15] Sherif Abdellatif Omran, Waikong Lai, Michael Büchler, and Norbert Dillier. Semitone frequency mapping to improve music representation for nucleus cochlear implants. *Eurasip Journal on Audio, Speech, and Music Processing*, 2011(1):1–11, 2011.

[16] E. L. Saldanha and John F. Corso. Timbre Cues and the Identification of Musical Instruments. *The Journal of the Acoustical Society of America*, 36(11):2021–2026, 1964.

[17] Jay Schulkin and Greta B. Raglan. The evolution of music and human social capability. *Frontiers in Neuroscience*, 8(SEP):1–13, 2014.

[18] John Thickstun, Zaid Harchaoui, and Sham M. Kakade. Learning features of music from scratch. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, pages 1–14, 2017.

[19] Peter M. Todd and D. Gareth Loy, editors. *Music and Connectionism*. The MIT Press, London, England, 1991.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem(Nips):5999–6009, 2017.

[21] Katrien Vermeire, David M. Landsberger, Paul H. Van de Heyning, Maurits Voormolen, Andrea Kleine Punte, Reinhold Schatzer, and Clemens Zierhofer. Frequency-place map for electrical stimulation in cochlear implants: Change over time. *Hearing Research*, 326:8–14, 2015.

[22] Wikipedia Contributors. Filter (signal processing).

[23] Wikipedia Contributors. Overtone.