Department of Physics and Astronomy

Heidelberg University

Bachelor thesis in Physics

submitted by

Josch Hagedorn

born in Jever (Germany)

2021

Investigation of classical filter and deep learning methods for the denoising of ECGs under various amounts and types of noise

This Bachelor thesis has been carried out by Josch Hagedorn at the

Interdisciplinary Center for Scientific Computing (IWR)

under the supervision of

Prof. Dr. Jürgen Hesser

Zusammenfassung:

Um EKG-Entrauschungsmethoden (klassischer Filter und Deep-Learning-Methoden) zu vergleichen, werden rauschfreie EKG-Proben künstlich mit den drei Rauscharten Grundlinienwanderung (BW), Muskelartefakt (MA) und Elektrodenbewegung (EM) verrauscht. Es werden verschiedene Experimente durchgeführt: EKG-Proben die einen einzelnen Herzschlag enthalten werden mit unterschiedlichen Mengen der einzelnen Rauscharten versehen, EKG-Proben die einen einzelnen Herzschlag enthalten werden mit verschiedenen Kombinationen der Rauscharten versehen und EKG-Proben die mehrere Herzschläge enthalten werden mit unterschiedlichen Mengen der einzelnen Rauscharten versehen. Die Leistung der Entrauschungsmethoden wird mit zwei verschiedenen Metriken gemessen. Klassische Filter sind für beliebig lange EKG-Proben ausreichend, aber nur für BW-Rauschen. Die Deep-Learning-Methoden werden für MA- und EM-Rauschen benötigt. Hier ist es erforderlich, dass die verwendeten EKG-Proben nur einen Herzschlag enthalten. Die Ergebnisse des Experiments für gemischtes Rauschen lassen sich nicht aus den Experimenten für die einzelnen Rauscharten ableiten.

Abstract:

In order to compare ECG denoising methods (classical filter and deep learning methods), noise free ECG samples are artificially corrupted with noise using the three noise types baseline wander (BW), muscle artifact (MA), and electrode motion (EM). Different experiments are set up: varying amounts of the separate noise types are added to ECG samples containing a single heart beat, various combinations of the noise types are added to ECG samples containing a single heart beat, and varying amounts of the separate noise types are added to ECG samples containing multiple heart beats. The performance of the denoising methods is measured using two different metrics. Classical filters perform sufficient for arbitrary lengths of ECG samples but only for BW noise. The deep learning methods are required for MA and EM noise. Here it is necessary that used ECG samples only contain one heart beat. The results of the mixed noise experiment can not be derived from the separate noise types experiments.

Contents

List of Abbreviations

BW Baseline wander. 3, 8–10, 19, 23, 27–29, 31–36, 38, 41, 55, 56

CNN Convolutional neural network. 16

CosSim Cosine similarity. 27–29, 31, 34–36, 40

DRNN Deep recurrent neural networks. 15, 27, 28, 30, 32–36

ECG Electrocardiagraph. 3–16, 18–25, 27–36, 38–41, 55–58

EM Electrode motion. 3, 9, 10, 19, 23, 28–33, 35, 36, 38, 41, 56

FCN Fully convolutional network. 16

FCN-DAE Fully convolutional denoising autoencoder. 16, 31–36, 39, 56

FIR Finite impulse response. 11–13, 34–36, 56

IIR Infinite impulse response. 12, 13, 34

LSTM Long short-term memory. 15

MA Muscle artifact. 3, 8, 10, 19, 23, 27–30, 32–36, 38, 41, 56

MAD Maximum absolute distance. 17

MKLNL Multi kernel linear and non-linear. 16

MSE Mean squared error. 15, 16

ReLU Rectified linear unit. 14, 16

SNR Signal to noise ratio. 19–23, 25, 29–33, 37, 55

SSD Sum of squared distance. 17, 31–36, 40, 56

1 Introduction

Electrocardiographs (ECGs) are of great significance in clinical diagnostics. To maximise the information an ECG recording can transmit, denoising methods can be applied. This is not only relevant for physicians but also for algorithms detecting anomalies in ECG recordings. [1] Denoising ECGs is especially important when the recordings are made in non-professional settings, recently made popular with the widespread usage and technical advancements of personal devices like smartwatches. [2]

Three of the most dominant noise types in ECG recordings are baseline wander (BW) noise, muscle artifact (MA) noise, and electrode motion (EM) noise. [3] Those noise types are artificially added to noise-free ECG recordings. Compared to using noisy ECG recordings directly, the chosen approach allows the comparison of the denoised ECG to the noise-free ECG.

Resulting decades of technical advancements, countless ECG denoising methods are available. Five of those who are expected to perform well are examined in several experiments. In those the amount of noise, the different combinations of noise, and the length of the ECG samples is varied.

With publications typically focusing either on one denoising method, one noise type or a fixed ECG sample length, the design of variable experiments yields insights into how the different factors influence the outcome.

2 Theoretical Background

2.1 ECG

Introduced in 1887, human electrocardiography allows the recording of an electrocardiogram (ECG) in order to analyse cardiac rhythm. The electrical activity of the heart is measured using body surface electrodes. Those recorded potentials are the spatial and temporal summation over the muscle cell activity in the heard. [4]

2.1.1 Electrical conduction system of the heart

To understand the mechanisms behind cardiac rhythm, the electrical conduction system of the heart shall be further discussed. In general, cells of the human body are separated from the extracellular space by a semipermeable membrane. The voltage over that membrane (i.e. one electrode on each side of the membrane and directly adjacent to the membrane) stems from different ion concentrations on each side. The ions are able to pass the membrane through channels and pumps. Channels only allow passage along the chemical and electrical gradient, while pumps are able to transport ions against the gradient. [5] Without external stimulus, the voltage over the membrane can be expected to be constant for most cells. If voltage is applied to the cell, the behavior of the channels and pumps in the membrane changes, leading to further change of the voltage. This characteristic change in voltage is able to travel along the membrane by effecting downstream channels and pumps. This behavior is called an action potential and is present in several types of cells, such as neurons or muscle cells. For cardiac cells two types of action potentials occur, namely slow and fast respond action potentials. Cells exhibiting the slow respond action potential do not posses a true resting potential, especially compared to cells exhibiting the fast respond action potential. With no resting potential, the cells are subjected to spontaneous depolarization as the channels and pumps are affected by the voltage changes. With these repetitive voltage changes and without external stimulus the cells act as pacemaker cells. Using the latter cells as stimulus the fast

respond action potential exhibiting cells are able to "pass along" the action potential. [6]
Three-dimensional models of the electrical activity of the heart allow a better understanding of the mechanisms defining a healthy heart compared to a unhealthy heart. [7], [8]

2.1.2 Electrode placement

With the action potential traveling through the heart in waves, the voltage measured by electrodes around the heart is not constant. This can be explained by the following example: In a two electrode system, one electrode might be close to an area of the heart where cells are already depolarized, while the action potential has not yet reached the cells near the other electrode. Hence an electric potential difference between the two electrodes is present. As per convention, when the action potential approaches the positive electrode, a positive signal will be recorded. When it travels away from the positive electrode, a negative signal will be recorded. [6] For this reason, an ECG is dependent on the vector through the heard along which the voltage is measured. The propagation of the action potentials through the heart results in a variable dipole moment. With the human body being a conducting medium, the electric potential of the dipole can be measured by placing electrodes on the skin. The electric potential difference between two electrodes is dependent on their orientation with respect to the direction of the dipole moment. [9] In order to make ECG recordings comparable, the placement of electrodes has

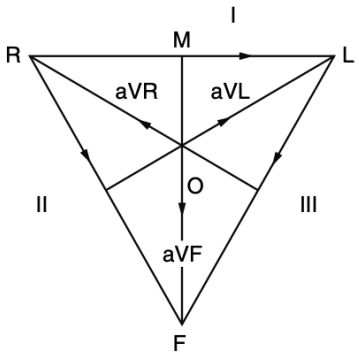

Figure 2.1: ECG leads in the vertical plane and their relation to one another. Source: [4]

been standardized. This standard records twelve electric potential differences along

different vectors through the heart. Six of these vectors lie in the vertical plane and six in the horizontal plane, using an upright body as a reference. The six potential differences in the vertical plane are measured using three limb electrodes, one on each arm and one on either leg. The six leads that can be recording using those electrodes are I, II, III, aVF, aVR, and aVL. Their direction and relation to one another is displayed in figure 2.1. Knowing two of these six leads, the other four can be derived using vector analysis (e.g. aVF = II - I/2). The six leads in the horizontal plane are recorded using six electrodes placed at specific locations around the chest. The recordings V_1 - V_6 use each of the six chest electrodes respectively and the sum of the three limb electrodes as a reference. [4]

2.1.3 Normal ECG

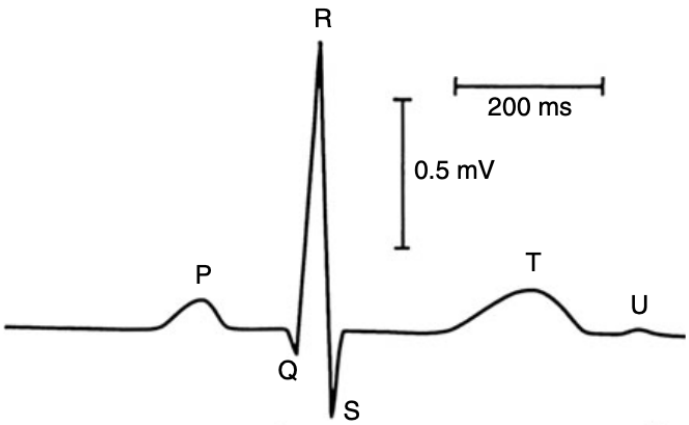

Figure 2.2: Lead II ECG, Source: [4]

In figure 2.2 the shape and different waves of a lead II ECG are shown. The amplitude of the different waves P, Q, R, S, T, and U (arbitrarily named) varies throughout different leads. The order of the waves however stays the same. With the P and U wave marking the beginning and end of a normal heart beat, the different waves represent activity in different parts of the heart during one cardiac cycle. The time in between the different waves is not constant. The QRS complex has a duration of about 120 ms. To measure the heart rate significant points in the ECG, like the R wave, are measured throughout several heart beats. Normal heart rates lie between 60 and 100 beats per minute. [10] The frequency spectrum of an ECG can be determined using a Fourier transform. The frequencies that make up the normal

ECG lie mostly below 30 Hz, with some in the range between 30 Hz and 100 Hz. [11]

2.1.4 Clinical relevance

With the definition of a "normal" the analysis of an ECG can be used as an indication for health-related complications. Defining limits after which an ECG is considered abnormal is not at all trivial. [4] Data to consider includes the following: heart rate, rhythm (correct order of waves), intervals in between waves (e.g. duration of QRS complex), and voltages (high, low, expected direction). [6] ECG abnormalities are often related to cardiac diseases. An example would be a ST-segment elevation which can be caused by myocardial infarcation (i.e. heart attack). Changes in the ECG do not imply cardiac diseases. Noncardiac conditions that result in changes in the ECG include the following: electrolyte disorders (e.g. calcium alters the duration of the QT interval), pulmonary embolisms (blockage of an artery in the lungs causes changes in morphology of different waves), cerebrovascular disorders (e.g. stroke leads to depressed ST segments), Hypothermia (low body temperature causes an additional wave between the QRS complex and the T wave), poisons (carbon monoxide causes alterations of the ST segment and the T wave), and electrical injuries (leads to changes in the ST segment and the T wave). [12]
Next to professional ECG recordings, modern wearable devices like smart watches allow amateur ECG recordings to be done directly by the patient. [13] Watches can record an ECG by having one electrode on the back of the watch against the arm. A second electrode located typically on the side of the watch is touched by the opposite arm. This allows the recording of a lead I ECG. [14]

2.2 ECG associated noise

In order to extract the aforementioned information from an ECG, the discussion of ECG-associated noise is indispensable. For this purpose, noise can be categorized incorporating both its origin and frequency spectrum. It is expected that a real ECG recording is subjected to a combination of noise types. The correct identification of noise types is therefore neither trivial nor absolute. [15]

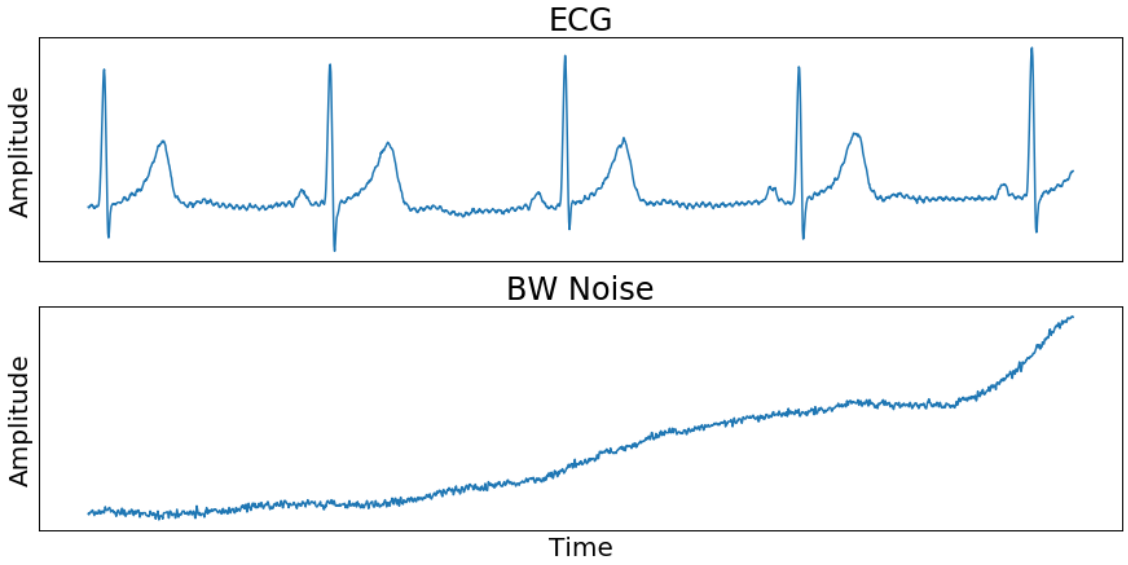

Figure 2.3: Unscaled BW noise in comparison to an ECG

2.2.1 Baseline Wander (BW)

Since an ECG is usually recorded with electrodes placed on the skin, changes in electrode-skin-contact lead to variations in impedance. This results in a variant baseline in the ECG. Electrode-skin-contact changes for several reasons, including slowly decaying adhesive durability, body movements, and respiration. The frequency spectrum is mostly located in the range between 0.05 Hz and 1 Hz. Figure 2.3 shows an unscaled recording of BW noise in comparison to an ECG recording.

2.2.2 Muscle artefacts (MA)

Similar to the muscles in the heart, electrical activity is present throughout all muscles in the human body. Muscles are not only required for physical activity but also for seemingly resting positions. Merely standing or sitting activates muscles in order to stabilize the position. Also small head movements, eye movements, swallowing, etc. require electrical activity and are therefor incorporated in the ECG. The frequency of that noise depends on the type of muscle activity but usually lies in the range between 20 Hz and 1000 Hz. At rest and without major contractions, the frequency tends to by much higher than the one of an ECG. Figure 2.4 shows an unscaled recording of MA noise in comparison to an ECG recording.

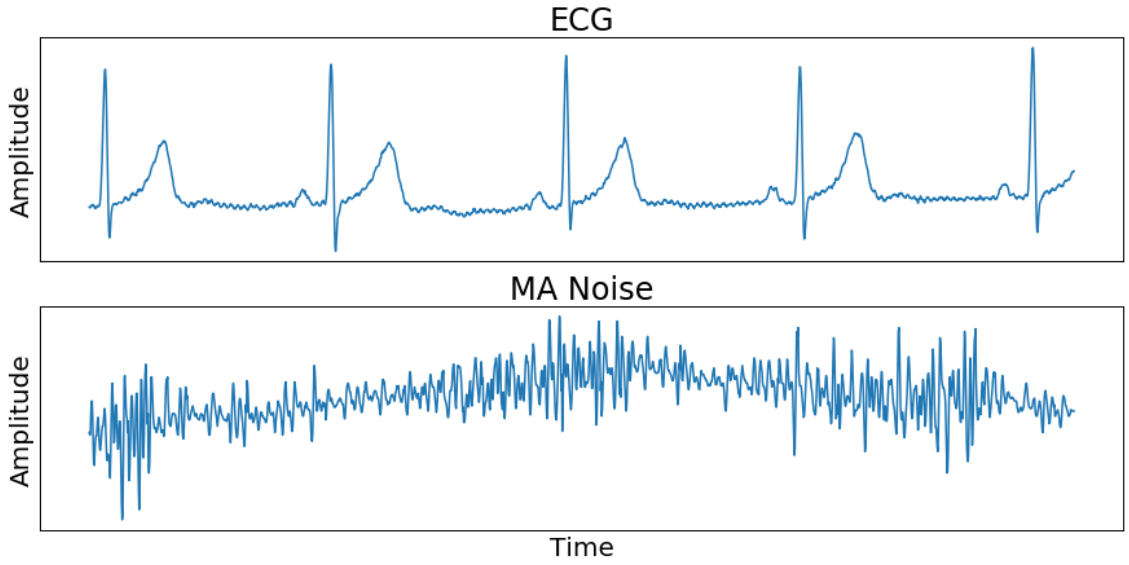

Figure 2.4: Unscaled MA noise in comparison to an ECG

2.2.3 Electrode motion artefacts (EM)

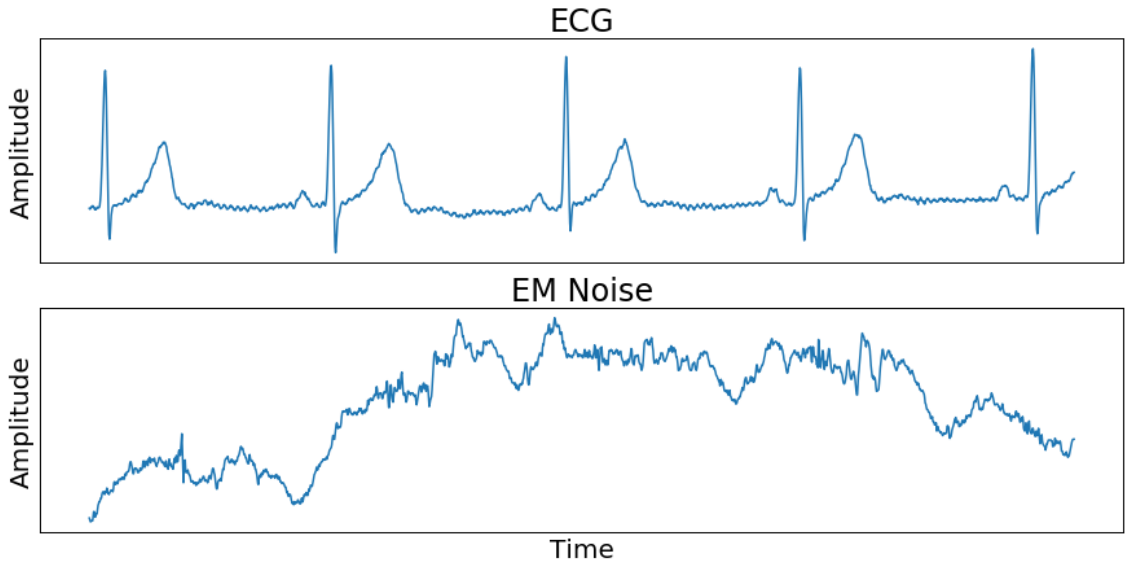

Figure 2.5: Unscaled EM noise in comparison to an ECG

Like BW noise, EM noise stems from changes in the electrode-skin-contact, but extends to much higher frequencies. Possible scenarios include physical activity during the recording of an ECG and recording systems that rely on the placement of hands or fingers onto an electrode. Especially problematic is the potential high

amplitude of EM noise and an overlapping frequency spectrum with the ECG. Figure 2.5 shows an unscaled recording of EM noise in comparison to an ECG recording.

2.2.4 Miscellaneous noise

Next to the above-mentioned there are a number of other noise types.
One of which is power line interference. That noise is only relevant if the recording device is connected to the power grid. The frequency of the noise matches the power line frequency, which is typically 50/60 Hz. Poor channel conditions are another source of noise. Channel noise is comparable to white Gaussian noise and is responsible for the "thickness" of the lines in an ECG recording.

Even more noise types have been introduced in numerous publications and as recording techniques evolve, so will the noise associated with the recordings. BW noise, MA noise, and EM noise however are arguably the most relevant noise types. [16], [17]

3 State of the art

3.1 ECG denoising

The purpose of ECG denoising as a pre-processing step is to reduce the amount of noise with which an ECG recording is corrupted. Doing so successfully is of interest in the field of medical science as more information can be drawn from noise-free ECG recordings. [16] As technical advancements continue to evolve, there are countless methods for denoising ECGs. Five especially promising different ECG denoising techniques are introduced in the following.

3.2 Digital filters

The general purpose of a traditional digital filter is to alter a signals frequency spectrum, i.e. allow certain frequencies to pass while restricting others.

3.2.1 Finite impulse response (FIR) filter

The first type of digital filters is the FIR filter. The output is defined as

$$y(n) = \sum_{k=0}^{N-1} h(k)\, x(n-k) \;=\; h(k) * x(n). \tag{3.1}$$

Using ECG denoising as an example $y(n)$ would be the denoised ECG, $x(n)$ the noise-corrupted ECG, h the filter impulse response coefficients and N the number of filter coefficients. The symbol $*$ is revering to the convolution of $h(k)$ and $x(n)$. The signals frequency spectrum can now be altered using that the discrete Fourier transform of the right side of equation 3.1 is equal to the product of the frequency spectrum of the signal and the discrete Fourier transform of the impulse response coefficients. The product of the frequency spectrum of the signal and the discrete Fourier transform of the impulse response coefficients is then the frequency spectrum of the denoised ECG. Trying to filter out low frequency noise, one can imagine the

discrete Fourier transform of the impulse response coefficients to be a function with low values around 0 and values around 1 everywhere else. That would make the FIR filter to behave as a high pass filter as only frequencies around 0 Hz would be restricted. In a similar fashion a FIR filter that behaves like a low pass filter that can be configured to filter out high frequency noise. [18]

The further used implementation of a FIR filter uses the work of van Alste and collaborators in [19] while incorporating the changes proposed by Romero and collaborators in [20]. The resulting filter incorporates the combination of a low and high pass filter with cutoff frequencies of 0.67 Hz and 150 Hz respectively.

3.2.2 Infinite impulse response (IIR) filter

Another form of digital filters is the IIR filter. While FIR filters depend only on the part of the input signal that lies before the sample point of interest, IIR filters depend on the part of the input signal that lies before the sample point of interest and on the output signal that lies before the sample point of interest. With that characteristic the IIR filter's output could extend endlessly with nonzero values even while the input has long become zeros. Compared to FIR filters, IIR filters require less multiplications and are therefore more efficient. The output of the IIR filter is defined as

$$y(n) = \sum_{k=0}^{N} b(k)\, x(n-k) \; + \; \sum_{k=1}^{M} a(k)\, y(n-k). \tag{3.2}$$

Using ECG denoising as an example, $y(n)$ would be the denoised ECG, $x(n)$ the noise corrupted ECG, b the feedforward filter coefficients, N the feedforward filter order, a the feedback filter coefficients, and M the feedback filter order. Instead of using the discrete Fourier transform, as for FIR filters, IIR filters require the Z-transform to be used. The Z-transform converts the expression 3.2, which is a discrete-time signal, into the complex frequency-domain 3.3.

$$Y(z) = X(z) \sum_{k=0}^{N} b(k)\, z^{-k} \; + \; Y(z) \sum_{k=1}^{M} a(k)\, z^{-k}. \tag{3.3}$$

Separating Y and X yields to

$$Y(z)\left(1 - \sum_{k=1}^{M} a(k)\, z^{-k}\right) = X(z) \sum_{k=0}^{N} b(k)\, z^{-k}. \tag{3.4}$$

The FIR filter's z-domain transfer function is defined as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{N} b(k)\, z^{-k}}{1 - \sum_{k=1}^{M} a(k)\, z^{-k}}. \tag{3.5}$$

Finally the FIR filter's frequency response is

$$H(\omega) = H(z)\big|_{z=e^{j\omega}} = \frac{\sum_{k=0}^{N} b(k)\, e^{-jk\omega}}{1 - \sum_{k=1}^{M} a(k)\, e^{-jk\omega}}. \tag{3.6}$$

Just as for the FIR filter, the IIR filter's frequency response can now be modeled to filter certain frequencies. For a low pass filter, $H(\omega)$ would need to be a function with values of 1 around 0 and values around 0 everywhere else. The opposite behaviour would result in a high pass filter. [18]

The further used implementation of an IIR filter uses the work of Kumar and collaborators in [21] while incorporating the changes proposed by Romero and collaborators in [20]. Just as the FIR filter, the resulting IIR filter incorporates the combination of a low and high pass filter with cutoff frequencies of 0.67 Hz and 150 Hz respectively.

3.3 Deep learning

Next to classical digital filters, deep learning offers a second, fundamentally different approach to denoise ECGs. At the heart of deep learning lies the usage of neural networks. Figure 3.1 shows an arbitrary neural network that will be further explained using the denoising of ECGs as an example. The red input neurons display the noise-corrupted ECG. Each neuron would be one sample of the signal. The blue neurons display the denoised ECG. Since the noise-corrupted ECG and the denoised ECG should have the same number of samples, the number of red and blue neurons is identical. In between the input and the output layer lies the hidden layers of defined dimension, which means a chosen number of layers with a chosen number of neurons each. The "deep" in deep neural networks is referring the number of hidden layers. Anything with more than one hidden layer can be considered a deep

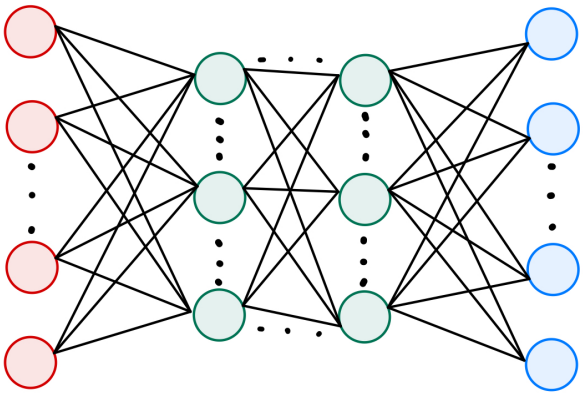

Figure 3.1: Arbitrary neural network where the red neurons represent the input layer, the green neurons the hidden layer, and the blue neurons the output layer

neural network. In a feedforward neural network, each neuron is connected with each neuron of the next layer. In the case of ECGs, for example, each neuron in the input layer holds the voltage value of the corresponding sample. The values of the neurons in the next layer are calculated as follows: Let i be the index of the neurons in the first layer and j be the index of the neurons the the second layer. Then each connection between the first and second layer has the weight $w_{i\,j}$ associated with it. With a_i and b_j being the the neurons of the first and second layer respectively, then the value of b_j is

$$b_j = f(x_j)\,,\ x_j = \sum w_{i\,j}\, a_i. \tag{3.7}$$

f is a function to be defined for each layer. An example for such a function is the rectified linear unit (ReLU), defined as $f(x) = max(0,\, x)$. The same procedure is applied for each layer. That is how the neurons of the output layer receive their values. Those values are then compared to their true value, i.e. the denoised ECG is compared to the corresponding noise-free ECG. The comparison is done using an error function that has to be defined. The error function $E(w)$ depends on all the weights in the neural network and has to be minimized. This is done using the gradient decent method. The local minimum of $E(w)$ is found by starting at a random point and then following the opposite direction of the gradient at that point. The gradient is calculated using backpropagation. Here, the gradient is calculated starting with the output layer incorporating the chain rule. The purpose of changing

the weights is to allow the network to train and find the set of weights that ensure the output, i.e. the denoised ECG to be as close as possible to the noise-free ECG. [22], [23]

3.3.1 Deep recurrent neural networks (DRNN)

Recurrent neural networks extend from feedforward neural networks altering the connection structure between the neurons. For feedforward neural networks, each neuron is connected to the neurons of the next layer. For recurrent neural networks, neurons are not only connected to neurons in the next layer but also to neurons in past layers and using loops and a delays also to themselves. [24]

The further used implementation of a DRNN uses the work of Antczak in [25] while incorporating the changes proposed by Romero and collaborators in [20]. The DRNN contains long short-term memory (LSTM) units. Those units are able to remember their states for different periods of time. One LSTM unit consists out of a memory cell, an input, an output, and a forget gate. The memory cell is able to store values while the input, output and forget gate manage the values by forwarding, updating or erasing them. Incorporating LSTM units into DRNNs solves their vanishing gradient problem. The problem states that backpropagation, using the chain rule, can lead to small partial derivatives towards the end of the backpropagation. This leads to weights in the first layers of the network to change very little compared to weights in the final layers of the network. [26] The used DRNN uses a mean squared error (MSE) metric for training the network.

3.3.2 Fully convolutional denoising autoencoder (FCN-DAE)

Another usage of neural networks are autoencoders. Considering a similar network structure as in figure 3.1, an autoencoder consists out of three neural layers. The first and last layer being the input and output of same size and the middle layer being the encoded representation of the input. The values of the neurons of the middle and last layer are dependent on the weights associated with the neural connections and the used activation function (see equation 3.7). Using ECG denoising as an example, a denoising autoencoder takes a noise-corrupted ECG as an input and encodes the data using weights and a defined function onto the middle layer. The decoder uses different weights and a defined function to map the encoded data

onto the third layer, the output. Here the denoised ECG is compared to the corresponding noise free ECG and the weights are changed to minimize the difference between the denoised ECG and the noise free ECG. [27]

Convolutional neural networks (CNNs) extend feedforward neural networks by introducing convolutional layers. Convolutional layers consist out of a set of filters and a defined dot product. Using the example of ECG denoising, the filters would be one-dimensional with a defined size. The filter moves sample by sample across the input, calculating the defined dot product each step of the way. Doing so, features of the input can be extracted. [28] CNNs contain layers where each neuron is connected to all neurons of the previous layer (just like in figure 3.1). For fully convolutional networks (FCNs) that is not the case. That reduces the number of parameters and increases performance. [29]

The further used implementation of a FCN-DAE uses the work of Chiang and collaborators in [27]. The encoder and decoder have 13 convolutional layers each, using between 1 and 40 filters of length 16. The MSE metric is used for training the model.

3.3.3 DeepFilter

Introduced by Romero and collaborators in [20], DeepFilter is a deep learning, FCN-based ECG denoising model. It stacks several convolutional layers on top of each other. Not only the weights are changed during backpropagation but also one of the stacked convolutional layers is chosen. For FCNs or CNNs, the filter or kernel size has to be defined. Rather than doing so in DeepFilter, each stacked layer has a kernel size of either 3, 5, 9, or 15. Next to different kernel sizes, two functions determining the value of each neuron are implemented, linear activation being one of them and ReLU the other. With four different kernel sizes and two different activation functions, eight convolutional layers are stacked. Six of these multi kernel linear and non-linear (MKLNL) filter modules are used following each other. The number of filters varies throughout the modules. Three of the MKLNL filter modules enlarge the receptive field size of the convolution while keeping the kernel size the same. Introducing a dilation rate, the spacing in between the fields of the kernel are increased. A dilation rate of one increases the the receptive field size from three for a kernel size of three to a receptive field size of five. A dilation rate of two increases the the receptive field size from three for a kernel size of three to a receptive field size

of seven. Using those non-consecutive kernels the performance of the method does not suffer while increasing the receptive field size. For the training of the model, a combination of the sum of squared distance (SSD) and the maximum absolute distance (MAD) is used.

4 Material and Methods

In order to investigate the above-mentioned state of the art methods, several experiments are performed. The general idea is:

1. Artificially add noise to noise-free ECG samples
2. Apply denoising methods
3. Compare denoised ECG samples to their corresponding noise free ECG samples
4. Quantify results after varying the denoising method, the length of the samples, and the applied noise type and noise strength

4.1 Producing noised ECG records

4.1.1 Noise free ECG recordings

As a source of noise-free ECG recordings, the QT Database [30] as published on PhysioNet [31] is used. Each of the 105 excerpts of recordings has a duration of 15 minutes, with a sample rate of 250 Hz, and uses two different leads. The leads vary from recording to recording. The recordings where chosen from several other databases, guaranteeing its variability. Because of this, the recordings show different anomalies and arrhythmia. At least 30 beats in each record are annotated. This is important, as the denoised ECG should not be a textbook ECG but the "true" ECG with potential anomalies still detectable. If a denoising algorithm were to remove all anomalies, the recording would lose its medical relevance, since a physician would falsely identify the recording to be normal.

Following the experiment design used by Romero and collaborators in [20], the ECG recordings are resampled from 250 Hz to 360 Hz and heartbeat-wise split. Fixing the number of samples per beat to 512 or about 1.42 seconds, results in a total

number of about 85000 separate beats. An alternative to the heartbeat-wise split is used as well. Here the 15 minutes recordings are divided so that each split has a length of 8190 samples or 22.75 seconds. This results in a total number of about 8200 separate ECG samples.

4.1.2 ECG Noise

For the noise recordings, the MIT-BIH Noise Stress Test Database [3] as published on PhysioNet [31] is used. The Database focuses on the three noise types baseline wander (BW), muscle artefacts (MA), and electrode motion artefacts (EM) as described above. The noises are found in three different, two-channel recordings, with a duration of 30 minutes and a sample rate of 360 Hz. To detect noise without interference from the heart, the electrodes where placed in such a way that the ECG is not visible.

The noise records are split to match the length of the ECG samples, i.e. a length of either 512 or 8192 samples.

4.1.3 Adding noise to ECGs

For each ECG sample, a random amount of noise out of a fixed range is added to a noise free ECG using the following scheme [3]:

$$ECG_{noised}(t) = ECG_{clean}(t) + \sum_i a_i \cdot Noise_i(t) \tag{4.1}$$

$ECG_{noised}(t)$, $ECG_{clean}(t)$, and $Noise_i(t)$ are time series arrays of identical length. The sum over i represents the option of adding different noise types (BW, MA, and/or EM). The calculation is done for each time step t. Factor $a \in IR^+$ determines the amount of noise added and is commonly defined by the signal to noise ratio (SNR) in decibel. Figure 4.1 shows the ratio of signal power to noise power as a function of the same ratio but expressed in decibel.

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \tag{4.2}$$

Where P_{signal} and P_{noise} are the power of the noise free ECG and noise respectively. The power of the noise is calculated before a potential scaling. If the noise array is

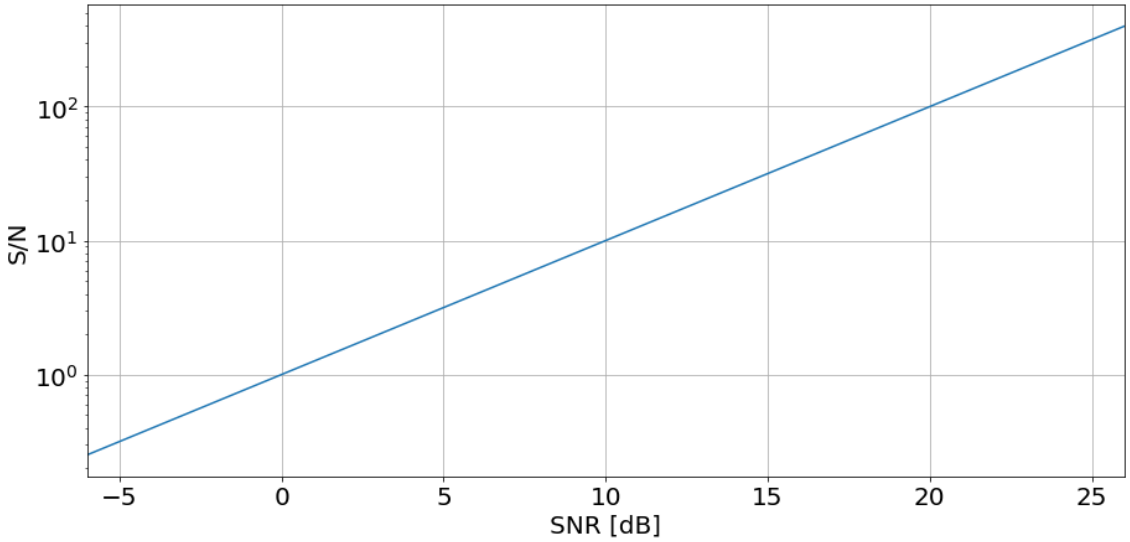

Figure 4.1: Visualisation of the signal to noise ratio in decibel. The higher the SNR value, the less noise with respect to the signal is used.

now scaled by the variable a, that factor is included in the SNR as such:

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{a^2 \, P_{noise}} \right) \tag{4.3}$$

For a desired SNR and a prior calculated P_{signal} and P_{noise}, a calculates to

$$a = \sqrt{\frac{10^{\frac{-SNR_{dB}}{10}} \, P_{signal}}{P_{noise}}} \tag{4.4}$$

The variable a therefore allows the noise-free ECG signal to be subjected to a quantifiable amount of noise.

The calculation of a signal or noise power applied to an ECG is not trivial and done separately for the two different sample lengths. For the samples of about 1.42 seconds, an approach similar to the one suggested in the WFDB Software Package is used [32]. The power of the noise-free signal is determined by first calculating the range of the signal. Then the result is squared and divided by eight. For the power of the noise, the mean amplitude is calculated first. Then the rms difference between the amplitude and the mean amplitude is determined while the largest and smallest 5% are discarded. After calculating the mean, the results are finally squared.
For the samples of 22.75 seconds, changes are made to the above-explained approach. For the power of the noise-free signal, the QRS complexes are found first. Then a 50

millisecond window around the R wave is calculated of which the range is calculated. The largest and smallest 5% are discarded. The mean is calculated before the result is squared and divided by eight. For the power of the noise, the samples are divided into one-second chunks. For each chunk, the mean amplitude and the rms difference between the amplitude and the mean amplitude is calculated. Again, the largest and smallest 5% are discarded. After calculating the mean, the result is squared. The approaches assure that the ECG power is not dependent on the heart rate and the noise power is not overestimating low frequency noise. [32]

4.1.4 SNR ranges

Noise is added to noise-free ECG using a SNR range of including 0 to excluding 20 dB. Since the amount of noise in reference to the signal is dependent on the calculation of the noise and signal power, the SNR value throughout publications is difficult to compare. For that purpose a visual representation of the SNR range for the different noise types is reasonable. Figures 4.2 to 4.4 show the noise range for

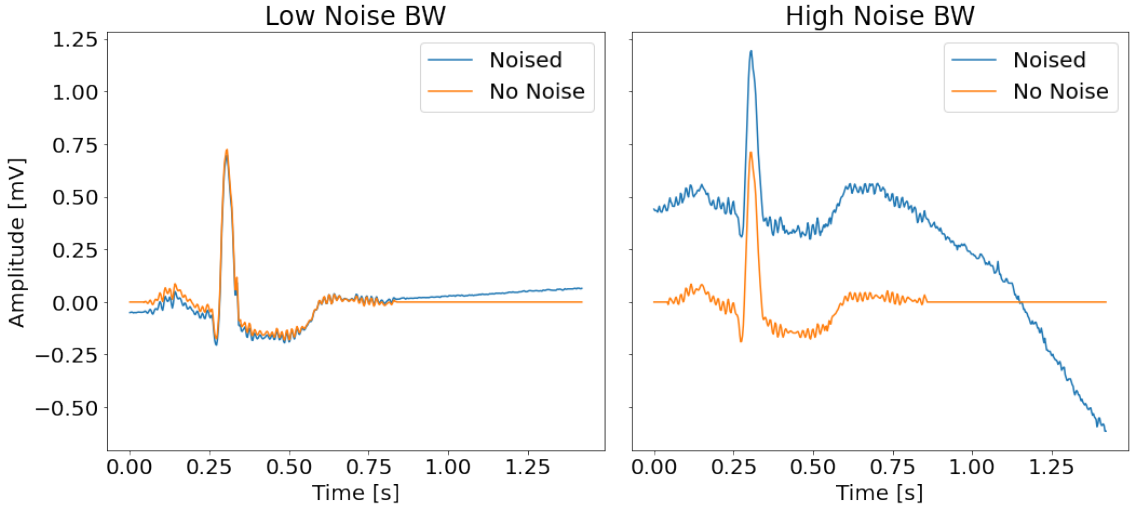

Figure 4.2: Visualisation of the noise range for BW

the three noise types using randomly chosen samples.

4.1.5 Mixed Noise

The quantification of mixed noise is more complex than for the separate noise types. The idea is that for each ECG sample three SNR values are chosen at random. One

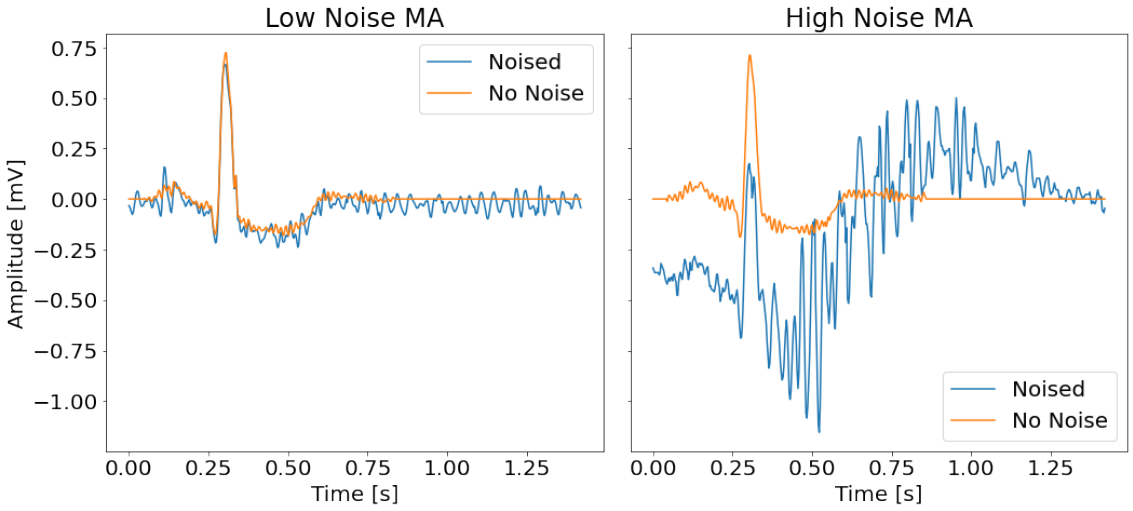

Figure 4.3: Visualisation of the noise range for MA

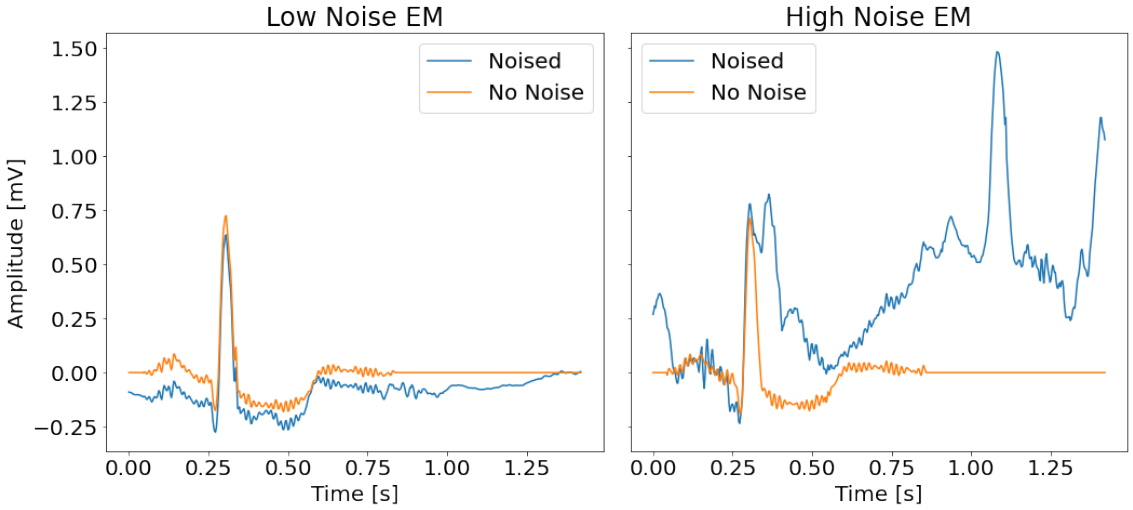

Figure 4.4: Visualisation of the noise range for EM

for each noise type. Following this, the three scaling factors are calculated using equation 4.4, the determined signal power, and the three noise powers. The scaled noises are then added to the ECG using equation 4.1. The total signal to noise ratio then calculates to

$$SNR_{dB,\,tot} = 10\,\log_{10}\left(\frac{P_{signal}}{a_1^2\,P_{noise,\,1} + a_2^2\,P_{noise,\,2} + a_3^2\,P_{noise,\,3}}\right) \tag{4.5}$$

If the SNRs for the three noise types are drawn from the same range as for the separate noise types, the total SNR would no longer be in that range, as the lowest

and highest amount of noise possible is now higher. To make the results comparable to the ones from the separate noise types, the initial SNRs that are combined to the total SNR are drawn from a different range. The range is calculated as follows:

$$SNR_{dB,i} \in [0dB + 10 \cdot log(3),\ 20dB + 10 \cdot log(3)) \tag{4.6}$$

4.2 ECG denoising experiments

Using the digital filters, finite impulse response (FIR) and infinite impulse response (IIR), as described in section 3.2, as well as the deep learning methods deep recurrent neural networks (DRNN), fully concolutional denoising autoencoder (FCN-DAE) and DeepFilter, as described in section 3.3, an experiment setup building on the work of Romero and collaborators in [20] and their GitHub repository [33] is used. The repository contains code allowing the comparison of their suggested denoising method DeepFilter against different state-of-the-art methods for denoising single beat ECG samples corrupted with BW noise.

4.2.1 Experiment setup

The five denoising methods are investigated in seven experiments:

- ECG samples containing one beat corrupted with BW noise
- ECG samples containing one beat corrupted with MA noise
- ECG samples containing one beat corrupted with EM noise
- ECG samples containing one beat corrupted with mixed (BW + MA + EM) noise
- ECG samples containing several beats corrupted with BW noise
- ECG samples containing several beats corrupted with MA noise
- ECG samples containing several beats corrupted with EM noise

The five denoising methods of one experiment receive identically noised ECG samples. The experiment setup is visualized in figure 4.5. The experiments are chosen to allow the best possible insight. First the denoising methods can be compared for different noise types. Varying the ECG sample length allows findings regarding the

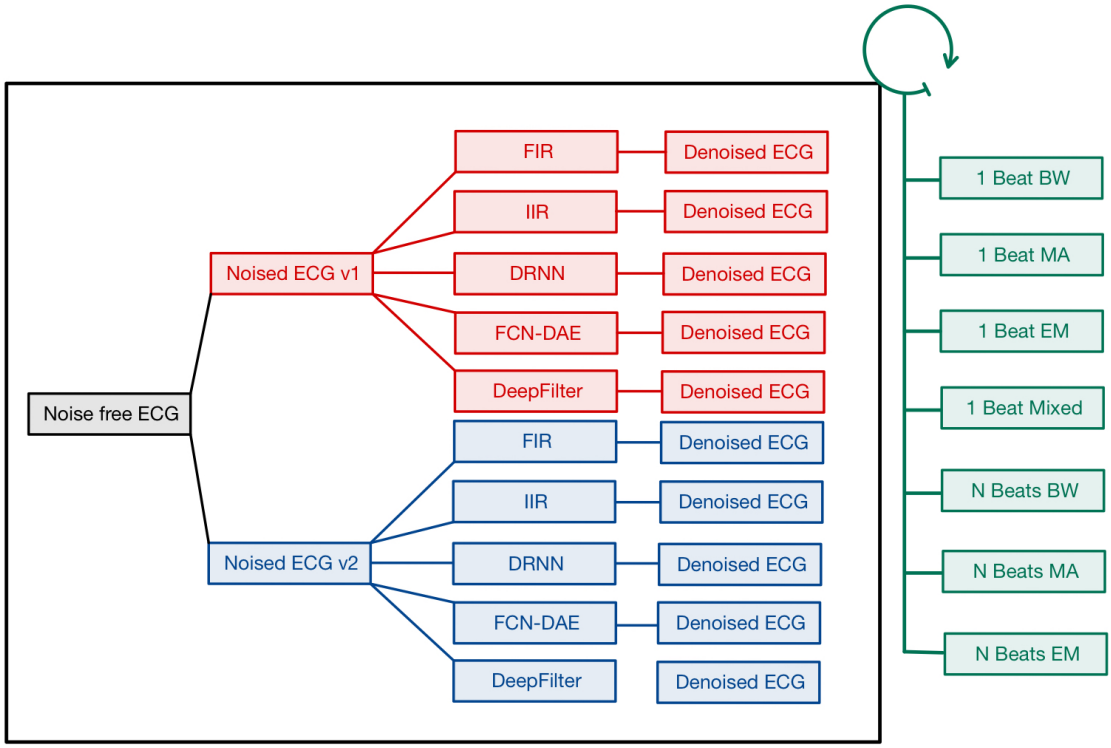

Figure 4.5: There are seven superior experiments (green boxes) where the length of the ECG samples and the used noise is varied. For each experiment the noise-free ECG samples are noised using two noise versions (red and blue path), which are then both denoised using five denoising methods.

dependency of denoising methods on the ECG sample length. Incorporating a mixed noise experiment allows conclusions to be made about the correlation between the performance of the denoising methods for separate noise types and for different combinations of noise types. Varying the amount of used noise allows the investigation of the functional dependencies between the performance of the denoising methods and the amount of noise.

4.2.2 Test, train and validation split

The performance of the different methods is measured using a test data set. From the QT database containing 105 recordings from 7 sub-databases, 14 recordings are chosen, two from each sub-database to assure variable pathologies. The remaining samples are divided into the validation set (30%) and the train set (70%). Since all recordings contain two channels, the recordings are concatenated. This includes both channels and makes a further distinction obsolete. [20]

4.2.3 Noise split

The noise recordings, being 30 minutes long and containing two channels, are split in the following manner: Two noise version are created. One with the first half of the first channel for the train set and the second half of the second channel test set and the other with the second half of the first channel and the first half of the second channel. The experiments are then done for each noise version, before the results are combined. [20]

4.3 Quantifying performance

The ECG samples are not randomly noised with continuous SNR values. For all experiments, excluding mixed noise, the SNR range is using discrete steps of 0.25 dB. This means that an ECG sample is noised using one of 80 SNRs. For mixed noise, each ECG sample is noised using only one of four SNRs, but for every noise type. This results in 64 combinations of mixed noise. Since the test data set contains about 27.000 single beat ECG samples or about 1100 several beats ECG samples, each SNR value is used several times. This allows the calculation of means and errors for the performance of a denoising method for a certain SNR value.

The output of each denoising method is the following: an ECG sample from the test set, be it of about 1.42 seconds or 22.75 seconds, will exist in five adaptations. The first one is the plain ECG recording. The next two are the noise corrupted adaptation of the plain ECG, using the two noise versions. The last two are the denoised adaptation of the noise corrupted ECG, one for each noise version. All the denoised ECGs of one method (combining both noise versions) will be compared to their noise free counterpart. Doing so one can quantify how well the denoising worked using different metrics.

4.3.1 Metrics

Two metrics are used to quantify the performance of the denoising methods. The metrics compare the denoised ECG sample with the corresponding noise free ECG sample (original ECG sample before noise was added). The first one is the sum of

the square of the distances (SSD). [20], [34]

$$SSD(s_1, s_2) = \sum_{n=1}^{N} (s_2(n) - s_1(n))^2 \ , \ 1 \leq n \leq N \tag{4.7}$$

The second metric is cosine similarity (CosSim). [20], [35]

$$CosSim(s_1, s_2) = \frac{\langle \overrightarrow{s_1}, \overrightarrow{s_2} \rangle}{\|\overrightarrow{s_1}\| \|\overrightarrow{s_2}\|} = \frac{\sum_{n=1}^{N} s_1(n) s_2(n)}{\sqrt{\sum_{n=1}^{N} s_1^2(n)} \sqrt{\sum_{n=1}^{N} s_2^2(n)}} \ , \ 1 \leq n \leq N \tag{4.8}$$

5 Results

5.1 One beat ECGs corrupted with separate noise

	BW		MA		EM	
Method	**SSD**	**CosSim**	**SSD**	**CosSim**	**SSD**	**CosSim**
FIR	21.1 ±0.3	0.8230 ±0.0011	47.3 ±0.6	0.7039 ±0.0013	53.8 ±0.6	0.6797 ±0.0014
IIR	16.45 ±0.27	0.8493 ±0.0010	51.8 ±0.7	0.6974 ±0.0013	48.9 ±0.6	0.6941 ±0.0014
DRNN	8.92 ±0.08	0.8222 ±0.0009	6.60 ±0.06	0.8733 ±0.0008	**9.78 ±0.18**	**0.8149 ±0.0010**
FCN-DAE	4.96 ±0.05	0.9073 ±0.0006	5.8 ±0.06	0.8894 ±0.0007	14.04 ±0.18	0.8005 ±0.0013
DeepFilter	**3.51 ±0.04**	**0.9287 ±0.0005**	**4.72 ±0.04**	**0.9067 ±0.0006**	12.69 ±0.14	0.8100 ±0.0011

Table 5.1: Results of the experiments using ECG samples with one heartbeat and separate noise types. The mean value and its standard error are given for the two metrics. The best mean values for each column are bold.

In table 5.1 the results of the one beat ECGs corrupted with separate noise types are displayed. The values show the mean of the calculated metrics. Their error is given by the standard error of the mean. The bold values mark the best values for each column. For all three noise types the best performing denoising method determined by the SSD metric is also the best performing one regarding the CosSim metric.
DeepFilter performs best for baseline wander (BW) noise. For the SSD metric all deep learning denoising methods outperform the classical filters. For the CosSim metric that is not the case, as the two classical filter methods perform better than the DRNN.
For muscle artifact (MA) noise, the trends are comparable to BW. DeepFilter performs best considering both metrics. The deep learning methods outperform the classical filter methods not only considering the SSD metric, but also the CosSim metric. Compared to the mean values from BW, the values for MA are worse,

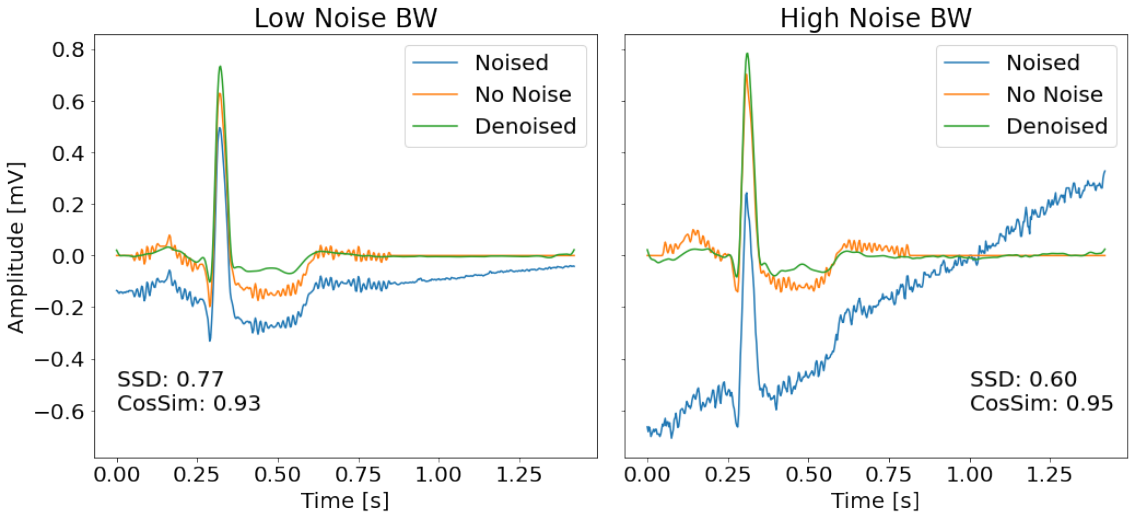

Figure 5.1: Exemplary visualisation of the performance of the best performing denoising method DeepFilter for BW noise for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

meaning that the SSD values are higher and the ones for the CosSim lower. It is noticeable that the performance of the deep learning methods between the two noise types is smaller than for the classical filter methods.

For electrode motion (EM) noise, DRNN is the best performing method. The deep learning methods perform better than the classical filter methods considering both metrics. While the metric values for the classical filter methods are similar to the ones from MA noise, the values for the deep learning methods have gotten worse.

To conclude table 5.1: DeepFilter is the best performing denoising method for both BW and MA noise. DRNN is the best one for EM noise. Denoising works best for BW noise, followed by MA and EM noise. The SSD and CosSim metric deliver coherent information.

To provide an exemplary visualisation of the performance of the best performing denoising methods for the respective noise types, figures 5.1 for BW noise and A.1 and A.1 for MA and EM noise in the appendix A show a noise free ECG sample, its noised version, and the denoised ECG sample. This is done for a low noise (20 dB) and a high noise (0 dB) example. Using those figures provides better context to the metric values presented in table 5.1.

To see how the denoising methods perform for varying amounts of noise, figures

5.2 - 5.4 have to be consulted. For every possible signal to noise ratio (SNR) value (discrete values using steps of 0.25 dB), the SSD metric mean values and its standard error are shown. Looking at the individual points, rather than for example a line connecting the mean values, allows a better discussion of trends as the variance and error of the mean values can be better incorporated. In the appendix A, next to the SSD metric, the CosSim metrics behaviour is compared to the one of the SSD metric while varying the SNR (figures A.3 - A.5). It can be seen that the separate discussion of the SSD metric and the CosSim metric is redundant for the case at hand.

For BW noise 5.2 it shows that DeepFilter is the best performing denoising method across the spectrum of tested SNR values. With lower SNR values the performance of the classical filter methods becomes increasingly worse compared to the deep learning methods. As indicated by the "straight line" behaviour in the Semi-log plot, an exponential function describing the dependency of the mean SSD value on the SNR value can be assumed. Looking at the slopes reveals that the deep learning methods seem to differ from the classical filter methods in their exponential constant. Next the MA noise in figure 5.3 is discussed. Many similarities to the

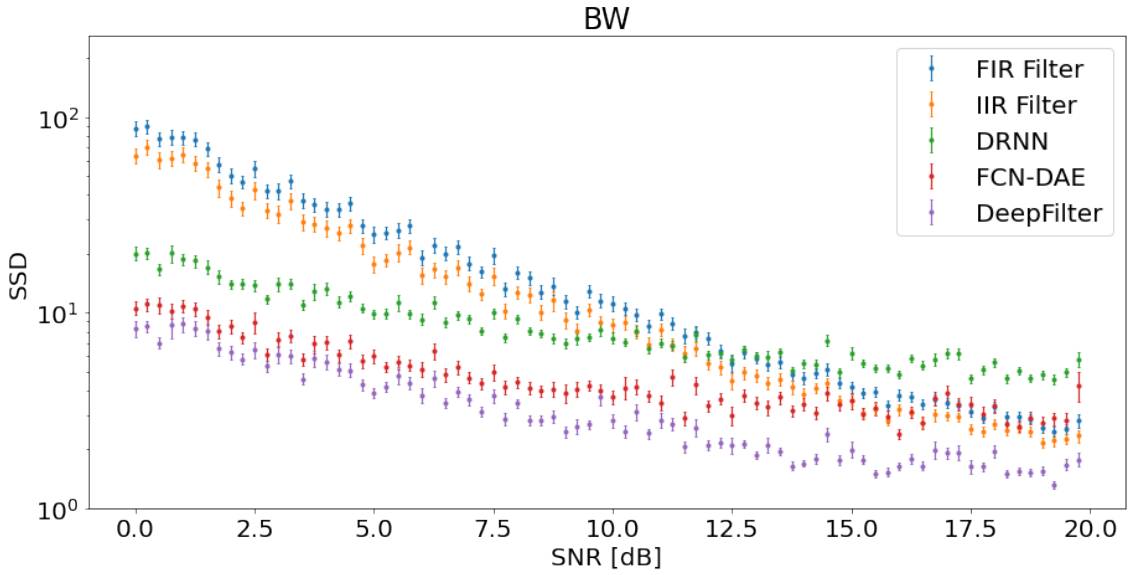

Figure 5.2: Performance of the different denoising methods for one beat ECGs corrupted with BW noise measured using the SSD metric

BW noise can be found. The exponential dependency is applicable and DeepFilter is the best performing denoising method. Different exponential constants for the deep learning methods and the classical filter methods are necessary. For EM noise

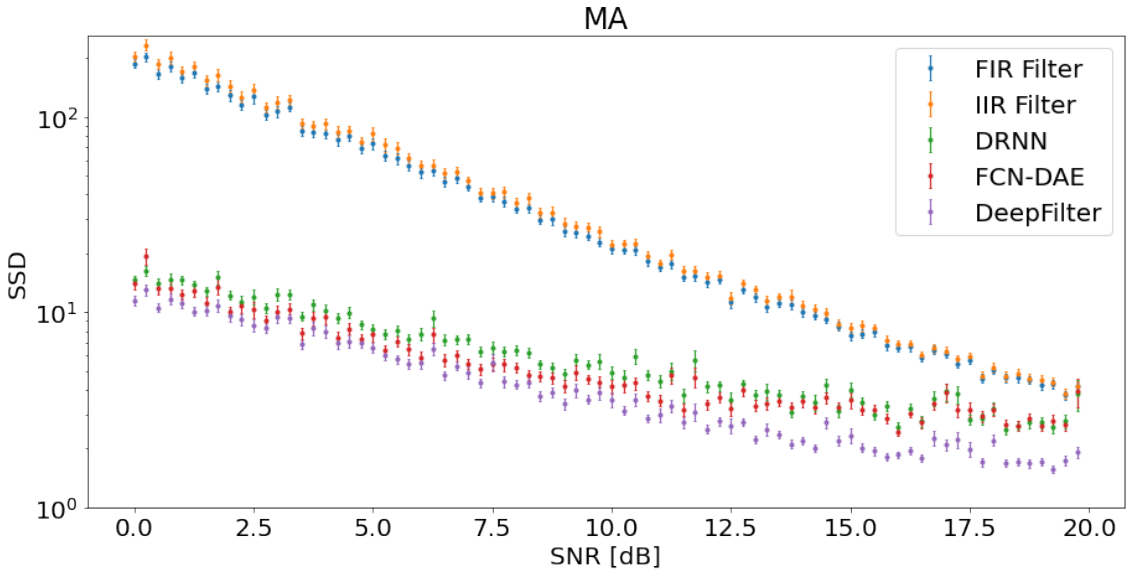

Figure 5.3: Performance of the different denoising methods for one beat ECGs corrupted with MA noise measured using the SSD metric

in figure 5.4 that is applicable as well. Interesting is here that the best denoising method is dependent on the SNR. Between about 0 dB and 10 dB DRNN is the best performing method. Between 10 dB and 20 dB DeepFilter is the best performing method.

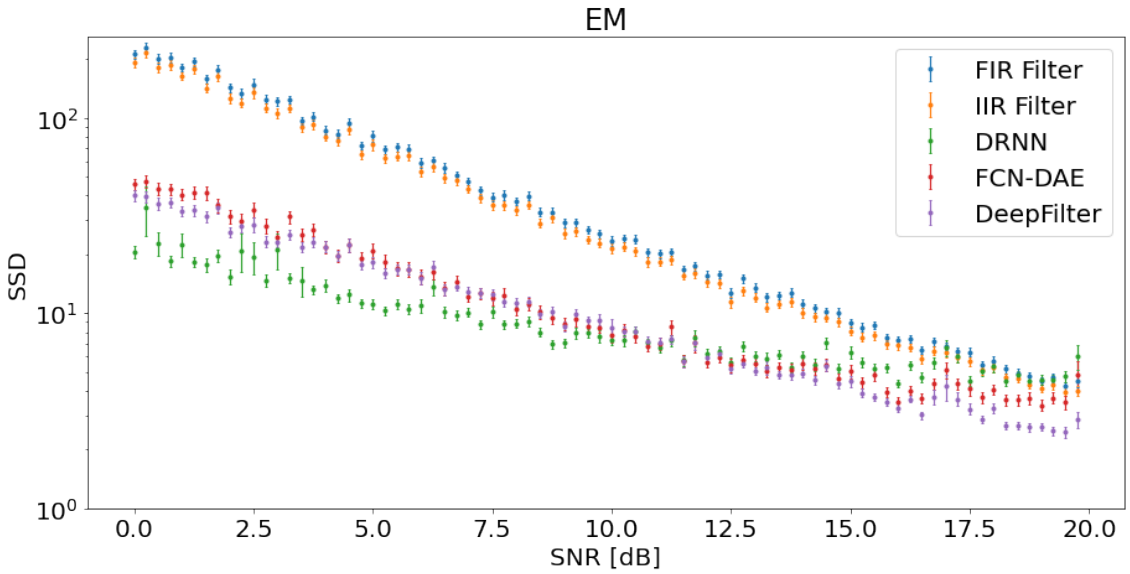

Figure 5.4: Performance of the different denoising methods for one beat ECGs corrupted with EM noise measured using the SSD metric

5.2 One beat ECGs corrupted with mixed noise

	Mixed	
Method	**SSD**	**CosSim**
FIR	58.1±0.5	0.6329±0.0013
IIR	55.5±0.5	0.6427±0.0013
DRNN	9.84±0.08	0.8048±0.0010
FCN-DAE	**7.97±0.07**	**0.8483±0.009**
DeepFilter	9.4±0.09	0.8347±0.009

Table 5.2: Results of the experiment using ECG samples with one heartbeat and mixed noise types. The mean value and its standard error are given for the two metrics. The best mean values for each column are bold.

For one beat ECGs corrupted with mixed noise a similar analysis can be done. Here not only the SNR is varied but also the fractions of the different noise types.
Table 5.2 shows that FCN-DAE performs best considering both metrics. The SSD and CosSim metric again deliver coherent information. The performance of the deep learning methods for mixed noise lies between the performance of BW noise and the performance for EM noise (see table 5.1). For the classical filters however the performance for mixed noise is worse than every separate noise.

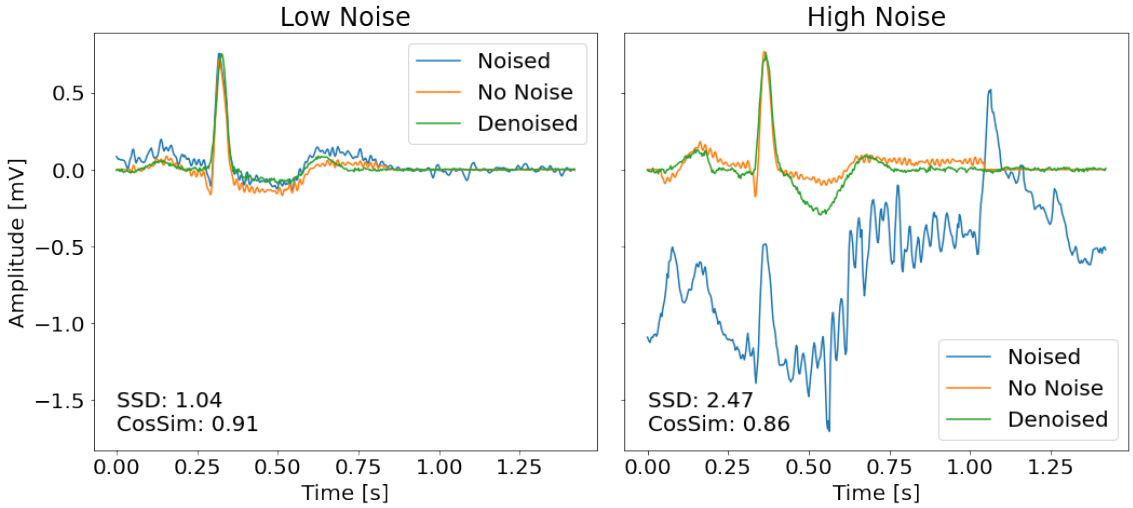

Figure 5.5: Exemplary visualisation of the performance of the best performing denoising method FCN-DAE for mixed noise for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

To provide an exemplary visualisation of the performance of the best performing

denoising method for mixed noise, figure 5.5 shows a noise-free ECG sample, its noised version, and the denoised ECG sample. This is done for a low noise (20 dB) and a high noise (0 dB) example.

To see the behaviour of the different methods for a varying SNR of mixed noise, a total SNR, which includes the noise strengths for all three noise types, is calculated using equation 4.5. The four discrete SNR values for each noise type result in the combined SNR spectrum given in figure 5.6. Here the SSD metric is used as well. It can be seen that FCN-DAE performs best only for SNRs lower than 10 dB. For higher SNRs DeepFilter performs best. The classical filter methods performance depends on the SNR in a similar fashion as already seen for BW, MA, and EM noise. However a different behaviour can be seen for the deep learning methods. For SNR values lower than 10 dB the exponential constant seems to be different to the one for SNR values higher than 10 dB. Noticeable is that by the fact that the SSD metric increases slower with decreasing SNR values in between 20 dB and 10 dB compared to SNR values in between 10 dB and 0 dB. This accounts especially for FCN-DAE and DRNN. Their performance can be considered constant between 20 dB and 10 dB.

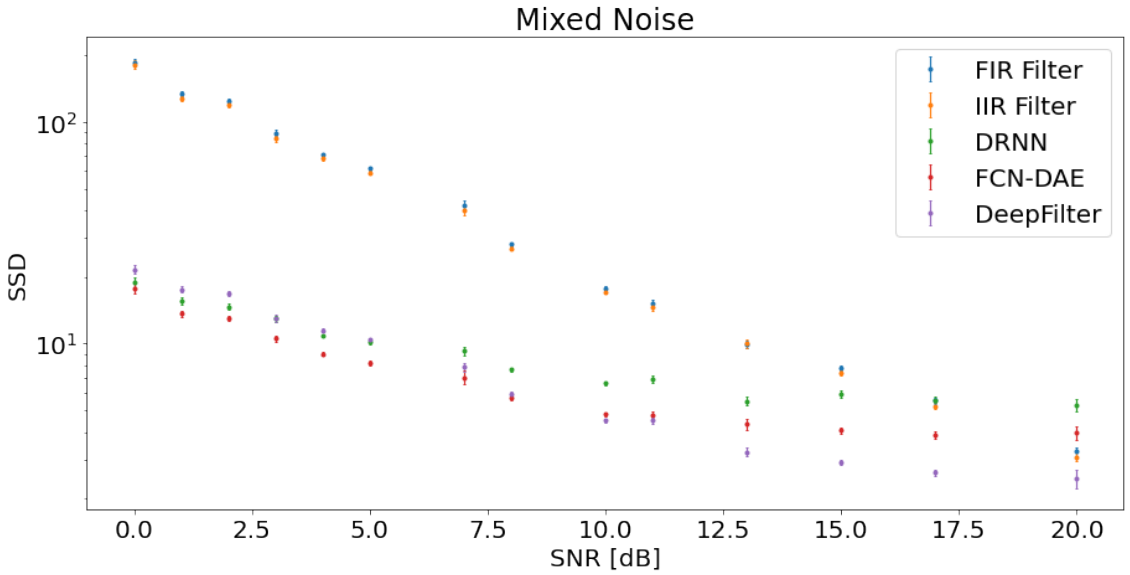

Figure 5.6: Performance of the different denoising methods for one beat ECGs corrupted with mixed noise measured using the SSD metric

Since only the total SNR is considered so far, no statements can be made about how respective noise strengths of the different noise types effects the performance

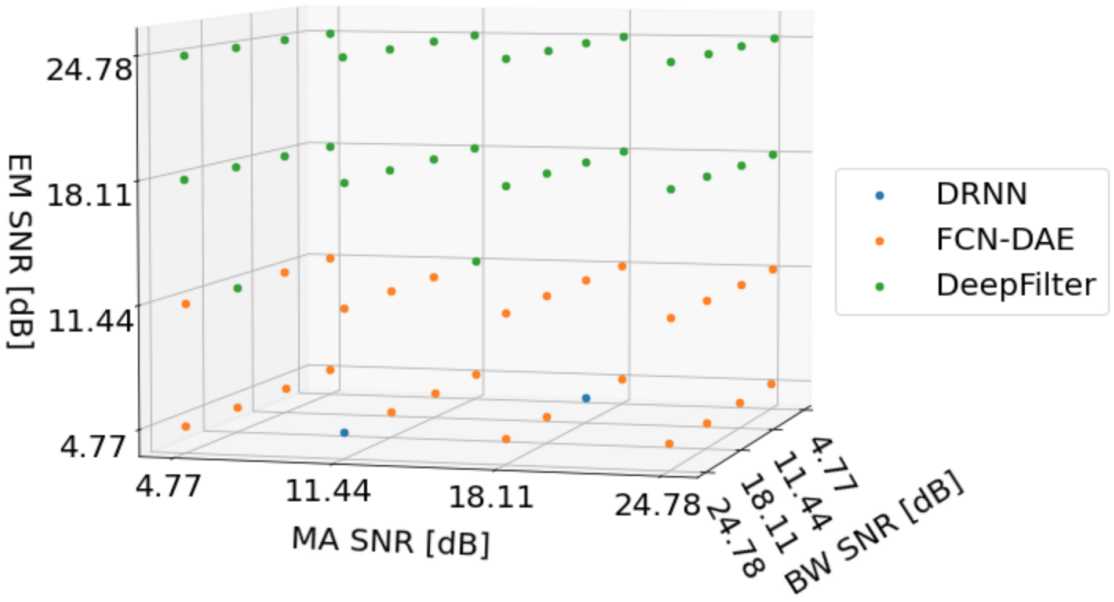

Figure 5.7: SNR values of the single noise types that make up mixed noise. The best performing denoising method for each combination is labeled.

of the denoising methods. For this, the three dimensional figure 5.7 can be consulted. Here the three axes show the SNR of the respective noise types. Since each noise type uses discrete SNR values (four different SNR values between 4.77 dB and 24.78 to allow a combined SNR between 0 dB and 20 dB), there are 64 different combinations of mixed noise. Since there are about 27.000 ECG samples, each of the 64 mixed noise combinations is used several times, allowing the calculation of mean performance metric values. For each of the 64 mixed noise combinations, the best performing denoising method is determined by finding the method with the lowest mean SSD metric value. In figure 5.7 the result is shown by color coding the best performing denoising method for each of the 64 combinations of mixed noise. It turns out that only the deep learning methods are included, as the classical filter methods are never the best performing method. Most represented are FCN-DAE and DeepFilter with only two mixes noise combinations where DRNN is the best performing method. It also becomes clear, that the best performing denoising method for mixed noise is independent on the SNR values for MA and BW noise. For EM noise however, FCN-DAE tends to be the best performing method for low SNR values (4.77 dB and 11.44 dB) and DeepFilter is the best performing method for high SNR values (18.11 dB and 24.78 dB). This is shown in the figure by noticing

that the bottom half of the "cube" is labeled predominantly "FCN-DAE" and the top half "DeepFilter".

5.3 Several beats ECGs corrupted with separate noise

	BW		MA		EM	
Method	**SSD**	**CosSim**	**SSD**	**CosSim**	**SSD**	**CosSim**
FIR	**4400 ±400**	**0.716 ±0.009**	5200 ±400	**0.599 ±0.009**	5000 ±400	0.615 ±0.009
IIR	**4400 ±400**	0.715 ±0.009	5200 ±400	0.598 ±0.009	5000 ±400	0.615 ±0.009
DRNN	10600 ±600	0.479 ±0.008	4000 ±400	0.592 ±0.008	2050 ±140	**0.665 ±0.007**
FCN-DAE	4620 ±190	0.319 ±0.009	**3370 ±180**	0.33 ±0.010	**1880 ±80**	0.385 ±0.009
DeepFilter	11100 ±700	0.335 ±0.009	3400 ±280	0.555 ±0.009	2520 ±150	0.470 ±0.009

Table 5.3: Results of the experiments using ECG samples with several heartbeats and separate noise types. The mean value and its standard error are given for the two metrics. The best mean values for each column are bold.

The results are more complicated for several heart beats. One reason for that is that there are fewer ECG samples available (due to the longer duration of the samples). This results in larger uncertainties when calculating mean values. Another complication is that the SSD metric depends on the length of sample. This makes comparing the SSD metric values of the single beat ECGs to the several beats ECGs difficult. In table 5.3 the SSD and the CosSim metric no longer provide coherent information when trying to find the best performing denoising method.
For BW noise FIR and IIR perform best with very similar values for both metrics. This is interesting as the deep learning methods are now outperformed by the classical filter methods. For BW noise using single ECG samples that was the other way around. Looking at only the deep learning methods, FCN-DAE delivers best results considering the SSD metric and DRNN considering the CosSim metric.
For MA noise FCN-DAE performs best considering the SSD metric (while performing the worst considering the CosSim metric) and FIR performs best considering the

CosSim metric (while performing the worst considering the SSD metric). The deep learning methods outperform the classical filter methods when considering the SSD metric. For the CosSim metric however the classical filter methods outperfrom the deep learning methods. Compared to BW noise the deep learning methods perform better considering both metrics. The opposite was the case when using single ECG samples. For the classical filter methods the SSD metric values improve moving from BW to MA noise, while the CosSim metric values are worse.
For EM noise, FCN-DAE performs best, considering the SSD metric and DRNN perform best considering the cosine similarity. The deep learning methods outperform the classical filter methods considering the SSD metric. The CosSim metric delivers mixed results so that neither the deep learning methods nor the classical filter methods perform definitely better.
To provide an exemplary visualisation of the performance of the best performing denoising methods for the respective noise types and to further investigate the incoherent information from the two performance metrics, figures 5.8 for BW noise and A.6 and A.9 for MA and EM noise in the appendix A show a noise-free ECG sample, its noised version, and the denoised ECG sample. In order to enhance readability not the entire sample is shown but a four second excerpt. This is done for a low noise (20 dB) and a high noise (0 dB) example. Looking at an example of the best performing method considering both the SSD and CosSim metric allows the incoherent information to be put in context.
A qualitative comparison of the denoising performance of the best performing denoising method BW noise considering several beats ECG samples in figure 5.8 and single beat ECG samples in figure 5.1 shows that BW noise can be removed from ECG samples with similar success.
For MA noise and EM noise this is not the case. Neither the FIR method (best performing considering the CosSim metric for MA) nor the FCN-DAE method (best performing considering the SSD metric for MA) make any attempt to remove MA noise in several beat ECG samples. While the classical filter method FIR slightly changes the noised ECG, the deep learning method FCN-DAE deforms the sample immensely. The resulting denoised output neither resembles the noise-free ECG sample nor the noised ECG sample. Comparing the best performing denoising method for MA, considering the SSD and CosSim metric, results in choosing the CosSim metric to be objectively better for determining the best (or least worst) performing method. This is based on the possibility to identify the different waves in figures A.6

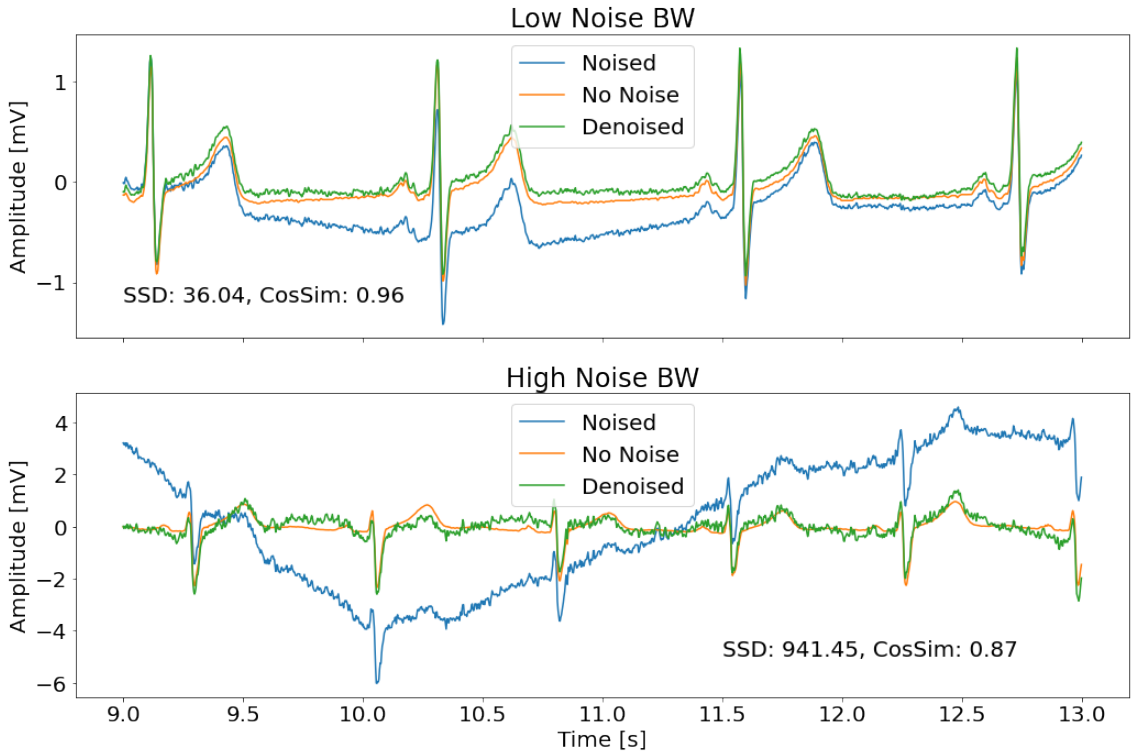

Figure 5.8: Exemplary visualisation of the performance of the best performing denoising method FIR for BW using several beats ECG samples for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

and A.7. Especially for low MA noise, waves of high amplitude (like the R wave) are visible in the denoised output of the FIR method (best performing method considering the CosSim metric) while not being visible in the denoised output of the FCN-DAE method (best performing method considering the SSD metric). The results of EM noise and MA noise are similar. Neither the best performing denoising method considering the CosSim metric (DRNN) nor the one considering the SSD metric (FCN-DAE) adequately remove EM noise from several beats ECG samples. The denoised output of the FCN-DAE method is comparable to the one from MA noise. The samples are immensely deformed and no waves of the noise-free ECG sample are visible. At least for DRNN, the R waves for low noise are detectable. Again, comparing the best performing denoising method for EM, considering the SSD and CosSim metric, results in choosing the CosSim metric to be better for determining the best method. This is again based on the possibility to indentify the different waves in figures A.8 and A.9.

The analysis of the denoising methods for different SNR values does not provide additional information. The corresponding figures A.10 - A.12 can be found in the appendix A. The variance of the mean values and their errors are too large to allow the determination of trends.

6 Discussion

Reviewing the separate experiments shows that the performance evaluation of single beat ECG samples with separate noise types, single beat ECG samples with mixed noise, and several beats ECG samples with separate noise types, is valid as the results are fundamentally different. The denoising method performance using single beat ECG samples is significantly better than when using several beats ECG samples. Only the classical filter methods yield acceptable results and only when denoising baseline wander (BW) noise. This can be explained by the filters' functionalities as a bandpass filter (i.e. filtering too low and too high frequencies leaving only the native ECG frequencies). Only the frequency spectrum of BW noise is, with its low frequency spectrum, different enough so that classical filter methods are able to remove the noise while leaving the ECG signal intact. For muscle artifact (MA) noise and electrode motion (EM) noise, the noise frequency spectrum overlaps significantly with one of an ECG signal. Filters can therefore not remove this noise while leaving the ECG signal intact. With the apprehensive functionality of the classical filter, filtering certain frequencies, the results can be justified. With this the performance of the filter methods can be considered to be independent of the length of the used ECG samples.

Justifying the results of the deep learning methods can not be done to the same extent as for the classical filter methods. This is due to their increased complexity. During the training process for the deep learning methods, the algorithms do not provide any feedback or reasoning explaining why and what the different layers of the networks try to achieve. This only leaves speculation and the suggestion of further experiments that aim to explain the behaviour of deep learning denoising methods while varying the ECG sample length. For ECG samples containing only one beat, the waves are at relatively constant positions within the sample, for ECG samples containing several beats that is not the case. Not only can the entire sample be shifted (i.e. first heart beat sooner of later in the recording) but also the heart rate changes the spacing between ECG waves of the same type (for instance the spacing

between the R waves) throughout the ECG samples and even within one sample if the heart rate fluctuates during the recording. The effect of those variations could be examined in a further experiment where the heart rate and the position of the waves in the ECG sample is fixed. Those results could then be compared to the results given here. Considering the experimental setup as used here, where the heart rate and relative positions of the waves is variable, the performance of the deep learning denoising methods is found to be highly dependent on the ECG sample length.

Something important to consider are the practical applications of the denoising methods and the experimental results. Of course ECG recordings usually extend over several seconds recording multiple heart beats. The classical filter methods can be assumed to function over an arbitrary duration, however for the deep learning methods that is not the case. In order to match the performance presented here, the ECG recording would have to be divided so that each fraction contains exactly one heart beat. While that might be possible for low noise scenarios where the R wave is still detectable, problems arise for high noise. Here the R wave might not be detectable making it impossible to divide the ECG recording into its single heart beats.

The results show that the outcome of the mixed noise experiment can not be derived from the experiments using separate noise. Trying to do so could involve the calculation of the mean metric value of the three noise types for each denoising method. Comparing those mean metric values would cause DeepFilter to be the best expected denoising method for mixed noise. The mixed noise experiment however shows that FCN-DAE is the best method, a method that yields best denoising performance for neither one of the three separate noise types. This proves that an independent discussion of denoising performance for separate noise types and for mixed noise is necessary as no model allowing to connect the findings of the two experiments can be found. Incorporating the best performing denoising method for different combinations of noise in figure 5.7 further complicates the requirements of such a model.

Another problem discovered is the definition of denoising performance. The two used metrics, trying to define the performance quantitatively, are expected to deliver coherent results. While for the ECG samples containing a single heart beat this can be confirmed, for ECG samples containing several heart beats contradicting results arise. This means that at least one of the two metrics describes the

performance of denoising methods insufficient. This is especially problematic since the deep learning methods require performance metrics to train their models. If a metric happens to be chosen that objectively calculates the denoising performance poorly, the overall performance of that method is threatened. Comparing the best performing denoising method considering the SSD and CosSim metric shows that the CosSim metric better finds the objectively better denoising method. Since deep learning methods are often trained using traditional distance metrics, like the SSD metric, another possible reason for the poor performance of the deep learning methods for ECG samples containing several beats is found. To further examine this, an experiment could be designed where deep learning denoising methods are trained using both a distance metric and the CosSim metric.

7 Conclusion

While classical filter methods prove their independence on the length of the used ECG samples, only baseline wander (BW) noise can be adequately removed. Deep learning denoising methods can improve ECG recordings containing muscle artifact (MA) noise and electrode motion (EM) noise, but require the ECG recording to be split into single heart beats. Mixed noise and the separate noise types should both be included when evaluating denoising methods as the denoising success for mixed noise can not be derived from the separate noise types. The metrics used to define denoising performance are critical, both for the training of the deep learning denoising methods and the measuring of the performance.

Part I

Appendix

A Additional figures

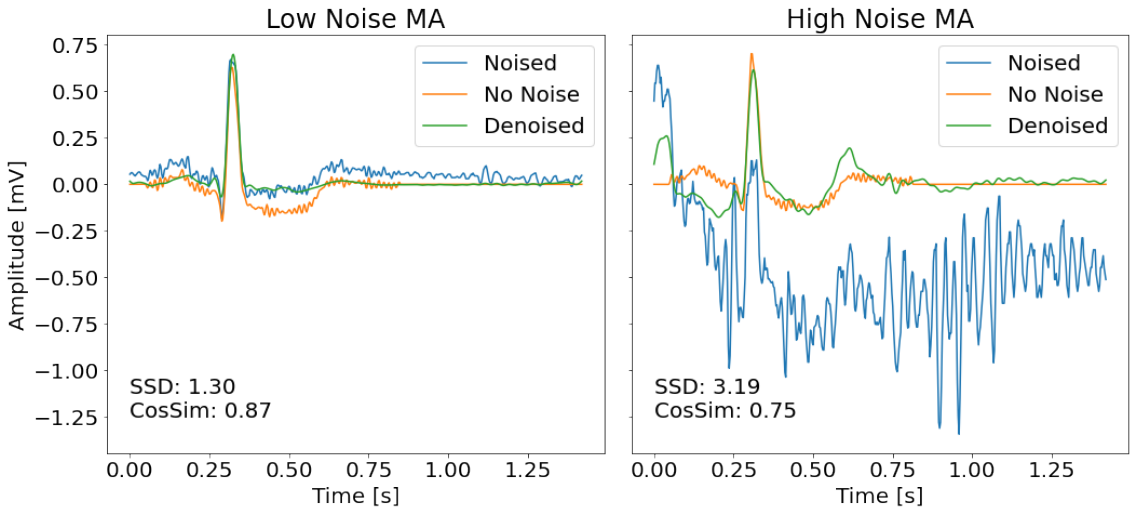

Figure A.1: Exemplary visualisation of the performance of the best performing denoising method DeepFilter for MA noise for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

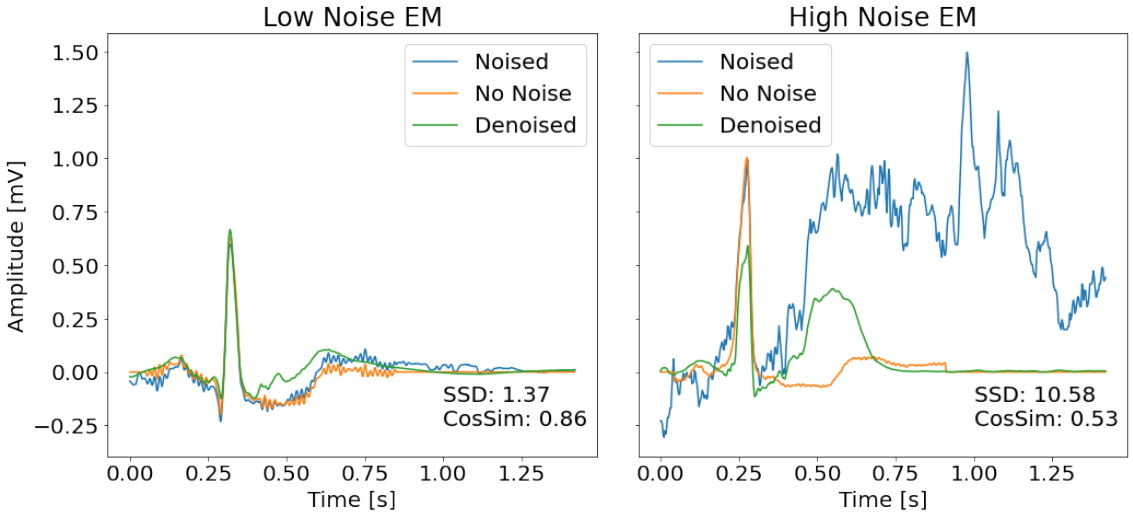

Figure A.2: Exemplary visualisation of the performance of the best performing denoising method DRNN for EM noise for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

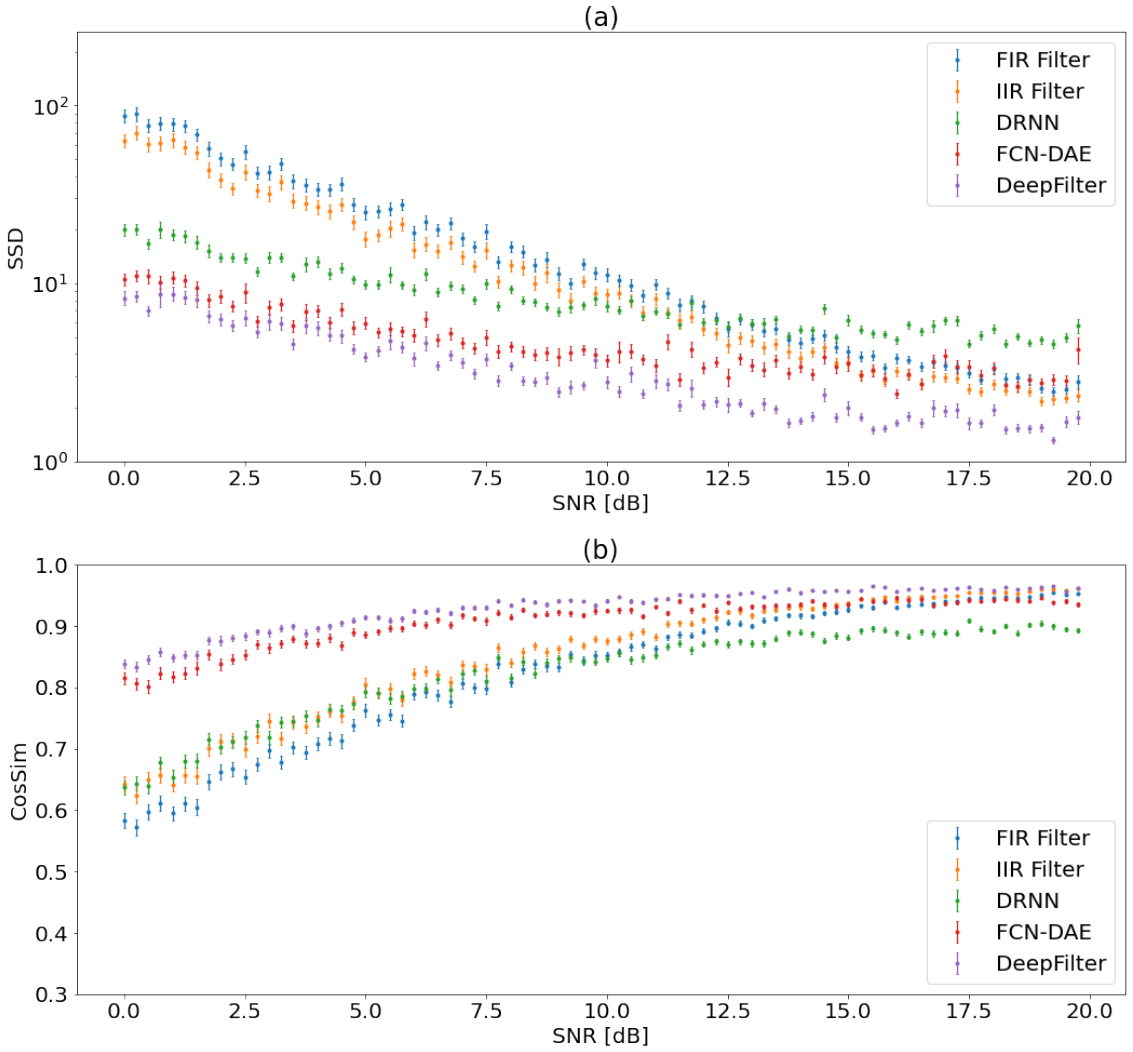

Figure A.3: Performance of the different denoising methods for one beat ECGs corrupted with BW noise measured using the SSD (a) and cosine similarity (b) metrics

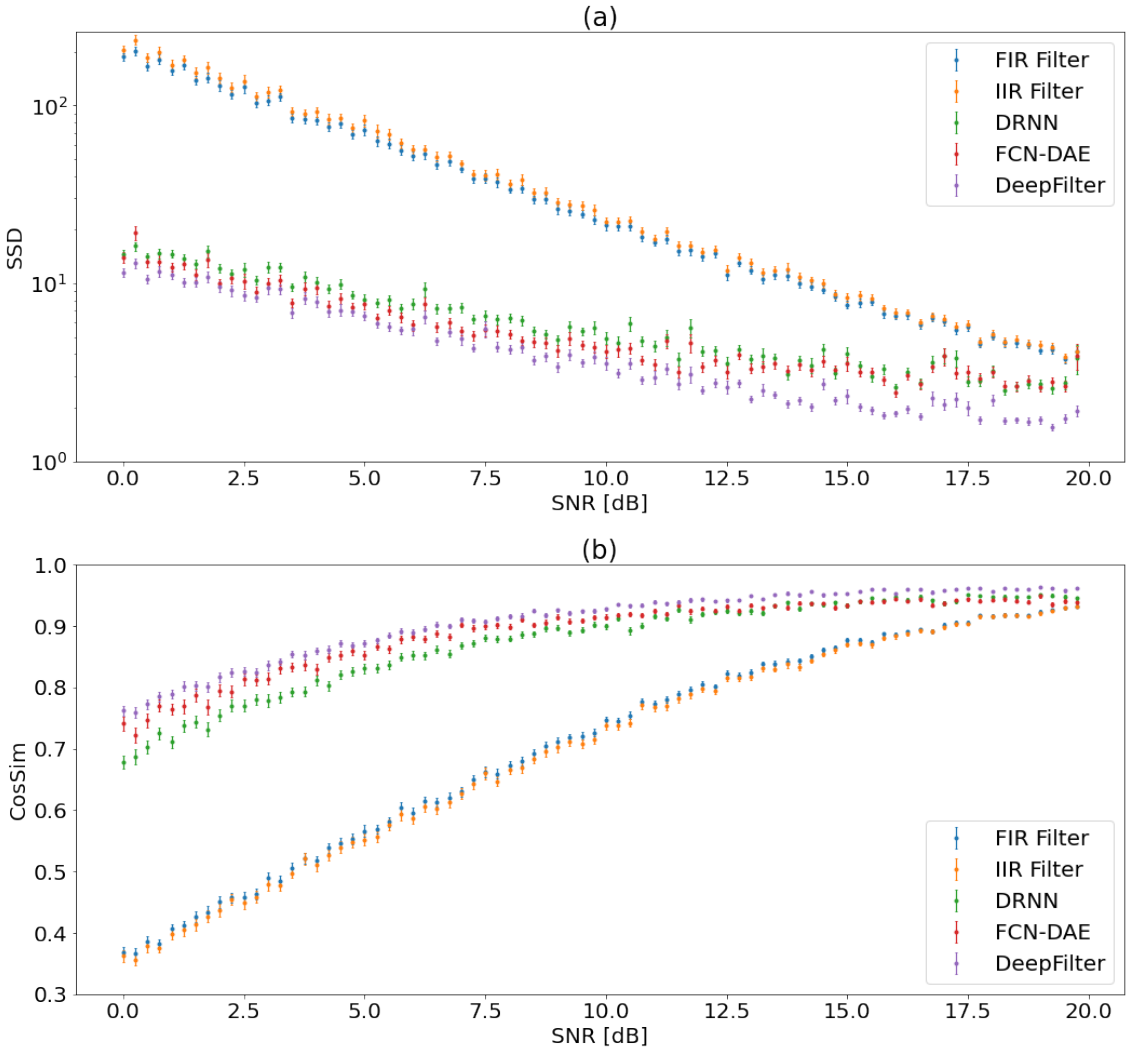

Figure A.4: Performance of the different denoising methods for one beat ECGs corrupted with MA noise measured using the SSD (a) and cosine similarity (b) metrics

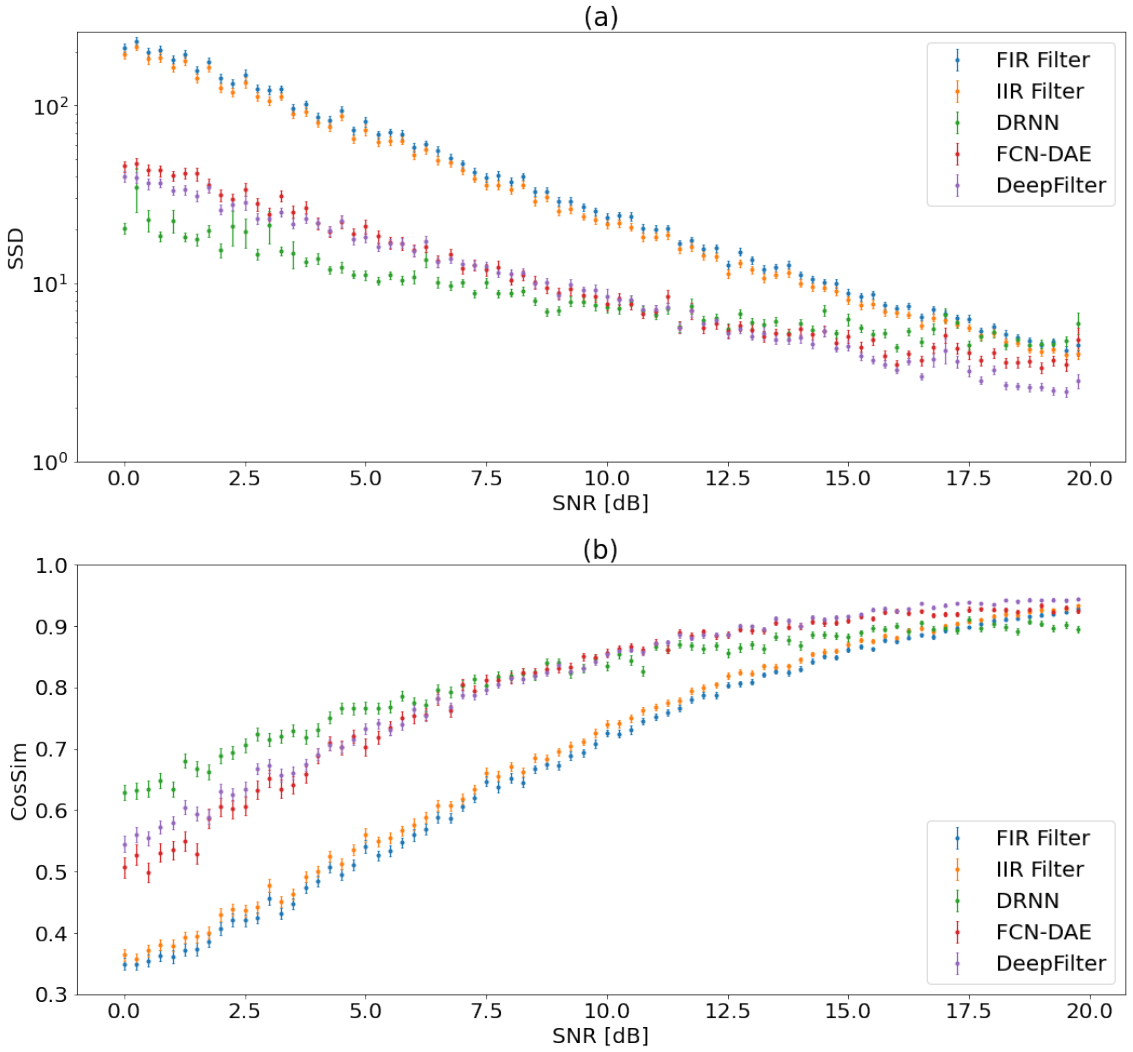

Figure A.5: Performance of the different denoising methods for one beat ECGs corrupted with EM noise measured using the SSD (a) and cosine similarity (b) metrics

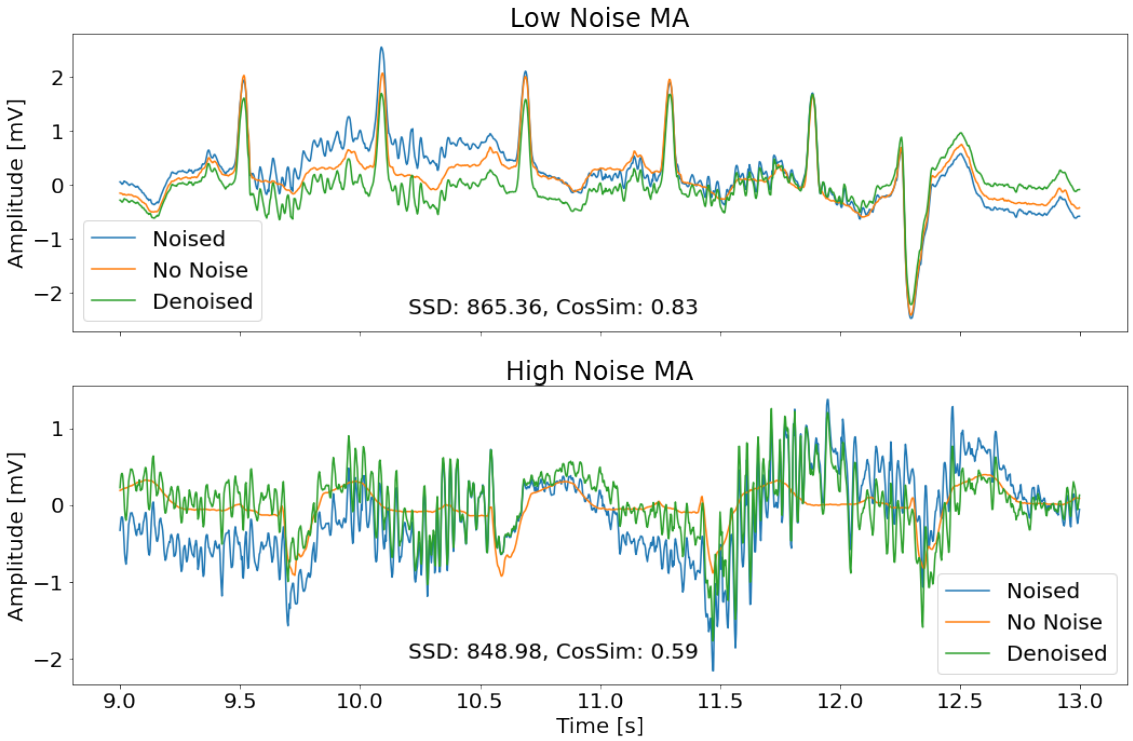

Figure A.6: Exemplary visualisation of the performance of the best performing denoising method considering the CosSim metric FIR for MA using several beats ECG samples for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

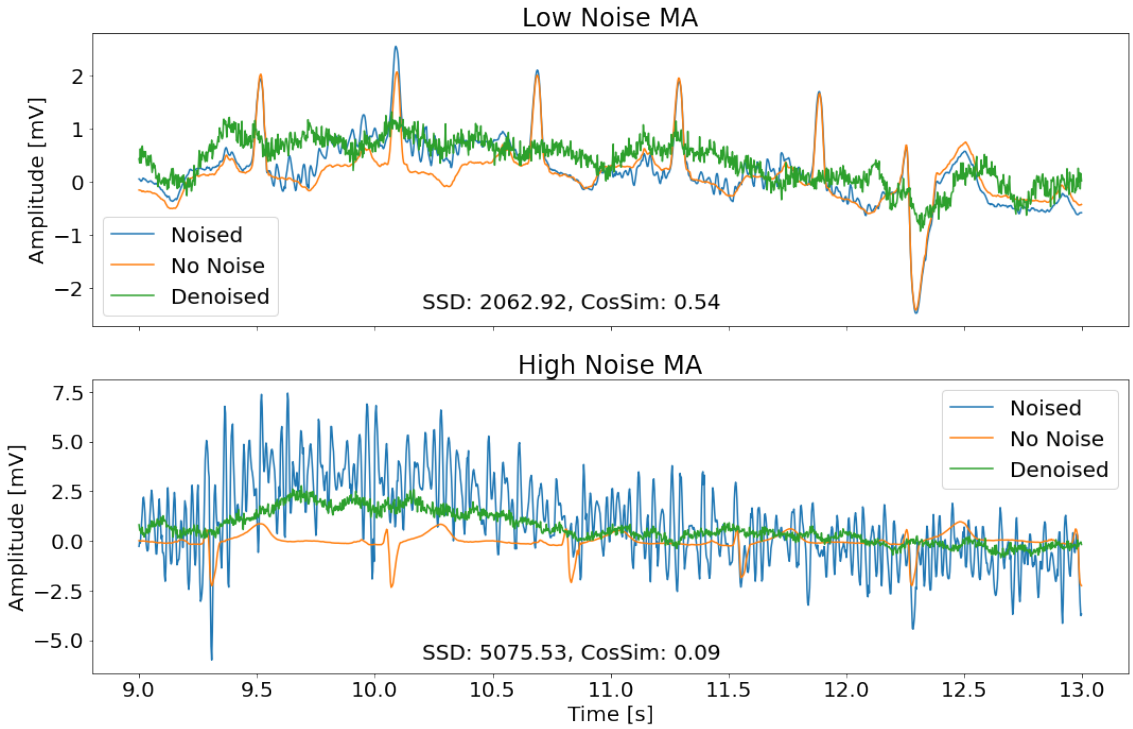

Figure A.7: Exemplary visualisation of the performance of the best performing denoising method considering the SSD metric FCN-DAE for MA using several beats ECG samples for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

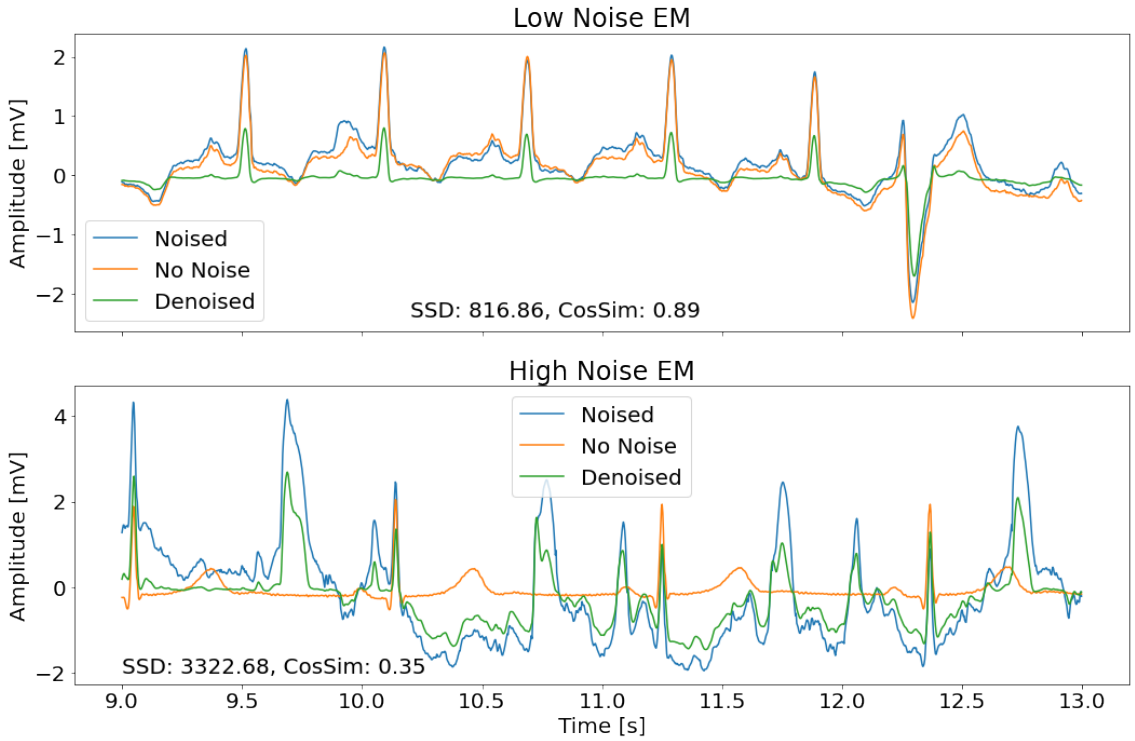

Figure A.8: Exemplary visualisation of the performance of the best performing denoising method considering the CosSim metric DRNN for EM using several beats ECG samples for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

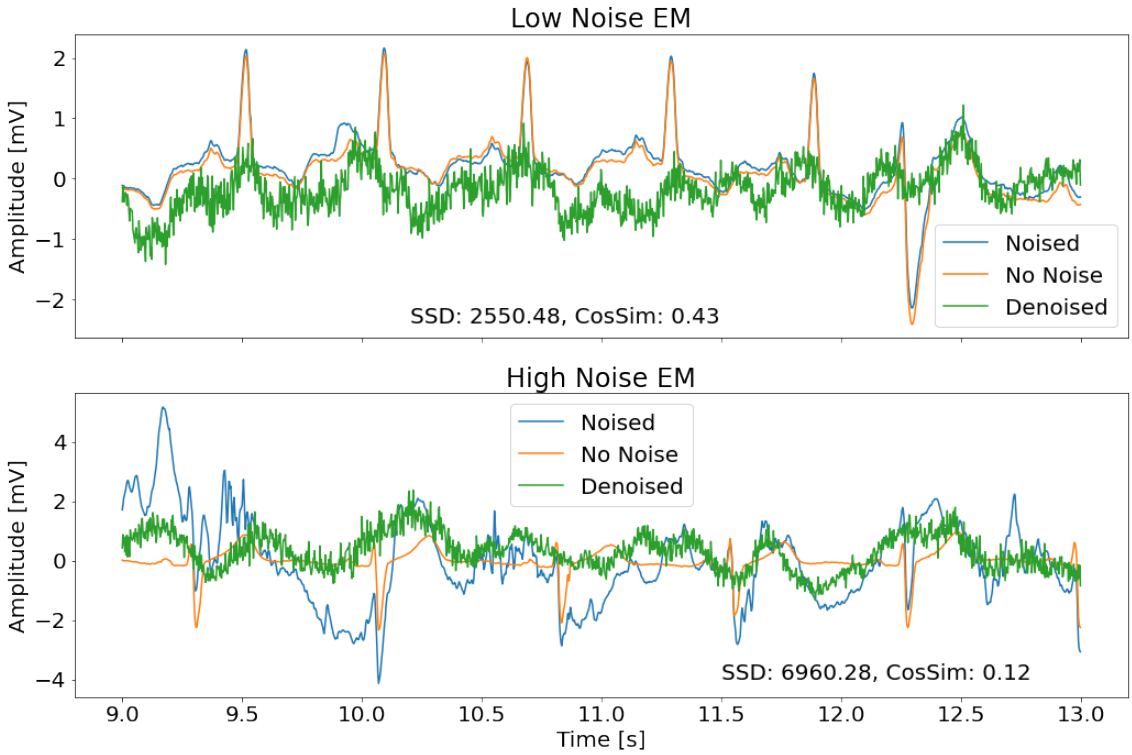

Figure A.9: Exemplary visualisation of the performance of the best performing denoising method considering the SSD metric FCN-DAE for EM using several beats ECG samples for both low noise (20 dB) and high noise (0 dB). The respective results of the performance metrics are printed.

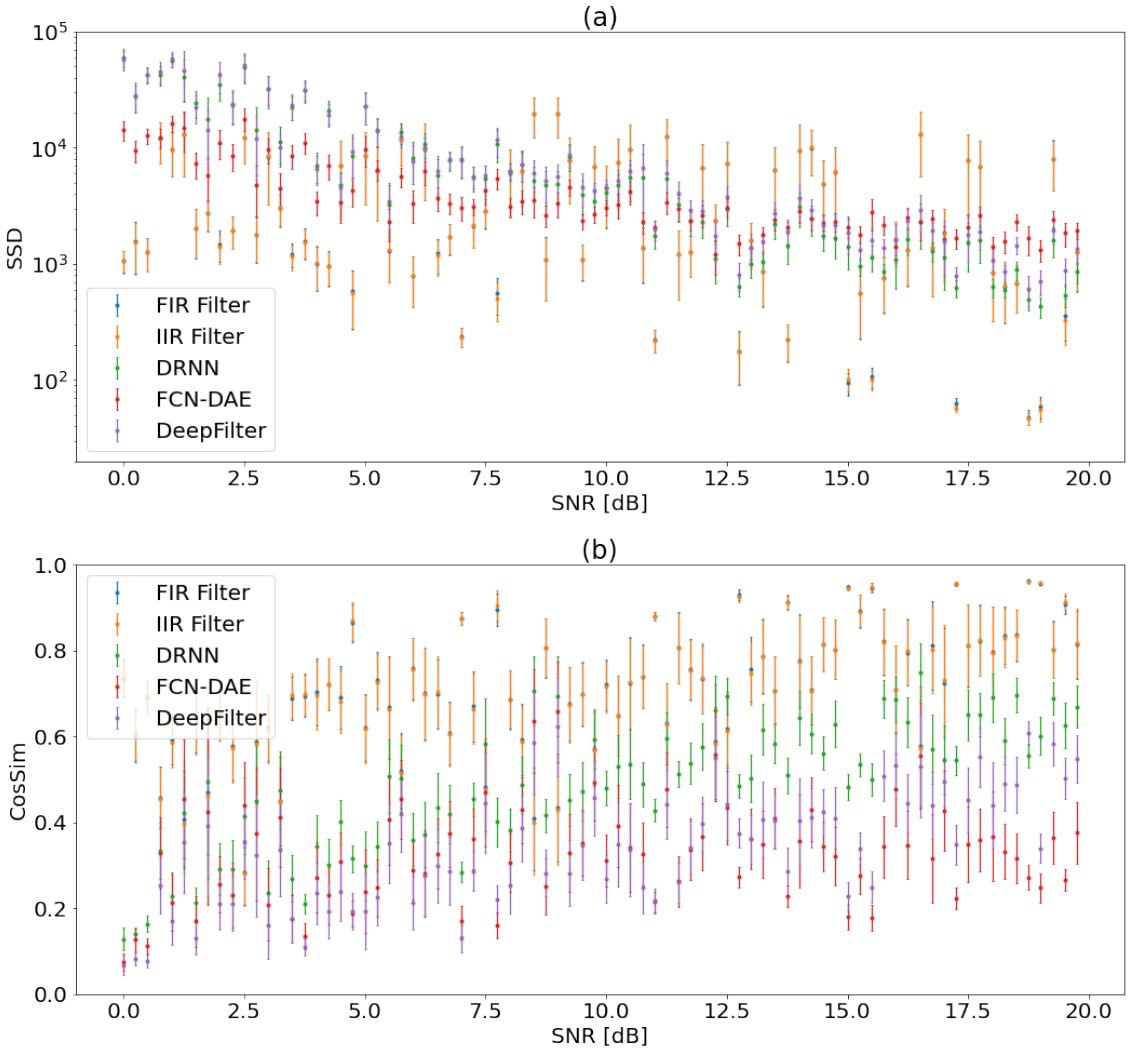

Figure A.10: Performance of the different denoising methods for several beats ECGs corrupted with BW noise measured using the SSD (a) and cosine similarity (b) metrics

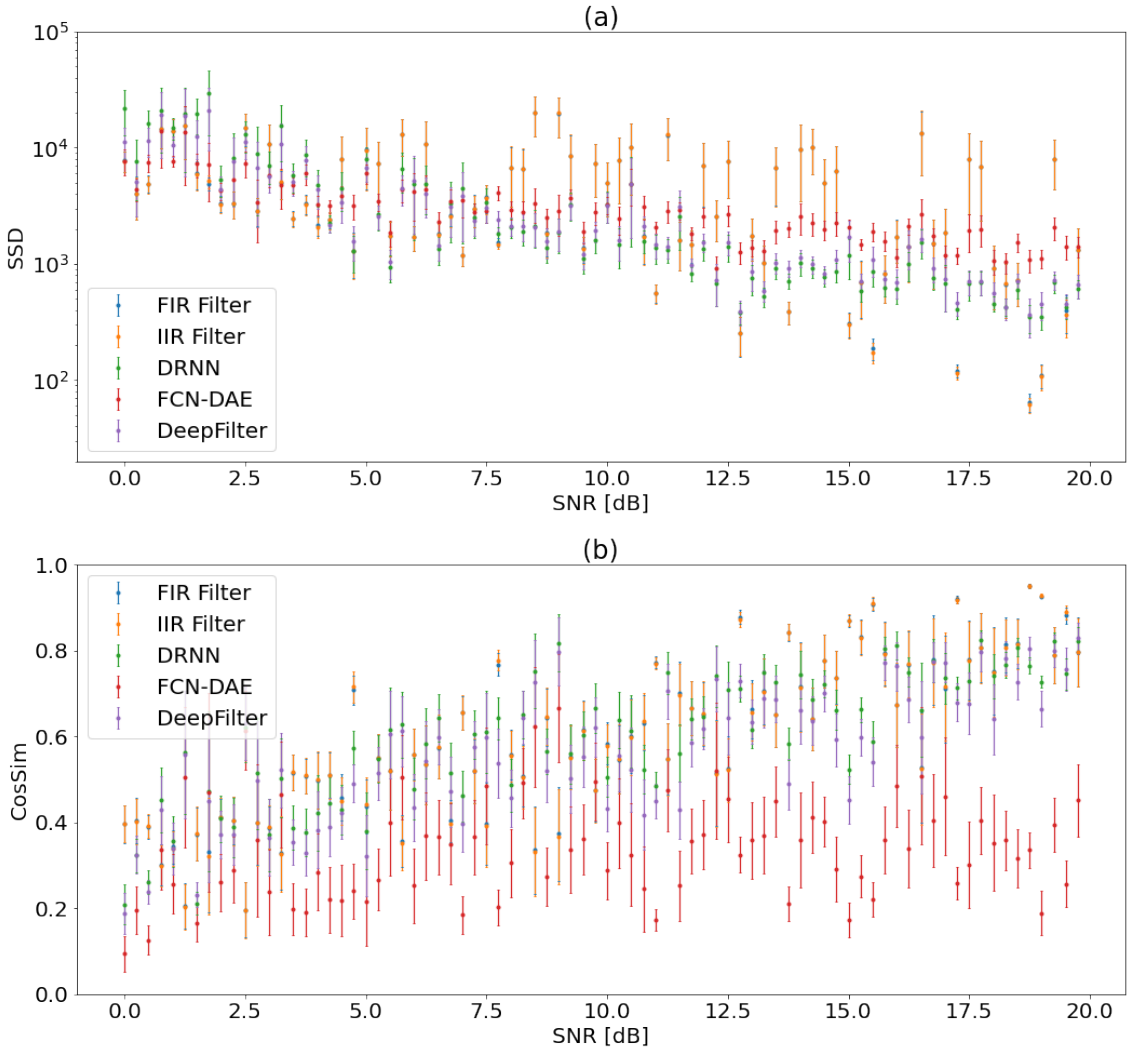

Figure A.11: Performance of the different denoising methods for several beats ECGs corrupted with MA noise measured using the SSD (a) and cosine similarity (b) metrics

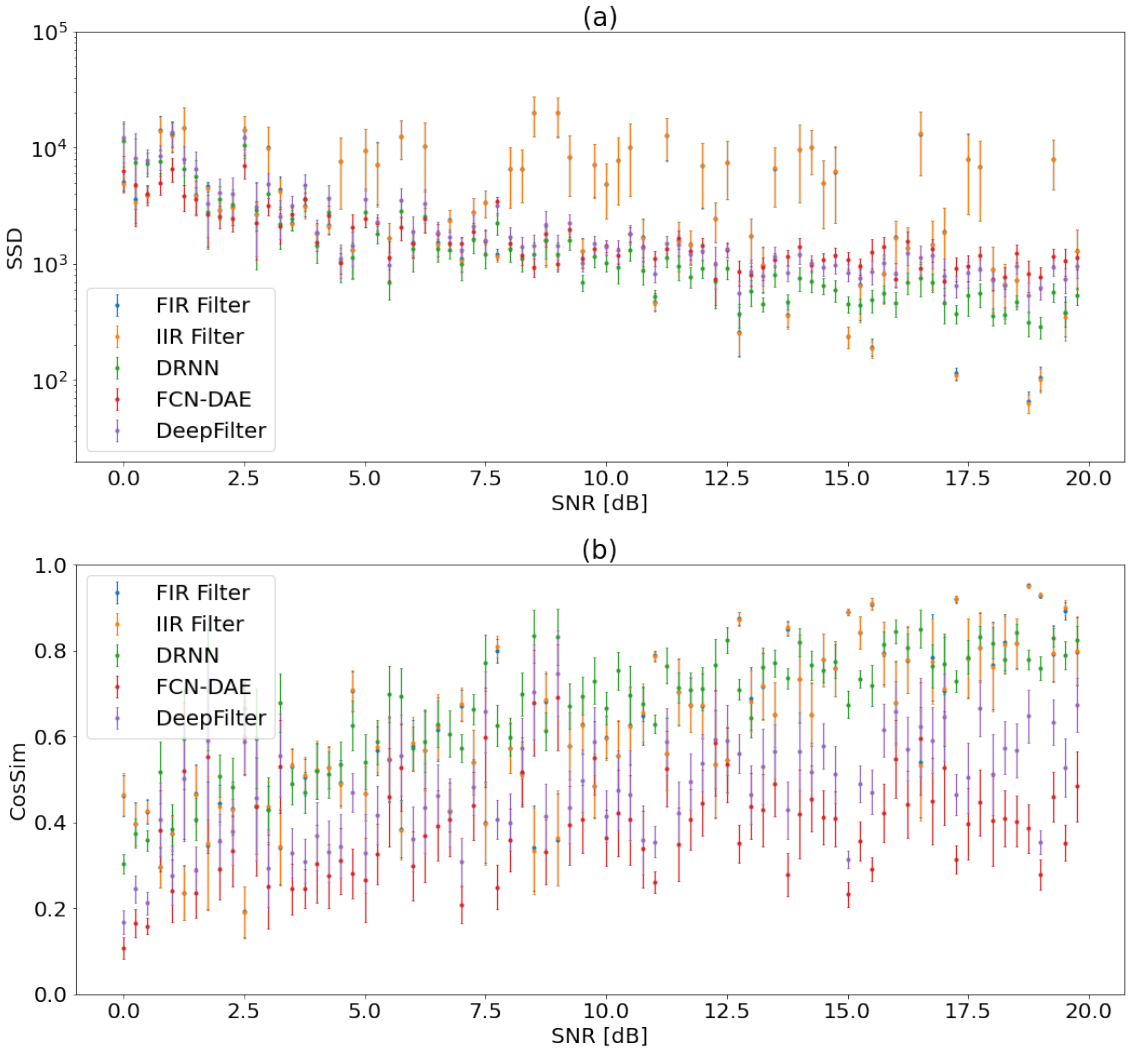

Figure A.12: Performance of the different denoising methods for several beats ECGs corrupted with EM noise measured using the SSD (a) and cosine similarity (b) metrics

B Lists

B.1 List of Figures

B.2 List of Tables

C Bibliography

[1] Nicos Maglaveras et al. "ECG pattern recognition and classification using non-linear transformations and neural networks: A review". In: *International journal of medical informatics* 52.1-3 (1998), pp. 191–208.

[2] Théo Caillol et al. "Accuracy of a Smartwatch-Derived ECG for Diagnosing Bradyarrhythmias, Tachyarrhythmias, and Cardiac Ischemia". In: *Circulation: Arrhythmia and Electrophysiology* 14.1 (2021), e009260.

[3] George B Moody, Roger G Mark, and W E Muldrow. "A noise stress test for arrhythmia detectors". In: *Computers in Cardiology* 11 (1984), pp. 381–384.

[4] Peter W Macfarlane et al. *Comprehensive electrocardiology*. Springer Science & Business Media, 2010.

[5] Bruce Alberts et al. *Molecular biology of the cell*. WW Norton & Company, 2017.

[6] Fred Kusumoto. *ECG interpretation: from pathophysiology to clinical application*. Springer Nature, 2020.

[7] Siniša Sovilj et al. "A simplified 3D model of whole heart electrical activity and 12-lead ECG generation". In: *Computational and mathematical methods in medicine* 2013 (2013).

[8] Niccoló Biasi and Alessandro Tognetti. "3D Model of the Heart Electrical Activity with Heterogeneous Ventricular Action Potentials". In: *2020 Computing in Cardiology*. IEEE. 2020, pp. 1–4.

[9] Anthony Dupre, Sarah Vincent, and Paul A Iaizzo. "Basic ECG theory, recordings, and interpretation". In: *Handbook of cardiac anatomy, physiology, and devices*. Springer, 2005, pp. 191–201.

[10] John R Hampton and Joanna Hampton. *The ECG made easy e-book*. Elsevier Health Sciences, 2019.

[11] Vrudhula K Murthy et al. “Clinical usefulness of ECG frequency spectrum analysis”. In: *The Second Annual Symposium on Computer Application in Medical Care, 1978. Proceedings.* IEEE. 1978, pp. 610–612.

[12] Carlos Van Mieghem, Marc Sabbe, and Daniel Knockaert. “The clinical value of the ECG in noncardiac conditions”. In: *Chest* 125.4 (2004), pp. 1561–1576.

[13] Jeffrey Chan, Andrew C Miller, and Emily B Fox. “Representing and Denoising Wearable ECG Recordings”. In: *arXiv preprint arXiv:2012.00110* (2020).

[14] Nino Isakadze and Seth S Martin. “How useful is the smartwatch ECG?” In: *Trends in cardiovascular medicine* 30.7 (2020), pp. 442–448.

[15] Hrishikesh Limaye and VV Deshmukh. “ECG noise sources and various noise removal techniques: a survey”. In: *International Journal of Application or Innovation in Engineering & Management* 5.2 (2016), pp. 86–92.

[16] Shubhojeet Chatterjee et al. “Review of noise removal techniques in ECG signals”. In: *IET Signal Processing* 14.9 (2020), pp. 569–590.

[17] Aswathy Velayudhan and Soniya Peter. “Noise analysis and different denoising techniques of ECG signal-a survey”. In: *IOSR journal of electronics and communication engineering* 1.1 (2016), pp. 40–44.

[18] Richard G Lyons. *Understanding digital signal processing, 3/E*. Pearson Education India, 2004.

[19] Jan A Van Alste and TS Schilder. “Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps”. In: *IEEE transactions on biomedical engineering* 12 (1985), pp. 1052–1060.

[20] Francisco Perdigon Romero, David Castro Piñol, and Carlos Román Vázquez Seisdedos. “DeepFilter: an ECG baseline wander removal filter using deep learning techniques”. In: *arXiv preprint arXiv:2101.03423* (2021).

[21] K Sravan Kumar, Babak Yazdanpanah, and P Rajesh Kumar. “Removal of noise from electrocardiogram using digital FIR and IIR filters with various methods”. In: *2015 International conference on communications and signal processing (ICCSP)*. IEEE. 2015, pp. 0157–0162.

[22] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.

[23] Michael A Nielsen. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, 2015.

[24] Mike Schuster and Kuldip K Paliwal. "Bidirectional recurrent neural networks". In: *IEEE transactions on Signal Processing* 45.11 (1997), pp. 2673–2681.

[25] Karol Antczak. "Deep recurrent neural networks for ECG signal denoising". In: *arXiv preprint arXiv:1807.11551* (2018).

[26] Sepp Hochreiter. "The vanishing gradient problem during learning recurrent neural nets and problem solutions". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02 (1998), pp. 107–116.

[27] Hsin-Tien Chiang et al. "Noise reduction in ECG signals using fully convolutional denoising autoencoders". In: *IEEE Access* 7 (2019), pp. 60806–60813.

[28] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. Ieee. 2017, pp. 1–6.

[29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[30] Pablo Laguna et al. "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG". In: *Computers in cardiology 1997*. IEEE. 1997, pp. 673–676.

[31] George B Moody, Roger G Mark, and Ary L Goldberger. "PhysioNet: a web-based resource for the study of physiologic signals". In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 70–75.

[32] George Moody, Tom Pollard, and Benjamin Moody. *WFDB Software Package*. https://doi.org/10.13026/zzpx-h016. (version 10.6.2), PhysioNet. 2021.

[33] Francisco Perdigón Romero, David Castro Piñol, and Carlos Román Vázquez-Seisdedos. *DeepFilter*. https://github.com/fperdigon/DeepFilter. 2021.

[34] Ranveig Nygaard, Gerry Melnikov, and Aggelos K Katsaggelos. "A rate distortion optimal ECG coding algorithm". In: *IEEE Transactions on biomedical engineering* 48.1 (2001), pp. 28–40.

[35] Peipei Xia, Li Zhang, and Fanzhang Li. “Learning similarity with cosine similarity ensemble”. In: *Information Sciences* 307 (2015), pp. 39–52.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 27.12.2021 .....