ROBUST REAL-TIME PPG-BASED HEART RATE MONITORING

Ewelina Nowak, Mateusz Kraiński, Maciej Rubiński, Marta Pazderska, Stanisław A. Raczyński

Gdańsk University of Technology
Faculty of Electronics, Telecommunications and Informatics,
Department of Automatic Control and Department of Multimedia Systems,
ul. G. Narutowicza 11/12, 80-233 Gdańsk,
email: {ewenowak1,matkrain}@student.pg.gda.pl
{macrubin,marpazde}@student.pg.gda.pl
stanislaw.raczynski@pg.gda.pl

ABSTRACT

Many existing methods for tracking heart rate (HR) from photoplethysmographic (PPG) signals can be found in the literature, but they are tested only in static scenarios and fail when motion artifacts are strong. Recently, an algorithm called TROIKA was proposed for robust HR tracking, but the computational complexity of that algorithm is very high which makes it difficult to implement in small embedded devices, such as wrist-wearable HR monitors. In this article we present a new, fast family of methods robust against very strong motion artifacts using spectral subtraction or nonnegative matrix factorization (NMF) for signal enhancement and MA removal and online Viterbi decoding or particle filtering for HR tracking. On our test data set we obtain an average error of 1.3%.

Index Terms— heart rate, PPG, photoplethysmography, particle filter, Viterbi algorithm, spectral subtraction, NMF

1. INTRODUCTION

Nowadays there is a range of products available for monitoring heart rate (HR), *e.g.*, Samsung Gear Fit, Atlas Fitness Tracker and Mio Alpha Heart Rate Sport Watch. These devices estimate HR in real time from photoplethysmographic (PPG) signals recorded at wearer's wrist using a light-emitting diode (LED) that measures the intensity changes in the light reflected by skin, forming a PPG signal [1]. The device is worn on the wrist during normal daily activities, but primarily during sport or other exercises. Heart rate monitoring is crucial for exercisers to monitor their training load and metabolism [2]. Unfortunately, PPG signals are strongly influenced by motion artifacts (MA) resulting from body (especially hand) movements, which strongly interfere with HR monitoring during exercise. Many existing methods for tracking heart rate from PPG can be found in the literature for clinical, static scenarios, but they fail when MA are strong [3]. Zhang *et al.* proposed a robust HR tracking method called TROIKA [3], but computational complexity of that algorithm is very high (mostly due to the usage of FOCUSS), which makes it difficult to implement in small embedded devices such as wrist-wearable HR monitors. The time factor define the device as real time or not, so it crucial to develop a low-latency tracking algorithm.

In this work we propose a robust against very strong motion artifact noise and fast method for signal enhancement and MA removal. We assume that we have access to accelerometric data, which can be used to enhance the PPG signal by means of spectral subtraction or nonnegative matrix factorization (NMF). We track the HR using online Viterbi decoding or particle filtering.

2. SIGNAL ENHANCEMENT

In our signal enhancement methods we assume that the device is equipped with accelerometers and that we receive 5 channels: 2 PPG signals, measured in different places and 3 signals from the acceleration sensor which show the acceleration for the x, y and z axes, which is the structure of data collected in [3]. The MA distortions in the PPG signals have the same frequencies as the changes in the accelerometer signal and in [3] Zhang *et al.* have shown that the singular spectrum analysis (SSA) with further removal of components with peaks at similar frequencies as those in the accelerometer spectrum was a very successful method for MA filtering. Here, we propose to use the faster NMF and a much simpler spectral subtraction.

All signals are divided into frames $x_{i,t}^k$, where $k = 1 \dots K$, $i = 1 \dots M$, $t = 1 \dots L$ indexes the time, $K = 5$ is the number of channels and $M = 1000$ is the number of samples per frame, which correspond to 8 seconds of recording with 125 Hz sampling rate, with 750-sample overlap, which means that the results are computed every 2 seconds for the last 8 seconds of the recording, as in [3].

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

2.1. Non-negative Matrix Factorization

Reduction of the influence of MA and other noise on the PPG signal can be obtained using non-negative matrix factorization (NMF). This approach decomposes a nonnegative (having only nonnegative elements) magnitude or power spectrogram matrix $\mathbf{Y}$ into a product of two, also nonnegative, matrices $\mathbf{A}$ and $\mathbf{S}$.

$$\mathbf{Y} \approx \mathbf{AS} \tag{1}$$

In the decomposition step, a distortion measure between the data $\mathbf{Y}$ and its approximation $\mathbf{AS}$ is minimized. The original NMF algorithm was designed to minimize Euclidean distance or I-divergence. Non-negative Matrix Factorization approximates each column of the data matrix $\mathbf{y}_t$ as a linear combination of basis vectors $\mathbf{a}_n$ (columns of $\mathbf{A}$):

$$\mathbf{y}_t = \sum_n s_{n,t}\mathbf{a}_n \tag{2}$$

where n is the basis vector number and t is the time index.

The matrices $\mathbf{A}$ and $\mathbf{S}$ are nonnegative. Only additive mixtures of nonnegative basis vectors are possible. The coefficient matrix S is called the weighting matrix.

Let us denote the FFT of the k-th channel ($k \in [1, 5]$) as $\mathbf{y}^k$. Magnitude spectrogram $\mathbf{Y}$ is created and each column of spectrogram is normalized:

$$\mathbf{y}^k = \frac{\mathbf{y}^k}{\max(\mathbf{y}^k)}. \tag{3}$$

Magnitude spectrogram $\mathbf{Y}$ consists of concatenated spectrograms of all 5 channels: $\mathbf{Y} = \left[\mathbf{Y}^1\,\mathbf{Y}^2\,\mathbf{Y}^3\,\mathbf{Y}^4\,\mathbf{Y}^5\right]$. $\mathbf{A}$ is initialized with random variables. $\mathbf{Y}$ is factorized by NMF with $r \in (0.1, 2.5)$ and a weight sparsity regularization with $\mu = 0.1$ and $p = 1.5$ [4]. The resulting weights $\mathbf{S}$ are split back into five submatrices which correspond to the five channels: $\mathbf{S} = \left[\mathbf{S}^1\mathbf{S}^2\mathbf{S}^3\mathbf{S}^4\mathbf{S}^5\right]$. For each of the matrices, its L^2-norm is computed and $\mathbf{s}$ is modified with weights on each channel:

if $||\mathbf{s}^1||_2 > ||\mathbf{s}^2||_2$

$\mathbf{s}_{out} = w_2\mathbf{s}^1 + w_1\mathbf{s}^2 - w_3\mathbf{s}^3 - w_4\mathbf{s}^4 - w_5\mathbf{s}^5$

$$w_i = \frac{||\mathbf{s}^i||_2}{||\mathbf{s}^i||_2 + ||\mathbf{s}^j||_2}, \quad \text{where } j, i \in \{1, 2\}$$

$$w_k = \frac{||w_2\mathbf{s}^1 + w_1\mathbf{s}^2||_2}{||3\mathbf{s}^k||_2}, \quad \text{where } k \in \{3, 4, 5\}$$

if $||\mathbf{s}^1||_2 < ||\mathbf{s}^2|_2$

$\mathbf{s}_{out} = w_1\mathbf{s}^1 + w_2\mathbf{s}^2 - w_3\mathbf{s}^3 - w_4\mathbf{s}^4 - w_5\mathbf{s}^5$

$$w_i = \frac{||\mathbf{s}^i||_2}{||\mathbf{s}^i||_2 + ||\mathbf{s}^j||_2}, \quad \text{where } j, i \in \{1, 2\}$$

$$w_k = \frac{||w_1\mathbf{s}^1 + w_2\mathbf{s}^2||}{||3\mathbf{s}^k||}, \quad \text{where } k \in \{3, 4, 5\}$$

else

$\mathbf{s}_{out} = \text{mean}(\mathbf{s}^1, \mathbf{s}^2) - w(\max(\mathbf{s}^3, \mathbf{s}^4, \mathbf{s}^5))$

$w = 0.55$.

The last step consists of reconstruction of the spectrogram $\mathbf{Y}$ by the product of $\mathbf{A}$ and $\mathbf{S}_{out}$.

$$\mathbf{Y}_{\text{rec}} = \mathbf{AS}_{\text{out}} \tag{4}$$

We use an online version of NMF, where the basis $\mathbf{A}$ is found beforehand and the $\mathbf{y}_{\text{rec}}$ is found on a frame-by-frame basis.

2.2. Spectral Subtraction

Spectral subtraction is a common method for signal enhancement [9]. We have found that the best scaling factor between the spectra is the sum of the samples of the spectrum. The enhanced signal can therefore be calculated as:

$$\mathbf{y} = R\Big(\frac{\mathbf{y_{ppg}}}{\sum \mathbf{y_{ppg}}} - \frac{\mathbf{y_{acc}}}{\sum \mathbf{y_{acc}}}\Big), \tag{5}$$

where $\mathbf{y_{acc}} = \mathbf{y_{acc_x}} + \mathbf{y_{acc_y}} + \mathbf{y_{acc_z}}$, $\mathbf{y_{acc_x}}$, $\mathbf{y_{acc_y}}$, $\mathbf{y_{acc_z}}$ are the amplitude spectra of the corresponding x, y and z accelerometer channels, $\mathbf{y_{ppg}}$ is the amplitude spectrum of the PPG signal and $R(x)$ is the ramp function defined as

$$R(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \le 0 \end{cases} \tag{6}$$

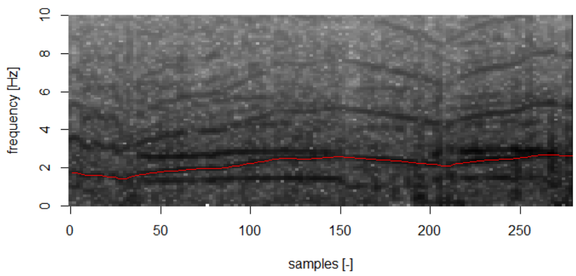
In Fig. 1, an example spectrograph is shown of the PPG signal before and after spectral subtraction [5]. As one can notice, the motion artifacts have been removed. It is also clear to notice that there is a visible HR trajectory and its first harmonic. In order to even further enhance the heart rate signal, we use the presence of its first and second harmonics by applying the following modification of the spectrum:

$$\mathbf{y}_k = a_1\mathbf{y}_n + a_2\mathbf{y}_{2n} + a_3\mathbf{y}_{3n}, \tag{7}$$

where $\mathbf{y}_k$ is the k-th element of the $\mathbf{y}$ vector. For this report we have used $a_1 = 1$, $a_2 = 0.66$ and $a_3 = 0.33$, which had been chosen experimentally.

Having two PPG signal channels we have had to determine which one to use. To determine it, we analyze the variance of their spectra. The spectrum with lower variance is chosen for further analysis.

The final step for preparing the signal before using the heart beat tracking algorithms, was to multiply the spectrum with a probability density function determined by the histogram of the training BPM values. The skewed normal distribution with mean value of 130 BPM, standard deviation of 30 BPM and skewness of 0.6 has been chosen as a parametric approximation of the histogram.



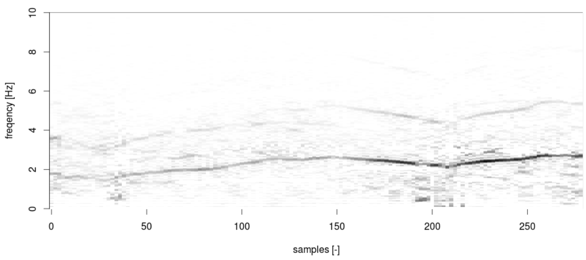


Fig. 1. Example of spectral subtraction on a PPG signal: raw PPG spectrogram with real HR shown as a red line (top) and PPG spectrogram enhanced using spectral subtraction (bottom). The strongest harmonics are not fully attenuated, but the resulting signal is very clean.

3. HEART BEAT TRACKING

3.1. Particle filtering

We use the original particle filter called sequential importance resampling (SIR) that has been presented in [6]. In our case, the particles represent the estimated heart rate values. The algorithm consists of the following steps:

1. generate N random particles $x_0^{(L)}$ and assign the weight of each as $w_0^{(L)} = \frac{1}{N}$,
2. estimate the next state of each particle based on the state model,
3. add random noise (here: Gaussian with variance v) to each particle,
4. to each of the particles, update its weight $w_k^{(L)}$ based on the distribution $p(x_k^{(L)}|y_k)$, $L \in (1:N)$,
5. if the effective number of particles N_{eff} is less than the desired threshold, N_{thr} perform resampling,
6. new signal state estimate can be computed as a weighted mean of the particle states
7. if not done, go to 2.

We have performed simulations for $N = 5000$, although in general increasing the number of particles should give better results. We have modeled the HR signal with the autoregressive (AR) model. The parameters of the model have been determined for the training data set using the Neadler-Mead minimization of the mean square prediction error. The model order was set to $q = 8$.

The variance v of the additive system noise is dynamically determined on the basis of the noise in the filtered PPG spectrum. The more noise is in the channel, the less we trust the measurement, the more we trust the autoregressive model. Tests have shown that the best quantitative measure of the noise in a channel for heart rate analysis is the L^p norm of the spectrum samples divided by the first few frames spectrum L^p norm average:

$$S = \sum_{f=0}^{f=\infty} \mathbf{y}(f)^p \tag{8}$$

$$v = R\Big(\frac{v_0 - aS}{S_{avg}}\Big), \tag{9}$$

where S is the L^p norm of the filtered PPG spectrum samples, S_{avg} is the average of first few frames spectrum L^p norm, $R(x)$ is the ramp function (6), v_0 is the maximum variance, and a is a parameter that binds the S value with the variance. Our experiments have shown that parameters $p = 0.6$, $v_0 = 4.7$ and $a = 1.6$ give the best results.

The $p(x_k^{(L)}|y_k)$ probability density funcion is approximated with the filtered, normalized PPG spectrum of one of the PPG channels.

The assigning process of probability weight $w_k^{(L)}$ for particle L in step k can be defined as:

$$\hat{w}_k^{(L)} = w_{k-1}^{(L)} p(x_k^{(L)}|y_k) \tag{10}$$

$$w_k^{(L)} = \frac{\hat{w}_k^{(L)}}{\sum_{J=1}^{N} \hat{w}_k^{(J)}} \tag{11}$$

Having the weights, we can estimate the number of effective particles N_{eff} as:

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{L=1}^{N} \big(w_k^{(L)}\big)^2} \tag{12}$$

If $\hat{N}_{\text{eff}}$ is less than a given threshold N_{thr} a sequential importance resampling has to be done:

1. draw a new set of N particles from the current particle set with probabilities equal to their weights.
2. set $w_k^{(L)} = \frac{1}{N}$ for $L \in (1:N)$

We have chosen $N_{\text{thr}} = 1000$.

The estimated value of the observed signal for step k can be computed as:

$$\hat{x}(k) = \sum_{L=1}^{N} w_k^{(L)} x_k^{(L)} \tag{13}$$

The major advantage of this method is that thanks to the broad range of search it is very likely to find the right

No.	$\mathbf{E}_{\text{ssvit}}$	$\mathbf{E}_{\text{nmfvit}}$	$\mathbf{E}_{\text{sspf}}$	$\mathbf{E}_{\text{nmfpf}}$	$\mathbf{E}_{\text{TROIKA}}$
1	3.60	8.08	3.11	5.10	**1.90**
2	1.85	1.72	**1.45**	2.31	1.87
3	0.80	1.07	**0.55**	2.51	1.66
4	**1.40**	1.57	1.64	2.42	1.82
5	0.76	1.09	**0.63**	2.57	1.49
6	1.18	1.55	**0.92**	2.59	2.25
7	**1.20**	1.40	1.30	2.30	1.26
8	**0.57**	0.95	1.57	1.60	1.62
9	**0.54**	1.04	0.64	1.70	1.59
10	14.25	10.04	3.60	5.58	**2.93**
11	0.80	2.89	**0.54**	2.49	1.15
12	0.85	1.29	**0.76**	1.71	1.99
mean	2.32	2.72	**1.30**	2.74	1.79

Table 1. Average absolute percentage error for individual experiments. E_{ssvit} for SS + Viterbi, E_{nmfvit} for NMF + Viterbi, E_{sspf} for SS + PF, E_{nmfpf} for NMF + PF and E_{TROIKA} (SSA+FOCUSS+Vrf) for the reference TROIKA from [3].

peak again after it has been lost. The PPG channel selection method is a big advantage. Even though in most of the recordings the first channel was the better one, it was not a rule.

3.2. Viterbi decoding

The second method of peak tracking is an online version of the Viterbi algorithm [7]. The aim of the method is to obtain the most probable sequence of HR values given the entire PPG signal recorded up to now. We define the cumulative weight as

$$p_{t+1} = p(y_{t+1}|x_{t+1} = i)p(x_{t+1} = j|x_t = i) \tag{14}$$

where x_{t+1} is the estimated new HR value, x_t is an HR value in the previous frame and y_{t+1} is a spectrum observation for frame $t+1$. Consequently, $p(x_{t+1} = j|x_t = i)$ is the probability of reaching $x_{t+1} = j$ if the last chosen value is $x_t = i$ represented by the the transition matrix $\mathbf{A}$, which contains probabilities of transition from state i to state j. It's elements are given by

$$a_{i,j} = \mathcal{N}(|i-j|, \mu, \sigma), \tag{15}$$

where $\mathcal{N}$ is the normal probability distribution with mean $\mu = 0$ and standard deviation $\sigma = 20$. $p(y_{t+1}|x_{t+1})$ is approximated as

$$p(y_{t+1}|x_{t+1}) = \frac{x_{t,f}}{\sum_f x_{t,f}}. \tag{16}$$

Furthermore, $\mathbf{s}_{init}$ is the initial probability vector which is used to find the initial value of the pulse. The prior probability for every pulse value is a normal distribution, with mean value $\mu = 93.37$ which is an average value for the first ground-truth HR values taken from the reference data and standard deviation σ arbitrary set to 20.

It eliminates potential errors caused by low- and high-frequency noises and protect the computed pulse track from unnatural changes. $\mathbf{e}_t$ is emission vector built by equation:

$$\mathbf{e}_t = \log\Big(\frac{\mathbf{y}_t}{\sum_{i=1}^{n} y_t(i)}\Big) \tag{17}$$

where $\mathbf{y}_t$ is symbols vector for the t-th data frame.

For every HR value which is possible to be chosen by described algorithm the probability is determined. The values are collected in the following matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm}, \end{bmatrix}$$

where $t = 1, 2, ..., T$ indexes data frames, $i = 1, 2, ..., M$ indexes the potential HR values. Every value in this matrix is computed by following equation:

$$p_{i,t} = \begin{cases} \mathbf{s}_{\text{init}} + \mathbf{e}_t, & \text{if } t = 1 \\ \max(\mathbf{p}_{t-1} + \mathbf{a}_i + e_{i,t}), & \text{if } t > 1 \end{cases} \tag{18}$$

where $\mathbf{p}_{t-1}$ is vector of probability values for $t-1$ frame and $\mathbf{a}_i$ is the $i-$th column of the $\mathbf{A}$ matrix.

Viterbi algorithm finds the best HR path through the set of all possible model states for every element of the observation sequence [8]. The last value of the path is the new pulse value in consecutive sequence, as follows

$$\hat{x}(t) = \arg\max(\mathbf{p}_t) \tag{19}$$

4. RESULTS

Data used for tests was the data from the 2015 Signal Processing Cup, which was also used in [3]. We have used data from 12 experiments involving human subjects during exercise. Subjects were aged 18–35 and performed running exercises. During data recording, each person ran on a treadmill with changing speeds. The first and second type of exercises consists of following stages:

1. rest (30s) – 8 km/h (1min) – 15 km/h (1min) – 8 km/h (1min) – 15 km/h (1min) – rest (30s),
2. rest (30s) – 6 km/h (1min) – 12 km/h (1min) – 6 km/h (1min) – 12 km/h (1min) – rest (30s).

In the experiments only the first data is used the first type of exercises. Rest of the data was based on the second type of exercises. We have tested all four combinations of signal

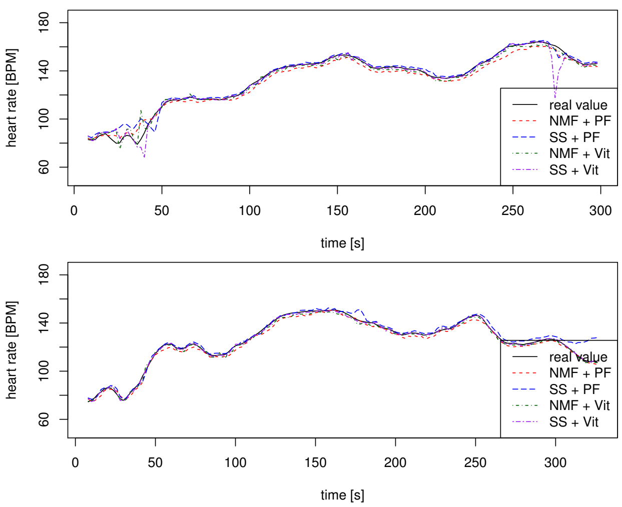


Fig. 2. Example results.

enhancement and peak tracking. Tracking error is defined as the average absolute error:

$$E = \frac{1}{W}\sum_{t=1}^{W}\frac{|\hat{x}(t) - x(t)|}{x(t)}, \tag{20}$$

where W is the number of frames, $\hat{x}(t)$ is the estimated heart rate and the $x(t)$ is the true heart rate value at time t. The obtained error values, along with the results from [3], are shown in Tab. 1. Fig. 2 presents example results for 3 experiments for all four combinations. The Viterbi decoder was prone to sudden changes in the estimated HR, but these could be removed with an additional smoothing filter as post-processing.

5. CONCLUSION

In this work we have proposed framework for heart rate estimation with a pulse oximeter and an accelerometer, which are assumingly embedded in a wristband. The approach have been tested and the results have been presented in section 4. More work and tests are required to determine the best parameter set. The best results were obtained using a combination of spectral subtraction and particle filtering. The simplicity of the spectral subtraction makes the algorithm very fast when it comes to computational complexity. Furthermore, the computational load also depends on the number of particles in the particle filter, which acts as a parameter controlling the trade-off between the speed and precision. In our experiments (a prototype GNU R implementation), the particle filtering with spectral subtraction analyzes a single frame in 0.2 s on average on a single core of an Intel Core i5 processor. There is a considerable difference in execution time of each method. For NMF+TROIKA the average execution time for one recording was about 25–35 minutes, while for NMF with particle filter about 10-15 minutes. We have found that this is a few orders of magnitude faster than the FOCUSS algorithm used in the original [3]. Using Spectral Substraction with Viterbi decoding method made the most satisfactory time results and the execution time for this method was several seconds per recording.

REFERENCES

[1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, R1, 2007.

[2] J. Achten, A. E. Jeukendrup. "Heart rate monitoring." *Sports medicine* 33.7, 2003, pp.517-538.

[3] Z. Zhang, Z. Pi, B. Liu, "TROIKA: a general framework for heart rate monitoring using wrist-type photoplethysmographic (PPG) signals during intensive physical exercise," *IEEE Trans. Biomedical Engineering*, vol. 62, no. 2, February 2015, pp. 522–531.

[4] S. A. Raczyński, N. Ono; S. Sagayama, "Extending nonnegative matrix factorization–a discussion in the context of multiple frequency estimation of musical signals," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2009, pp. 934–938.

[5] M. Dorbecker, S. Ernst, "Combination of Two-Channel Spectral Subtraction and Adaptive Wiener Post-Filtering for Noise Reduction and Dereverberation," in *Proc. European Signal Processing Conference (EUSIPCO)*, 1996.

[6] N.J. Gordon, D.J. Salmond, A.FM Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, 1993, pp. 107–113.

[7] L. V. Bao, S. Garcia-Salicetti, B. Dorizzi, "On Using the Viterbi Path Along With HMM Likelihood Information for Online Signature Verification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 5, 2007, pp. 1237–1247.

[8] R.M. Stern, G.G. Cano, S.A. Brille, "An approach to cardiac arrhythmia analysis using hidden Markov models," *IEEE Trans. Biomedical Engineering*, vol. 37, no. 9, 1990, pp. 826–836.

[9] Y. Ephraim, M. and David. "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator." *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, 1984, pp. 1109–1121.