Using Consumer Camera and Custom Firmware to Monitor Heart Rate in Terminally Ill Children during Music Therapy

Maurice Rohr[1], Monika Hoog Antink[2,3], Sebastian Dill[1], Christoph Hoog Antink[1]

[1] KIS*MED - AI Systems in Medicine, Technische Universität Darmstadt, Darmstadt, Germany
[2] Department of Psychiatry and Psychotherapy, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
[3] Institute for Music Therapy, Hochschule für Musik und Theater, Hamburg, Germany

Abstract

Photoplethysmography Imaging (PPGI) is considered a prospective non-contact replacement for existing methods of heart rate (HR) and HR variability monitoring. However, "in the wild" studies are rare. Music therapy for children with severe disabilities aims to alleviate pain and increase their well-being. As they are not able to communicate via spoken language, changes in facial expressions and vital signs may be used to quantitatively assess effects of music therapy. Here, cameras can be used to capture both, hints of the emotional expressions and, via PPGI, HR. From a technical perspective, the high variability in HR, the interaction with the therapist, as well as natural lighting form an ideal scenario to stress-test existing algorithms. We used a consumer camera with custom firmware to record raw videos of children during music therapy. We employ the often used pipeline of face detection, skin segmentation, pulse signal extraction and heart rate estimation. To enforce temporal coherence, we apply temporal filtering to the skin masks inside a 5-frame window. We evaluate the estimated HR using a reference ECG. For the full 30 minute therapy sessions we achieved a mean absolute / root-mean-squared error of 3.98 / 7.11 BPM. Finally, PPGI methods show promising results for music therapy and can provide accurate HR over a wide range.

1. Introduction

Music therapy is the clinical use of music based interventions within a therapeutic relationship to promote well-being. The intervention used in this study is based on improvisation, utilizing different instruments and the voice to create songs, rhythmical synchronization, movement and sensory stimulation and takes place in the Theodorus Kindes-Tageshospiz Hamburg. The patients are children with severe and multiple disabilities (SMD) that imply impeded communication and movement. One of the difficulties of evaluating the efficacy of the therapy is that the children are not able to communicate their feelings/condition due to medication, illness and physical weakness. Furthermore, Orita et al. [1] and others could show that heart rate (HR) and its variability are possible markers for the effectiveness of the therapy and that music therapy might suppress activities of the parasympathetic nervous system in children with SMD. Also there are indications that some children show changes in HR after music therapy without any other clinical observation, which makes HR analysis a promising tool. Moreover, music therapy has been shown to also stabilize vital signs [2]. On the other hand, the children in this study tend to remove attached cables or sensors due to their condition willfully but they can also detach because of rapid movement (e.g. seizures, myoclonus). In some cases it is not possible to attach cables, e.g. when a child is lying on their stomach to assist with breathing. Thus, the goal during monitoring is to minimize any invasive procedure or unnecessary disturbance and achieve a "natural" setting.

Photoplethysmography-Imaging (PPGI) employs a camera, often RGB, to measure a photoplethysmogram from afar. There exists a pool of signal-processing and machine learning based algorithms that are tested on a growing number of openly available datasets [3]. However, the datasets consist predominantly of healthy adults. Accordingly, there are studies missing that answer whether these algorithms apply well to children/people with conditions and if the measurement scenario is realistic for real world application. On the other hand, setups used in research are often both not cheap and not easy to use for a technically inexperienced practitioner. Hence we present a setup using a consumer camera with custom firmware to capture videos for later HR extraction from PPGI that is cheap and in principle easy to use. We evaluate the performance of state of the art signal-processing based algorithms on 30 minute therapy sessions of these children.

Finally we can show that HR can be estimated reliably

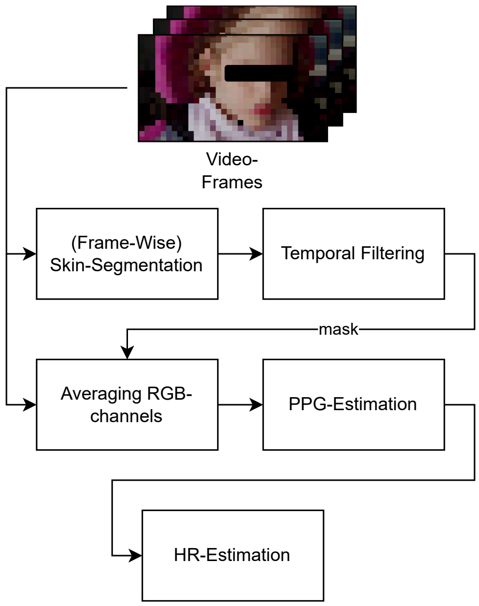

Figure 1. PPGI processing pipeline

unobtrusively in this setting, even under difficult, "natural" illumination.

2. Methods

2.1. Experimental Setup

9 patients with age ranging from 8-15 (light skin) were recorded 5 minutes before, 30 minutes during and 5 minutes after music therapy sessions. During the music therapy sessions the children usually sit upwards in a wheel chair. More details about the music therapy study can be found in [4]. After discarding erroneous recordings (e.g. bad camera configuration such as out of focus video or too much occlusion, no synchronization between camera and patient monitor), recordings from 16 sessions of 6 patients were used for the study in this paper.

The setup consists of a camera (Canon EOS 100D, Magic Lantern)[1] with custom firmware[2] that was directed to the patient with approximately 1 m distance such that the face mostly filled the sensor. Camera-settings, e.g. focus and exposure time, were set to automatic. Common biosignals such as ECG, PPG, spO2 were recorded with a patient monitor (Vitaguard, Vitawin Software) for reference.

2.2. Video Processing

The raw video is debayered and stored as 10bit RGB-video in a lossless format. Then as depicted in Figure 1, Face detection with BlazeFace and Convolutional Neural Network based deeplab [5] skin-segmentation model is used to segment the facial skin frame-wise. After that, temporal filtering of the segmentation is done by using the and-operator on pixels inside a window of length k. The masks created by this segmentation are used to select the pixels that are averaged per channel of each frame. To estimate the PPGI signal from the resulting RGB-signal the following commonly used and robust methods are used: CHROM [6], POS [7] and LGI [8]. We also show results for the green channel (GREEN) as a baseline to highlight influences on the signal quality.

2.3. Evaluation

In order to ensure consistency between the reference and predicted HR, both are computed in the same manner. First, the reference HR is computed by extracting the intervals from a QRS detection and interval correction from the ECG. For estimating PPGI-HR two methods are used for comparison: Spectrogram-based and interval-based estimation. The spectrogram is computed using 30 second windows with 15 second overlap and the HR is estimated using MATLAB's ridge detection on the bandpass filtered signal (cutoff-frequencies 0.4-4 Hz). For the interval estimation, the first guess of the main HR, $\mathrm{HR}_{\mathrm{guess}}$, of a recording is computed as the average HR after ridge detection. The PPGI signal is IIR bandpass filtered around the HR (cutoff frequencies $[0.8\mathrm{HR}_{\mathrm{guess}}, 1.2\mathrm{HR}_{\mathrm{guess}}]$, steepness 0.85) before simple peak-detection was used to estimate the intervals. Based on the intervals from the ECG as well as PPGI the momentary HR is computed and averaged in 30 second windows. The resulting PPGI-signals are evaluated using mean absolute error (MAE) and root mean squared error (RMSE).

The videos show changing light conditions with e.g. sun light that comes through the blinds and produces sharp edges on the face (Figure 6). We specifically investigated the bright and dark/shadowed regions produced by that phenomenon for six selected clips of 30 s duration. Due to the sharp edges it can be assumed that the brightness of the skin follows a bimodal distribution. Therefore we applied a Gaussian mixture model to the segmented skin pixels for each to separate them into bright and dark pixels where a pixel was considered to be inside the group if its brightness v fulfilled $\mu_i - 1.65\sigma_i < v < \mu_i + 1.65\sigma_i$, which equals a 90% confidence interval, for the respective Gaussian with parameters μ_i, σ_i. Inside these regions the signal-to-noise ratio (SNR) as defined in [6] is evaluated.

3. Results

The estimated HR (POS,interval-based) generally follows the reference precisely in low movement scenarios

[1] `https://magiclantern.fm/`
[2] Build: Magic Lantern Nightly.2018Jun06.100D101

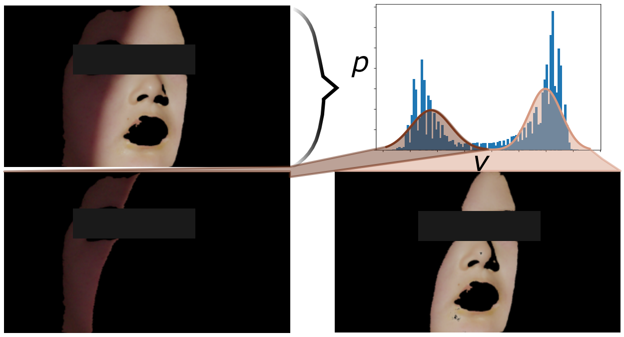

Figure 2. Split regions with sharp edges from illumination before processing based on Gaussian Mixture

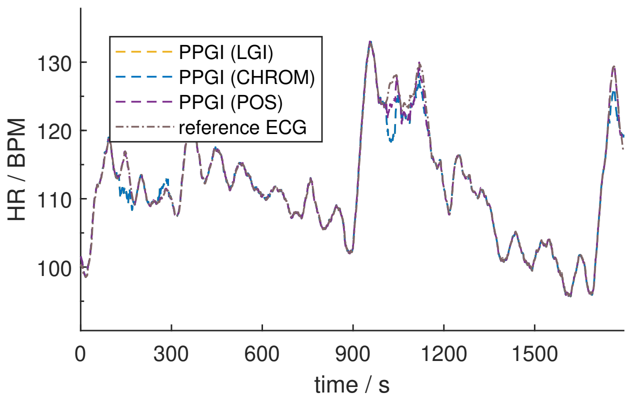

Figure 3. Estimated POS, CHROM and LGI signals of Patient P06 follow reference

(Figure 3 and 4). Whereas the spectrum-based variant is sometimes more robust but most of the times not as close to the reference as the interval-based. We can also see from Figure 5 that CHROM achieves the lowest MAE/RMSE with 0.3/0.7 BPM, while POS and LGI perform better in the median. Using the Green channel alone produces significantly worse results in all cases. The resulting MAE align with the performance of comparable adult databases in Table 1.

With regards to separating dark and bright regions, the 6 selected clips in Figure 6 contain the trend that using only dark regions results in a slightly higher SNR compared to

Table 1. We compare MAE values achieved under similar settings (natural light, rotations) on datasets containing healthy adults PURE [9,10] and UBFC [10,11] as reported by the authors with those achieved on the therapy data.

MAE (BPM)	PURE	UBFC	MUSIC
CHROM	7.29	3.13	6.30
GREEN	8.48	9.32	55.77
LGI	10.61	5.64	4.74
POS	7.29	2.79	3.98

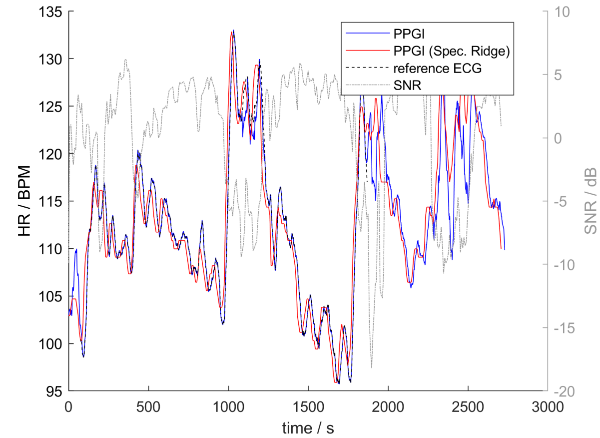

Figure 4. Comparison of interval- and spectrogram-based HR estimation and SNR of PPGI for POS

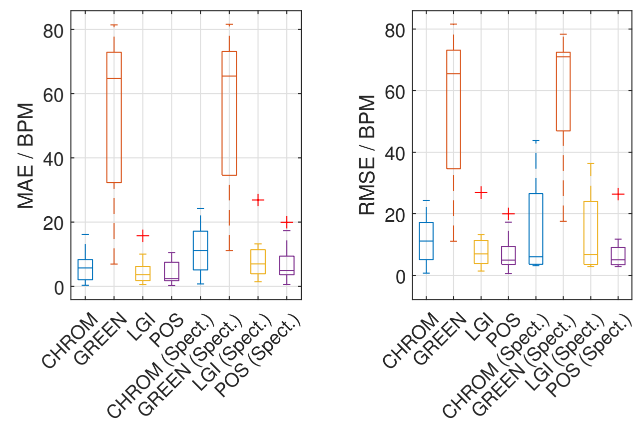

Figure 5. Boxplot comparing four commonly used PPGI methods that use interval-based and spectrogram-based HR estimation

the complete facial skin for POS and CHROM and significantly improves SNR if only the green channel is used. The bright regions generally result in a lower SNR.

4. Discussion

LGI, POS and CHROM yield similar results with a slight edge for POS, which confirms recent analysis by Van Es et al. [12]. Which method performs better depends slightly on the light reflection conditions. We also analysed the effect of the debayering, but the pipeline we present in this study does not seem to yield better results using the raw frames at 10 bit compared to converting and debayering the frames to RGB frames with 8 bit color depth. Only under perfect conditions and carefully selected region, we saw a clean PPGI signal in the green channel. Employing the intervals instead of estimating the HR from

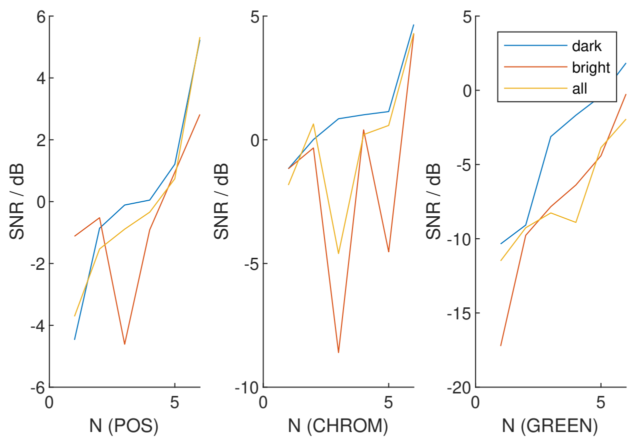

Figure 6. Influence of bright regions with direct sunlight

the spectogram directly resulted in lower MAE and RMSE as well as minimal errors under good conditions. The temporal filtering showed to be only useful for removing jitter from skin masks created based on BlazeFace. In most cases it seems that additional skin pixels are more useful to reduce the noise through averaging than it is important that the respective pixels show the same skin consistently. Originally, we split differently illuminated skin regions in order to facilitate the development of signal fusion algorithms that can improve SNR. However, we found that the darker, less illuminated region, not affected by direct sunlight, shows consistently higher SNR for all methods. The processing with CHROM/POS and their internal normalization obviously helps to reduce the effect of illumination compared to the green channel by filtering the specular reflection. Still, direct sunlight deteriorates signal quality in all cases and skin parts affected can be removed efficiently. Overall the methodology and setup presented in this paper is well suited for estimating HR on ill children over a whole music therapy session.

5. Conclusion

Music therapy is an ideal scenario for the application of PPGI. We have shown that under real-world conditions, the estimation of the average heart-rate of children with severe disabilities is feasible for 30 minute recordings with high variability using commonly used algorithms. Significant deterioration of the PPGI signals often stems from caretaker-child interaction (e.g. hand in front of camera) and ill-configured camera. In future research we will focus on estimating HR variability parameters to facilitate the evaluation of music therapy efficacy in an undisturbed, “natural” setting.

Acknowledgments

We would like to thank the staff and foremost the children of the Theodorus Kinder-Tageshospiz Hamburg for their contribution and support.

References

[1] Orita M, Hayashida N, Shinkawa T, Kudo T, Koga M, Togo M, Katayama S, Hiramatsu K, Mori S, Takamura N. Monitoring the Autonomic Nervous Activity as the Objective Evaluation of Music Therapy for Severely and Multiply Disabled Children. The Tohoku journal of experimental medicine 2012;227(3):185–189.

[2] Kobus S, Diezel M, Dewan MV, Huening B, Dathe AK, Felderhoff-Mueser U, Bruns N. Music Therapy Is Effective during Sleep in Preterm Infants. International journal of environmental research and public health 2021;18(16).

[3] McDuff D. Camera Measurement of Physiological Vital Signs. ACM Computing Surveys 2023;55(9):1–40.

[4] Monika Hoog Antink. Musiktherapie für Kinder mit Komplexer Behinderung: Eine explorative Beobachtungsstudie. Musiktherapeutische Umschau 2022;(43):74–76.

[5] Matthieu Scherpf, Hannes Ernst, Leo Misera, Hagen Malberg, Martin Schmidt. Skin Segmentation for Imaging Photoplethysmography Using a Specialized Deep Learning Approach. In Computing in Cardiology. 2021; .

[6] de Haan G, Jeanne V. Robust Pulse Rate from Chrominance-based rPPG. IEEE transactions on bio medical engineering 2013;60(10):2878–2886.

[7] Wang W, den Brinker AC, Stuijk S, de Haan G. Algorithmic Principles of Remote PPG. IEEE Transactions on Biomedical Engineering 2017;64(7):1479–1491.

[8] Pilz CS, Zaunseder S, Krajewski J, Blazek V. Local group invariance for heart rate estimation from face videos in the wild. In CVPRW. IEEE, 2018; 1335–13358.

[9] Stricker R, Muller S, Gross HM. Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot. In The 23rd IEEE International Symposium on Robot and Human Interactive Communication. 2014; 1056–1062.

[10] Liu X, Zhang X, Narayanswamy G, Zhang Y, Wang Y, Patel S, McDuff D. Deep Physiological Sensing Toolbox. arXiv 2022;.

[11] Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J. Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography. Pattern Recognition Letters 2019; 124:82–90.

[12] van Es VAA, Lopata RG, Scilingo EP, Nardelli M. Contactless Cardiovascular Assessment by Imaging Photoplethysmography: A Comparison with Wearable Monitoring. Sensors 2023;23(3):1505.

Address for correspondence:

Maurice Rohr
Merckstraße 25, 64283 Darmstadt, Germany
rohr@kismed.tu-darmstadt.de