Neonatal Video Database and Annotations for Vital Sign Extraction and Monitoring

Hussein Sharafeddin[1], Lama Charafeddine[2], Jamila Khalaf[1], Ibrahim Kanj[1] and Fadi A Zaraket[2]
[1]*Lebanese University, Beirut, Lebanon*
[2]*American University of Beirut, Beirut, Lebanon*

Keywords: Video Database, Neonatal Monitoring Data Set, Noninvasive Monitoring.

Abstract: **Background:** The end goal of this project is to detect early signs of physiological disorders in term and preterm babies at the Neonatal Intensive Care Unit using real time camera-based non-contact vital signs monitoring technology. The contact sensors technology currently in use might cause stress, pain, and damage to the fragile skin of extremely preterm infants. Realization of the proposed camera based method might complement and eventually replace current technology. Non-invasive early detection of heart rate variability might allow earlier intervention, improve outcome, and decrease hospital stay. This study constructed a curated set of videos annotated with accurate and reliable measurements of the monitored vital parameters such as heart and respiratory rates so that further analysis of the curated data set lead towards the end goal. **Body:** The data collection process included 56 total hours of recording in 127 videos of 27 enlisted neonates. The video annotations include (1) vital signs acquired from bedside patient monitors at second based intervals, (2) the neonate state of health entered and manually reviewed by a healthcare provider, (3) region of interest in video frames for heart rate detection extracted semi-automatically, and (4) the anonymized and clipped region of interest videos. **Conclusion:** The paper presents a curated data set of 127 video recordings of deidentified neonate foreheads annotated with vital signs, and health state in XML format. The paper also presents a utility study that shows accurate results in estimating the heart rate of term and preterm neonates. We hypothesize that the data set we collected is beneficial for improving state of the art monitoring techniques. Its timely dissemination may help lead to techniques that detect anomalies earlier, hence, leading to earlier treatment and improved outcome.

1 INTRODUCTION

Current medical and technological advancements are increasingly leading to higher survival rates of *term and premature neonates*. Of all live births at least ten percent need some intervention after birth to help them adapt to the extra-uterine life and close to one percent will need further management because of different conditions and illnesses. Typically, all infants admitted to the *neonatal intensive care unit* (NICU) require continuous cardiopulmonary monitoring which is essential for early detection of signs of illness and for tracking changes in physiologic state (Liu et al., 2012; Lozano et al., 2012). Since the earliest sign of physiologic disturbances in neonates is a change in heart rate (HR) followed by the respiratory rate (RR), it is important to have a reliable continuous monitoring system that permits early detection and prompt intervention as early as possible to prevent treatment delay and avoid potential morbidity or mortality (Group, 2008).

In neonates, watchful waiting until more obvious signs are revealed may already be too late to achieve a complete recovery. On the other hand, over treatment may be detrimental as it is the case of excessive use of antibiotics that increases the likelihood of selecting resistant bacteria hence rendering future treatments more difficult (Edmond and Zaidi, 2010). Current monitoring tools used in the NICU require direct contact with the patient which might cause harm at times or, at least, might interfere with patient care or parent-infant bonding. This is particularly true for extremely premature infants who have thin gelatinous skin that is prone to sloughing in case of strong bond or erroneous signal due to poor adherence.

For the above reasons, minimally invasive monitoring tools using "non-contact" electrodes have been the subject of extensive analytical and data oriented research (Wu et al., 2012; Poh et al., 2010; Alghoul et al., 2017; Christinaki et al., 2014) that requires

Sharafeddin, H., Charafeddine, L., Khalaf, J., Kanj, I. and Zaraket, F.
Neonatal Video Database and Annotations for Vital Sign Extraction and Monitoring.
DOI: 10.5220/0011637800003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 767-774
ISBN: 978-989-758-626-2; ISSN: 2184-4313

collecting data sets for training, learning and evaluation. We review these techniques and others in the Related Work section 3. In this work, we present a video database annotated with vital signs for neonate infants. The work makes the database available to the research community aims to advance research and development efforts towards non-invasive monitoring techniques for neonates.

2 CONSTRUCTION AND CONTENT

The following describes our work to collect video recordings of neonates while in care and curate them with heart rate, respiratory rate and other health care relevant annotations. The data collection process proceeded as follows.

- We designed our video capture environment with the help of the NICU staff and director. Our target was to allow studying the possibility of establishing norms for each neonate population and to establish research data resources for public access.
- Simultaneously, we designed a method to collect synchronized data from bedside monitors connected to the neonates through electrodes. We used the bedside monitors to obtain automatic heart rate and respiratory rate annotations. We bridged and synchronized the existing bedside monitoring sensory devices with the same machine capturing the video recordings. We were interested in capturing the change in heart rate and heart rate variability (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996; Sztajzel, 2004) as we would like to study their correlation with skin color variations.
- We constructed our annotation terms and notes based on discussions with the NICU healthcare team. The terms concern (i) medical state of the neonate, and (2) technical conditions such as reasons for obstructions to recordings or lighting conditions.
- We drafted a consent form to collect consent from the parents. The form describe to parents the aims of the study and the possible risks involved including how we planned to preserve privacy and anonymity.
- We drafted our design as proposal to the Institutional Review Board (IRB) of the American University of Beirut and obtained their approval to begin the study. Institutional Review Board (IRB) who also approved the sharing of the anonymised videos as an open access repository for further research in this area.

Table 1: Sample demographics.

Variables	Value
Number of infants enrolled	27
Birth weight (g) Mean (±sd)	1302(576)
Gestational age (wk) Mean (±sd)	32.33(4.98)
Total hours recorded	55:34:28
Number of videos	127
Infants with bradycardia events	11
Infants with apnea events	2

- We trained our research assistants on how to approach parents for consent, and how to setup the video recording environment to be compliant with the objective of the studies.
- The research assistants approached the parents for consent and once they received written informed consent from the parents, they started the video recordings.
- To date, we recorded, annotated and analyzed over 55 hours of videos from 27 term and preterm neonates.

We evaluated the utility of our collected data set by running existing techniques (Wu et al., 2012; Alghoul et al., 2017) on the data. Knowing that one of the first signs of physiologic disturbances in newborn is the change in heart rate and heart rate variability (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996; Sztajzel, 2004), we focused on correlating heart rate with skin color changes. The results we obtained are promising and we discuss them further in Section 4.

2.1 Video Database Content

Table 1 shows that the data set contains 56 hours in 127 recordings of 27 enrolled infants. Eleven infants had bradycardia events and two infants had Apnea events.

Table 3 shows the detailed schematic of the annotations. We implemented the annotations in XML format and associated one XML file with each video. The published video files capture the *forehead* region of interest (ROI) that is instrumental in determining the vital signs. Table 2 illustrated statistics of the video recordings.

Table 2: Recording Statistics.

Recording Time (minutes)	Total all videos Overall Average(SD) Average per patient (SD)	3334 26.2(14.1) 120.5(131)
Age at first recording (days)	Min - Max	1 - 185
Recording Place (% time)	Open incubator Closed incubator	61% 39%
Infant status (average % time)	Awake Deep Sleep Light Sleep	31% 8% 61%
Infant motion (average % time)	At rest Minimal Agitated Agitated & Crying	8% 63% 25% 5%
Detection (average % time)	Detected Not detected	81% 19%

Figure 1: RECORDING SETUP: Tripod with camera mounted near incubator.

2.2 Video Acquisition

Video acquisition started November 2016 and continued until April 2018. The videos feature infants admitted to the NICU at the AUBMC after parental written and informed consent. A total of 160 videos were captured using the LogitecC920 high definition camera with 1920×1080 pixel resolution at 30 frames per second. The camera was mounted on a tripod 40 to $60cm$ away from the baby incubator as depicted in Figure 1 showing the camera mounted facing the baby incubator in the NICU. We excluded 33 out of the 160 videos due to technical issues or to failure of forehead detection. The forehead regions extracted from the remaining 127 videos are all admitted to the database with their annotations.

2.3 Data De-Identification

We performed data anonymization to clear patient names, dates of birth and identifying facial features. We replaced patient names and dates of birth with a unique patient ID number and a day index while preserving order, respectively. We only publish the detected forehead region of the videos that is instrumental for vital sign extraction. Consequently, this removes identifying facial features such as the eyes, nose and lips. We reviewed the automatically detected ROI videos and made sure no identifying facial features exist in them.

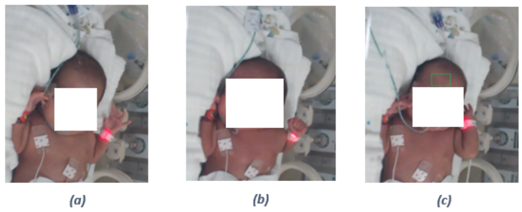

Figure 2: CHALLENGES IN FACE DETECTION AND TRACKING. FRAMES A-C: (a) Patient ID 08, sleeping at rest. Medical equipment occluding face and challenging face detection. (b) Baby moving and turning; challenging continuous ROI localization. (c) face detection and forehead (ROI) localization (note green superimposed box). .

Figure 3: TWO PATIENT SAMPLES EACH REQUIRING ITS SPECIALIZED INITIAL TEMPLATE FOR MATCHING. FRAMES A-D: (a) Patient P09 in video1. (b) Patient P08 in video2. (c) Template for video1. (d) template for video2.

2.4 Forehead Detection

Typically, forehead detection proceeds by face detection followed by facial feature tracking, and then segmenting the forehead portion. However, inter- and intra- video variability in patient appearance and state complicated this task. For example, the presence of medical tape and nasal tubes affected the quality of features necessary for automatic face detection. Hence, forehead detection robustness suffered and we needed semi-automatic detection. Figure 2 shows an example of such challenges for one baby. The baby face is continuously partially occluded with medical dressing. Occasionally more occlusion occurs due to baby face or hand motion. This sometimes leads to lost detection.

We created several image templates that cover different scenarios, and for each video, we manually identified the specific template that best matches the nose and surrounding area. Once there is a match, the forehead region to be extracted is the rectangular region above the template. Figure 3 shows two different patient presentations together with their identified templates and extracted regions.

The forehead tracking proceeded frame-by-frame. Once we matched the template in an initial frame, we automatically track the region across subsequent frames measuring similarity mostly with no need for human intervention. When a significant scene change occurs such as nurse intervention or heavy baby movement, the tracking fails. We remedy this by restarting the forehead detection and skipping frames where detection fails. We manually inspected the videos to make sure that forehead tracking failure did not result in identifying facial features leaking into the published videos.

2.5 Video Organization and Annotation

We organized the videos of each patient into a separate directory. The videos of patient 01 and the associated annotation files go into directory `P01`. Each patient directory contains a number of videos with an associated annotation file in XML format.

File `RecodingList` includes metadata records about all patient videos. Each record has the name of the video file, the number of frames per second, the day of the recording in preserved order, the length of the video, the age of the baby at recording since inception, and the status of the incubator.

It might also have temporal annotations indicating when the baby was in states such as deep sleep, light sleep, agitated, bradycardia, apnea and rest. Other temporal annotations indicate when the region of interest was detected, re-detected or lost.

Each video in a patient directory comes with an associated xml video annotation file. The video annotation file includes two sections. The first section provides metadata about the file including its name, duration, number of frames per second, width and height of the captured region of interest, the status of the incubator, feeding method, whether an event happened during the recording and whether a nutritive feed happened during the recording.

The second section provides a time stamp from the start of recording, the heart and respiratory rates. For this purpose, we configured a dual MIB/RS232 serial cord to extract realtime vital signs from the MP40-70 Philips Intellivue monitor. The second section also has the status of the baby, whether an action happened, and the location of the captured ROI with respect to the video frame. It might also have some notes taken by the healthcare providers. Table 3 describes the entire annotation scheme with examples for each field.

3 RELATED WORK

Recently, several methods for extracting HR data from video recordings of adult subjects have been published. Eulerian Video Magnification (Wu et al., 2012) (EVM) utilizes minute skin color variations and low amplitude motion that are magnified to reveal signals of interest which reflect physiologic changes at the cellular level such as changes in skin perfusion, temperature and heart rate. Poh et al. (Poh et al., 2010) used Blind Source Separation (BSS) based on independent component analysis (ICA) of the red, green, and blue (RGB) intensity channels in facial videos for HR measurement. Alghoul et al. (Alghoul et al., 2017) compared between EVM and ICA and found approaches based on ICA to deal better with lighting-related noise; however, approaches based on EVM performed better with motion-related noise. A comparison of three BSS-ICA based methods is found in (Christinaki et al., 2014) where different statistical transformations aimed at enhancing component separability for extracting the HR.

EVM and BSS only require video recordings from a close distance without any close contact with the patient hence they are non-invasive. In (Wu et al., 2012), colour changes are tracked over time, thus permitting analysis of physiological state changes such as heart rate and subsequently perfusion. Those changes would then be correlated with particular condition or disease states for the purpose of automatic diagnosis and alarm issuance.

4 UTILITY AND DISCUSSION

We performed video analysis to extract vital signs using two methods reported in the literature to validate the sanity of our data set and establish its utility. The video analysis proceeds in two steps: face detection and heart rate extraction.

Face Detection: Baby faces maybe often partially covered with medical equipment and dressing while in the incubator. That led to failure of off-the-shelf face detection algorithms. Thus for each video, and in the first frame, we selected an initial region of interest (ROI) containing the face. We set this as a face template and used ROI tracking to capture it in the rest of the video. This resulted in successful detection throughout the recordings.

HR Extraction: We tested two methods to recover the PPG signal from the detected faces. One is based on analyzing the frequency content of the green channel(Wu et al., 2012), and the other is based on independent component analysis (ICA)(Alghoul et al.,

Table 3: Description of annotation scheme.

Label	Description	Examples
Patient ID	The code given to the patient in chronological order	P15
VideoName	combination of patient ID, date index and video sequence number	P25-Day0505-V25-1.avi
Video porperties	recording parameters	
Frame rate	how many image frames are captured per second (fps) of recording	30 fps
Duration	the video length in h:m:s	1:18:49
Width	frame width in pixels	90
Height	frame height in pixels	50
Recording Info	environment-related information	
Incubator	incubator status being open or closed	Closed
Position	the baby's position being supine or prone	Supine
Non-nutritive feed	whether non-nutritive feeding (pacifier) is being used at recording time	yes
Recording Annotations	Instantaneous heart rate, Status, Action, Detection, Coordinates, Notes	
HR	Instantaneous heart rate extracted from the Intellivue Philips Monitor every second	167
Status	baby's status in three states : • deep sleep (absolutely no motion except for breathing related) • light sleep (possible slow movement of hand, face, etc..) • awake (open eyes, possible movements or interaction with environment)	deep sleep
Action	We identified four states : • rest (usually coinciding with deep sleep) • minimal (light or slow movement of face, limbs, etc.) • agitated (heavy / fast movements) • agitated with crying	agitated
Detection	whether detection has occured at this second which means forehead availability	Detected
X,Y	coordinates of the forehead in the initial frame • x is the top left corner • y is the bottom right corner	759,250
Notes	Any activity observed during recording such as nurse care, changes in ambient light, etc.	nurse care

2017). Both methods aim to extract rate of the cardiovascular pulse wave which circulates throughout the body when the heart beats.

The green channel method proceeds as follows.

a. R=Select ROI

b. Obtain raw signal s from each frame of selection R

c. Extract $\langle g \rangle$ the green channel from s

d. Compute discrete signal $\langle d_g \rangle$ = zero-mean, windowed $\langle g \rangle$.

e. Select the best component $d_{best} = MAXPEAK_{60}^{240} FFT(d_g)$ by applying Fast Fourier Transform (FFT) on on $\langle d_g \rangle$ and choose the resulting FFT component with the highest peak within the range of 60 and 240 beats per minute (bpm).

The ICA method follows the (a,b) steps from above and proceeds as follows.

c. Extract $\langle r,g,b \rangle$ the red, green and blue (RGB) channels from s

d. Compute zero-mean, unit-variance normalized discrete signals $\langle d_r, d_g, d_b \rangle$ from $\langle r,g,b \rangle$.

e. Apply ICA to compute three independent source signals $\langle i_1, i_2, i_3 \rangle = JADE(s,r,g,b)$. The signals break up the raw elements $\langle r,g,b \rangle$ into three independent source signals (i_1, i_2, i_3) using the *joint approximate diagonalization of Eigen* matrices (JADE) algorithm.

f. Apply FFT on i_1, i_2, and i_3.

g. Compute the best FFT component with the highest peak within the 40 and 240 bpm range.

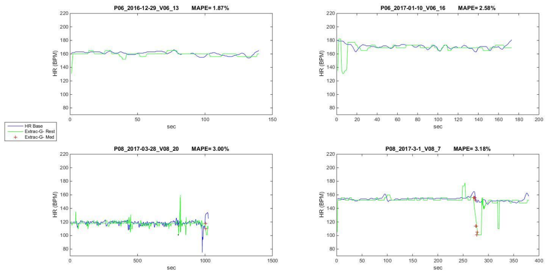

Figure 4: HEART RATE EXTRACTION EXAMPLES: Four examples of HR extraction by the Green Channel method. The title of each subplot has the source video name and the MAPE. The HR monitor output (considered the ground truth) is the blue line. The extracted HR is shown in green when the baby is at ”rest” and in red plus when the baby is ”agitated” - as defined in Table 3.

Results: Both methods were tested on a sample of 10 recordings and gave acceptable heart rate estimates. We compared the estimated heart rate with that of the baby monitor, considered as ground truth, on a second-by-second basis. The sample had a mean absolute percentage error (MAPE) range of 3% to 8% and standard deviation of 6% to 8% with increased error during agitated (medium) baby motion. Whenever our algorithm could not detect the baby face, we considered that the signal has been lost. This usually happens when the baby turns more than 45 degrees away from the camera or when there is total occlusion.

Figure 4 shows four sample results from the green channel method with MAPE between 1.87% and 3.18%. The red crosses are the estimated HR at instances when the baby is mildly agitated. The blue and the green lines show the heart rates from the HR monitor (the ground truth) and from the green channel method respectively.

Figure 5 shows a sample comparison between the green channel and the ICA methods. The mean absolute percentage error for the green channel method (MAPEG) is 2.1%, slightly less that that of the ICA method (MAPEICA), 3.33%. We note that the challenge in the green channel method is its sensitivity to the bandwidth selection to which the green signal

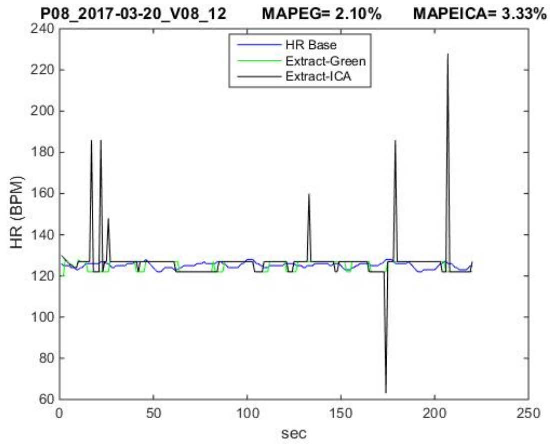

Figure 5: GREEN CHANNEL VS ICA HR EXTRACTION: An example juxtaposing the ground truth HR (blue line) from the attached monitor with both HR extraction methods, the Green Channel (green line) and the ICA method (black line). The mean absolute percentage error for the green channel method (MAPEG) is slightly less that that of the ICA method (MAPEICA). In this recording of abour 3 minutes the baby is at ”rest” (see Table 3).

5 CONCLUSION

Automated early detection of physiological disturbances in newborns is essential for initiating therapy as soon as possible. A contactless and non-invasive system would be particularly helpful in vulnerable populations such as sick neonates. This is a cross sectional, observational study where all infants admitted to the NICU are eligible. After parent consent, 127 video recordings of infants (total of 56hours of recording) were obtained and analyzed. To establish the utility of the data set, the study performed blind source separation (BSS) for detecting HR and HR variability of infants admitted to the intensive care unit. The estimated HR was compared to the “ground truth” values of the regular monitors used in the NICU. The testing results showed feasible heart rate measurements on premature babies. There was an average absolute error range of 3% to 8% and standard deviation of 6% to 8% with increased error during baby motion. As a conclusion, the proposed method proves the utility of the dataset for vital sign extraction. This technique proved beneficial for patient monitoring and we hope further work presents techniques for early detection of diseases, leading to earlier treatment and possibly improved outcome.

6 LIST OF ABBREVIATIONS

- **NICU:** Neonatal intensive care unit
- **HR:** Heart rate
- **BSS:** Blind source separation
- **AUBMC** American University of Beirut Medical Center
- **ROI:** Region of interest

7 DECLARATIONS

Ethics Approval and Consent to Participate

All research procedures have been reviewed and approved by the institutional review board (IRB) at the American University of Beirut Medical Center (AUBMC). We obtained *written* consent statements from the investigators and the participants. The investigator and participant consent statements follow.

Investigator's Statement

I have reviewed and explained in detail, the informed consent document for this research study with [name of participant, legal representative or guardian/ parent if the participant is a minor or is unable to sign] the nature and purpose of the study and its risks and benefits. I have answered to all his/her questions clearly to the best of my ability. I will inform the participant in case of any changes of this study or its negative impacts or benefits in the event of their occurrence.

Consent to Participate

I have read this consent form and understood its content. All my questions have been answered. Accordingly, I agree, with my own free will, to be part of this research study, and I know that I can contact Dr. Lama Charafeddine at 01-350000 Ext: 5874 or any of her designee/assistant involved in the study in case of any questions. If I feel that my questions have not been answered or need further clarification I can contact one of the members of the Institutional Review Board for human rights or its chair Dr. Fouad Zyadeh at 01-350000 Ext: 5445.

I understand that I am free to withdraw my consent from this study and discontinue my participation at any time, even after signing this form without prejudice to the medical care provided to me. I know that I will receive a copy of this signed informed consent.

- I agree to participate in this study and authorize the investigator and her designee the access to my child's AUBMC medical records
- I authorize the investigator and her designee to contact me at a later stage for future follow-up if needed.

Consent for Publication

We have obtained consent to publish de-identified and anonymized versions of the data. We are only sharing publicly areas of foreheads which do not include any identification features.

Availability of Data and Material

The dataset supporting the conclusions of this article is available in the American University of Beirut research repository via the link: http://research-fadi.aub.edu.lb/neonates/home.html.

Funding

Authors (LC, HS, and FZ) used Lebanese National Council for Scientific Research Awards and Lebanese Univeristy Research Funding Awards to hire research assistants to complete the work. The funders did not intervene in the process of the research work.

Authors' Contributions

Authors JK and IK helped define and implement the curation process. Authors HS and FZ helped define and implement the curation process and the signal processing process. Author LC defined and implemented the clinical aspects and processes. All authors contributed equally to the ideation and organization of the manuscript, revised and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the baby parents for their consents in recording their newborns and recording them on video, the engineers at the AUBMC for their cooperation and help in setting up the medical monitor equipment for data extraction and the nurses for their patience and cooperation in the data collection process.

REFERENCES

Alghoul, K., Alharthi, S., Al Osman, H., and El Saddik, A. (2017). Heart rate variability extraction from videos signals: Ica vs. evm comparison. *IEEE Access*, 5:4711–4719.

Christinaki, E., Giannakakis, G., Chiarugi, F., Pediaditis, M., Iatraki, G., Manousos, D., Marias, K., and Tsiknakis, M. (2014). Comparison of blind source separation algorithms for optical heart rate monitoring. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pages 339–342.

Edmond, K. and Zaidi, A. (2010). New approaches to preventing, diagnosing, and treating neonatal sepsis. *PLOS Medicine*, 7(3):1–8.

Group, Y. I. C. S. S. (2008). Clinical signs that predict severe illness in children under age 2 months: a multicentre study. *The Lancet*, 371(9607):135–142.

Liu, L., Johnson, H. L., et al. (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, 379(9832):2151–2161.

Lozano, R., Naghavi, M., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380(9859):2095–2128.

Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774.

Sztajzel, J. (2004). Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly*, pages 514–522.

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology (1996). Heart rate variability. task force of the european society of cardiology the north american society of pacing electrophysiology. *Circulation*, 93(5):1043–1065.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.*, 31(4):65:1–65:8.