Amplification of Heart Rate in Multi-Subject Videos

Matt Estrada
Department of Mechanical Engineering
Stanford University
Stanford, California 94305
Email: estrada1@stanford.edu

Amanda Stowers
Department of Mechanical Engineering
Stanford University
Stanford, California 94305
Email: astowers@stanford.edu

***Abstract*—The goal of this project is to create a visualization of heart rate in videos with multiple subjects. Previous work [1] has created visualizations of heart rate for single subject videos. Other work has automatically detected heart rate from recordings of single faces [2]. Here we attempt to combine these into the ability to detect multiple faces and amplify them individually around a narrow range to emphasize each individuals heart rate, and then combine the videos back together to produce a visualization where relative heart rates are visible.**

I. Introduction

To visualize pulse in subjects, some researchers have applied Eulerian Video Magnification [1]. This technique amplifies color differences within a particular frequency spectrum to enable visualization of otherwise imperceptible features. This works well on a single subject with a known heart rate. However, when applied across a broad range of frequencies this results in significant noise in the rest of the spectrum. This is particularly detrimental when you either do not know the original heart rate, or have multiple subjects with different heart rates present in your video.

People have also used multiple methods to identify heart rate within the videos. One method is by performing principle component analysis on the motions of the head [3], which was shown to work with detection of 0-3.4 bpm on a set of 18 subjects. Other people have used similar methods to extract out principle components and correlate them with physiological signals [4]. The method we are using here is based off a slightly different method involving detecting the variations in color over time and extracting a pulse out of that [2].

Here we aim to combine detection of heart rate with amplifying it for visualization. To do this we first locate the subjects of the video, then split the video into subvideos for each subject and detect their individual heart rate. Then we amplify the desired frequency range in each video and recombine the videos to form a multi-frequency amplified result.

II. Face Identification and Tracking

To identify individual pulse rates, we first identify and track the faces between frames. MATLAB has a built in face detector in the computer vision library which we initially used. However, this proved to be insufficiently accurate to distinguish the face from the background, especially since it would only report a rectangle bounding box around the face. Therefore, we decided to use this as our initial guess as to where the face was in the first frame and then search a subsection of the video corresponding to an enlarged version of that box in other frames. The bounding box has sides 60% larger than the originally reported box and is centered at the same location.

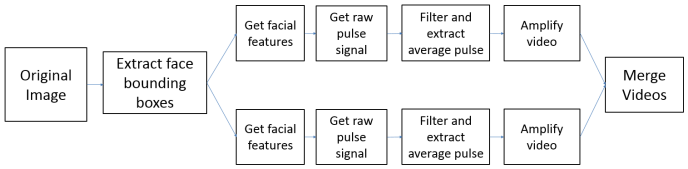

Figure 1. Flowchart of processing for each video. Videos are split up into subvideos for each face and then recombined at the end to form a multi-frequency amplified video.

A. Face Keypoint Detection

For greater accuracy, we decided to track the face itself. We chose an algorithm which would deliver 66 points around the face, each with a unique ID [5]. For these points, we used 17 of them which defined the boundary along the chin and lower face in order to create a convex hull corresponding to a relatively flat area of the face.

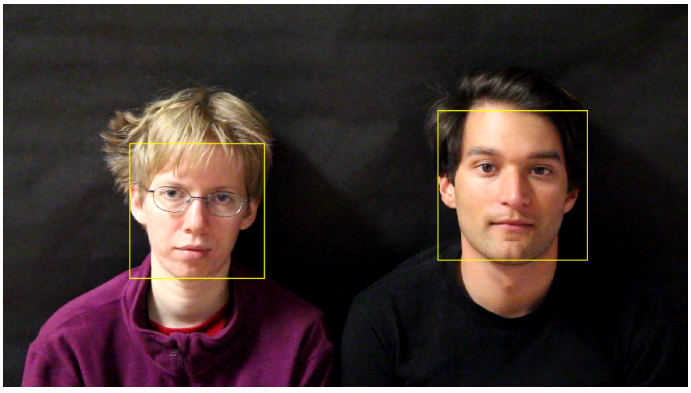

Figure 2. Faces detected by Matlab recognition from first frame of video

B. Interframe Face Tracking

The algorithm could return an estimate of the face position in each frame [5]. However, it would occasionally get the

face wrong which would lead to large undesirable spikes in our heart rate signal (Figure 3). Therefore, during the initial video processing we recorded the values of each of the 17 facial keypoints we were interested in. We then median filtered these points and passed them into another pass through of the video in order to get a better estimation of the face position. This eliminated errors where a single frame had an incorrect face recognition while still giving us fairly accurate face and background areas to use for heart rate detection.

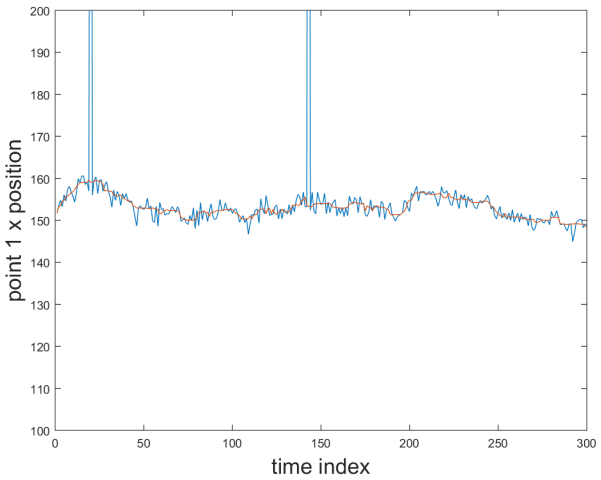

Figure 3. Sample trace of point location over time. Misdetections are seen as large spikes in the data. Median filtering removes the large spikes and also smooths the curve helping to eliminate high frequency noise.

III. Heart Rate Detection and Filtering

To detect the heart rate, we divided each subvideo up into three parts - the face region, the background region, and a neutral unused region around the edge of the face. For each of the face and background region, we computed the average green value [6] in order to extract the pulse from it and compare them with the background. We chose green because there is the most variation in it for heart rate because of properties of hemoglobin [6]. We then applied filtering to the face and background signals in order to extract a heart rate estimate.

A. Region Segmentation

For each video, we assign a region to each face as described above. These subvideos are allowed to interact, however the algorithm prefers they do not so as to avoid conflating two individuals pulse signals. Within that subvideo, SV, we assign a face area, a background area and a neutral area based off the tracked facial keypoints. To get the raw pulse signal, p, we use the convex hull, F, and erode it by a structuring element, SE, which is a disk with a radius of 5% of the original bounding box size. We then mask the face with this hull and take the mean of the values within the masked area as the raw pulse signal. An improved algorithm could also remove the forehead from the background signal, perhaps by fitting an ellipsoid through the chin points.

$$p_{raw} = mean(mask(erode(F, SE), SV)) \tag{1}$$

To get the background signal bg, we take the convex hull and dilate it with the same structuring element. We then mask the image with the inverse of this and take the mean of that area for the background.

$$bg = mean(mask(1 - dilate(F, SE), SV)) \tag{2}$$

This gives us an initial estimate of the pulse signal as well as of the amount that signal would change based just off lighting variations in the room. We discard the neutral area as it contains both foreground and background signals.

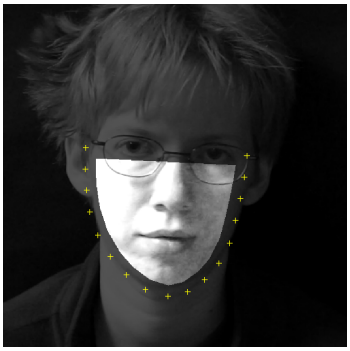

Figure 4. Sample subvideo wiht detected keypoints and convex hull for extracting signal

B. Filtering to Approximate Pulse Signal

The signals that we receive from the mean values of the segmented areas are still very noisy. To better estimate the pulse we first subtract off the background from the face area. This allows us to better remove variations that are due to things other than heart rate such as lighting variations. To do this, we create a new signal

$$p_{nobg} = p_{raw} - h * bg \tag{3}$$

where h is a constant determined iteratively from the two signals in order to create the best value for subtraction based off the combination of the setting and the subject's facial characteristics [2]. Specifically, we assign an initial value of h and then update it according to the following equation (from [2])

$$h(n+1) = h(n) + \mu * \frac{(p_{raw} - h(n) * bg) * bg}{bg^H * bg} \tag{4}$$

where μ is a parameter controlling rate of convergence which we selected to give convergence within approximately 10 iterations.

After adjusting based off the background, we then fit a low order polynomial to the data and then subtracted off that trendline in order to try to correct for extremely low varying

effects of lighting. We tried polynomials of degrees 1-3, with there being slight improvement with higher order polynomials, so we kept the third degree polynomial (P, computed with MATLAB's polyfit command) for the final setup.

$$p_{detrend} = p_{nobg} - polyval(P, p_{no_bg}) \quad (5)$$

After correcting for slow varying trends, we used a Butterworth filter to filter out frequencies which would not be able to correspond to generally plausible human heart rates. Specifically, we had a band pass filter from 0.4 to 4 Hz, corresponding to 24-240 bpm. With the exception of some rare disorders that allow higher heart rates, this should allow most human heart rates to pass through. For 20-39 year olds, 98% of resting heart rates fall between 48 bpm and 103 bpm [7], so this is a sufficiently wide range.

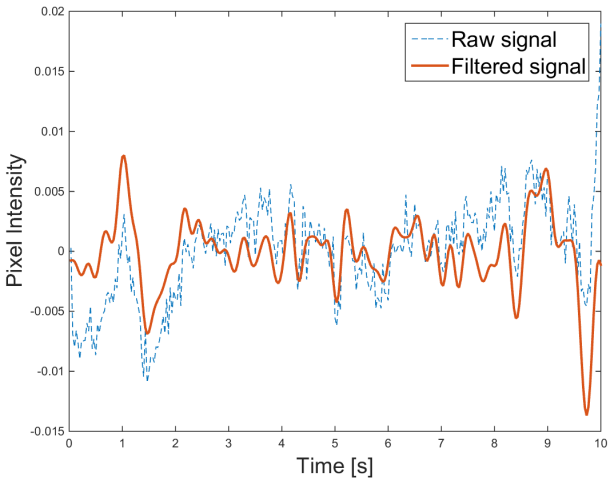

Figure 5. Raw and filtered signals used for heartbeat recognition

C. Estimation of Primary Heart Rate

Human heart rates do not actually proceed exactly as a sine wave with a single frequency. Even accounting for their non-sinusoidal shapes, adjacent heartbeats are not necessarily identical lengths. A measure of this is referred to as the heart rate variability, and for healthy adults tends to have a peak frequency of 0.1-0.4 Hz [8], tending towards the lower side. Therefore, a difference of 6-24 bpm might be created during adjacent heartbeats. This is solved by averaging over several beats. However, our total video length is only approximately 10 seconds due to processing time and so there can still be a fair bit of variation. This is especially true in videos where we had to get up in order to turn on the camera and then sit back down.

To compute the heart rate, we took the Fourier transform of the final filtered pulse signal. The heart rate we chose was the maximum peak in the desired range (60-120 bpm). However, there were often multiple peaks in the area and the highest was not necessarily the one which corresponded to heart rate.This produced a spectrum from which we could extract a primary

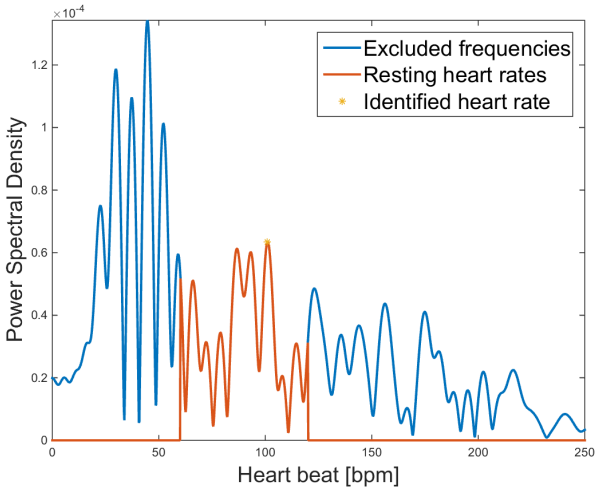

Figure 6. Fourier transform of the heartbeat signal previously shown. Excluded and allowed heartbeat ranges are marked.

frequency. We extracted only frequencies lying between 60 and 120 beats per minute in order to minimize errors and corroborate with most adult resting heart rates (approx. 10th to 99th percentiles [7]). This also helped in removing breathing rates, which fall in a similar spectrum to heart rates in terms of order of magnitude, and can alias in into the heart rate signal itself. Two potentially relevant signals for interference are blinking (about 17 bpm and its aliases [9]) breathing (about 8-16 bpm [10]). These alias into the signal and can cause later peaks at multiples of their frequency.

As an example, we tested signals consisting of sinusoids at a breathing and heart rate corresponding to reasonable values (breathing rate 10 bpm, heart rate 75 bpm), with the breathing rate being approximately 10x stronger of a signal than the heart rate (this appeared to give the closest comparison when visually inspecting the signals and power spectrums). This indicates that to get better results we need to either better eliminate the breathing rate signal by tracking better, reduce the noise, or generate an estimate of the breathing rate and filter that out (or have people hold their breath during the video as some authors have done). Even with a bandpass filter around the breathing rate, this still did not provide acceptable results (Figure7).

We also tried using the Welch's power spectral density, but this did not produce good results and was unable to define heart rates beyond a very coarse estimation. The coarseness was primarily due to only having 10 seconds or 300 frames in each video.

Unfortunately our actual heart rate results were not as accurate as we had hoped (Figure 8). For several videos, we generated both the heart rate estimated by the video as well as the heart rate measured with a pulse oximeter which we took as 'ground truth' data. It should be noted that the oximeter was a $20 one off Amazon and its calibration was unknown. Also, during most of the videos the oximeter value varied by

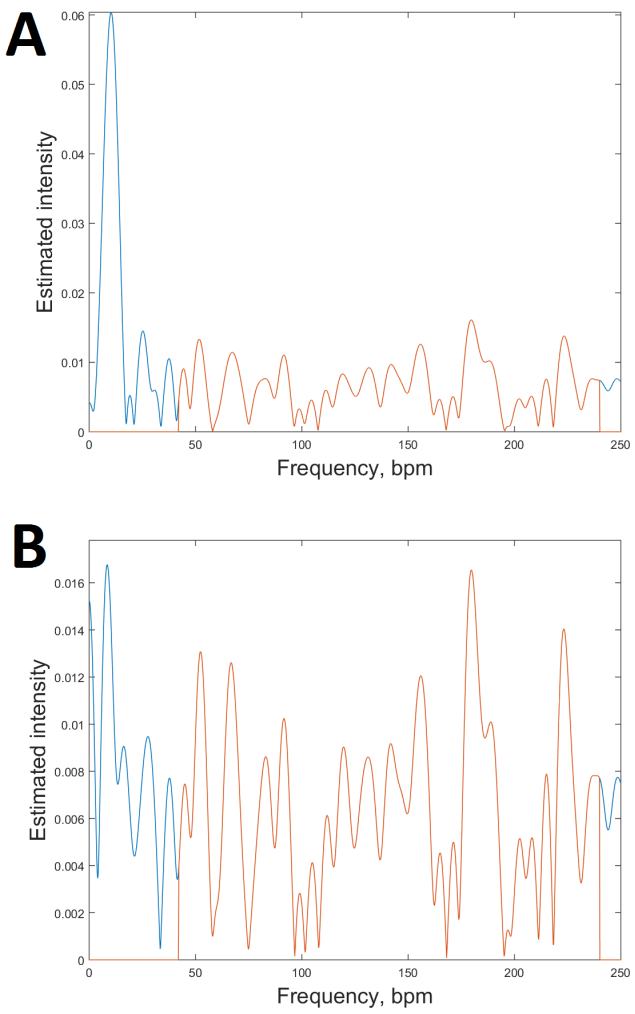

Figure 7. Sample generated heart rate+breathing rate signal with noise added and its power spectrum (A). With the breathing rate bandpassed out (B), it is still seen that there is significant noise and the dominant frequency is not the heart rate frequency, which should be 75 bpm.

Table I
CORRELATION COEFFICIENTS FOR ALL VIDEOS AND SEPARATED BY VIDEO SUBJECT

Video Set	Correlation Coefficient
All	0.13
Matt	-0.29
Amanda	0.56

multiple beats per minute (mean 5, median 4.5 bpm). Our correlation values for both of us together and each invidually are listed in Table I.

Overall we had a slight positive correlation. We did detect who had the higher heart rate in each video, which is useful for amplification. Compared to previous studies, our results are not too surprising - on databases which would be similar to ours, correlation coefficients ranged from 0.08 to 0.81 in one study implementing several algorithms [5]. However, with controlled environments (varying from lighting controls to head steadiness to breath holding) some of these algorithms performed with correlation coefficients of 0.98 to 0.99. We implemented a variation on the algorithm which performed best in their study, but did not have as successful of results. We also have a fairly small sample size (8 detections for each person) due to time constraints.

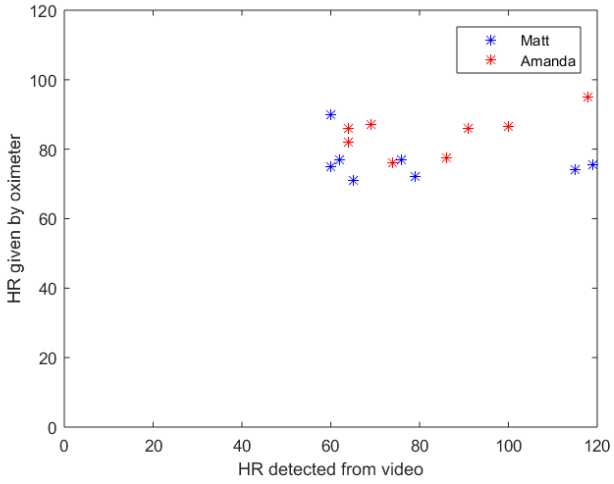

Figure 8. Scatterplot of heartrates based off detection in video compared to measurement with pulse oximeter. Different colors denote different people.

The recordings of Amanda had somewhat better predictive value than those from Matt (whose recordings actually had negative correlation coefficient). This probably has to do with a combination of her heart rate varying more over the course of the experiments and differences in skin color. We did try to improve videos by eliminating background variation by using a plain black background, but this did not really provide any benefit over using any constant background.

IV. VIDEO MAGNIFICATION

To visualize the effects of changing heart rate, we used a technique called Eulerian Video Magnification [1]. This technique had been used before to amplify heart rate in videos where the heart rate was already known - the demo video they have includes a subject where they magnify only a 10 bpm range. Larger ranges lead to more noise, so we used our heart rate detection in order to create a more narrow range for each subvideo where a face was located.

The way this technique works is by taking how much each pixel has changed in color compared to the original input pixel and amplifying this with a frequency bounded amplification. This is Eulerian because it does not attempt to track individual features, but instead amplifies everything which comes through a given point in space. By adjusting the amount of magnification desired as well as the frequencies chosen, videos can be made which amplify relevant effects.

A drawback of this method is that it is trying to amplify color changes regardless of their source. When we use high amounts of magnification this leads to magnifying noise and

compression artifacts making stray colors appear where we did not intend them to. The magnification was somewhat clearer on the darker face in each image as it was less obvious when colors were being added than in the lighter face. Generally we chose the same magnification amount for each video and left only the frequency detection to be done automatically, although future work could included determining the maximum magnification reasonable before the video would become visually unpleasant.

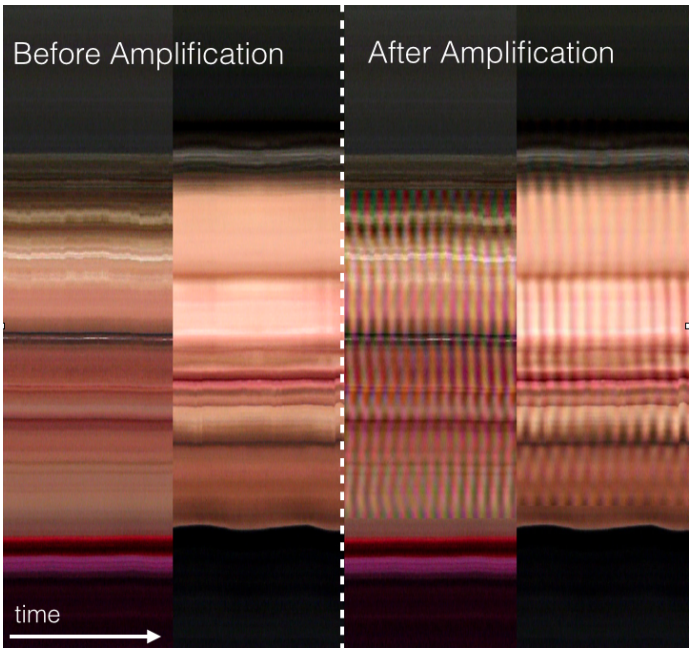

Figure 9. Visual representation of video amplification for two subjects. A column of pixels is taken at each frame and plotted left to right. Fluctuations can be seen with the human eye after processing.

V. Combining the Videos

After each segment around an individual's face was amplified, the video was simply stitched together by replacing the pixels from the original video with each sub-video. Since each person remained relatively stationary and the background was a uniform black, we implemented no blending at the boundary of each stitched video. The most visible artifacts can be seen at the neck of each person, where amplified and non-amplified skin come together.

These effects could be mitigated with more advanced stitching algorithms such as blurring the edges together or other methods of blending.

VI. Conclusion and Future Work

The most difficult part of extracting heartbeats proved to be discerning the actual signal from other forms of noise. We suspect other physiological signals present within roughly the same frequency played a large part in this. Most notably, our breathing and blinking both fell within these ranges, especially when aliased. By eliminating potential noise sources, we could remedy each accordingly and move to more uncontrolled situations.

Figure 10. Stitched video frame. Around the lower neck, combination artifacts are visible. They are also slightly visible on the left hand size of the hair where the hair was not all in the magnification subvideo.

Towards practical applications, more work could be done to explore the extent to which this could be applied to larger crowds. Investigating the ability to perform these operations on a surveillance camera would be of great interest. Specifically, we would explore how performance drops off with lowering resolution, or bits of information per pixel. Additional challenges would be to explore the functionality of algorithms on partially occluded faces.

Finally, concerning visualization, we feel that other forms of amplification could be explored. Since the MIT Eulerian algorithm amplifies both color and displacement variations, other methods may prove to be more intuitive for conveying pulse. For instance, changing the coloration or contrast of an individual's face could work.

Acknowledgment

The authors would like to thank the teaching Staff for EE368 since every member provided useful feedback and encouragement on our project. Additionally, our team mentor Matt Yu, who was always available to meet and consult when needed.

We also acknowledge use of the MIT Eulerian Video Magnification code and of a DRMF based face detection code.

Appendix

A. Videos

Manual heart rate ID (this was taken as a best case scenario with ground truth heart rates from the oximeter put in): https://www.youtube.com/watch?v=55-Lco4cwGg

Automatic heart rate ID (our experimental results): https://www.youtube.com/watch?v=k-CsB41JSZI

B. Work Distribution

Coding was done jointly between partners, with Amanda leading. The report and poster were written together.

REFERENCES

[1] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world." *ACM Trans. Graph.*, vol. 31, no. 4, p. 65, 2012.

[2] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 4264–4271.

[3] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3430–3437.

[4] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.

[5] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3444–3451.

[6] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[7] R. Yechiam Ostchega, Ph.D., M. Kathryn S. Porter, M.D., M. Jeffery Hughes, P. Charles F. Dillon, M.D., and D. Tatiana Nwankwo, M.S., "Resting pulse rate reference data for children, adolescents, and adults: United states, 1999 to 2008," National Health Statistics Report, Tech. Rep., Aug. 2011.

[8] G. G. Berntson, J. T. Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone *et al.*, "Heart rate variability: origins, methods, and interpretive caveats." *Psychophysiology*, no. 34, pp. 623–48, 1997.

[9] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of blink rate patterns in normal subjects," *Movement Disorders*, vol. 12, no. 6, pp. 1028–1034, 1997.

[10] D. C. Dugdale. (2013) Rapid shallow breathing. [Online]. Available: http://www.nlm.nih.gov/medlineplus/ency/article/007198.htm