Effect of the Initial Condition on SCG Clustering using Unsupervised Machine Learning

S. Ahdy[1,3], T. Hassan[1], B. Rahman[1,4], R. Sandler[1,2] and H. Mansy[1,2]

1. Biomedical Acoustic Research Lab, University of Central Florida, Orlando, Florida, USA
2. Biomedical Acoustic Research Company, Orlando, Florida, USA
3. Mechanical Power Engineering Department, Zagazig University, Zagazig, Egypt
4. College of Medicine, Yale University, New Haven, Connecticut, USA

sherif.farahat@ucf.edu, tanhassan31@gmail.com, badrun.rahman@yale.edu, rhsandler@gmail.com, Hansen.Mansy@ucf.edu

INTRODUCTION

Seismocardiographic (SCG) signals refer to chest surface vibrations induced by the cardiac cycle [1, 2]. SCG may be generated by valve closure, blood movement and cardiac muscle contraction [3-5]. SCG may serve as a non-invasive clinical tool to detect heart disease [6] and sleep apnea [7] and may be used to determine cardiac time intervals [8]. However, SCG signal variability can hinder accurate extraction of clinically useful SCG features [9], such as cardiac time intervals [10]. SCG variability can be reduced by clustering SCG beats [1, 9]. Respiration was found to be a source of SCG variability [1, 11]. For example, respiration changes the heart rate (a phenomenon known as respiratory sinus arrhythmia) which affects the length of the cardiac cycle [1, 12]. Respiration can also modulate the cardiac preload and afterload which can affect the intensity of heart sounds and cardiac time intervals [11, 13, 14]. Therefore, early efforts of SCG clustering grouped SCG beats based on the respiratory phase, e.g., high/low lung volume (HLV/LLV) or inspiration/expiration (INS/EXP) [15]. In recent studies, unsupervised machine learning (e.g., K-medoid) was used as it does not require any prior assumptions about SCG clusters [9, 16]. In the K-medoid algorithm, cluster medoids are initially chosen either randomly [9] or based on a physiological basis (e.g., HLV/LLV or INS/EXP) [1], then SCG beats are assigned to closest medoid, and cluster medoids are updated; the process is repeated until cluster assignments converge (i.e., stop changing) [1]. The quality of the clustering solution (i.e., final cluster assignment) can be assessed by calculating the intra-cluster and inter-cluster variabilities [1, 16]. However, based on the choice of the initial medoids and cluster assignment (i.e., the initial condition), different clustering solutions could be obtained [9].

The effect of the initial condition on the clustering solution has not been sufficiently studied. There is one recent study that addressed it, by considering only HLV/LLV and INS/EXP [1]. However, these two initial conditions may not guarantee that the optimum clustering solution is attained as suggested by the current study results. The situation might be even worse when random initial conditions are employed. Therefore, the objective of the current research is to study the effect of the initial condition on the clustering solution and recommend an efficient procedure for finding the optimum clustering solution.

METHODS

After Institutional Review Board (IRB) approval, electrocardiogram (ECG) and SCG were collected from seven healthy adult subjects for one minute. The sampling rate was 500 Hz. SCG (dorsoventral component) was acquired near the left lower sternal border by a uniaxial accelerometer attached at the fourth intercostal space (ICS). The baseline wander of SCG (0.15-0.5 Hz) was integrated once and twice to get a measure of the low-frequency chest surface velocity and displacement, which would correlate with the breathing flow rate and lung volume signals, respectively. SCG recording was segmented into beats; each beat started 100 ms prior to the timing of the ECG R wave. SCG beats were clustered using the K-Medoid algorithm [9, 17]. Two clusters were chosen following recommendations of previous studies [9]. The dissimilarity (i.e., the distance) between SCG beats was calculated using dynamic time warping (DTW) method [1]. All possible combinations of initial medoid pairs (from a data subset) were considered; the data subset size was varied from 5% to 100% of the total number of beats. Choosing a smaller percentage of the data, as initial conditions, helps reduce processing time. The data included in the subset was not chosen randomly. Rather, we avoided including nearby beats (as these may have similar morphology) when forming the subset. For example, when the subset has 10% of the total data, beats #1, #11, and #21, etc. were used to build the subset (where all beat combinations were used as initial conditions). For each initial condition, the final cluster assignment quality was assessed by calculating the intra-cluster variability and inter-cluster variability, Equations 1 and 2 respectively [18]. In Equations 1 and 2, dtw, C_j, x_{i_j}, and n_j refer to the DTW distance, the j[th] cluster medoid, the i[th] beat in the j[th] cluster, and the number of beats in the j[th] cluster, respectively.

$$Intra-cluster\ Variability = \frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1} dtw\left(C_1, x_{i_1}\right) + \sum_{j=1}^{n_2} dtw\left(C_2, x_{j_2}\right)\right) \quad (1)$$

$$Inter-cluster\ Variability = \frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1} dtw\left(C_2, x_{i_1}\right) + \sum_{j=1}^{n_2} dtw\left(C_1, x_{j_2}\right)\right) \quad (2)$$

Results and Discussion

Typical clustering solutions of three different randomly chosen initial conditions are shown in Figure 1. Here the coordinates are the normalized lung volume and breathing flow rate. The dots represent SCG beats at the ECG R wave timing. The two clusters were colored in red and blue. The two numbers (V1 and V2) shown above each subplot in Figure 1 refer to the intra-cluster variability and inter-cluster variability, respectively. Figure 1 shows how the choice of the initial condition can lead to different clustering solutions, including solutions with sparse clusters as in the rightmost subplot in Figure 1. In the current study, the optimum clustering solution was defined as the one having the lowest intra-cluster to inter-cluster variability ratio and not having a sparse cluster.

For a data subset size from 5% to100%, it was found that many different initial conditions resulted in the same clustering solution. Therefore, it was possible to attain the optimum clustering solution while using a data subset when choosing initial conditions. Testing all possible initial conditions from 100% of the data can be time consuming especially when analyzing long recordings. It was found, in the current study, that choosing the initial medoids from a subset of SCG beats (10% in six subjects and 50% in one subject) leads to solutions that included the optimum clustering solution. The study also considered lung volume and flow rate based initial conditions (i.e., HLV/LLV and INS/EXP respectively). The results showed that these two choices do not always converge into the optimal clustering solution. In three of the study subjects, the initial condition based on the flow rate resulted in the optimum solution. The lung volume-based initial condition resulted in the optimum solution in three subjects and in a clustering solution that had a sparse cluster in one subject.

Conclusions

The current study showed that the initial condition is determinantal to the quality of SCG clustering solution and described a procedure to attain the optimum clustering solution by considering all possible initial conditions from a data subset. Optimum clustering will help reduce variability and extract accurate and reliable waveform features. This may leverage the clinical utility of SCG in monitoring heart failure patients and other cardiac conditions. Future work will include analyzing more subjects, diverse ethnicity and longer recordings, to confirm the study findings, and testing other techniques for selecting the initial condition.

Acknowledgments

This publication is based on work supported by the National Science Foundation under Grant No. FW-HTF-P 2026516. The research reported in this publication was also supported by the National Heart and Lung Institute of the National Institutes of Health under award number R44 HL099053. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or National Institute of Health.

COI statement: Richard H. Sandler and Hansen A. Mansy are part owners of Biomedical Acoustics Research Company, which is the primary recipient of the NIH grant R44HL099053, as such they may benefit financially because of the outcomes of the research work reported in this publication.

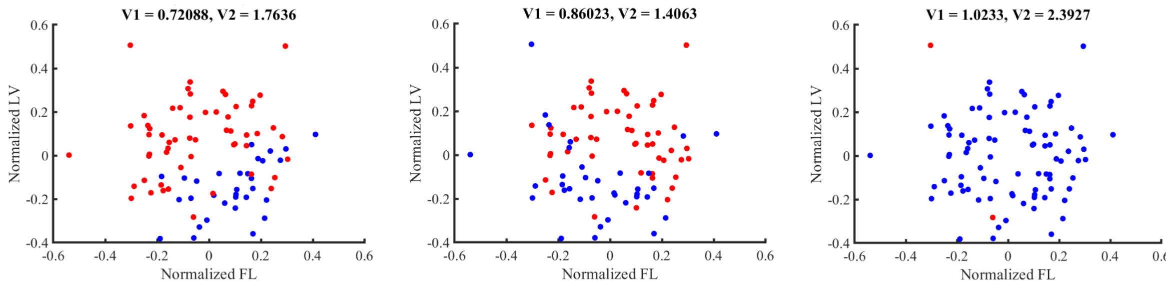

Figure 1. Samples of different clustering solutions obtained after implementing the K-Medoid algorithm using different initial conditions. The dots represent SCG beats at the ECG R wave timing. The two clusters are colored in red and blue. The two numbers (V1 and V2) shown above each subplot refer to the intra-cluster variability and inter-cluster variability respectively.

REFERENCES

[1] S. Ahdy, M. K. Azad, R. H. Sandler, N. Raval, and H. A. Mansy, "Spatial Distribution of Seismocardiographic Signal Clustering," in *Signal Processing in Medicine and Biology: Innovations in Big Data Processing*, I. Obeid, J. Picone, and I. Selesnick Eds. Cham: Springer International Publishing, 2023, pp. 87-119.

[2] A. Taebi, B. E. Solar, A. J. Bomar, R. H. Sandler, and H. A. Mansy, "Recent Advances in Seismocardiography," *Vibration,* vol. 2, no. 1, pp. 64-86, January 14, 2019.

[3] R. S. Crow, P. Hannan, D. Jacobs, L. Hedquist, and D. M. Salerno, "Relationship between Seismocardiogram and Echocardiogram for Events in the Cardiac Cycle," *American Journal of Noninvasive Cardiology,* vol. 8, pp. 39-46, 1994.

[4] D. M. Salerno, J. M. Zanetti, L. A. Green, M. R. Mooney, J. D. Madison, and R. A. Van Tassel, "Seismocardiographic changes associated with obstruction of coronary blood flow during balloon angioplasty," *The American Journal of Cardiology,* vol. 68, no. 2, pp. 201-207, July 15, 1991.

[5] D. M. Salerno and J. Zanetti, "Seismocardiography for monitoring changes in left ventricular function during ischemia," (in eng), *CHEST,* vol. 100, no. **4**, pp. 991-3, October 1991.

[6] C. Yang, N. D. Aranoff, P. Green, and N. Tavassolian, "Classification of Aortic Stenosis Using Time–Frequency Features From Chest Cardio-Mechanical Signals," *IEEE Transactions on Biomedical Engineering,* vol. 67, no. 6, pp. 1672-1683, June 2020.

[7] D. S. Morillo, J. L. R. Ojeda, L. F. C. Foix, and A. L. Jiménez, "An Accelerometer-Based Device for Sleep Apnea Screening," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 2, pp. 491-499, 2010.

[8] K. Tavakolian, A. P. Blaber, B. Ngai, and B. Kaminska, "Estimation of hemodynamic parameters from seismocardiogram," in *2010 Computing in Cardiology*, 2010: IEEE, pp. 1055-1058.

[9] P. T. Gamage, M. K. Azad, A. Taebi, R. H. Sandler, and H. A. Mansy, "Clustering of SCG Events Using Unsupervised Machine Learning," in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, I. Obeid, I. Selesnick, and J. Picone Eds. Cham: Springer International Publishing, 2020, pp. 205-233.

[10] P. T. Gamage, "Seismocardiography - Genesis, And Utilization of Machine Learning for Variability Reduction and Improved Cardiac Health Monitoring," Ph.D., Department of Mechanical and Aerospace Engineering, University of Central Florida, 2020.

[11] K. Pandia, "Seismocardiographic Assessment of Cardiopulmonary Health," Ph.D., Stanford University, Ann Arbor, 28115540, 2017.

[12] F. Yasuma and J.-i. Hayano, "Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize With Respiratory Rhythm?," *Chest,* vol. 125, no. 2, pp. 683-690, 2004/02/01/ 2004.

[13] K. Pandia, O. T. Inan, and G. T. A. Kovacs, "A frequency domain analysis of respiratory variations in the seismocardiogram signal," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3-7 July 2013 2013, pp. 6881-6884.

[14] S. Lara and B. Desmond, "Cardiovascular Effects of Mechanical Ventilation," *Archives of Disease in Childhood,* vol. 80, no. 5, p. 475, May 1, 1999.

[15] A. Taebi and H. A. Mansy, "Grouping similar seismocardiographic signals using respiratory information," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2-2 Dec. 2017 2017, pp. 1-6.

[16] M. K. Azad, P. T. Gamage, R. H. Sandler, N. Raval, and H. A. Mansy, "Seismocardiographic Signal Variability During Regular Breathing and Breath Hold in Healthy Adults," in *2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 7-7 Dec. 2019 2019, pp. 1-7.

[17] S. Ahdy, M. Azad, R. Sandler, N. Raval, and H. Mansy, "Spatial Distribution of Seismocardiographic Signal Clustering," in *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 4-4 Dec. 2021 2021, pp. 1-3.

[18] M. K. Azad, "Seismocardiographic Signal Variability and Pulmonary Phase Detection in Adults" Ph.D., Department of Mechanical and Aerospace Engineering, University of Central Florida, 2020.

BIOMEDICAL ACOUSTICS RESEARCH LAB

Effect of the Initial Condition on SCG Clustering using Unsupervised Machine Learning

S. Ahdy[1,3], T. Hassan[1], B. Rahman[1,4], R. Sandler[1,2] and H. Mansy[1,2]

1. Biomedical Acoustic Research Lab, University of Central Florida, Orlando, Florida, USA
2. Biomedical Acoustic Research Company, Orlando, Florida, USA
3. Mechanical Power Engineering Department, Zagazig University, Egypt
4. College of Medicine, Yale University, New Haven, CT

UCF College of Engineering and Computer Science

Introcution

- Seismocardiography (SCG) is cardiac-induced chest wall vibration that is often measured non-invasively by accelerometers.
- SCG is thought to be related to myocardial contractions, valve closures and changes in blood momentum during the cardiac cycle.
- SCG is clinically valuable; it was reported to assess myocardial contractility and detect valvular heart diseases.
- SCG variability (e.g., due to breathing) can hinder accurate extraction of clinically useful SCG features. It can be reduced by clustering based on respiratory phase or using unsupervised machine learning (e.g., K-Medoid).
- The effect of the initial condition of unsupervised machine learning on the clustering solution has not been sufficiently studied.

Objectives

- study the effect of the initial condition on the clustering solution and recommend an efficient procedure for finding the optimum clustering solution.

Methodology

- After IRB approval, 7 healthy subjects were recruited.
- SCG (at 4th ICS) and ECG were simultaneously acquired for one minute.
- Data was collected during normal breathing (NB).
- The baseline wander of SCG (0.15-0.5 Hz) was integrated once and twice to get a measure of the low-frequency chest surface velocity and displacement, which would correlate with the breathing flow rate and lung volume signals, respectively.
- SCG was segmented into beats using the ECG-R waves such that each SCG beat starts 100 ms before the corresponding ECG-R wave.
- SCG beats were clustered using the K-Medoid algorithm (2 clusters). The distance between SCG beats was measured using dynamic time warping (DTW).
- All possible combinations of initial medoid pairs (from a data subset) were considered; the data subset size was varied from 5% to 100% of the total number of beats.
- The data included in the subset were not chosen randomly. Rather, we avoided including nearby beats (as these may have similar morphology) when forming the subset. For example, when the subset has 10% of the total data, beats #1, #11, and #21, etc. were used to build the subset (where all beat combinations were used as initial conditions).
- For each initial condition, the final cluster assignment quality was assessed by calculating the intra-cluster variability and inter-cluster variability.
- The optimum clustering solution was defined as the one having the lowest intra-cluster to inter-cluster variability ratio and not having a sparse cluster.

Results

- Typical clustering solutions of three different randomly chosen initial conditions are shown in Figure 1.

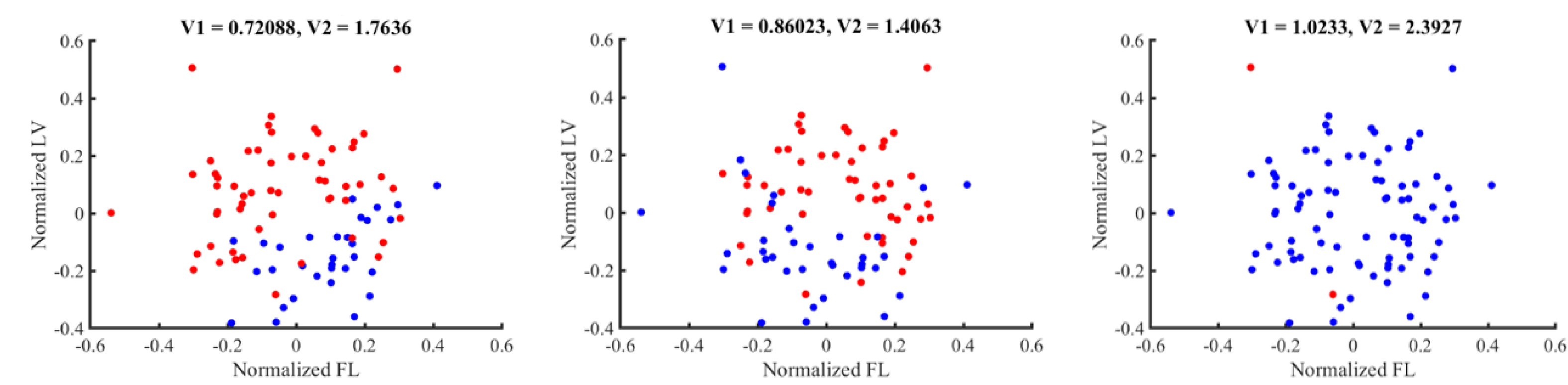

Figure 1. Samples of different clustering solutions obtained after implementing the K-Medoid algorithm using different initial conditions. The dots represent SCG beats at the ECG R wave timing. The two clusters are colored in red and blue. The two numbers (V1 and V2) shown above each subplot refer to the intra-cluster variability and inter-cluster variability respectively.

- Figure 1 shows how the choice of the initial condition can lead to different clustering solutions, including solutions with sparse clusters as in the rightmost subplot.
- For a data subset size from 5% to100%, it was found that many different initial conditions resulted in the same clustering solution.
- Therefore, it was possible to attain the optimum clustering solution while using a data subset when choosing initial conditions.
- Testing all possible initial conditions from 100% of the data can be time consuming especially when analyzing long recordings.
- It was found, in the current study, that choosing the initial medoids from a subset of SCG beats (10% in six subjects and 50% in one subject) leads to solutions that included the optimum clustering solution.
- The study also considered lung volume and flow rate based initial conditions (i.e., HLV/LLV and INS/EXP respectively).The results showed that these two choices do not always converge into the optimal clustering solution.
- In three of the study subjects, the initial condition based on the flow rate resulted in the optimum solution.
- The lung volume-based initial condition resulted in the optimum solution in three subjects and in a clustering solution that had a sparse cluster in one subject.

Conclusions

- The current study showed that the initial condition is determinantal to the quality of SCG clustering solution
- It described a procedure to attain the optimum clustering solution by considering all possible initial conditions from a data subset.
- Optimum clustering will help reduce variability and extract accurate and reliable waveform features.
- This may leverage the clinical utility of SCG in monitoring heart failure patients and other cardiac conditions.
- Future work will include analyzing more subjects, diverse ethnicity and longer recordings, to confirm the study findings, and testing other techniques for selecting the initial condition.

Acknowledgments

- This material is based upon work supported by the National Science Foundation under Grant No. FW-HTF-P 2026516. The research reported in this publication was also supported by the National Heart and Lung Institute of the National Institutes of Health under award number R44 HL099053. Any opinions, findings, and conclusions or recommend-dations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or National Institute of Health.

COI statement

- Richard H. Sandler and Hansen A. Mansy are part owners of Biomedical Acoustics Research Company, which is the primary recipient of the NIH grant R44HL099053, as such they may benefit financially because of the outcomes of the research work reported in this publication.

References

[1] S. Ahdy, M. K. Azad, R. H. Sandler, N. Raval, and H. A. Mansy, "Spatial Distribution of Seismocardiographic Signal Clustering," in *Signal Processing in Medicine and Biology: Innovations in Big Data Processing*, I. Obeid, J. Picone, and I. Selesnick Eds. Cham: Springer International Publishing, 2023, pp. 87-119.

[2] A. Taebi, B. E. Solar, A. J. Bomar, R. H. Sandler, and H. A. Mansy, "Recent Advances in Seismocardiography," *Vibration,* vol. 2, no. 1, pp. 64-86, January 14, 2019.

[3] R. S. Crow, P. Hannan, D. Jacobs, L. Hedquist, and D. M. Salerno, "Relationship between Seismocardiogram and Echocardiogram for Events in the Cardiac Cycle," *American Journal of Noninvasive Cardiology,* vol. 8, pp. 39-46, 1994.

[4] D. M. Salerno, J. M. Zanetti, L. A. Green, M. R. Mooney, J. D. Madison, and R. A. Van Tassel, "Seismocardiographic changes associated with obstruction of coronary blood flow during balloon angioplasty," *The American Journal of Cardiology,* vol. 68, no. 2, pp. 201-207, July 15, 1991.

[5] D. M. Salerno and J. Zanetti, "Seismocardiography for monitoring changes in left ventricular function during ischemia," (in eng), *CHEST,* vol. 100, no. 4, pp. 991-3, October 1991.

[6] C. Yang, N. D. Aranoff, P. Green, and N. Tavassolian, "Classification of Aortic Stenosis Using Time–Frequency Features From Chest Cardio-Mechanical Signals," *IEEE Transactions on Biomedical Engineering,* vol. 67, no. 6, pp. 1672-1683, June 2020.

[7] D. S. Morillo, J. L. R. Ojeda, L. F. C. Foix, and A. L. Jiménez, "An Accelerometer-Based Device for Sleep Apnea Screening," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 2, pp. 491-499, 2010.

[8] K. Tavakolian, A. P. Blaber, B. Ngai, and B. Kaminska, "Estimation of hemodynamic parameters from seismocardiogram," in *2010 Computing in Cardiology*, 2010: IEEE, pp. 1055-1058.

[9] P. T. Gamage, M. K. Azad, A. Taebi, R. H. Sandler, and H. A. Mansy, "Clustering of SCG Events Using Unsupervised Machine Learning," in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, I. Obeid, I. Selesnick, and J. Picone Eds. Cham: Springer International Publishing, 2020, pp. 205-233.

[10] P. T. Gamage, "Seismocardiography - Genesis, And Utilization of Machine Learning for Variability Reduction and Improved Cardiac Health Monitoring," Ph.D., Department of Mechanical and Aerospace Engineering, University of Central Florida, 2020.

[11] K. Pandia, "Seismocardiographic Assessment of Cardiopulmonary Health," Ph.D., Stanford University, Ann Arbor, 28115540, 2017.

[12] F. Yasuma and J.-i. Hayano, "Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize With Respiratory Rhythm?," *Chest,* vol. 125, no. 2, pp. 683-690, 2004/02/01/ 2004.

[13] K. Pandia, O. T. Inan, and G. T. A. Kovacs, "A frequency domain analysis of respiratory variations in the seismocardiogram signal," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3-7 July 2013 2013, pp. 6881-6884.

[14] S. Lara and B. Desmond, "Cardiovascular Effects of Mechanical Ventilation," *Archives of Disease in Childhood,* vol. 80, no. 5, p. 475, May 1, 1999.

[15] A. Taebi and H. A. Mansy, "Grouping similar seismocardiographic signals using respiratory information," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2-2 Dec. 2017 2017, pp. 1-6.

[16] M. K. Azad, P. T. Gamage, R. H. Sandler, N. Raval, and H. A. Mansy, "Seismocardiographic Signal Variability During Regular Breathing and Breath Hold in Healthy Adults," in *2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 7-7 Dec. 2019 2019, pp. 1-7.

[17] S. Ahdy, M. Azad, R. Sandler, N. Raval, and H. Mansy, "Spatial Distribution of Seismocardiographic Signal Clustering," in *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 4-4 Dec. 2021 2021, pp. 1-3.

[18] M. K. Azad, "Seismocardiographic Signal Variability and Pulmonary Phase Detection in Adults" Ph.D., Department of Mechanical and Aerospace Engineering, University of Central Florida, 2020.