Modeling Consistent Dynamics of Cardiogenic Vibrations in Low-Dimensional Subspace

Jonathan Zia, *Student Member, IEEE*, Jacob Kimball, *Student Member, IEEE*, Sinan Hersek, and Omer T. Inan, *Senior Member, IEEE*

***Abstract*—The seismocardiogram (SCG) measures the movement of the chest wall in response to underlying cardiovascular events. Though this signal contains clinically-relevant information, its morphology is both patient-specific and highly transient. In light of recent work suggesting the existence of population-level patterns in SCG signals, the objective of this study is to develop a method which harnesses these patterns to enable robust signal processing despite morphological variability. Specifically, we introduce seismocardiogram generative factor encoding (SGFE), which models the SCG waveform as a stochastic sample from a low-dimensional subspace defined by a unified set of generative factors. We then demonstrate that during dynamic processes such as exercise-recovery, learned factors correlate strongly with known generative factors including aortic opening (AO) and closing (AC), following consistent trajectories in subspace despite morphological differences. Furthermore, we found that changes in sensor location affect the perceived underlying dynamic process in predictable ways, thereby enabling algorithmic compensation for sensor misplacement during generative factor inference. Mapping these trajectories to AO and AC yielded R^2 values from 0.81–0.90 for AO and 0.72–0.83 for AC respectively across five sensor positions. Identification of consistent behavior of SCG signals in low dimensions corroborates the existence of population-level patterns in these signals; SGFE may also serve as a harbinger for processing methods that are abstracted from the time domain, which may ultimately improve the feasibility of SCG utilization in ambulatory and outpatient settings.**

***Index Terms*—Seismocardiogram, dimensionality reduction, autoencoder, cardiac monitoring, generative modeling.**

I. Introduction

ADVANCES in wearable sensing for outpatient monitoring are revolutionizing both healthcare delivery and our understanding and treatment of disease. In particular, there are now myriad ways to monitor heart health outside the clinic using wearable sensors. Among these, the seismocardiogram (SCG) holds promise, particularly in monitoring diseases or conditions affecting the mechanical aspects of cardiovascular health and performance. The SCG measures the movement of the chest wall in response to underlying cardiovascular events [1]. Most notably, valvular events such as aortic opening (AO) and closing (AC) have been shown to occur concurrently with SCG features, with high correlations established between cardiac timing intervals measured with the SCG compared to reference standards [2], [3]. These correlations enable inference of key indicators of cardiomechanical function which derive from AO and AC such as pre-ejection period (PEP), left-ventricular ejection time (LVET), and pulse transit time (PTT) [4], [5]. Notably, the role of such indicators in the diagnosis and management of cardiovascular diseases including hypertention [6], heart failure [7], [8], and coronary artery disease [9] has been well-studied.

Typically captured using a tri-axial accelerometer mounted to the chest wall with concurrent ECG [10], [11], the application of SCG in ambulatory and at-home environments has been limited. By its nature, the morphology of the waveform is highly transient in the time domain, influenced by the coupling of the vascular system with the chest wall, the chest wall with the sensing system, and by the patient's physiological state. Consequently, morphological variability poses a significant challenge in SCG processing [12]. Furthermore, prior literature has shown that SCG morphology varies with sensor position as well, requiring the sensor to be placed properly to accurately estimate cardiomechanical indicators [12], [13].

The ultimate goal of this work is to develop a method of SCG processing which adapts to the patient's anatomy and physiology as well as the position of the sensor for accurate assessment of cardiomechanical indicators, namely rAO and rAC — or the duration between the ECG R-peak and AO and AC respectively. Doing so would not only improve the robustness of SCG processing algorithms, but usability as well by not requiring the user to move the sensor. Toward this goal, this work proposes a new method of modeling SCG signals which is summarized in Fig. 1.

To develop this approach, we begin with the perspective that the cardiovascular system — governed by closed-loop autonomic feedback — follows simple dynamic processes in response to individual stimuli [14]. A dynamic process is one that is governed by a set of rules, such that future states of the system may be predicted from past states and the system's inputs [15]. Consider a patient undergoing an exercise stress test;

Manuscript received August 9, 2019; revised January 24, 2020 and March 7, 2020; accepted March 11, 2020. Date of publication March 16, 2020; date of current version July 2, 2020. This work was supported in part by the National Institutes of Health under Grant 1R01HL130619-A1 and in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award UL1TR002378. *(Corresponding author: Jonathan Zia.)*

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: zia@gatech.edu; jacob.kimball@gatech.edu; shersek3@gatech.edu; omer.inan@ece.gatech.edu).

Digital Object Identifier 10.1109/JBHI.2020.2980979

2168-2194 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

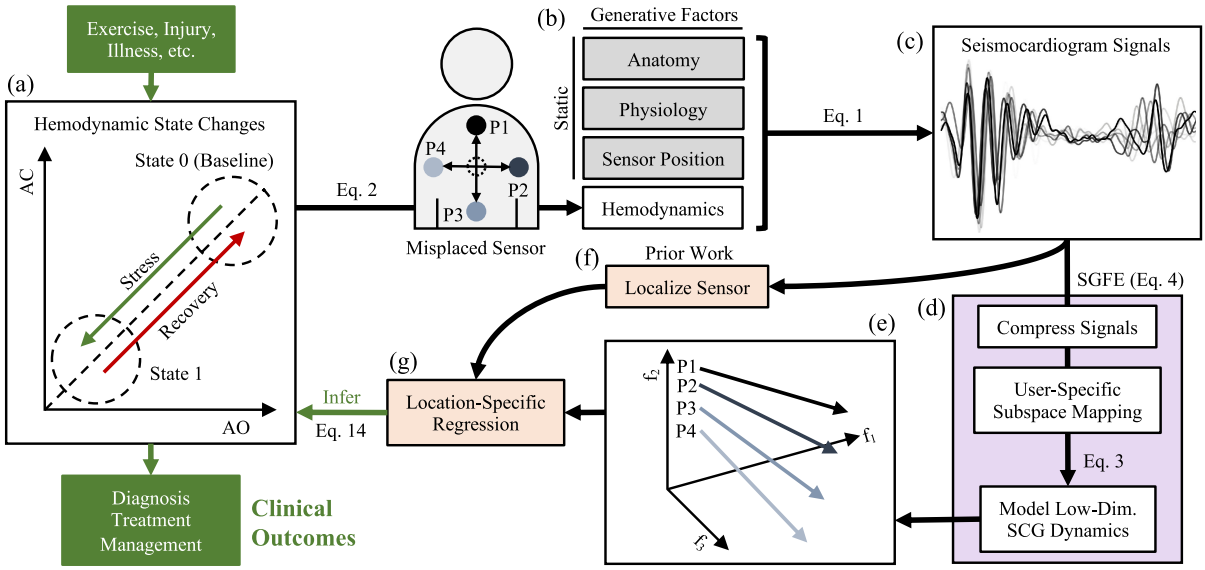

Fig. 1. (a) Illustration of the consistent dynamics of the rAO and rAC interval during an exercise stress test. (b) Hemodynamic factors such as rAO and rAC are among the generative factors of SCG signals. Other factors reflect the particular anatomy and physiology of the patient and sensor position, which are static factors and do not change over time. (c) The SCG may be modeled as a stochastic sample from these underlying generative factors. (d) The proposed SGFE maps SCG signals to a low-dimensional subspace by modeling them in this manner. (e) SCG signals exhibit consistent dynamics in this learned subspace, however observed dynamics are dependent on sensor position. (f) Prior work has demonstrated that SCG sensor position on the chest wall may be localized. (g) By applying position-specific regression to the learned subspace, the hemodynamic factors rAO and rAC may be inferred independently from the other factors. Purple boxes indicate an unsupervised model while orange boxes indicate a supervised mdoel. Equation numbers correspond to those in the text.

after beginning in a baseline resting state, the patient transitions to a new equilibrium state upon the onset of exercise. When the test is complete, the patient returns to their baseline state. Fig. 1(a) illustrates this process in a state space defined by rAO and rAC, which both decrease during exercise and increase during recovery [16]. While the particular trajectory in this state space in response to exercise may be patient-specific, the dynamic behavior is largely preserved.

In this work, we model SCG signals as a stochastic sample from an underlying dynamic process. Consider the process above; if rAO and rAC were the only factors influencing the SCG, this waveform could be losslessly-encoded by the two-dimensional subspace of Fig. 1(a). In reality, the subspace which defines SCG is likely dependent upon a variety of other cardiogenic factors, requiring additional dimensions to achieve lossless encoding [1], [12]. Furthermore, observed signals sampled from this subspace may also be affected by other factors such as the patient's anatomy and physiology and sensor location on the chest wall [17]. As shown in Fig. 1(b), the factors which influence the generation of SCG signals are known as *generative factors* [18], [19].

Though SCG morphology is highly variable, its hemodynamic generative factors, such as rAO and rAC, follow consistent dynamics; these observed signals may therefore exhibit consistent behavior in subspaces defined by these factors [16] as in Fig. 1(a). Mapping signals into these subspaces may thereby enable analysis methods that are robust to morphological variability. To do so, this work introduces the seismocardiogram generative factor encoder (SGFE), which maps SCG signals into a learned low-dimensional subspace (latent space) as illustrated in Fig. 1(c)–(e). As will be shown, SCG signals exhibit consistent behavior in this subspace despite morphological variability, though they follow trajectories that are dependent on sensor position. It is then shown that if sensor position is known, a position-specific linear regression model can be applied to the learned subspace of Fig. 1(e) to accurately estimate the known generative factors rAO and rAC. With this approach, one may estimate the hemodynamics underlying the SCG signal independently from the other generative factors which affect SCG morphology.

To enable this work, prior literature has demonstrated that sensor location on the chest wall may be inferred from SCG signals without user-calibration [20]; therefore, in this work it is assumed that sensor position is known. Regarding SCG modeling, previous studies have proposed principal component analysis (PCA), independent component analysis (ICA), and eigenvector decomposition as possible subspace mapping methods for SCG processing [21]–[23]. Similar methods have also been employed for other cardiovascular signals including ECG and PPG for the purposes of noise reduction and feature extraction [24], [25]. Notably, though, such methods do not incorporate the dynamic behavior of these signals.

The purpose of this work is to formulate the SGFE and analyze its ability to encode the known hemodynamic generative factors rAO and rAC. In the following section, we introduce the SGFE, first illustrating that SCG signals yield consistent trajectories in the low-dimensional latent space of this model despite morphological variability. Subsequently, we will analyze whether this subspace encodes useful information by characterizing its ability to estimate changes in the rAO and rAC intervals. Finally, we will show that consistent changes in subspace behavior due to sensor placement enables algorithmic compensation when inferring AO

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

and AC event timing using this subspace. The contributions of this work include:

1) Introducing the SGFE as a method of inferring seismocardiogram generative factors
2) Demonstrating that SCG waveforms follow consistent patterns in low-dimensional subspace
3) Demonstrating algorithmic correction for sensor misplacement for generative factor inference.

II. Methods

A. Notation

For brevity in the following sections, shorthand will be used when describing matrices and vectors. Matrices in this work are collections of row-wise vectors containing data from subsequent observations in the time interval $\mathscr{T} := [1, T]$. Consider an example T-by-M matrix of real numbers $\boldsymbol{U} \in \mathbb{R}^{T \times M}$. It can be assumed that, $\boldsymbol{U} := [\boldsymbol{u}_1^\top \ldots \boldsymbol{u}_T^\top]^\top$ where $\boldsymbol{u}_t \in \mathbb{R}^M \; \forall t \in \mathscr{T}$. In other words, $\boldsymbol{U}$ is composed of T vectors of length M, where each vector $\boldsymbol{u}_t$ is an observation at time t. Since this notation is used frequently, the shorthand $\boldsymbol{U} := \{\boldsymbol{u}^{(M)}\}_{\mathscr{T}}$ is used. In any such matrix, $u^{(i,j)}$ refers to the element of $\boldsymbol{U}$ in the i^{th} row and j^{th} column while $u_t^{(i)}$ refers to the i^{th} element in vector $\boldsymbol{u}_t$.

Tuples, which are ordered sequences of objects, are indicated by lists of variables enclosed by parentheses. For example, the notation $\boldsymbol{V} := (\boldsymbol{U}, \boldsymbol{w})$ is used to define the variable $\boldsymbol{V}$ as a tuple of the matrix $\boldsymbol{U}$ and vector $\boldsymbol{w}$.

B. Mathematical Framework

Since the SCG derives from the chest wall's response to underlying events, we can abstract this signal as

$$\boldsymbol{F} \longrightarrow \boxed{R_{\boldsymbol{\Phi}}(\boldsymbol{F} \mid P)} \longrightarrow \boldsymbol{X}_P \tag{1}$$

where $\boldsymbol{F} := \{\boldsymbol{f}^{(D)}\}_{\mathscr{T}}$ represents the hemodynamic generative factors of the signal, R is a response function that generates the waveform, and $\boldsymbol{X}_P := \{\boldsymbol{x}^{(M)}\}_{\mathscr{T}}$ is the set of observed SCG vectors from position P. The response function R is parameterized by $\boldsymbol{\Phi}$, which represents the static generative factors related to the patient's anatomy and physiology (Fig. 1(b)), and is conditioned on the sensor position P. Under the assumption that hemodynamic factors vary dynamically according to the state of the cardiovascular system, the factors at each timestep may be described as

$$(\boldsymbol{s}_0, \boldsymbol{\Delta}) \longrightarrow \boxed{G(\boldsymbol{s}_0, \boldsymbol{\Delta})} \longrightarrow \boldsymbol{F} \tag{2}$$

where $\boldsymbol{s}_0 \in \mathbb{R}^K$ is an initial state vector, $\boldsymbol{\Delta} := \{\boldsymbol{\delta}^{(L)}\}_{\mathscr{T}}$ represents changes in state at each point in the time period $\mathscr{T}$, and G is a generator function that produces hemodynamic generative factors using this state information. Though the dimensionality of $\boldsymbol{F}$ and the state variables $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$ are in reality unknown, acceptable values for D, K, and L in a computational model may be inferred, as will be subsequently described. The implications of modeling the SCG in this manner is that there may exist an encoder function E such that

$$\boldsymbol{X}_P \longrightarrow \boxed{E_{\boldsymbol{\Phi}}(\boldsymbol{X}_P \mid P)} \longrightarrow (\boldsymbol{s}_0, \boldsymbol{\Delta}). \tag{3}$$

Now consider that, given a set of observations $\boldsymbol{X}_P$ generated with Equation 1, we wish to approximate the factors $\boldsymbol{F}$ that yielded these signals. Using Equations 2 and 3, this may be accomplished via

$$\boldsymbol{X}_P \rightarrow \boxed{E_{\boldsymbol{\Phi}}(\boldsymbol{X}_P \mid P)} \rightarrow (\boldsymbol{s}_0, \boldsymbol{\Delta}) \rightarrow \boxed{G(\boldsymbol{s}_0, \boldsymbol{\Delta})} \rightarrow \boldsymbol{F}. \tag{4}$$

While the functions E and G are unknown, learning functions experimentally that approximate this behavior may allow inference of hemodynamic generative factors.

C. Model Architecture

This formulation naturally parallels the architecture of a sequence-to-sequence VAE [26]. The proposed model for this study is shown in Fig. 2(a). The input to the model is a sequence $\boldsymbol{X} := \{\boldsymbol{x}^{(M)}\}_{\mathscr{T}}$ of T consecutive heartbeat-separated SCG signals with length M. Note that, for simplicity, the effect of sensor position is omitted for the time being and will subsequently be re-introduced.

To compress the signals, each signal $\boldsymbol{x}_i \in \boldsymbol{X}$ is processed with the multi-layer convolutional network shown in Fig. 2(b). This network is composed of $N = 6$ convolution blocks in series, which convolve the signal with each of k_n filters (kernels) of length ℓ_n in the n^{th} block with unit step. Convolutional networks are commonly used in cardiovascular signal processing due to the temporal dependence of time-series data [27]. The outputs of each convolution layer are normalized before application of an exponential linear unit (ELU) activation function [28], [29]. As was performed in [30], dropout regularization with a rate of 0.2 is imposed on the output of the activation function [31]. Dimensionality reduction is induced by gradually decreasing the number of filters ($k_n = [64, 64, 32, 32, 16, 16]$) and max pooling, which down-samples each signal by a factor of two. To accommodate for shorter signals, the kernel length is also decreased ($\ell_n = [7, 5, 5, 3, 3, 2]$). These parameters were derived heuristically from [30], which explored dimensionality reduction of SCG signals with convolutional networks. Like most of the model, the layers in this network are time-distributed, meaning the same operation is performed for each signal $\boldsymbol{x}_i \in \boldsymbol{X}$.

Before modeling the dynamics present in $\boldsymbol{X}$, the outputs of the compression network are flattened and passed through a dense "read-in" layer with 64 input units, $2 \times (K + L)$ output units, and rectified linear unit (ReLU) activation with dropout regularization at a rate of 0.2. The read-in and read-out layers — also called projection layers — exist because generative factors may present differently as signal features across patients. Thus, though the subspace defined by the generative factors may be conserved, mapping into and out of this subspace may require compensation for signal heterogeneity by fitting these layers on a session-specific basis. In other words, the projection layers capture anatomical and physiological differences — represented by $\boldsymbol{\Phi}$ in Equations 1, 3, and 4 — so that the dynamic model can focus on inferring factors that are common to the population.

Modeling dynamics requires estimation of the initial state $\boldsymbol{s}_0$ and change in state at each timestep $\boldsymbol{\Delta}$. As shown in Fig. 2(c), the former is computed with a bi-directional long short-term memory (LSTM) network E_Z, where the output is the

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

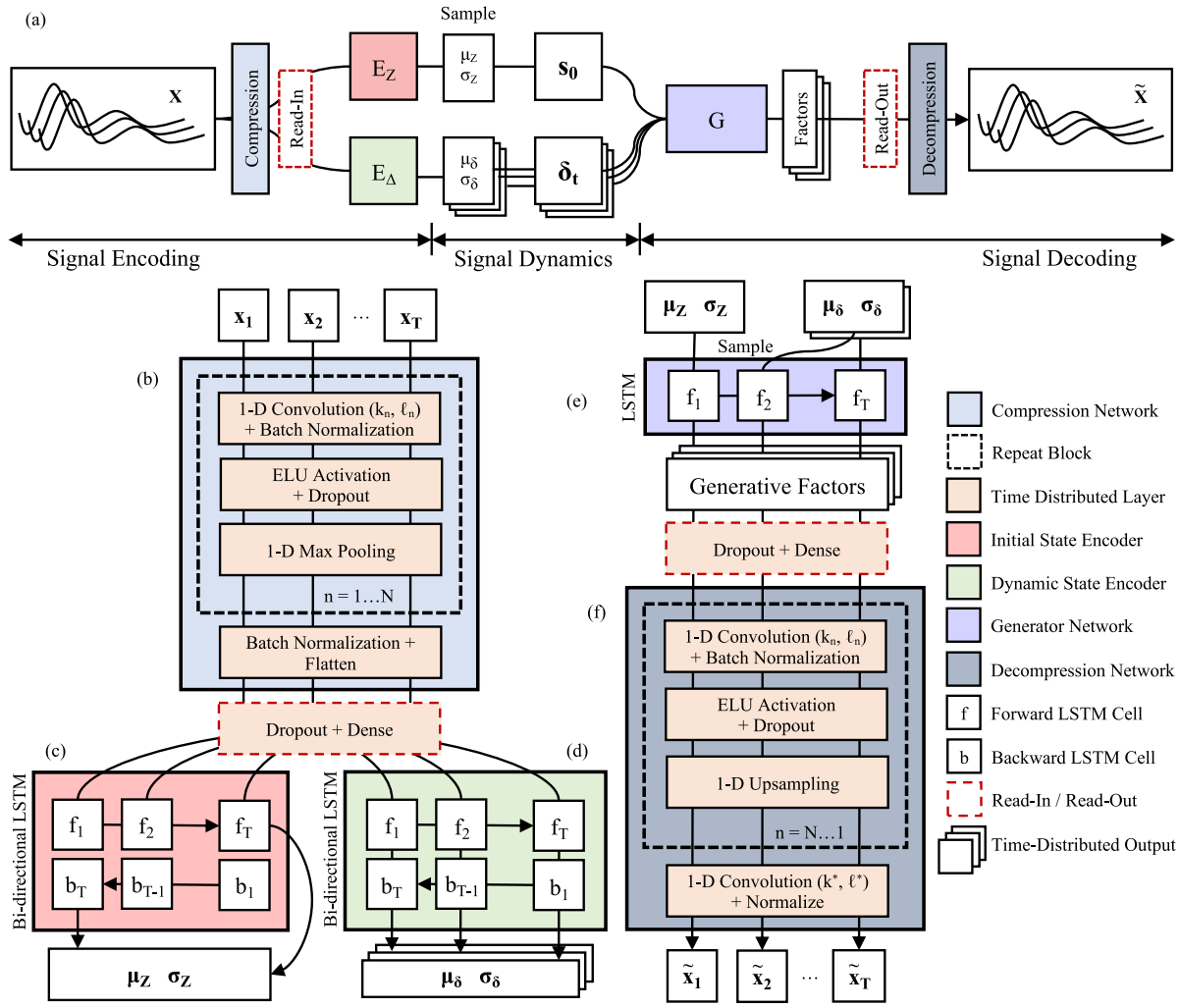

Fig. 2. (a) Proposed seismocardiogram generative factor encoder (SGFE). Detailed descriptions are provided in the text. (b) The input $\boldsymbol{X}$ is first processed by a compression network, which uses a series of N time-distributed convolution blocks to compress the input vector. Each 1-D convolution layer n has k_n kernels with length ℓ_n. A read-in layer encodes the resultant vector as inputs to the dynamic model. Two bi-directional LSTM networks encode (c) the initial state of the system $\boldsymbol{s}_0$ and (d) the change in state with each timestep $\boldsymbol{\delta}_t$. (e) The generator network is an LSTM network that outputs estimates of the generative factors at each timestep. The factors are passed through a read-out layer, which is used to construct the estimate $\tilde{\boldsymbol{X}}$ of the original input. (f) This is achieved with a decompression network, a mirror-image of the compression network. 1-D = one-dimensional.

average between the final outputs of the forward and backward layers [32]. Shown in Fig. 2(d), the latter is also computed with a bi-directional LSTM network E_Δ, where an output $\boldsymbol{\delta}_t$ is produced at each timestep as the average output between the forward and backward cells. However, since this is a VAE instantiation, these values are not evaluated explicitly; rather, they are drawn from a Gaussian distribution, the parameters of which are explicitly evaluated. Thus, the output of E_Z is a tuple $(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)$, $\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0 \in \mathbb{R}^K$. The output of E_Δ at each timestep $t \in \mathscr{T}$ is a tuple $(\boldsymbol{\mu}_{\delta,t}, \boldsymbol{\sigma}_{\delta,t})$, $\boldsymbol{\mu}_{\delta,t}, \boldsymbol{\sigma}_{\delta,t} \in \mathbb{R}^L$. The i^{th} element of the initial state vector $\boldsymbol{s}_0$ is then sampled from

$$s_0^{(i)} \sim \mathcal{N}\left(\mu_0^{(i)}, \sigma_0^{(i)}\right) \forall i \in [1, K]. \tag{5}$$

where $\mathcal{N}(\mu, \sigma)$ is a Gaussian distribution with mean μ and standard deviation σ. Similarly, at each timestep t, the j^{th} element of the state change vector $\boldsymbol{\delta}_t$ is sampled from

$$\delta_t^{(j)} \sim \mathcal{N}\left(\mu_{\delta,t}^{(j)}, \sigma_{\delta,t}^{(j)}\right) \forall j \in [1, L], \quad t \in \mathscr{T}. \tag{6}$$

Note that each element in $\boldsymbol{s}_0$ and $\boldsymbol{\delta}_t$ is drawn independently. The probabilistic nature of the VAE yields a structured latent space, as nearby points will produce inherently similar outputs.

As shown in Fig. 2(e), the generator network estimates the generative factors at each timestep based on the system state. The generator is a uni-directional LSTM network, outputting a vector of factors $\boldsymbol{f}_t \in \mathbb{R}^D$ at each step t. As before, these factors are passed through a read-out dense layer with D inputs and 64 outputs, which maps the generative factors to corresponding signal features. Like the read-in layer, this mapping is learned on a session-specific basis to account for changes in factor manifestation as signal features.

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

Finally, the translated factors are used to construct the output signals $\tilde{\boldsymbol{X}} := \{\tilde{\boldsymbol{x}}^{(M)}\}_{\mathscr{T}}$ with the decompression network shown in Fig. 2(f). This is a mirror-image of the compression network of Fig. 2(b), with the number and length of kernels applied in the reverse order and up-sampling by a factor of two rather than max pooling. The output of the decompression network is a convolution layer with a single filter ($k^\star = 1$) with length $\ell^\star = \ell_1$ such that the output is a single vector at each timestep.

D. Human-Subject Experimental Protocol

Experimental data used in this study was collected under two protocols approved by the Georgia Institute of Technology Institutional Review Board (IRB). In the first protocol, SCG data was collected from different locations on the chest wall during exercise-recovery. In the second, SCG sensors were located on the mid-sternum only, however the protocol featured a large cohort of subjects. The latter was therefore used to train the dynamic model and tune hyperparameters while the former was used to test model performance.

1) Protocol 1: This protocol, explained in detail in [17], included 10 healthy subjects (5 male, 5 female; age 24.7 $\pm$ 2.3 years; weight 70 $\pm$ 10.5 kg; height 170 $\pm$ 11.6 cm) and was performed on two consecutive days. During the sessions, electrocardiogram (ECG), impedance cardiogram (ICG), and SCG signals were collected concurrently. On the first day, individual accelerometers for SCG data collection were placed on the mid-sternum, 7.5 cm to the right, and 7.5 cm to the left. On the second day, SCG sensors were placed on the mid-sternum, 5 cm above, and 5 cm below. For each session, the subject stood motionless for a 60 second rest period, followed by a stepping exercise for 60 seconds, and concluding with a five-minute recovery period during which the subject stood upright and motionless. For consistency in this study, only data from the first of the two sessions was used for SCG data from the central sensor location. Furthermore, this study uses the notation C, L, R, T, B to refer to the center, left, right, top, and bottom sensor locations respectively.

2) Protocol 2: This protocol, explained in detail in [33], included 36 healthy subjects (21 male, 15 female; age 24.7 $\pm$ 3.4 years; weight 68.5 $\pm$ 13.6 kg; height 170.9 $\pm$ 9.5 cm). SCG was recorded with an accelerometer on the mid-sternum along with reference ECG and ICG signals. As with the previous protocol, the subjects began by standing upright and motionless for a five-minute rest period; they then performed three minutes of walking at 4.83 km/h on a treadmill followed by 90 seconds of a squatting exercise; the protocol then concluded with the subject again standing upright and motionless for a five-minute recovery period.

E. Signal Pre-Processing

1) Noise Reduction: All signals were filtered with a band-pass finite impulse response (FIR) filter with Kaiser window. Cutoff frequencies were 0.5–40 Hz for the ECG, 1–30 Hz for ICG, and 1–40 Hz for SCG [33]. During data collection, these signals were sampled at 2000 Hz. For the SCG signals, only the dorsoventral axis (z-axis) acceleration was used to minimize network complexity, as this is considered the most useful axis for SCG processing [4]. The signals were heartbeat-separated using the R-peaks of the concurrent ECG signal as a reference. It should be noted that the results in this work suppose access to concurrent ECG, though prior work in this field has explored ECG-free SCG segmentation [34]. All signal segments were then abbreviated to a length of 800 samples (400 ms) before being down-sampled to $M = 256$ samples using linear interpolation with an anti-aliasing filter. Note that a signal length of 400 ms was sufficient to capture AC for this dataset due to its focus on exercise recovery, during which LVET is low; this may not hold true for other datasets, and signal length should be adjusted accordingly. For each protocol and for each subject, the dataset was windowed using a sliding window of 50 signal segments with 50% overlap such that $T = 50$. All signal segments were then normalized to have zero mean and unit variance. As the final step of processing, ICG and SCG signal segments were smoothed using a rolling-window ensemble average of five heartbeats to remove aberrant noise.

2) AO and AC Estimation: Reference values for AO and AC were obtained from ICG B- and X-points respectively. The B-point was computed as the point of maximum second derivative occurring before the global maximum of the waveform; the X-point was computed as the lowest signal minimum following the global maximum [35]. While ICG is commonly used for this purpose, the gold-standard for AO and AC estimation is the echocardiogram; for this reason, the reference values obtained from ICG are intended for use in this study as AO and AC correlates rather than ground-truth measurements [36]. All timing intervals were computed in reference to the respective ECG R-peak for each heartbeat. Thus, rAO (PEP) and rAC refer to the time in milliseconds between the ECG R-peak and AO and AC respectively.

F. Loss Function and Training Protocol

The goal of training was to minimize the loss function

$$\mathscr{L} = \alpha\,\mathrm{MSE}\left(\boldsymbol{X}, \tilde{\boldsymbol{X}}\right) + \beta\left[\mathrm{D}_0 + \frac{1}{T}\sum_{t=1}^{T}\mathrm{D}_t\right]. \tag{7}$$

The MSE operator computes the mean square error between $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, specifically

$$\mathrm{MSE}\left(\boldsymbol{X}, \tilde{\boldsymbol{X}}\right) = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}(x^{(t,m)} - \tilde{x}^{(t,m)})^2. \tag{8}$$

When calculating the reconstruction error, each target vector $\boldsymbol{x}_i \in \boldsymbol{X}$ and output vector $\tilde{\boldsymbol{x}}_j \in \tilde{\boldsymbol{X}}$ was normalized as will be described below. The variables D_0 and D_t in Equation 7 represent the Kullback-Leibler (KL) divergence, which is a measure of similarity between two probability distributions. For distributions P and Q, The KL divergence is given by

$$\mathrm{D}(P\|Q) = -\sum_{x} P(x)\log\left(\frac{Q(x)}{P(x)}\right). \tag{9}$$

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

In Equation 7, the variable D_0 is given by

$$\mathrm{D}_0 = \sum_{k=1}^{K} \mathrm{D}\left(\mathcal{N}(0,1) \| \mathcal{N}\left(\mu_0^{(k)}, \sigma_0^{(k)}\right)\right) \tag{10}$$

and the variable D_t is similarly given by

$$\mathrm{D}_t = \sum_{\ell=1}^{L} \mathrm{D}\left(\mathcal{N}(0,1) \| \mathcal{N}\left(\mu_{\delta,t}^{(\ell)}, \sigma_{\delta,t}^{(\ell)}\right)\right). \tag{11}$$

While the MSE term represents the reconstruction error, the divergence terms impose a penalty on the distributions from which $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$ are sampled. This has two benefits for the model. First, the size of the state space defined by $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$ is limited, as divergence from a zero-centered distribution with unity variance will increase the KL divergence; this increases the continuity of the latent space, as it is disadvantageous for inputs from different sessions to cluster in different locations of the state space. Second, this serves to disentangle the dimensions of the state space, since redundancy in information encoded by each variable may increase the KL divergence as well [37]. Increases in KL divergence are tolerated only if they lead to a sufficient decrease in reconstruction error.

The variables α and β in Equation 7 are scalars computed during the first training step which normalize the value of each term to 0.5. This serves to equalize the contribution of both terms and express the loss at each epoch as a percentage of initial error with random network weights.

Since the AO-related features in the first half of the signal generally have a higher SNR than the AC-related features in the second half, the first and second halves of each signal vector were normalized separately with zero mean and unit variance. If this normalization was not performed, the decrease in MSE resulting from modeling AC-related features did not surpass the increase in KL divergence penalty for doing so. Though this method produced a discontinuity in the middle of each signal, it has the benefit of not increasing the number of hyperparameters parameters of the model as would be the case with other solutions such as using a true β-variational scheme [37] or weighing the MSE differently at each sample point. Furthermore, normalizing the amplitude features has the benefit of preventing the model from encoding amplitude features, which are not of interest in this model [30].

The model was implemented in Keras with Tensorflow backend. The hardware setup was based on a 3.6 GHz Intel Core i7 7820X processor with a GeForce GTX 1080 Ti GPU. Training was performed using mini-batch stochastic gradient descent [38]. At the beginning of each epoch — which represents a group of training steps in which all training samples are incorporated — the training samples were randomized and split into batches of 32 samples for each gradient computation. The ADAM optimizer was used to compute gradient updates, with initial learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1.0 \times 10^{-7}$, which are the standard hyperparameters for this optimizer [39]. The learning rate was decayed by a factor of 0.5 after each set of 10 consecutive epochs without achieving a new minimum validation loss. Training was terminated after 30 such consecutive epochs. This model required 95 minutes to train using 9.3×10^6 training samples.

During training, a simplifying assumption was made whereby a single pair of projection layers was trained for all sessions in the training set. Thus, data from all sessions was mixed together at the beginning of each epoch. Subsequently, during testing, session-specific projection layers were learned by freezing all network weights besides those in the projection layers and repeating the same training protocol separately for each session in the testing set. Learning session-specific projection layers for the training set greatly increased computational complexity and did not yield corresponding improvements in model performance, so this was only performed during testing.

G. Dimensionality Estimation

Before modeling SCG dynamics, proper dimensionality for the state variables $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$ was estimated. The model in Fig. 2 was fitted with recovery-period data from the 36 subjects of Protocol 2, training on 16 subjects, validating on 10, and testing on 10. The value of β in Equation 7 was set to zero such that the latent space was not arbitrarily regularized. As a starting point, the values of K and L were both set to 10, and D was set to 20. In this study, D was always set to $K + L$ such that the generator network did not additionally perform dimensionality reduction or expansion.

After training, the vector $\boldsymbol{s}_0$ and matrix $\boldsymbol{\Delta}$ were computed for each sample in the testing set. Concatenating the former across testing samples yielded a matrix $\boldsymbol{S}_0 \in \mathbb{R}^{N \times K}$ where N is the number of testing samples. The dimensionality of the initial state was estimated by performing PCA on the matrix $\boldsymbol{S}_0$ and returning the variance explained by each resultant PCA dimension, of which there were K [40].

Note that $\boldsymbol{\Delta}$ returns a vector at $\boldsymbol{\delta}_t$ each timestep $t \in \mathcal{T}$, and thus $\boldsymbol{\Delta} \in \mathbb{R}^{T \times L}$ for each testing sample. Therefore, for each timestep t, the vector $\boldsymbol{\delta}_t$ was concatenated across testing samples to yield T matrices $\boldsymbol{\Delta}_t \in \mathbb{R}^{N \times L}$. PCA was performed on each matrix $\boldsymbol{\Delta}_t$ and the variance explained by each PCA dimension was calculated. For each dimension, the variance explained was averaged across each timestep to compute the mean variance explained across time. To determine the dimensionality of $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$ used in this study, a cutoff of 10% variance explained was used, as additional dimensions would increase the complexity of the model without yielding significant increases in explained variance.

H. Training and Testing Dynamic Model

The model in Fig. 2 was trained using the exercise-recovery period data from each of the 36 subjects in Protocol 2. To focus the modeling on dynamic processes, resting period data was not used. A total of 10 subjects in the training set were selected at random for validation and thereby removed from the training set. Based on results from the previous section, the dimensionality parameters K and L were set to 4 and the parameter D was therefore set to 8.

After training, all network weights save for those in the projection layer were frozen. The model was then trained separately

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

on data from each subject and sensor position in the testing set. This consisted of data from the 10 subjects in Protocol 1 with five position-specific sessions each, leading to 50 session-specific pairs of projection layers with universal compression, dynamic, and decompression networks. Therefore, though the subspace defined by generative factors remained constant, projection into and out of this subspace was learned on a session-specific basis. For each testing sample, data collected included $\boldsymbol{s}_0$, $\boldsymbol{\Delta}$, $\boldsymbol{F}$, and $\tilde{\boldsymbol{X}}$.

Held-out validation was not used for learning session-specific projection layers in the testing set. This is because the SGFE is a fully-unsupervised model, meaning that for practical implementation, it is a reasonable assumption that data collected from the patient may be used to update the model and infer generative factors concurrently. Furthermore, since the projection layers accounted for approximately 1% of network parameters (1096 of 103169 total), this enabled rapid training of the session-specific projections, supporting that this approach is reasonable for quasi-real-time feedback systems.

I. Visualizing Behavior of Subspace Projections

For visual analysis of subspace behavior, the goal of the following method was to identify the pair of dimensions in the learned subspace $\boldsymbol{F}$ that encoded the most consistent linear trajectories. Linear trajectories were expected to arise in the latent space because, as will be illustrated, AO and AC were found experimentally to follow linear trends in exercise-recovery when plotted against one another.

To do so, for each session in the test set defined by the subject $S \in [1, 10]$ and sensor position $P \in \{\mathrm{C}, \mathrm{L}, \mathrm{R}, \mathrm{T}, \mathrm{B}\}$, the subspace projection $\boldsymbol{F} \in \mathbb{R}^{T \times D}$ for each of $N_{S,P}$ samples in the session was concatenated to form the matrix $\boldsymbol{F}_{S,P} \in \mathbb{R}^{TN_{S,P} \times D}$. In this manner, each matrix $\boldsymbol{F}_{S,P}$ contained the subspace encoding of all data for one of the 50 sessions in the test set. These matrices were further concatenated row-wise across all subjects to form the matrix $\boldsymbol{F}_P$ for each sensor position. Thus, $\boldsymbol{F}_P$ contained the subspace encoding of all data from sessions from a particular sensor position.

The following was then performed for all P. For each pair of column vectors $(\boldsymbol{f}_i, \boldsymbol{f}_j) \in \boldsymbol{F}_P, i \neq j$, linear regression was used to find the optimal linear fit between $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$. The pair i, j in which the coefficient of determination (R^2) of the linear fit averaged across all P was maximal was selected as the optimal axis pair for further analysis [41]. Subspace trajectories were visualized by plotting the resultant vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ against one another.

Though this method is useful in identifying hyperplanes in the learned subspace in which trajectories are consistent, this does not necessarily mean that the information encoded in the hyperplane is useful and that the two dimensions simply covary despite attempts at disentanglement. Therefore, a second qualitative analysis was performed to determine whether the identified dimensions may contain useful information about the known generative factors AO and AC. For five of the 10 subjects in the testing set chosen at random, the ICG-derived rAO interval was plotted against the rAC interval on a scatter plot for the first of the two recording sessions. Best-fit lines were then overlaid on data from each subject to better visualize the trajectories of these intervals. For the same subjects, the subspace projections $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ from the same session for the central sensor location were plotted on a scatter plot. Best-fit lines were again overlaid on the subspace encoding for each patient in order to observe whether changes in rAO/rAC trajectories may be reflected by the identified dimensions.

J. Visualizing Sensor Location Effect on Observed Dynamics

Though the hyperplane defined by f_1 and f_2 may be a suitable subspace in which to observe the consistent dynamics of SCG signals, it may be sub-optimal for visualizing the effects of changing sensor state on observed dynamics. To do so more effectively, PCA was used to find an informative three-dimensional representation of the the subspace $\boldsymbol{F}$, and the average trajectory for each of the five sensor positions was then plotted in these PCA dimensions for visualization.

To do so, the matrix $\boldsymbol{F}_P$ was concatenated across positions to form $\boldsymbol{F}_{\text{tot}} \in \mathbb{R}^{TN_{\text{tot}} \times D}$ where N_{tot} is the total number of samples in the testing set. The matrix $\boldsymbol{F}_{\text{tot}}$ thus contained the subspace projections for all samples in the testing set. PCA was then perfomed on $\boldsymbol{F}_{\text{tot}}$ to obtain the transformation $\boldsymbol{A} \in \mathbb{R}^{D \times D}$ mapping dimensions of $\boldsymbol{F}_{\text{tot}}$ into the orthogonal subspace defined by PCA dimensions.

The following was then performed for each matrix $\boldsymbol{F}_{S,P}$, which contained the subspace encoding for the session with subject S and position P. Each of the 10 matrices $\boldsymbol{F}_{S,P}, S \in [1, 10]$ was averaged elementwise to obtain a session-averaged matrix $\bar{\boldsymbol{F}}_P$. $\bar{\boldsymbol{F}}_P$ thereby contained the subspace encoding for position P averaged across all subjects. Subsequently, this matrix was transformed using the matrix $\boldsymbol{A}$ to obtain $\boldsymbol{A}_P = \bar{\boldsymbol{F}}_P \boldsymbol{A}$, the projection of $\bar{\boldsymbol{F}}_P$ in the PCA subspace. Finally, for each position, the first three dimensions of $\boldsymbol{A}_P$ were then plotted on a scatter plot for visualization.

K. Evaluating Generative Factor Inference

Based on the results of qualitative analysis, quantitative analysis was performed to determine the extent to which the learned subspace $\boldsymbol{F}$ encodes known generative factors derived from the ICG reference. Since VAE models are fully-unsupervised, generative factors may not necessarily correspond to the dimensions of the latent space in a one-to-one manner; rather, such factors may be encoded by combinations of dimensions. Because of this, we instead apply transformations to the latent space to better estimate generative factors.

In this work, linear regression was used to infer ICG-derived AO and AC event timing using the learned subspace dimensions. As shown in Fig. 1(g), this method identified a linear mapping from the dimensions of $\boldsymbol{F}$ to known generative factors. To begin with, a separate linear mapping was learned for each sensor position P and with each of the 10 subjects in the testing set held-out. To do so, least-squares regression was used to solve

$$\boldsymbol{X}_{P,\bar{S}} = \underset{\boldsymbol{X}}{\operatorname{argmin}} \|\boldsymbol{Y}_{P,\bar{S}} - \boldsymbol{F}_{P,\bar{S}} \boldsymbol{X}\|_2^2 \tag{12}$$

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

where $\boldsymbol{F}_{P,\bar{S}}$ is the matrix $\boldsymbol{F}_P$ with the subject S held-out and $\boldsymbol{Y}_{P,\bar{S}}$ is a matrix where each column is a vector of known generative factor values corresponding to each row of $\boldsymbol{F}_{P,\bar{S}}$. The columns of $\boldsymbol{Y}_{P,\bar{S}}$ thus contained the ICG-derived rAO and rAC intervals respectively. This process was performed for each of the five sensor positions and with each of the 10 subjects held-out. Once the mapping $\boldsymbol{X}_{P,\bar{S}}$ was learned for each held-out subject, it was used to obtain predictions from the held-out subject such that

$$\tilde{\boldsymbol{Y}}_{P,S} = \boldsymbol{F}_{P,S}\boldsymbol{X}_{P,\bar{S}} \tag{13}$$

where $\tilde{\boldsymbol{Y}}_{P,S}$ is a vector of predicted generative factors for subject S with sensor position P. The R^2 and root-mean-square error (RMSE) were obtained for the predicted factors $\tilde{\boldsymbol{Y}}_{P,S}$ versus the known generative factors $\boldsymbol{Y}_{P,S}$ after each session, and the performance results were plotted for each sensor position [40].

L. Quantifying Sensor Location Effect on Subspace Encoding

If alterations in sensor state have predictable effects on observed dynamics, this would mean that the mapping from the latent space $\boldsymbol{F}$ to the generative factors would perform strongly for signals from a single position, but sub-optimally for others. Consequently, if sensor placement was known, this would allow algorithmic compensation for sensor placement when inferring generative factors. To observe this effect, the following was calculated for each pair of positions $P_i, P_j \in P$ and subject S:

$$\tilde{\boldsymbol{Y}}_{(i,j),S} = \boldsymbol{F}_{P_i,S}\boldsymbol{X}_{P_j,\bar{S}} \tag{14}$$

where P_i is the position being tested and the mapping was trained using data from P_j. For each session — corresponding to subject S and sensor location P_i — the R^2 was obtained between $\tilde{\boldsymbol{Y}}_{(i,j),S}$ and $\boldsymbol{Y}_{P_i,S}$ for both the rAO and rAC intervals, where the former is the model's estimate and the latter is the ICG-derived reference values. The result was then averaged across subjects to yield the matrices $\tilde{\boldsymbol{Y}}^{\mathrm{AO}}, \tilde{\boldsymbol{Y}}^{\mathrm{AC}} \in \mathbb{R}^{5\times5}$, where each element $\tilde{y}^{(i,j)}$ was the average R^2 across subjects for sensor data from position P_i with a mapping trained using data from position P_j. The performance matrices $\tilde{\boldsymbol{Y}}^{\mathrm{AO}}$ and $\tilde{\boldsymbol{Y}}^{\mathrm{AC}}$ were then plotted as confusion matrices to visualize changes in performance when using different position-specific mappings for testing data from each position.

III. Results and Discussion

A. Dimensionality Estimation

Fig. 3 shows the variance explained by PCA dimensions for $\boldsymbol{s}_0$ and $\boldsymbol{\Delta}$. Notably, after the first four PCA dimensions, the variance explained by additional dimensions of $\boldsymbol{s}_0$ or $\boldsymbol{\Delta}$ does not exceed 10%. Therefore, by limiting the dimensionality of these vectors to 4, the complexity of the network is reduced without sacrificing the ability to encode information that substantially impacts signal reconstruction.

Dimensionality selection presents an essential trade-off in autoencoder architectures. Low dimensionality of the latent layers both reduces network complexity — limiting the number of parameters that must be learned while increasing generalizability — and compels each dimension to encode more useful attributes, in terms of variance explained. On the other hand, limiting dimensionality too severely may inhibit the network from adequately reconstructing the signal, and thus small variations that may nevertheless be important in encoding factors such as sensor state may not be represented in the latent space [42]. For this reason, the selected dimensionality may not generalize to applications in which encoding of more minute changes in SCG morphology is required.

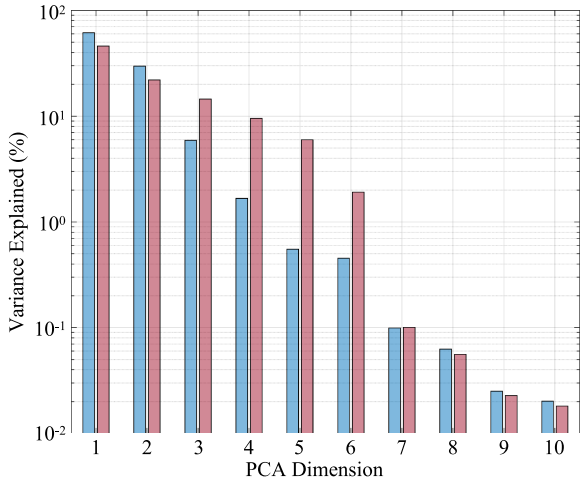

Fig. 3. Percent variance explained by each PCA dimension for model trained using $K = 10$, $L = 10$ with $\beta = 0$. Results are shown for initial state vector $\boldsymbol{s}_0$ (blue), and state change matrix $\boldsymbol{\Delta}$ (red) using logarithmic axis.

Along these lines, while the chosen dimensionality was adequate for sensor state encoding, the results in Fig. 3 do not necessarily indicate that the process underlying SCG generation is inherently low-dimensional. During the dynamic process of exercise-recovery explored in this work, variance in the SCG waveform is likely driven by key factors such as valvular event timing, which may lead the contribution of other factors to be understated. In other applications and during other processes, the dimensionality of the latent space for effective computational modeling may increase or decrease.

B. Visualizing Behavior of Subspace Projections

Subspace projections of SCG signals for two subjects during exercise-recovery are shown in Fig. 4. From the first and last columns of the figure, it is apparent that signal morphology between the subjects — and even at different sensor locations for the same subject — often varies substantially. This time-domain variability is juxtaposed with trajectories in the learned subspace which are largely conserved. Specifically, the subspace projection of the signal during this period follows an approximately linear trajectory in the reference frame defined by the selected subspace dimensions f_1 and f_2.

This consistency is essential because it suggests that this subspace encodes features that are common to SCG signals despite apparent morphological differences. As aforementioned,

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

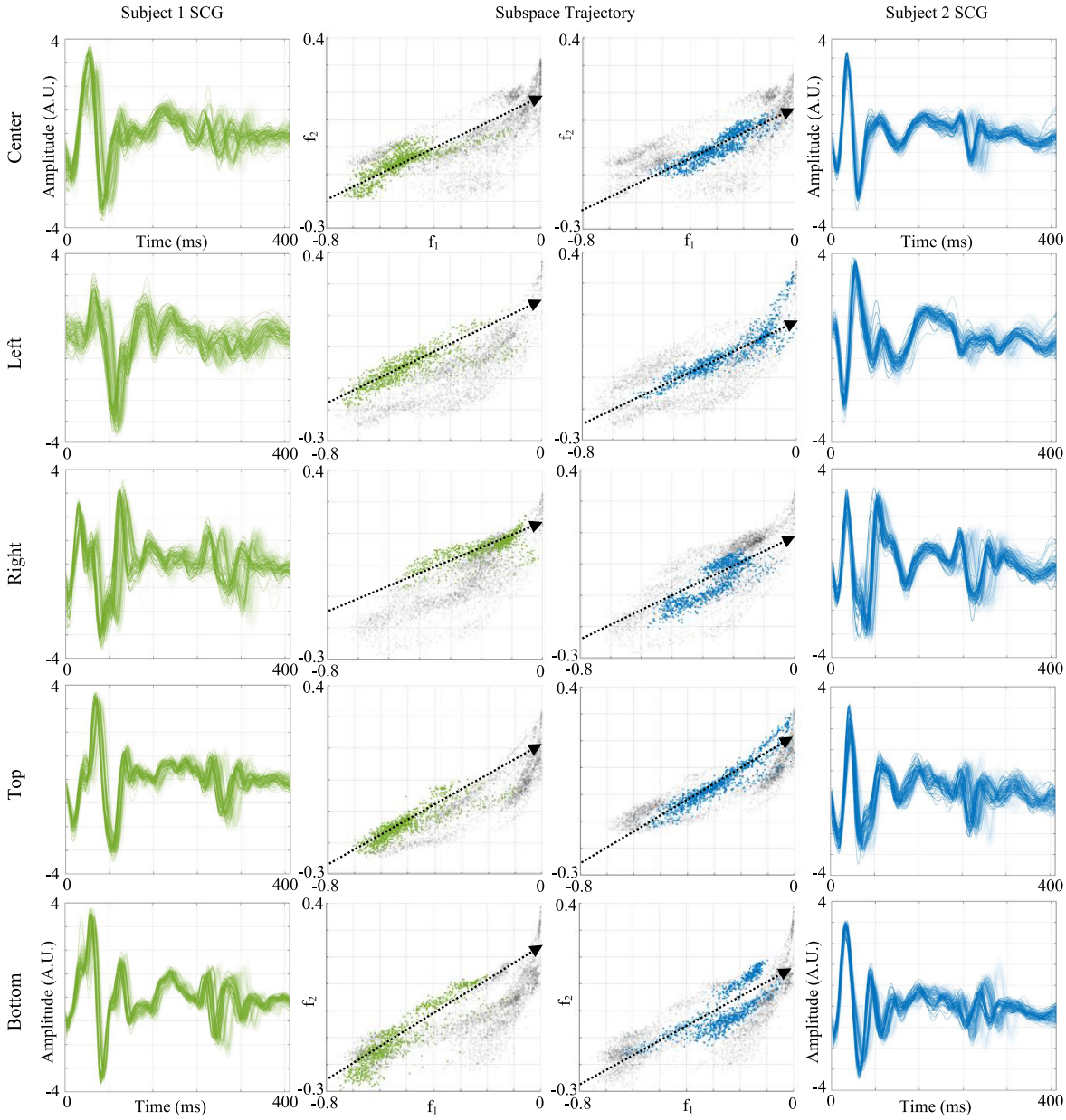

Fig. 4. Subspace projections of recovery-period SCG data for two subjects. The rows of the figure represent each of the five different sensor positions. The left and right columns show a subset of the amplitude-normalized SCG data from Subjects 1 and 2 respectively, with the second and third columns showing the corresponding subspace trajectories in green and blue respectively. The axes represent learned subspace dimensions f_1 and f_2; gray points in the figure represent subspace projections with the same sensor position from the remaining patients in the testing set. Trajectory directions are overlaid (black, dotted). A.U. = arbitrary units.

this is made possible by the session-specific projection layers, which encode the translation between estimated generative factors and time-domain signal features. In this manner, anatomical heterogeneity is captured by the projection into and out of this subspace, and not by the subspace itself. Such a result suggests that constructing models which incorporate rather than eschew patient-specific heterogeneity may consistently model underlying patterns.

With regards to practically implementing such a system, it is important to note that this subspace projection was learned in a *fully-unsupervised* manner. Therefore, it is reasonable to assume that such patient-specific tuning of the model for optimal performance will be feasible in practical systems: the projection may be learned passively without any labeled training data. Furthermore, approximately 1% of model parameters were contained by the projection layers, which may enable rapid training in quasi-real-time systems. While training the full model required approximately five hours with this dataset and hardware setup, fitting session-specific projection layers was typically achieved in less than three minutes.

An example of the relationship between ICG-derived rAO and rAC and the learned subspace dimensions f_1 and f_2 is shown in Figs. 5(a) and (b). Fig. 5(a) shows the trajectories in the subspace defined by rAO and rAC for each subject, while Fig. 5(b) shows

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

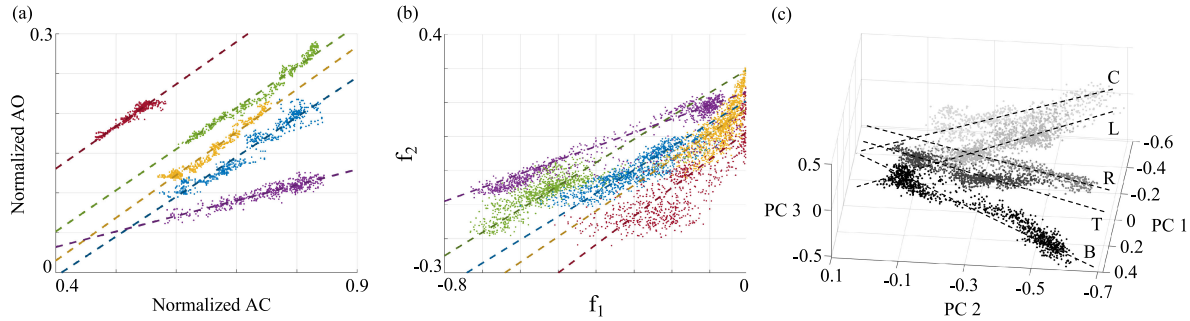

Fig. 5. (a) ICG-derived AO and AC points during exercise-recovery for five subjects in the test set, with each subject assigned a different color. AO and AC are shown as scatter points; best-fit lines for the scatter points are overlaid as dashed lines. (b) Subspace trajectories in dimensions f_1 and f_2 from centrally-placed sensors for the same subjects with the same color-coding as in (a). Subspace projections are shown as scatter points with best-fit lines overlaid as dashed lines. (c) Trajectories in PCA dimensions of $\boldsymbol{F}$ for SCG signals from each of the five sensor positions averaged across all subjects. From lightest to darkest shading, the positions include center, left, right, top, and bottom. The trajectories are also indicated with black dashed lines.

the corresponding trajectories in the subspace defined by f_1 and f_2. In Fig. 5(a), the linear dynamics are apparent; while the trajectories are similar for most patients, one of the patients in this set — shown in purple — has a trajectory which differs visibly from the others. This difference is reflected in Fig. 5(b), which shows a corresponding change in trajectory in the learned subspace. The qualitative results shown in Figs. 4 and 5(a)–(b) serve to visually demonstrate what will be shown quantitatively in the following sections. To enable robust generative factor inference, subspace trajectories for similar processes must be consistent, and changes in underlying generative factors must be reflected in the learned subspace.

C. Visualizing Sensor Location Effect on Observed Dynamics

While the dimensions f_1 and f_2 demonstrate consistent trajectories for all positions, they may not best illustrate changes in observed dynamics associated with sensor location. Fig. 5(c) shows the session-averaged trajectories for each of the five sensor positions in the first three PCA dimensions of $\boldsymbol{F}$. The figure illustrates that each of the sensor positions has a characteristic, distinguishable trajectory in the subspace. Changing the position of the SCG sensor is akin to altering the reference frame from which the underlying hemodynamic process is observed. This is reflected in Fig. 5(c): though the trajectories observed at each position are consistently linear, their direction varies with the change in reference frame. As will be shown, predictable changes in these trajectories allow for correcting the altered reference frame algorithmically when inferring generative factors, mitigating the effect of sensor position on observed dynamics.

D. Evaluating Generative Factor Inference

The performance of position-specific linear mappings for rAO and rAC inference from the learned subspace $\boldsymbol{F}$ is shown in Figs. 6(a) and (b). Fig. 6(a) shows that these mappings produced values that correlated strongly with ICG-derived intervals. Additionally, Fig. 6(b) shows the RMSE between estimated and reference values for the generative factors.

Notably, while the R^2 values for rAO only slightly exceed those for rAC, the RMSE of the estimated rAO is significantly lower than that of rAC. This indicates that while the learned subspace $\boldsymbol{F}$ effectively encoded changes in rAO and rAC, the precise value of rAC had a larger offset versus the ICG reference. This is unsurprising, since the signal features corresponding to AC generally have lower energy, often causing ambiguity for precise AC identification. Beyond demonstrating accurate asseessment of rAO and rAC, Figs. 6(a) and (b) demonstrate that the latent space of the SGFE model contains information on measurable physical phenomena.

The RMSE for rAO estimation shown in Fig. 6(b) is within acceptable limits for all sensor positions, which in prior work typically falls between 11–18 ms compared to ICG-derived reference values [33]. For instance, [13] used XGBoost regression on an ad hoc feature set to estimate rAO using SCG sensors in four different sensor locations, achieving RMSE values from 11.6($\pm$0.4) ms to 17.1($\pm$0.6)ms using z-axis acceleration. Recently, [33] used a similar method to achieve an RMSE of 11.46($\pm$0.32) ms from centrally-placed sensors fusing multiple accelerometer and gyroscope axes. As shown in Fig. 6(b), the RMSE for this task ranged from 7.23($\pm$1.54) ms to 10.53($\pm$1.11) ms in this work. Regarding rAC estimation, the RMSE was larger than for rAO when expressed in miliseconds; however, since the rAC interval is much longer than rAO, the error in rAC estimation relative to its magnitude was comparable to that of rAO. This is reflected in Fig. 6(a), which shows a more comparable R^2 between estimated and true rAO and rAC, with values in the range 0.81–0.90 for rAO and 0.72–0.83 for rAC.

Though the results in Figs. 6(a) and (b) show that some sensor locations achieved somewhat higher performance than others, it is important to note that the optimal sensor location for rAO and rAC estimation is likely an idiosyncracy dependent upon the processing method or perhaps even the dataset being used. For instance, Fig. 6(b) suggests that the lower-sternum sensor placement is optimal for rAO estimation while [13] achieved highest performance under the left clavicle. Finally, it is important to note that ICG is not the gold-standard reference for AO and AC event timing; therefore, the results in Fig. 6(b) do not necessarily reflect the true error of the estimated generative factors.

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

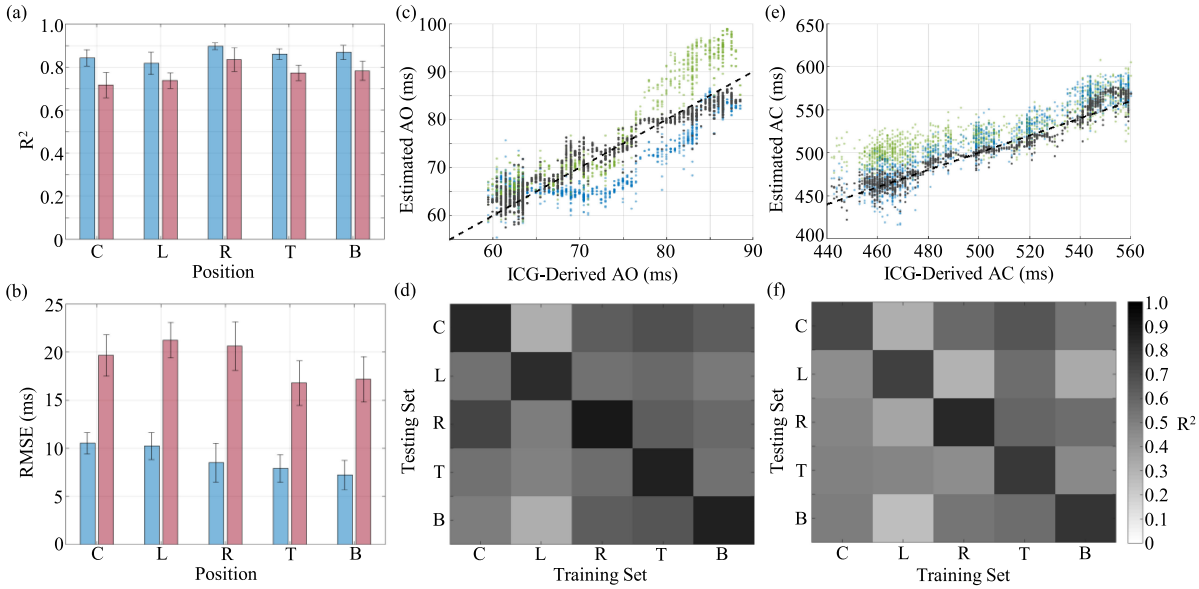

Fig. 6. (a) R^2 and (b) RMSE between ICG-derived rAO (blue) and rAC (red) and estimations from the learned subspace $\boldsymbol{F}$ using position-specific linear mappings (x-axis) on held-out subjects. (c) Scatter plot of ICG-derived vs. estimated rAO for one subject with sensors placed in the center (black), left (blue), and right (green) locations using the linear mapping trained on centrally-placed SCG data. 1:1 correspondence line is overlaid (black, dashed). (d) Confusion matrix of average R^2 for rAO estimation for all held-out subjects for a specific sensor position (y-axis) derived using linear mappings trained on a specific position (x-axis). Analogous results for rAC estimation are shown in (e) and (f).

E. Quantifying Sensor Location Effect on Subspace Encoding

Figs. 6(c)–(f) show the effect of sensor position on the encoding of known generative factors in the learned subspace. Figs. 6(c) and (e) show that rAO and rAC estimates correlate more consistently with the ICG-derived values when the proper position-specific mapping is used. Figs. 6(d) and (f) illustrate this effect for all subjects in the testing set and with all sensor position and linear mapping combinations. This result corroborates Fig. 5(c) in suggesting that subspace trajectories from a particular position are more similar to those from the same position than to others; therefore, if the position is known, the proper linear mapping $\boldsymbol{X}_P$ can be applied to the subspace $\boldsymbol{F}$ to obtain estimates of the generative factors. In effect, modeling sensor position as a generative factor as shown in Fig. 1(b) enables adaptation to sensor placement by removing the bias in observed dynamics introduced by the sensor's position.

Notably, mismatching the linear model to the true sensor position in Figs. 6(c) and (e) still yielded generative factor estimates that followed the same general trend, though the variance of these trends was higher. This may be because the linear mapping is primarily driven by dimensions in which dynamics are consistent such as f_1 and f_2 in Fig. 4 while the remaining dimensions are used for fine-tuning these estimates.

The above results demonstrate the final step for algorithmic correction of sensor misplacement for rAO and rAC inference. After reducing the dimensionality of SCG signals with SGFE, selecting a position-specific regression model between the latent space and rAO and rAC enables improved estimation of these parameters, as shown in Figs. 6(d) and (e). These results also highlight the clinical application of this work: by inferring these indicators in a manner that is robust to changes in SCG morphology and sensor position, the practicality of using SCG in healthcare settings may be improved.

F. Limitations and Future Work

To achieve the potential clinical applications of this work, future studies should first explore how to optimize this model for rAO and rAC estimation; as optimization of deep learning models is a complex process and largely dependent on the nature of the dataset, this procedure and discussion should be explored at length in future studies. As the focus of this work was model formulation rather than optimization, these hyperparameters were derived heuristically from the results in [30]. Future work should also compare the performance of SGFE-based models to existing methods of rAO and rAC estimation in outpatient and clinical environments and, if possible, employ echocardiography as a gold-standard reference in lieu of ICG. While the sample size of this study was designed for validation of the model, comparisons against other methods will require both optimization of the model and a larger cohort of subjects. More broadly, a key avenue of future work is exploring the role of SCG generative factor modeling in the diagnosis and assessment of disease states. In particular, the underlying dynamics of SCG signals may vary in heart failure patients compared to healthy controls. Elucidating differences in these dynamics may yield a deeper understanding of the effect of heart failure on SCG signals [7], [8].

IV. Conclusion

In seeking to improve the usability of SCG signals in clinical and outpatient environments, this work presented a new method of modeling SCG signals using dynamic and generative modeling. It was shown that SCG signals exhibit consistent behavior

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.

in low dimensions despite morphological variability. Harnessing this result enabled the inference of key cardiomechanical indicators while adapting to inter-subject variability and sensor misplacement. Ultimately, developing SCG processing methods which are robust to these factors may better enable the noninvasive assessment of cardiomechanical function for the diagnosis and management of cardiovascular disease.

REFERENCES

[1] V. Gurev, K. Tavakolian, J. Constantino, B. Kaminska, A. P. Blaber, and N. A. Trayanova, "Mechanisms underlying isovolumetric contraction and ejection peaks in seismocardiogram morphology," *J. Med. Biol. Eng.*, vol. 32, no. 2, pp. 103–110, 2012.

[2] R. Crow, P. J. Hannan, D. R. J. Jr, L. Hedquist, and D. Salerno, "Relationship between seismocardiogram and echocardiogram for events in the cardiac cycle," *Amer. J. Noninvasive Cardiology*, vol. 8, no. 1, pp. 39–46, 1994.

[3] K. Sørensen, S. E. Schmidt, A. S. Jensen, P. Søgaard, and J. J. Struijk, "Definition of fiducial points in the normal seismocardiogram," *Scientific Rep.*, vol. 8, no. 1, 2018, Art. no. 15455.

[4] O. T. Inan *et al.*, "Ballistocardiography and seismocardiography: A review of recent advances," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1414–1427, Jul. 2015.

[5] J. M. Zanetti and K. Tavakolian, "Seismocardiography: Past, present and future," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2013, pp. 7004–7007.

[6] C. Yang and N. Tavassolian, "Pulse transit time measurement using seismocardiogram, photoplethysmogram, and acoustic recordings: Evaluation and comparison," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 733–740, May 2017.

[7] O. T. Inan *et al.*, "Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients," *Circulation Heart Failure*, vol. 11, no. 1, 2018, Art. no. e004313.

[8] M. M. H. Shandhi, J. Fan, J. A. Heller, M. Etemadi, O. T. Inan, and L. Klein, "Seismocardiography and machine learning algorithms to assess clinical status of patients with heart failure in cardiopulmonary exercise testing," *J. Cardiac Failure*, vol. 25, no. 8, pp. S64–S65, 2019.

[9] R. A. Wilson, V. S. Bamrah, J. Lindsay Jr, M. Schwaiger, and J. Morganroth, "Diagnostic accuracy of seismocardiography compared with electrocardiography for the anatomic and physiologic diagnosis of coronary artery disease during exercise testing," *Amer. J. Cardiol.*, vol. 71, no. 7, pp. 536–545, 1993.

[10] M. Etemadi and O. T. Inan, "Wearable ballistocardiogram and seismocardiogram systems for health and performance," *J. Appl. Physiol.*, vol. 124, no. 2, pp. 452–461, 2018.

[11] M. Di Rienzo *et al.*, "Wearable seismocardiography: Towards a beat-by-beat assessment of cardiac mechanics in ambulant subjects," *Autonomic Neurosci.*, vol. 178, pp. 50–59, 2013.

[12] A. Taebi, B. Solar, A. Bomar, R. Sandler, and H. Mansy, "Recent advances in seismocardiography," *Vibration*, vol. 2, no. 1, pp. 64–86, 2019.

[13] H. Ashouri, S. Hersek, and O. T. Inan, "Universal pre-ejection period estimation using seismocardiography: Quantifying the effects of sensor placement and regression algorithms," *IEEE Sensors*, vol. 18, no. 4, pp. 1665–1674, Feb. 2017.

[14] M. C. Khoo, *Physiological Control Systems: Analysis, Simulation and Estimation.* Hoboken, NJ, USA: Wiley, 2000.

[15] L. Ljung and T. Glad, *Modeling of Dynamic Systems*. Englewood Cliffs, NJ, USA: Prentice Hall, 1994.

[16] Y. Miyamoto *et al.*, "Dynamics of cardiac output and systolic time intervals in supine and upright exercise," *J. Appl. Physiol.*, vol. 55, no. 6, pp. 1674–1681, 1983.

[17] H. Ashouri and O. T. Inan, "Automatic detection of seismocardiogram sensor misplacement for robust pre-ejection period estimation in unsupervised settings," *IEEE Sensors*, vol. 17, no. 12, pp. 3805–3813, Jun. 2017.

[18] C. Doersch, "Tutorial on variational autoencoders," Aug. 2016, *arXiv:1606.05908*.

[19] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Advances Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[20] J. Zia *et al.*, "A unified framework for quality indexing and classification of seismocardiogram signals," *IEEE J. Biomed. Health Informat.*, to be published.

[21] V. Zakeri *et al.*, "Preliminary results on quantification of seismocardiogram morphological changes, using principal component analysis," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2014, pp. 6092–6095.

[22] T. Choudhary, M. K. Bhuyan, and L. Sharma, "Orthogonal subspace projection based framework to extract heart cycles from SCG signal," *Biomed. Signal Process. Control*, vol. 50, pp. 45–51, 2019.

[23] C. Yang and N. Tavassolian, "An independent component analysis approach to motion noise cancelation of cardio-mechanical signals," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 784–793, Mar. 2019.

[24] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "Ensemble empirical mode decomposition with principal component analysis: A novel approach for extracting respiratory rate and heart rate from photoplethysmographic signal," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 766–774, May 2018.

[25] M. Chawla, "PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2216–2226, 2011.

[26] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," Dec. 2013, *arXiv: 1312.6114*.

[27] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*.

[29] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," Nov. 2015, *arXiv:1511.07289*.

[30] S. Hersek, B. Semiz, M. M. H. Shandhi, L. Orlandic, and O. T. Inan, "A globalized model for mapping wearable seismocardiogram signals to whole-body ballistocardiogram signals based on deep learning," *IEEE J. Biomed. Health Informat.*, to be published.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[33] M. M. H. Shandhi, B. Semiz, S. Hersek, N. Goller, F. Ayazi, and O. T. Inan, "Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2365–2374, Nov. 2019.

[34] J. Wahlström *et al.*, "A hidden markov model for seismocardiography," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2361–2372, Oct. 2017.

[35] A. Sherwood *et al.*, "Methodological guidelines for impedance cardiography," *Psychophysiology*, vol. 27, no. 1, pp. 1–23, 1990.

[36] P. Carvalho *et al.*, "Robust characteristic points for ICG-definition and comparative analysis," in *Proc. Biosignals*, 2011, pp. 161–168.

[37] I. Higgins *et al.*, "Beta-VAE: Learning basic visual concepts with a constrained variational framework," *Proc. Int. Conf. Learn. Representations*, 2017, vol. 2, p. 6.

[38] M. Li, T. Zhang, Y. Chen, and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 661–670.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*.

[40] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.

[41] W. J. Vincent and J. P. Weir, *Statistics in Kinesiology*, 4th ed. Champaign, IL, USA: Human Kinetics, 2012.

[42] Y. Bengio *et al.*, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

Authorized licensed use limited to: Emory University. Downloaded on September 25,2021 at 17:42:19 UTC from IEEE Xplore. Restrictions apply.