VABAM: Variational Autoencoder for Amplitude-based Biosignal Augmentation within Morphological Identities

Junetae Kim, Kyoungsuk Park, Kyunglim Kim

***Abstract*—Pulsatile physiological signals, characterized by rhythmic fluctuations, are vital for assessing health conditions and are widely used in wellness devices and medical equipment. Despite their significance, models addressing domain-specific unmet needs and considerations have not been developed as much as in other fields. Therefore, building on the foundation of variational autoencoders, we introduce VABAM, a novel model for the amplitude-based synthesis of pulsatile physiological signals. The uniqueness of VABAM lies in its ability to maintain the morphological identity of signals throughout the synthesis process, achieved by integrating pass filter effects within the variational autoencoder architecture. To assess the effectiveness of the model, we developed three novel metrics based on joint mutual information. These metrics were aimed at evaluating the disentanglement of latent spaces, influence of ancillary information on signal morphologies, and controllability of amplitude-based synthesis within morphological identities. Comparative analyses demonstrated that VABAM and its variants were notably effective at preserving morphological integrity, highlighting their potential to minimize morphological distortions in physiological signal processing and their compatibility with artificial intelligence models employing frequency and amplitude features. Additionally, the proposed metrics, compatible with probabilistic models, were empirically proven to capture the characteristics of various models from multiple perspectives.**

***Index Terms*—Frequency pass filters, generative models, mutual information, permutation entropy, pulsatile physiological signals, signal synthesis, variational autoencoders.**

I. INTRODUCTION

PULSATILE physiological signals, including arterial blood pressure (ABP), photoplethysmogram (PPG), and electrocardiogram (ECG), are vital biosignals that are characterized by their rhythmic fluctuations [1]–[3]. These signals are critical for assessing various health conditions, providing essential insights into the physiological activities of the heart and vascular system [1], cardiac function [2], and systemic hemodynamics [3]. Thus, they are employed in both wearable health monitors and medical devices for surgery and critical care [3]–[6]. However, these signals are frequently compromised by noise originating from electrical interference, motion artifacts, and extraneous physiological factors, thereby complicating the extraction of accurate diagnostic information [7], [8].

This challenge can be overcome by applying frequency pass filters that selectively enhance or diminish the amplitude of signals according to their frequency components [9]. For example, low-pass filters are instrumental in removing high-frequency noise [10], [11], whereas high-pass filters are key to eliminating low-frequency drifts and baseline wander [12], [13]. Furthermore, band-pass filters are crucial for isolating specific frequency bands [8], [14]. Therefore, each filter plays a pivotal role in improving signal clarity.

However, these filters often involve convolution, a process that can result in edge effects [9], [15], which lead to signal truncation, potentially compromising important information at the signal boundaries. Although padding strategies can mitigate these effects [16], their implementation can either extend the length of the original signal or cause variations in phase and horizontal shifts along the time-axis. From a physiological perspective, such distortions, including edge effects, phase alterations, and time-axis shifts, can skew the interpretation of physiological activities [9], [17]. Additionally, the design of pass filters with suboptimal frequency cutoffs can distort the morphological characteristics of pulsatile signals [9], [18], especially when the chosen cutoffs do not align precisely with the inherent frequency components of the signal, resulting in the attenuation or amplification of certain aspects of the signal critical for accurate interpretation [9], [18]. In conclusion, these alterations and distortions pose the risk of the misrepresentation of the patient's condition, resulting in inaccurate health assessments.

Thus, in the realm of pulsatile physiological signals, the use of pass filters serves as a necessary but potentially problematic tool: indispensable yet prone to issues if overused. Recognizing this, the development of a synthesis model with high controllability over pass filter effects can offer solutions to the aforementioned challenges from multiple perspectives. Furthermore, these synthesis models can be effectively integrated with rapidly advancing artificial intelligence models designed to utilize the frequency and amplitude-based features of physiological signals [6], [19]. Additionally, by preserving the essence of physiological signals while synthesizing rare events, these models can enhance training performance and contribute to research on physiological dynamics. However, despite considerable advancements that can be made in this field, models that offer precise controllability in amplitude-based biosignal synthesis are lacking, a technology gap

Manuscript received 31 March 2024. This work was supported in part by the National Cancer Center of Korea under Grant 2310800 and 24H1111-1. *(Corresponding author: Kyunglim Kim.)*

Junetae Kim is with the Graduate School of Cancer Science and Policy and the Healthcare AI Team, National Cancer Center, Goyang 10408, Republic of Korea (e-mail: lyjune0070@gmail.com)

Kyoungsuk Park is with the Graduate School of Cancer Science and Policy, National Cancer Center, Goyang 10408, Republic of Korea

Kyunglim Kim is with Samsung Research, Seoul 06765, Republic of Korea (e-mail: eugeniakkim@gmail.com)

This article has a supplementary document provided by the authors

The source code the reproducibility of this work is available at https://github.com/JunetaeKim/VABAM

addressed in this study.

The **main objectives** of this study are as follows:

- We developed a probabilistic deep learning model, named VABAM, that enables the synthesis of pulsatile physiological signals via pass filter effects (i.e., amplitude-based modulation) while ensuring the preservation of the morphological identity of the signals.
- We propose three novel metrics to establish a comprehensive evaluation of the synthesis and representation capabilities of the model: Disentanglement of Z in signal morphology, contribution of ancillary (i.e., conditional) information to signal morphology, and controllability of amplitude-based synthesis within morphological identities.

We define *amplitude-based* as the capability to modulate frequency-amplitudes, a direct consequence of employing frequency pass filters. Additionally, we define *morphological identity preservation* as the absence of phase shifts and horizontal movements along the time axis of the signal.

To achieve precise control in conditional synthesis, we developed VABAM by extending the capabilities of the representative probabilistic model, variational autoencoders (VAEs), which are renowned for their explicit latent space modeling and direct access to encodings [20]. Furthermore, we ensured that the probabilistic nature of VABAM was effectively assessed within a stochastic context by developing metrics based on joint mutual information (JMI) [21]. Moreover, we utilized widely recognized models within the VAE family as benchmarks or combined them with VABAM to develop its derivatives, thereby enabling a comparative analysis of the characteristics of each model.

The notable novelties of VABAM and **major contributions** of our study are outlined as follows:

- **Integration of pass filter methods:** Incorporating domain-critical pass filter effects within a probabilistic model facilitates controlled synthesis via amplitude-based augmentation while offering protection against horizontal shifts and phase alterations in signal morphology.
- **Theoretically robust foundation:** Embedding a probabilistic approach into the core tenets of domain knowledge, model development, and evaluation establishes a coherent and theoretically robust cornerstone, positioning VABAM as a potential foundational model for pulsatile physiological signals.
- **Novel representation of morphological characteristics:** Representing morphological characteristics as random variables via power spectral density (PSD) exemplifies the integration of physiological domain knowledge into probabilistic modeling and metrics, enriching the theoretical and practical applications of the field.
- **Development of a specialized metric for controllability:** Creating a metric to evaluate the controllability of ancillary (i.e., conditional) input-based synthesis lays a solid foundation for probabilistically quantifying the coherence of synthesized pulsatile physiological signals.

The structure of this paper is outlined as follows: Section II provides a comprehensive review of the literature, delineating both the distinctions and connections between our proposed model and existing methodologies. Section III introduces the preliminary concepts necessary for understanding our model, derives the evidence lower bound (ELBO), and discusses its components. Section IV presents three novel JMI-based metrics alongside an additional metric, offering a comprehensive evaluation framework for our study. Section V elaborates on the execution of benchmark tests and ablation studies, providing an assessment and interpretation of findings based on the proposed metrics. Section VI deliberates on key insights and potential areas for improvement, outlining directions for future research.

II. RELATED WORKS

A. Pass Filtering in Physiological Signal Processing

A diverse array of digital pass filters are widely used to refine pulsatile physiological signals. Notably, low-pass filters have been proven effective in mitigating high-frequency noise present in cardiac bioelectrical signals. These filters are typically employed to attenuate frequencies above a specified threshold (e.g., 10 Hz), enhancing the quality of ECGs [10], [11]. Similarly, high-pass filters used in electroencephalography (EEG) have demonstrated efficacy in eliminating low-frequency drifts or baseline wander. By blocking frequencies below a designated threshold (e.g., 1 Hz), these filters enhance the sharpness of brain wave patterns, thereby improving EEG analysis [12], [13]. Additionally, band-pass filters play a crucial role in photoplethysmogram (PPG)-based heart rate monitoring by isolating specific frequency bands. These filters effectively eliminate both high-frequency noise and low-frequency artifacts, ensuring accurate detection of heartbeats [8], [14].

Cascading filters, which extend the concept of frequency pass filters through a series of interconnected stages to systematically refine signals [22], have been instrumental in enhancing patient monitoring and diagnostic accuracy. Each stage in these filters builds upon the output of its predecessor [22], effectively reducing noise and improving signal clarity [5], [23]. Their effectiveness can be attributed to the decomposition of signals into specific subsets, enabling the precise removal of a wide range of unwanted frequencies. The importance of cascading filters extends beyond basic signal processing; their application can be pivotal in the development of medical devices and AI-driven health analytics [24], [25].

B. Physiological Signal Generative Model

In the domain of physiological signals, generative models such as generative adversarial networks (GANs) and VAEs have gained increasing significance. These models are predominantly used for tasks such as augmentation, denoising, and modality transfer. Augmentation models are essential for creating larger, more diverse training datasets, particularly for generating synthetic instances of rare conditions [26], [27]. Denoising models play a crucial role in removing various noise interferences from physiological signals, including baseline wander, muscle artifacts, and environmental noise [28], [29]. Modality transfers facilitate the integration of various signal types, such as converting PPG into ECG signals, which improves analysis and diagnostic capabilities [30].

Although the application of VAEs in this domain is relatively limited compared to that of GANs, they have potential applications in augmentation and imputation tasks. In augmentation, key applications involve extracting typical one-cardiac cycle shapes from ECGs [31], synthesizing ECGs using demographic details such as age and sex [27], and generating ECGs conditioned on four simplified classes, one of which is *low amplitude and high frequency* [32]. For imputation, one approach has effectively merged VAEs with Fourier neural networks for ECGs missing imputation [32]. Another approach has enhanced importance-weighted autoencoders with Gaussian process priors for latent variables, significantly improving the imputation of missing data [4].

C. Variational Autoencoders for Disentangled Representation

In the vast landscape of VAE derivatives, we focus on models distinguished by their efficiency at disentangling representations and exhibiting advanced control in the generative process. Thus, starting with the foundational principles of VAE, we delve into the evolution of related models, spotlighting their key differentiators.

VAE marks a pivotal advancement in generative modeling, introducing a probabilistic approach to encode input data into a latent space [33], [34]. This methodology empowers deep learning models to efficiently generate a diverse array of complex data. At the heart of VAE lies the optimization of the ELBO, which comprises a reconstruction loss and regularization term that encourages the latent space Z to adhere to a normal distribution [33], [34]. This balance between precise data reconstruction and effective latent space organization is paramount for ensuring the practical utility of a model across various applications.

***β*-VAE**, an extension of the VAE framework, introduces a modification to the ELBO by adjusting the weight of the regularization term in Z [35]. This adjustment is facilitated by the introduction of the β factor. Such a refinement in the β-VAE model is instrumental in fostering more distinct and interpretable data representations, particularly effective in disentangling complex latent variable structures within image datasets [35]. This enhances our understanding of feature representation and separation, aligning directly with the goals of precision and efficiency in generative modeling.

Conditional VAE (C-VAE) extends the VAE framework by integrating additional auxiliary data into the generative process [36]. This conditioning mechanism enables the model to generate outputs that closely align with specific attributes or conditions [36]. This model is particularly advantageous in scenarios where controlled generation is paramount, such as targeted data augmentation or the development of models for specific, scenario-based applications.

Total Correlation variational autoencoder (TC-VAE) represents an innovative extension within the β-VAE framework, focusing on minimizing the total correlation (TC), a generalization of MI, in the latent space [37]. This strategy is aimed at promoting the independence of latent variables, which can significantly advance the disentanglement process. By reducing inter-variable dependencies, TC-VAE enhances the clarity and distinction in data representation, effectively capturing the intrinsic structure of the dataset [37]. This improvement in TC-VAE underscores its contribution to improving the interpretability and utility of generative models.

FactorVAE (FAC-VAE) adopts a distinct strategy for disentanglement within the β-VAE framework [38]. While sharing objectives with TC-VAE, FAC-VAE introduces a specific term in the loss function to penalize the total correlation between latent variables. This approach, which requires an auxiliary discriminator network and the density ratio trick, effectively isolates independent factors in the data [38]. FAC-VAE thus contributes to a nuanced understanding of complex data structures, offering a unique method for disentanglement.

D. Mutual Information in Disentangled Representation

Mutual information (MI) stands as a pivotal metric in information theory, quantifying the knowledge acquired about a random variable by observing another, thereby effectively capturing their statistical dependence [21]. Within the realm of deep learning, MI plays a crucial role in developing disentangled representations, which endeavor to segregate distinct aspects of the data into independent components, and in assessing the extent of disentanglement among variables.

Among representative models stemming from generative adversarial networks (GAN), Deep InfoMax (DIM) harnesses a specialized estimation to maximize MI between input data and high-level representations. It offers an adjustable focus on global or local information to tailor representations for tasks such as classification or reconstruction [39]. Similarly, InfoGAN employs an adversarial framework to maximize MI between latent factors and generated data, thereby establishing a lower bound for efficient optimization, facilitating the generation of interpretable representations [40].

While the previously introduced FAC-VAE and TC-VAE are aimed at minimizing TC to foster disentangled representations [37], [38], TC-VAE advances this approach further. It decomposes Kullback–Leibler divergence (KLD) terms to not only isolate the TC term but also directly incorporate the index-code MI between the data and latent variable into the objective function [37]. Moreover, this study introduces the mutual information gap as a means to evaluate the degree of disentanglement.

III. METHODS

A. Preliminary

1) ELBO in β-VAE: VABAM diverges from β-VAE, embarking on a unique trajectory separate from FAC-VAE and TC-VAE. While VABAM shares similarities with C-VAE, its structure distinctly sets it apart. This differentiation necessitates a foundational grasp of the β-VAE ELBO, which encompasses the core tenets of the standard VAE ELBO. β-VAE introduces a hyperparameter β, enabling a balance between the adherence of Z to a normal distribution and reconstruction fidelity [35]. The ELBO in the context of β-VAE is formulated as follows:

$$\text{ELBO}_\beta = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \beta\text{KLD}(q_\phi(z|x)\|p(z)) \quad (1)$$

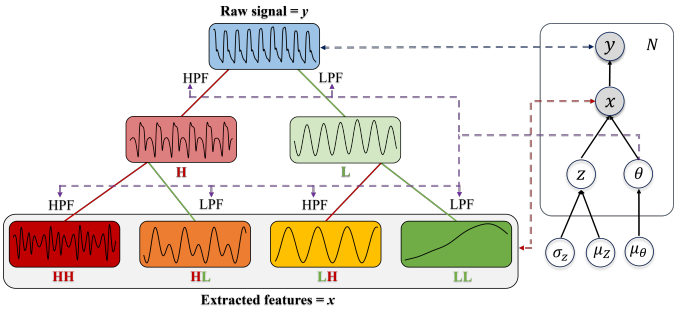

Fig. 1. Illustration of Cascading Filters Corresponding to Generative Process.

This formulation comprises two components: expected log-likelihood of the data given the latent variable, ensuring accurate data reconstruction, and a β-weighted KLD, thereby imposing a weighted regularization on the latent space.

2) Cascading Filters: This method begins with defining the low-pass filter (LPF) and high-pass filter (HPF) as follows:

$$\mathrm{LPF}[t] = \frac{S[t] \cdot w[t]}{\sum_{i=0}^{T-1} S[i] \cdot w[i]}, \tag{2}$$

$$\mathrm{HPF}[t] = \delta[t] - \mathrm{LPF}[t]. \tag{3}$$

Here, $S[t]$, $w[t]$, and $\delta[t]$ respectively denote the sinc filter, the Hamming window [41], and the discrete unit impulse function, defined as follows:

$$S[t] = \mathrm{sinc}\left(2\Theta\left(t - \frac{T-1}{2}\right)\right), \tag{4}$$

$$w[t] = 0.54 - 0.46\cos\left(\frac{2\pi t}{T}\right), \tag{5}$$

$$\delta[t] = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases}. \tag{6}$$

In these formulations, t and T correspond to the time and the filter size, respectively. Applying a Hamming window to a filter primarily aims to attenuate sidelobes in the frequency domain, yielding a smoother signal spectrum [41]. Θ, functioning as the cutoff frequency to filter out higher frequencies, represents a neural network weight that influences backpropagation [6].

The operation of these filters is illustrated in Figure 1 [22]. Essentially, VABAM employs a two-stage cascading filter framework to systematically break down the original signal into four sub-signals. This breakdown is achieved through convolution operations between the pass filters and the signals from preceding stages, with each resulting sub-signal undergoing specific frequency filtering.

B. Training and Generating Framework

VABAM comprises five integral components: Feature extractor, encoder, sampler, feature generator, and signal reconstructor (Fig. 2). Each component works in synergy to contribute to the overall functionality of the model, ensuring the effective synthesis of pulsatile physiological signals.

Feature Extractor $g_x(\cdot)$ applies cascading filters to the raw input signal y, segmenting it into four amplitude-modulated subsets $x \in \{x_{HH}, x_{HL}, x_{LH}, x_{LL}\}$. These subsets serve as learning targets for the feature generator.

Encoder $g_e(\cdot)$ is responsible for learning the parameters defining the distributions of the latent variable, Z and the cutoff frequency, Θ. It operates under two assumptions:

- $\theta_k \sim \mathcal{U}(0,1)$ for $k = 1, \ldots, 6$, representing six instances within the proposed model structure. This distribution is approximated by a Bernoulli distribution (refer to Eq.(20)).
- $z_j \sim \mathcal{N}(\mu_{z_j}, \sigma^2_{z_j})$ for each dimension j, with $j \in \{1, 2, \ldots, J\}$, where J is a hyperparameter determining the number of dimensions.

Sampler $g_z(\cdot)$ and $g_\theta(\cdot)$ facilitate a differentiable approximation of the data distribution. It employs the reparameterization trick for backpropagation, where z_j and θ_k are sampled to enable gradients to flow through the network node.

Feature Generator $g_{x'}(\cdot)$ utilizing parameters from the encoder, generates four feature signals that serve as inputs for the signal reconstructor. These sets are tailored to match the amplitude-modulated subsets processed by the feature extractor while integrating latent elements parameterized by z_j and θ_k.

Signal Reconstructor $g_y(\cdot)$ utilizes the generated feature subsets to reconstruct signals, preserving the fundamental aspects of the original signal.

C. Evidence Lower Bound (ELBO) Derivation

To clarify the ELBO, standard notation conventions are utilized, with lowercase letters denoting realized values and subscripts omitted for brevity. However, for seamless integration with the subsequent discussion on JMI, uppercase letters are employed to denote random variables or model parameters.

Based on the assumptions of the generative process in VABAM, we express the joint probability as follows:

$$p(y, x, z, \theta) = p(y \mid x)\, p(x \mid z, \theta)\, p(z)\, p(\theta). \tag{7}$$

The marginal likelihood $p(y)$ is derived by integration over variables x, z, and θ:

$$p(y) = \iiint_D p(y, x, z, \theta)\, dx\, dz\, d\theta. \tag{8}$$

To enable variational inference, we introduce a distribution $q(x, z, \theta \mid y)$ that approximates the true posterior distribution $p(x, z, \theta \mid y)$. Consequently, we get

$$p(y) = \iiint_D \frac{p(y, x, z, \theta)\, q(x, z, \theta \mid y)}{q(x, z, \theta \mid y)}\, dx\, dz\, d\theta. \tag{9}$$

Applying Jensen's inequality to the integral, we derive the evidence lower bound (ELBO) as follows:

$$\log p(y) \geq \mathbb{E}_{q(x,z,\theta \mid y)}\left[\log \frac{p(y, x, z, \theta)}{q(x, z, \theta \mid y)}\right] = ELBO(y). \tag{10}$$

According to our model structure and its generative process, we assume that the distribution $q(x, z, \theta \mid y)$ can be factorized as follows:

$$q(x, z, \theta \mid y) = q(x \mid y, \theta)\, q(z, \theta \mid y), \tag{11}$$

where $q(x \mid y, \theta)$ is related to the feature extractor, and $q(z, \theta \mid y)$ is associated with both the encoder and sampler.

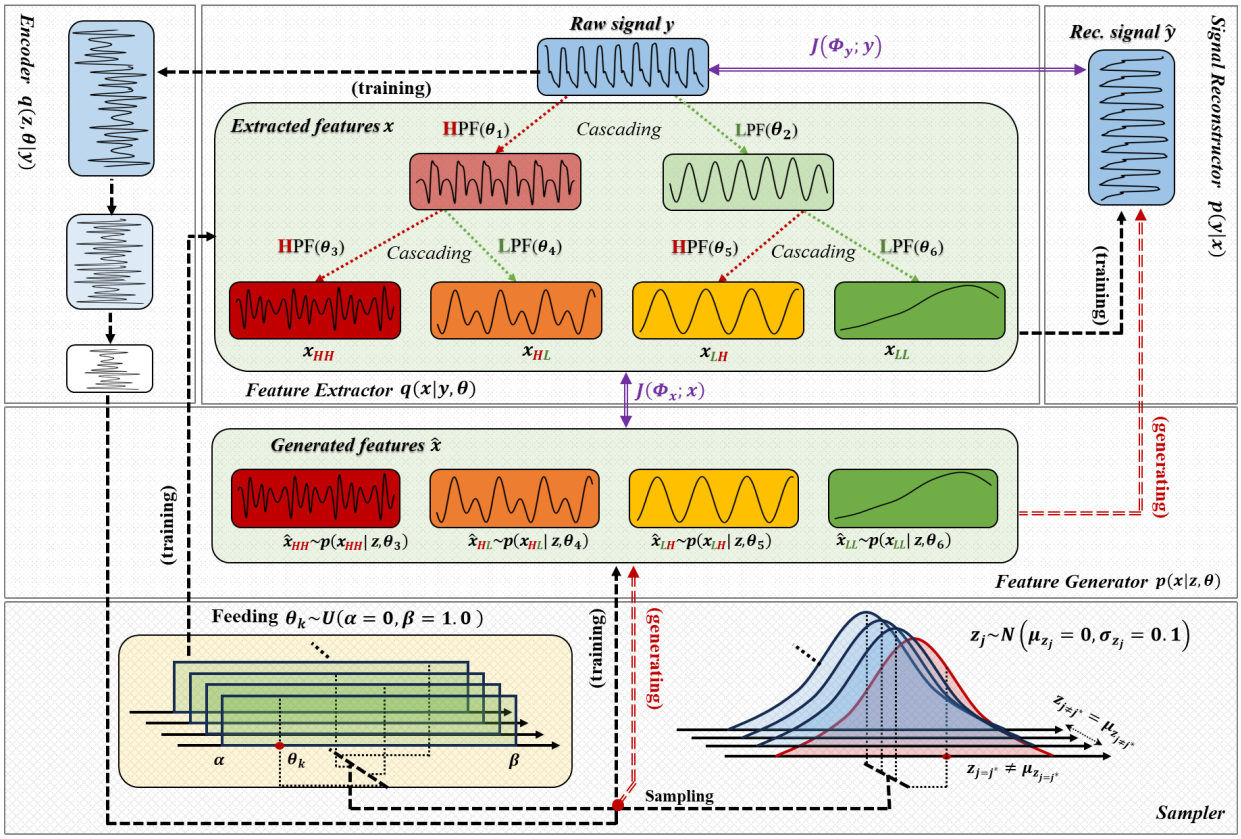

Fig. 2. Intuitive Illustration of Generation and Training Process.

This implies that feature signals x are modulated through the cascading filter, given the θ drawn from the sampler and the raw signal y (Fig.1).

The ELBO is then expressed as follows:

$$\begin{aligned} ELBO(y) = \mathbb{E}_{q(x,z,\theta\,|\,y)}\big[&\log p(y\,|\,x) + \log p(x\,|\,z,\theta) \\ &+ \log p(z) + \log p(\theta) - \log q(x\,|\,y,\theta) \\ &- \log q(z\,|\,y) - \log q(\theta\,|\,y)\big]. \end{aligned} \tag{12}$$

The final objective function is defined as follows:

$$\begin{aligned} J(\Phi; y, x, z, \theta) = \arg\min_{\Phi} \mathbb{E}_{q(x,z,\theta|y)} \Big[&-\log p(y|x) \\ &-\log \frac{p(x|z,\theta)}{q(x|y,\theta)} + \log \frac{q(z|y)}{p(z)} + \log \frac{q(\theta|y)}{p(\theta)} \Big], \end{aligned} \tag{13}$$

where Φ represents all trainable network weights.

D. ELBO Terms

1) Reconstruction of Raw Signals: The expected log-likelihood for the reconstruction of raw signals, y given reconstructed feature signals, x can be approximated using the Monte Carlo (MC) method as follows:

$$-\mathbb{E}_{q(x,z,\theta|y)} \left[\log p(y|x)\right] \approx -\frac{1}{L} \sum_{l=1}^{L} \log p(y|x^{(l)}), \tag{14}$$

where L is the number of MC samples. To ensure convergence stability, we minimized the mean squared error (MSE) rather than the negative log likelihood from Eq.(14) as follows:

$$J(\Phi_y; y) = \frac{1}{N} \sum_{n=1}^{N} \big(y_n - g_y(x_n)\big)^2, \tag{15}$$

with n and N denoting the sample index and the total number of training samples, respectively.

2) Reconstruction of Feature Signals: The reconstruction of feature signals involves computing the expected log ratio of the probabilities of the model and those of the variational posterior by minimizing the discrepancy between the two distributions using KLD, as expressed in

$$\begin{aligned} &-\mathbb{E}_{q(z,\theta|y)q(x|y,\theta)} \left[\log \frac{p(x|z,\theta)}{q(x|y,\theta)}\right] \\ &\approx \frac{1}{L} \sum_{l=1}^{L} \mathrm{KLD} \left(q(x|y,\theta^{(l)}) \,||\, p(x|z^{(l)},\theta^{(l)})\right), \end{aligned} \tag{16}$$

where $p(x|z,\theta)$ corresponds to the feature generator.

Considering x, which comprises of four amplitude-modulated sub-signals with real numbers, we posit that both $q(x|y,\theta^{(l)})$ and $p(x|z^{(l)},\theta^{(l)})$ follow Gaussian distributions. Thus, we approximate Eq.(16) using MSE as follows:

$$J(\Phi_x; x) = \frac{1}{N} \sum_{n=1}^{N} \big(g_x(y_n,\theta_n) - g_{x'}(z_n,\theta_n)\big)^2. \tag{17}$$

3) Variational Inference for $p(z)$: Optimization of the latent variable Z involves reducing the KLD relative to the prior:

$$\mathbb{E}_{q(z|y)q(\theta|y)q(x|y,\theta)}\left[\log\frac{q(z|y)}{p(z)}\right] = \text{KLD}(q(z|y)||p(z)). \quad (18)$$

For this optimization, we applied the KLD (denoted as $J(\Phi_z; z)$) specific to Gaussian distributions, along with a re-parametrization trick commonly used in VAEs [33], [34].

4) Variational Inference for $p(\theta)$: Similarly, the optimization of the cutoff frequency for the pass filter, Θ is achieved by minimizing the KLD:

$$\mathbb{E}_{q(z|y)q(\theta|y)q(x|y,\theta)}\left[\log\frac{q(\theta|y)}{p(\theta)}\right] = \text{KLD}(q(\theta|y)||p(\theta)). \quad (19)$$

While θ inherently follows uniform distributions, the KLD (denoted as $J(\Phi_\theta; \theta)$) is optimized under the presumption that θ adhere to Bernoulli distributions [42], [43], to execute the reparameterization trick, as stated in the following formula:

$$\begin{aligned}\text{KLD}(q(\theta|y)||p(\theta)) &\approx \text{KLD}(q(\theta|y)||\text{Bern}(\mu_\theta = 0.5)) \quad (20)\\ &= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{c\in\{0,1\}}\left(\mu_{\theta_{n,k}}^{c}(1-\mu_{\theta_{n,k}})^{1-c}\right.\right.\\ &\quad\left.\left.\times\ln\frac{\mu_{\theta_{n,k}}^{c}(1-\mu_{\theta_{n,k}})^{1-c}}{0.5^{c}0.5^{1-c}}\right)\right), \quad (21)\end{aligned}$$

with K set to 6 according to the structure of VABAM.

IV. Evaluation Metrics

A. Preliminary

1) Dimensional isolation in Z: We introduce a specialized approach to support the evaluation of disentanglement in Z, specifically designed to isolate a randomly selected dimension within the latent space. To this end, we define a set $z = \{z_j\}_{j=1}^{J}$, where z_j follows a normal distribution, $z_j \sim \mathcal{N}(\mu_j = 0, \sigma_j^2)$. Given the set z, we introduce a new set $\acute{z} = \{\acute{z}_j\}_{j=1}^{J}$, where each element $\acute{z}_j$ is defined as

$$\acute{z}_j = \begin{cases} z_{j^*}, & \text{if } j = j^* \\ \mathbb{E}_q[z_j|y], & \text{if } j \neq j^*. \end{cases} \quad (22)$$

Here, j is an index within the set $\{1, 2, \ldots, J\}$ and j^* is an index selected uniformly at random from the same range, with the probability $p(j = j^*) = \frac{1}{J}$. Considering the expected value $\mathbb{E}_q[z_j|y]$, articulated as

$$\mathbb{E}_q[z_j|y] \approx \frac{1}{M}\sum_{m=1}^{M} z_j^{(m)} \approx 0, \quad (23)$$

in the set $\acute{z}$, every element except for z_{j^*} approximates the mean of the Gaussian distribution, which is zero. However, the z_{j^*} retains its original value from the set z, thereby achieving its isolation within the latent space.

2) Power Spectral Density: We utilize PSD to represent signal morphology as random variables, thereby incorporating physiological characteristics into our proposed probabilistic metrics. PSD, which characterizes the distribution of power across various frequencies in a signal [44], helps identify patterns and irregularities, thereby facilitating the interpretation of its morphological features. It is defined as

$$\text{PSD}_{n,m}(v) = \frac{1}{T}\left|X_{n,m}(v)\right|^2, \quad (24)$$

where n, m, and T denote the sample index, generation index, and the time length of the signal, respectively. The term $X_{n,m}(v)$ represents the frequency domain representation of the signal, obtained through the FFT. We further refine this measure by normalizing PSD, denoted as $q_{n,m}(v)$, which provides a more nuanced understanding of power distribution via the equation below:

$$q_{n,m}(v) = \frac{\text{PSD}_{n,m}(v)}{\frac{1}{|V|}\sum_{v\in V}\text{PSD}_{n,m}(v)} \quad (25)$$

3) Permutation Density in Power Spectral Density: We introduce a novel metric to gauge the controllability of signal synthesis, building upon the previously discussed PSD. This metric evaluates the orderliness of synthesized signals by analyzing the permutation distribution of PSD (PD-PSD) values, particularly focusing on the level of structured generation influenced by organized ancillary inputs. The development of this metric follows the procedures outlined below.

① *Defining PSD series:* We initiate the procedure by revisiting and extending the previously defined PSD:

$$\text{PSD} = \{\text{PSD}_{n,v,1}, \text{PSD}_{n,v,2}, \ldots, \text{PSD}_{n,v,M}\},$$

where n, v, and M represent the sample index, the frequency index, and the generation size, respectively.

② *Constructing windowed vectors:* We proceed by constructing W-dimensional vectors for a selected window dimension $W \geq 2$ as follows:

$$\text{WPSD}_t = (\text{PSD}_{n,v,t}, \text{PSD}_{n,v,(t+1)}, \ldots, \text{PSD}_{n,v,(t+W-1)})$$
where $t \in \{1, \ldots, M - W + 1\}$.

③ *Generating permutation sets:* We consider permutation sets, which comprise $W!$ unique ordinal sequences. Each permutation π_s is defined as

$$\pi_s \in \{r|r \text{ is the } s\text{-th permutation of } \{1, 2, \ldots, W\}\}.$$

These sets are crucial for measuring the order of components within WPSD_t.

④ *Tallying occurrences:* We tally occurrences of each permutation π_s as $\Pi_n(s, v)$, by associating every WPSD_t with its respective permutation. This step captures the controllability of signal synthesis.

⑤ *Defining permutation density:* Finally, we define the permutation density as the joint probability of the

TABLE I
DESCRIPTION OF THREE JMI-BASED METRIC

Metric	Baseline	Key	Measure
$I(V;\acute{Z},Z)$	$I(V;Z)$	$I(V;\acute{Z}\|Z)$	Disentanglement of Z in morphology
$I(V;\acute{\Theta},\acute{Z})$	$I(V;\acute{Z})$	$I(V;\acute{\Theta}\|\acute{Z})$	Contribution of Θ to morphology
$I(S;\acute{\Theta},\acute{Z})$	$I(S;\acute{Z})$	$I(S;\acute{\Theta}\|\acute{Z})$	Controllability of Θ-based synthesis

occurrence $\Pi_n(s,v)$ according to v and s relative to the total count of all permutations as

$$q_n(s,v) = \frac{\Pi_n(s,v)}{\sum_{v=1}^{|V|}\sum_{s=1}^{W!}\Pi_n(s,v)}. \tag{26}$$

4) Joint Mutual Information: The concept of conditional mutual information (CMI) is defined as $I(X;Y|Z)$, quantifying the extent to which the information about X is enhanced by introducing Y into the context already established by Z [45]. Applying the chain rule, CMI can be expressed in the form of JMI as $I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$ [45].

This expansion allows for a more nuanced examination of MI, distinctly separating the shared information in X and Z from the information in X and the added Y given the existing Z. In this context, we designate $I(X;Z)$ as the baseline and $I(X;Y|Z)$ as the key indicator (Tab. I). Through this lens, we comprehensively evaluate the characteristics of our proposed model, alongside those of benchmark and ablation models. This approach not only enhances our understanding of the model's performance but also elucidates the individual contributions of both the baseline and the primary indicator to overall performance. Consequently, our analysis prioritizes the scrutiny of the specific terms $I(X;Z)$ and $I(X;Y|Z)$ over the combined metric $I(X;Y,Z)$.

B. Evaluation Metrics

We introduce novel JMI-based metrics to assess the disentanglement of the latent variable Z in signal morphology [37], [46], and to evaluate the contribution of Θ (i.e., the cutoff frequency in VABAM) to signal morphology by quantifying the information gain from the ancillary exogenous variable Θ to V (Tab. I). Additionally, we assess the controllability of amplitude-based synthesis within morphological identities by quantifying the orderliness of the PSD in signals generated from an ordered sequence of Θ, termed $\acute{\Theta}$ (Tab. I).

The following notations and concepts are crucial for understanding these JMI-based metrics: V represents random variables associated with frequency indices in the normalized (PSD); Z is a random variable within a Gaussian distribution; $\acute{Z}$ denotes its isolated variant; Θ and $\acute{\Theta}$ correspond to random variables for conditional input data and its ordered form, respectively. The corresponding lowercase variables signify the realizations of these random variables.

When calculating JMI-based metrics, it is necessary to approximate expectations. This was achieved by conducting Monte Carlo simulations (MCS) [37] through the VABAM sampler. Although the subsequent equations are presented from the perspective of individual samples, the empirical approach is structured at the batch level to enhance computational efficiency. The dimensions of key metrics V and S, and latent variables Z and Θ, vary among the six JMI-based metrics. Additionally, the sampling strategies are distinct for each metric. The supplementary document covers these details, encompassing the analytic derivation, approximations, dimensionalities, and sampling strategies.

1) $I(V;\acute{Z},Z)$: This metric, which captures how much information V contains about both $\acute{Z}$ and Z jointly, can be decomposed as $I(V;\acute{Z},Z) = \underbrace{I(V;Z)}_{\text{(i)}} + \underbrace{I(V;\acute{Z}|Z)}_{\text{(ii)}}$.

The term (i) quantifying the intrinsic dependency within the joint distribution $q(v,z)$ against the marginal distributions $q(v)$ and $q(z)$ under the independence assumption, is articulated as follows:

$$\text{(i)}\ I(V;Z) = \mathbb{E}_{q(v,z)}\left[\log\frac{q(v,z)}{q(v)q(z)}\right] \tag{27}$$

$$\approx \mathbb{E}_{q(z)}\left[\mathrm{KLD}\Big(q(v|z)\parallel p(v)\Big)\right]. \tag{28}$$

The term (ii) measures the disentanglement in Z, as it quantifies the information gained by V when the existing dimensions of Z are probabilistically isolated along one of their axes $\acute{Z}$. It can be quantified based on the ratio between the joint probability $q(v,\acute{z}|z)$ and the product of the conditional probabilities $q(v|z)$ and $q(\acute{z}|z)$ as follows:

$$\text{(ii)}\ I(V;\acute{Z}|Z) = \mathbb{E}_{q(v,\acute{z},z)}\left[\log\frac{q(v,\acute{z}|z)}{q(v|z)q(\acute{z}|z)}\right] \tag{29}$$

$$= \mathbb{E}_{q(\acute{z},z)}\left[\mathrm{KLD}\Big(q(v|\acute{z})\parallel q(v|z)\Big)\right]. \tag{30}$$

2) $I(V;\acute{Z},\acute{\Theta})$: This metric reflects the extent of information V holds jointly about $\acute{\Theta}$ and $\acute{Z}$, and is decomposable as follows: $I(V;\acute{Z},\acute{\Theta}) = \underbrace{I(V;\acute{Z})}_{\text{(iii)}} + \underbrace{I(V;\acute{\Theta}|\acute{Z})}_{\text{(iv)}}$.

The term (iii) centers on the intrinsic dependency present in the joint distribution $q(v,\acute{z})$ relative to the marginals $q(v)$ and $q(\acute{z})$, highlighting $\acute{Z}$ as an isolated dimension of Z.

$$\text{(iii)}\ I(V;\acute{Z}) = \mathbb{E}_{q(v,\acute{z})}\left[\log\frac{q(v,\acute{z})}{q(v)q(\acute{z})}\right] \tag{31}$$

$$\approx \mathbb{E}_{q(\acute{z})}\left[\mathrm{KLD}\Big(q(v|\acute{z})\parallel p(v)\Big)\right]. \tag{32}$$

The term (iv) focuses on the additional information about V, which is acquired by introducing $\acute{\Theta}$ to an existing $\acute{Z}$. This term is quantified based on the relationship between the joint probability $q(v,\acute{\theta}|\acute{z})$ and the product of the conditional probabilities $q(v|\acute{z})$ and $q(\acute{\theta}|\acute{z})$, as follows:

$$\text{(iv)}\ I(V;\acute{\Theta}|\acute{Z}) = \mathbb{E}_{q(v,\acute{\theta},\acute{z})}\left[\log\frac{q(v,\acute{\theta}|\acute{z})}{q(v|\acute{z})q(\acute{\theta}|\acute{z})}\right] \tag{33}$$

$$= \mathbb{E}_{q(\acute{\theta},\acute{z})}\left[\mathrm{KLD}\left(q(v|\acute{\theta},\acute{z})\parallel q(v|\acute{z})\right)\right]. \tag{34}$$

3) $I(S;\acute{Z},\acute{\Theta})$: This measure assesses the control of amplitude-based synthesis as affected by $\acute{Z}$ and the addition of $\acute{\Theta}$ to $\acute{Z}$, decomposed as follows: $I(S;\acute{Z},\acute{\Theta}) = \underbrace{I(S;\acute{Z})}_{\text{(v)}} + \underbrace{I(S;\acute{\Theta}|\acute{Z})}_{\text{(vi)}}$

The term (v), focused on evaluating the controllability of amplitude-based synthesis under the condition of provided $\acute{Z}$, given as follows:

$$\text{(v)}\ I(S;\acute{Z}) = \mathbb{E}_{q(s|\underline{\acute{z}})q(\underline{\acute{z}})}\left[\log\frac{q(s,\underline{\acute{z}})}{q(s)q(\underline{\acute{z}})}\right] \tag{35}$$

$$= \mathbb{E}_{q(\underline{\acute{z}})}\left[\mathrm{KLD}\Big(q(s|\underline{\acute{z}}) \parallel q(s)\Big)\right]. \tag{36}$$

Here, $\underline{\acute{z}}$ is conceptualized as a sequence of M realized instances of $\acute{z}$, represented by $\underline{\acute{z}} = \{\acute{z}_m\}_{m=1}^{M}$.

The term (vi) assesses the controllability of amplitude-based synthesis by modulating $\acute{\Theta}$ while holding $\acute{Z}$ fixed, implying that the synthesis is influenced by adjustments in the variable. This influence can be verified by observing the entropy changes in the PSD of signals synthesized upon introducing $\acute{\Theta}$ to $\acute{Z}$.

$$\text{(vi)}\ I(S;\acute{\Theta}|\acute{Z}) = \mathbb{E}_{q(s,\underline{\acute{\theta}},\underline{\acute{z}})}\left[\log\frac{q(s,\underline{\acute{\theta}}|\underline{\acute{z}})}{q(s|\underline{\acute{z}})q(\underline{\acute{\theta}}|\underline{\acute{z}})}\right] \tag{37}$$

$$= \mathbb{E}_{q(\underline{\acute{\theta}},\underline{\acute{z}})}\left[\mathrm{KLD}\left(q(s|\underline{\acute{\theta}},\underline{\acute{z}}) \parallel q(s|\underline{\acute{z}})\right)\right]. \tag{38}$$

Analogous to $\underline{\acute{z}}$, $\underline{\acute{\theta}}$ is defined as a series of M realized instances of $\acute{\theta}$, expressed as $\underline{\acute{\theta}} = \{\acute{\theta}_m\}_{m=1}^{M}$. Further details are provided in the supplementary document.

C. Filtering High-Quality Synthesis

Capturing and extracting realizations of Z and Θ, which contribute to high-quality synthesis, provide valuable information. Consequently, we devise specific metrics and procedures for this filtering process and subsequently verify the PSD similarity between the filtered synthesis signals and the original signals.

Building upon the preliminary steps for JMI-based metrics, we introduce the normalized PSD for $q(v|z^{(d,m)},\theta^{(d,m)})$ and $Q(v|y_{d'})$. Here, d' denotes a sample index within a batch D, while m represents the sampling index, with M being the total number of sampled instances. Additionally, Y signifies the set of the original signals, implying that $y_{d'}$ is an element of D such that $y_{d'} \in D \subseteq Y$.

The subsequent step involves quantifying the synthesis quality during the filtering process. This begins with the definition of the $\mathrm{KLD}_{d,m,d'}$ as follows:

$$\mathrm{KLD}_{d,m,d'}(Q(v)||q(v)) := \sum_{v\in V} Q(v|y_{d'})\log\left(\frac{Q(v|y_{d'})}{q(v|z^{(d,m)},\theta^{(d,m)})}\right). \tag{39}$$

Here, the indices d, m, and d' are aligned to traverse their respective dimensions, ensuring all indices d', d and m are matched correctly for the element-wise multiplication. Then, we define the key metric Ω_i as the minimum KLD across dimension d' to focus on identifying the best d' for fixed d, m, given as

$$\Omega_i = \min_{d'} \mathrm{KLD}_{d,m,d'}(Q(v)||q(v)) \tag{40}$$

where i combines indices d and m, represented as $i \in \{(d,m) \mid d \in \{1,2,\ldots,|D|\}, m \in \{1,2,\ldots,M\}\}$.

Subsequently, by leveraging Algorithm 1 with Ω and other parameters, we acquire the set of realized z, denoted by $\vec{z}$, and the set of observed θ, represented by $\vec{\theta}$, both of which collectively contribute to the generation of high-quality signals. Intuitively, the algorithm processes each batch of $|D|$ samples through M synthesis attempts, extracting only those realized z and θ pairs that demonstrate a distribution similarity above a specified threshold with the PSD of the original signal.

Algorithm 1 Selecting Ω, z, and θ

Input:

- $\Omega_i = \begin{cases}\Omega_i(v|z^{(d,m)},\theta^{(d,m)}) & \text{w/ ancillary data}\\ \Omega_i(v|z^{(d,m)}) & \text{w/o ancillary data}\end{cases}$
- $V_i^{max} = \begin{cases}\mathrm{argmax}_v\, q(v|z^{(d,m)},\theta^{(d,m)}) & \text{w/ ancillary data}\\ \mathrm{argmax}_v\, q(v|z^{(d,m)}) & \text{w/o ancillary data}\end{cases}$
- z_i^*: realized z values at the current MCS.
- θ_i^*: realized θ values at the current MCS.
- τ: threshold of Ω for selecting z_i^* and θ_i^*.

Initialization: $\vec{z},\vec{\theta} \leftarrow \emptyset$

Procedure:

for each $l \in \{1,2,\ldots,L\}$ (MCS Iteration) **do**
  $I \leftarrow \{(d,m)|d \in \{1,2,\ldots,|D|\}, m \in \{1,2,\ldots,M\}\}$
  for each v in the frequency domain **do**
    $I^v \leftarrow \{i \in I | V_i^{max} = v\}$
    $i^\Omega \leftarrow \mathrm{argmin}_{i\in I^v}\Omega_i$
    if $\Omega_{i=i^\Omega} < \tau$ **then**
      $\vec{z} \leftarrow \mathrm{CONCAT}(\vec{z}, z^*_{i=i^\Omega})$
      $\vec{\theta} \leftarrow \mathrm{CONCAT}(\vec{\theta}, \theta^*_{i=i^\Omega})$ w/ ancillary data
    end if
  end for
end for

Return $\vec{z}$ and $\vec{\theta}$

w/o ancillary data pertains benchmark models operating without ancillary (i.e., conditional) data inputs. In our experiments, τ was set to 1.

Ultimately, we evaluate the overall quality of the filtered-synthesized signals using the filtering quality index (FQI) proposed as follows:

$$FQI = \frac{1}{2}\Big(\mathrm{KLD}(P(v|y)||Q(v|\vec{z},\vec{\theta})) + \mathrm{KLD}(Q(v|\vec{z},\vec{\theta})||P(v|y))\Big). \tag{41}$$

Here, $P(v|y)$ represents the aggregated normalized PSD from the true signal, expressed as $P(v|y) = \frac{1}{N}\sum_{n=1}^{N} P(v|y_n)$, where y_n denotes the n-th original signal. Similarly, $Q(v|\vec{z},\vec{\theta})$ signifies the aggregated normalized PSD from the extracted $\vec{z}$ and $\vec{\theta}$, defined as $Q(v|\vec{z},\vec{\theta}) = \frac{1}{|\vec{z}|}\sum_{i=1}^{|\vec{z}|} Q(v|\vec{z}_i,\vec{\theta}_i)$.

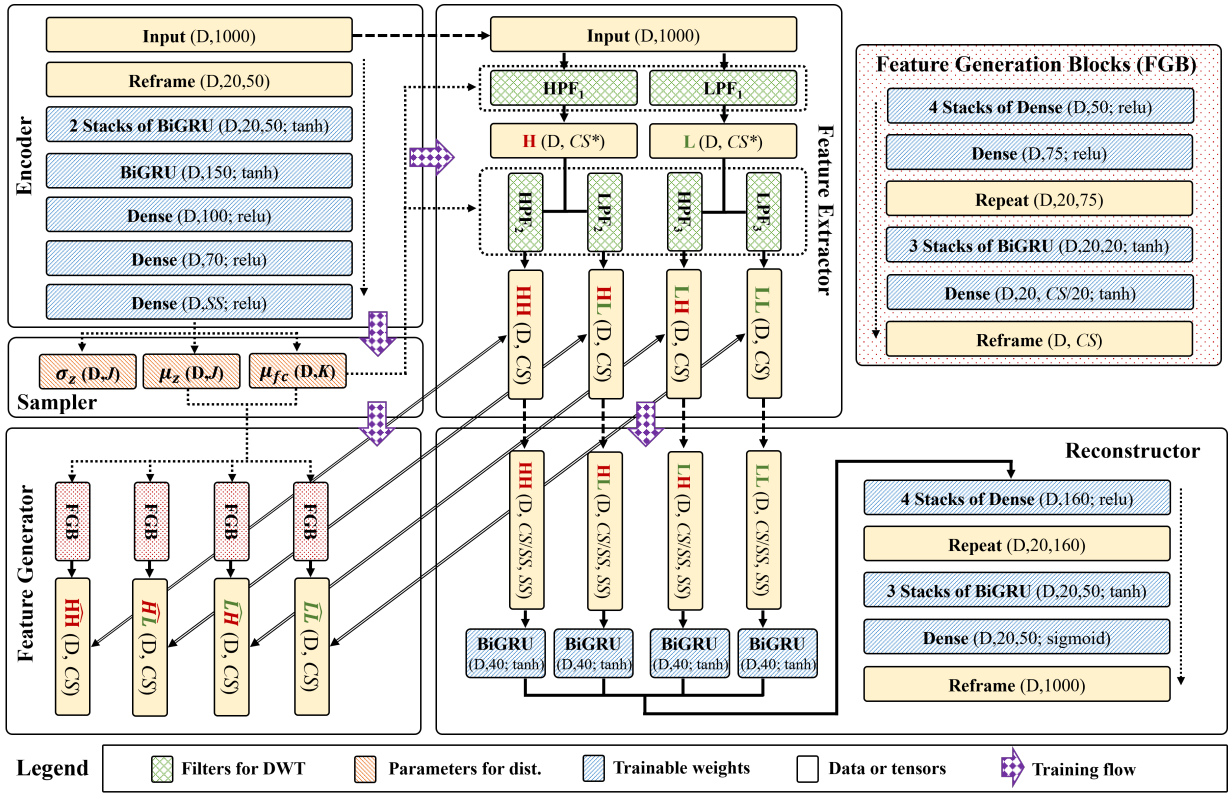

Fig. 3. Illustration of Training Model Structure; CS denotes the signal's compressed size, CS^* is defined as $CS^* = SL - \lfloor (SL - CS)/2 \rfloor + 1$, with SL representing the signal duration in seconds for a 100 Hz sampling rate, and SS the sliding size. Additionally, D, J, and K denote the data size in a batch, and the sizes of dimensions Z and θ, respectively.

V. EXPERIMENTS AND RESULTS

A. Quantitative Evaluation Approaches and Strategies

1) Evaluation Overview: We comprehensively evaluated VABAM's performance by comparing it to several benchmark models known for their expertise in disentanglement and conditional generation. Additionally, we included variants of VABAM, which were developed by integrating the objective functions from these benchmark models, in the analysis (Tab. II). Our evaluation also included ablation studies, primarily focusing on the dimensional size of the latent variables Z and the length of the decomposed feature x via cascading filters (i.e., the compression size).

2) Model Implementation and Adaptation: VABAM and its variants were implemented within a neural network architecture primarily composed of multiple bidirectional GRU layers (Fig. 3). The evaluation involved models tailored for two pulsatile signals: ABP and ECG. We configured the models to utilize 10 s data segments sampled at 100 Hz (100 samples per second) for each signal type. Subsequently, the benchmark models were restructured to align with the VABAM settings by adjusting their core objective functions. This adjustment ensured that the models were compatible with the VABAM framework, enabling a direct and fair comparison across all evaluated models. In total, 48 models were assessed, with 24 models evaluated for each signal type.

3) Ancillary Input Operations: Our models, along with C-VAE, distinguish themselves from other models by uniquely integrating ancillary inputs Θ. During the evaluation phase for our models, θ is sampled from a conditional uniform distribution expressed as $\theta \sim \mathcal{U}(0, 1 \,|\, y)$. These samples are then sorted in ascending order according to the synthesis index before being utilized as input. Similarly, in the case of C-VAE, the normalized PSD (Eq.(25)) of the original signal is computed during the training phase for the conditional input. For evaluation, this normalized PSD is arranged in ascending order by the synthesis index before being inputted.

4) Evaluation Metrics: Eight metrics were employed for each model (Tab. III): the seven previously proposed metrics along with the Mean absolute percentage error (MAPE). The JMI-based metrics were iteratively estimated through MCS using small batches (i.e., interval estimation) during the synthetic process (Fig. 1 in the supplementary document), whereas FQI and MAPE were estimated only once from the entire test dataset (i.e., point estimation). By default, the MCS was executed with 10 × 100 nested iterations on each batch, comprising 50 samples. Nonetheless, configuration variations stemmed from the model's inherent structure and metric calculation methods, with more details provided in the supplementary document.

B. Data Sources and Processing

ABP and ECG data were sourced from the MIMIC-III Waveform Database [47], which contains records from approximately 30,000 ICU patients. These records underwent multi-stage preprocessing to extract high-quality pulsatile physiological signals for model training. The data preprocessing steps were

TABLE II
MODEL SUMMARIES FOR BENCHMARK TESTING AND ABLATION STUDIES

Models		Objective functions					Ablations			
Type	Name	SKZ	DKZ	TC	MI	θ	J		C	
							30	50	5s	8s
Benchmarks	B-VAE	O	-	-	-	-	✓	✓	-	-
	C-VAE	O	-	-	-	Δ	✓	✓	-	-
	TC-VAE	-	O	O	O	-	✓	✓	-	-
	FAC-VAE	O	-	O	-	O	✓	✓	-	-
Ours & its variants	SKZ	O	-	-	-	Δ	✓	✓	✓	✓
	SKZFC	O	-	-	-	O	✓	✓	✓	✓
	TCMIDKZFC	-	O	O	O	O	✓	✓	✓	✓
	FACFC	O	-	O	-	O	✓	✓	✓	✓

SKZ, DKZ, TC, MI, and θ, respectively represent the standard KLD for $p(z)$, dimension-wise KLD for $p(z)$, total correlation, mutual information, and the ancillary (conditional) input. J denotes the dimensional size of Z, and C refers to the compression size of the signal given an initial duration of 10 s. C is only applicable to our model and its variants. Δ in θ signifies the existence of ancillary input without parameterizing its distribution. Objective functions were applied with an *and* basis, and ablation criteria with an *or* basis.

as follows: Initially, representative signals such as ABP and ECG were selected, reducing the dataset to 13,574 records from 1,285 patients. In the quality assessment phase, data cleaning procedures were applied to remove records with missing values and to confirm those within the acceptable value range for ABP. Subsequently, the selected signals were standardized, and peak analysis was performed to identify and analyze signal peaks [48]. Segments with irregular peak patterns or insufficient peaks, as well as those exhibiting significant deviations in initial or final peaks, or anomalies in the inverted ECG signal, were discarded. The next step involved resampling the data from 125 Hz to 100 Hz [49], resulting in a dataset comprising 3,063,004 ten-second segments from 895 patients. Finally, 300,000 segments were chosen from this refined dataset to form the training set, while another 60,000 and 10,000 segments were allocated to the validation and test sets, respectively.

C. Evaluation of Reconstruction and Synthesis Quality

Reconstruction and synthesis quality are evaluated using MAPE and FQI metrics (Tab. III, Fig. 4, and Fig. 5). Notably, the majority of SKZ models exhibited higher MAPE values compared to other models across all signals. This suggests a potential limitation in the SKZ design, wherein Θ does not adhere to a uniform distribution, potentially leading to a convergence on Θ that extracts only easily reconstructible features. Such a tendency could trap the model in local optima, hindering the accurate reconstruction of the original signal morphology. Furthermore, our experiments revealed that models compressing the original signal to 8 s via cascading filters demonstrated lower MAPE compared to those compressing it to 5 s, implying that excessive signal compression may result in information loss within features. Additionally, TC-VAE consistently exhibited high MAPE values across signals, indicating that adjustments in parameters such as $Beta$, TC, and MI could offer avenues for improvement [50], warranting further exploration.

Regarding FQI, SKZ displayed overall poor performance across signals. The calculation of FQI involves filtering synthesized signals based on the KLD between their PSD and that of their corresponding original signals, with a threshold τ set to 1. FQI is then calculated based on the KLD between the PSD of the filtered synthesized signals and that of the ground truth signals. Thus, this metric is designed to often yield positively skewed performance evaluations. Despite this inherent advantage, the significant underperformance of SKZ indicates its tendency to converge on local optima, thereby impeding the accurate reconstruction of the original signal morphology.

D. Evaluation of Z Disentanglement in Signal Morphology

Disentanglement of Z in signal morphology is evaluated through $I(V;Z)$ and $I(V;\acute{Z}|Z)$ (Tab. III, Fig. 4, and Fig. 5). A prominent result is that C-VAE exhibits very low values for both $I(V;Z)$ and $I(V;\acute{Z}|Z)$, thereby indicating that Z and $\acute{Z}$ poorly convey information about signal morphology. This can be attributed to the experiment's configuration, where PSD was utilized as the conditional input, causing the morphology representation to depend heavily on the PSD rather than on Z. Additionally, SKZ shows values for $I(V;Z)$ comparable to other models but tends to have lower values for $I(V;\acute{Z}|Z)$. This indicates that although Z contributes to morphology, the model struggles with effective disentanglement due to its inherent structural limitations that lead to convergence at local optima. These distinct results for C-VAE and SKZ, compared to other models, are further highlighted in cross-comparative analysis between baseline and key metrics (Fig. 6).

Other models exhibited higher $I(V;\acute{Z}|Z)$ values than $I(V;Z)$ across signal types, indicating that disentanglement was achieved to some extent, although the degree of disentanglement can vary. Furthermore, among ABP models, the performance was quite similar, but differences were observed in the ECG models, thereby suggesting the need for further attempts on optimizing parameters.

E. Evaluation of Θ's Contribution to Signal Morphology

Θ's contribution to signal morphology is evaluated through $I(V;\acute{Z})$ and $I(V;\acute{\Theta}|\acute{Z})$ (Tab. III, Fig. 4, and Fig. 5). Similar to the previous metric, C-VAE and SKZ display distinctive patterns compared to other models (Fig. 6). A notable discovery across both signals is that for C-VAE, $I(V;\acute{\Theta}|\acute{Z})$ markedly surpasses the baseline$I(V;\acute{Z})$. This observation corroborates earlier interpretations, indicating that morphology is minimally influenced by $\acute{Z}$ and predominantly dependent on the conditional input of the PSD.

For SKZ, $I(V;\acute{Z})$ values are notably higher in ABP compared to those of other models. This observation underscores the impact of parameterizing the distribution of Θ on the model's outcomes, given that SKZ shares identical settings with its primary model, SKZFC, except for the constraint imposed on the uniform distribution of Θ. Furthermore, considering the propensity of SKZ models to converge to local optima, the outcomes of $I(V;\acute{Z})$ suggests that Θ is trained to extract easily reconstructable features, thereby failing to generate a diverse range of morphologies that encompass the entire spectrum of ground truth signals. This limitation potentially contributes to the elevated KLD values observed.

TABLE III
MODEL PERFORMANCE COMPARISON BASED ON METRICS

Models			ABP								ECG							
Name	**J**	**C**	**FQI**	**MAPE**	**i**	**ii**	**iii**	**iv**	**v**	**vi**	**FQI**	**MAPE**	**i**	**ii**	**iii**	**iv**	**v**	**vi**
B-VAE	30	-	0.047	0.054	1.783	4.08	-	-	-	-	0.293	0.02	1.827	3.177	-	-	-	-
	50	-	0.078	0.057	1.765	4.084	-	-	-	-	0.192	0.019	1.787	3.594	-	-	-	-
C-VAE	30	-	0.063	0.055	0.145	0.066	0.107	0.981	0.0	0.255	0.121	**0.018**	0.225	0.173	0.188	**2.1**	0.0	0.281
	50	-	0.063	0.057	0.119	0.062	0.113	**1.245**	0.0	0.163	0.11	0.02	0.322	0.392	0.436	1.047	0.0	0.271
TC-VAE	30	-	0.049	0.073	1.736	3.94	-	-	-	-	**0.07**	0.022	1.907	3.368	-	-	-	-
	50	-	**0.038**	0.077	1.719	4.062	-	-	-	-	0.129	0.022	1.943	3.571	-	-	-	-
FAC-VAE	30	-	0.072	0.058	1.718	3.974	-	-	-	-	0.186	0.019	1.884	**3.833**	-	-	-	-
	50	-	0.043	0.055	1.599	3.746	-	-	-	-	0.255	0.019	1.891	3.631	-	-	-	-
SKZ	30	5	0.189	0.068	2.514	2.062	2.981	0.019	0.0	0.424	1.419	0.024	1.974	1.053	**2.032**	0.0	0.002	0.401
	50	5	0.133	0.091	1.663	1.766	1.337	0.13	0.0	0.487	1.118	0.025	1.864	1.908	1.483	0.013	0.0	0.469
	30	8	1.366	0.064	**3.137**	0.582	**3.352**	0.0	0.0	0.101	1.34	0.026	1.872	0.764	2.017	0.001	0.0	0.343
	50	8	0.372	0.065	2.734	1.264	2.889	0.004	0.0	0.231	1.193	0.028	1.754	0.838	1.721	0.002	0.0	0.134
SKZFC	30	5	0.073	0.067	1.93	4.089	1.561	0.086	0.0	0.33	0.235	0.019	2.07	2.587	1.973	0.232	0.0	0.358
	50	5	0.095	0.066	2.276	4.219	1.855	0.122	0.0	0.393	0.159	0.02	1.967	2.112	1.486	0.145	0.0	0.555
	30	8	0.044	**0.049**	1.707	3.925	1.658	0.068	0.0	0.451	0.153	0.019	1.882	2.659	1.295	0.125	0.0	0.42
	50	8	0.068	0.05	2.032	4.366	1.564	0.087	0.0	0.558	0.119	0.02	1.781	3.006	1.536	0.196	0.0	0.494
TCMIDKZFC	30	5	0.066	0.063	1.762	3.8	1.467	0.46	0.0	0.455	0.128	0.02	1.871	2.645	1.263	0.651	0.0	0.422
	50	5	0.051	0.064	1.806	3.795	1.38	0.124	0.0	**0.664**	0.139	0.02	**2.306**	3.165	1.685	0.285	0.0	0.594
	30	8	0.051	0.053	1.695	3.084	0.971	0.199	0.0	0.549	0.285	0.019	1.701	2.66	1.548	0.265	0.0	0.512
	50	8	0.049	0.05	1.712	3.712	1.461	0.195	0.0	0.54	0.141	0.019	1.885	2.65	1.206	0.234	0.0	0.539
FACFC	30	5	0.084	0.071	2.051	**4.432**	1.589	0.045	0.0	0.654	0.149	0.021	1.991	2.697	1.475	0.21	0.0	**0.711**
	50	5	0.061	0.065	1.892	4.228	1.52	0.088	0.0	0.512	0.136	0.021	1.961	2.871	1.747	0.133	0.0	0.437
	30	8	0.042	0.054	1.855	3.992	1.718	0.139	0.0	0.548	0.159	0.02	1.999	2.578	1.104	0.154	0.0	0.688
	50	8	0.066	0.052	1.971	4.106	1.557	0.111	0.0	0.402	0.144	0.02	1.834	2.355	0.875	0.145	0.0	0.442

J and C denote the dimensional size of Z and the compression size of the signal, respectively. The terms i to vi correspond to $I(V;Z)$, $I(V;\acute{Z} \mid Z)$, $I(V;\acute{Z})$, $I(V;\acute{\Theta} \mid \acute{Z})$, $I(S;\acute{Z})$, and $I(S;\acute{\Theta} \mid \acute{Z})$, respectively. $I(V;\acute{Z})$, $I(V;\acute{\Theta}|\acute{Z})$, $I(S;\acute{Z})$, and $I(S;\acute{\Theta}|\acute{Z})$ are applicable exclusively to models with ancillary inputs. Lower FQI and MAPE, and higher values for other metrics, signify superior performance, with boldface highlighting the best outcomes.

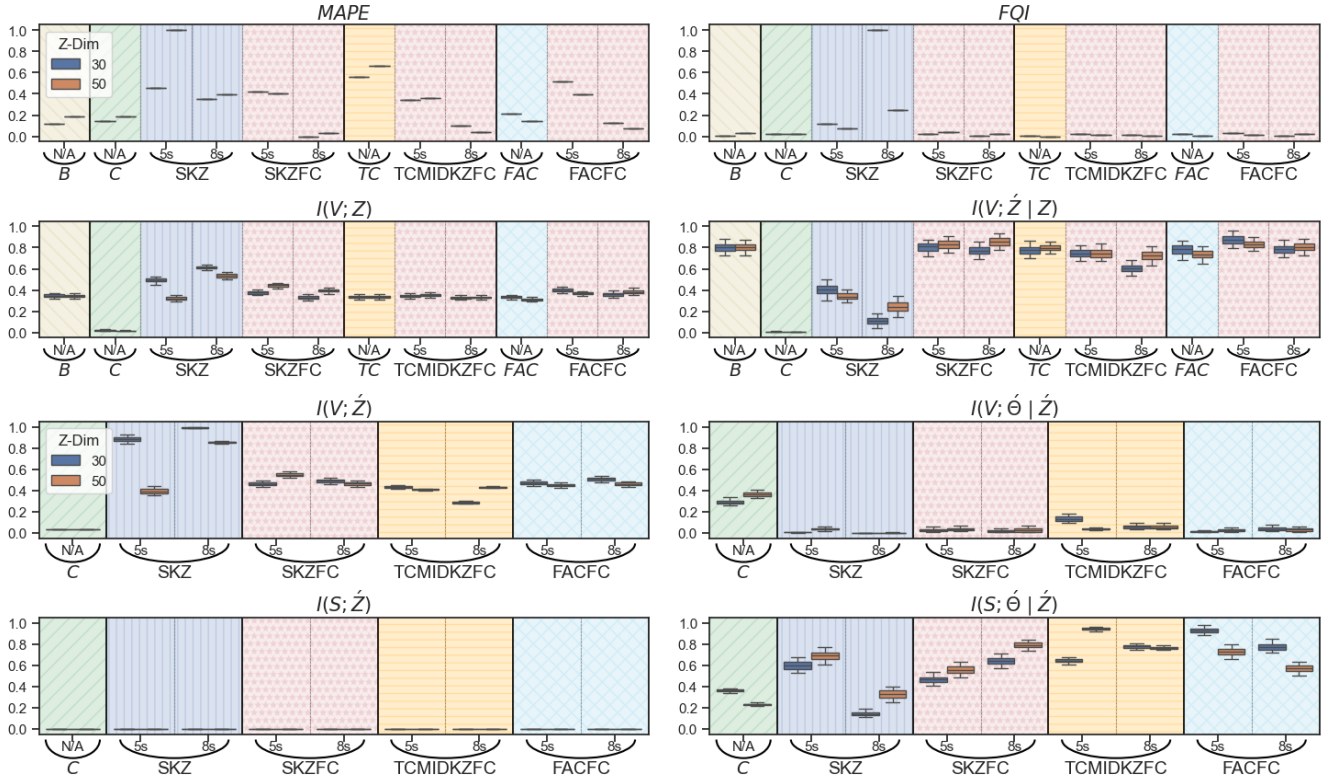

Fig. 4. Performance Comparison on ABP Data through Metric Distributions; $MAPE$ and FQI are point estimates, whereas other metrics are interval estimates. $MAPE$ and FQI values are normalized between 0 and 1 for each respective metric, while other metrics are normalized across both baseline and key metrics.

F. Evaluation of Controllability in Θ-based Synthesis

The controllability of Θ-based synthesis is evaluated through $I(S;\acute{Z})$ and $I(S;\acute{\Theta}|\acute{Z})$ (Tab. III, Fig. 4, and Fig. 5). A notable finding is that for both signals, the contribution of $\acute{Z}$ to the orderliness of the synthesized signals, as measured by S, is minor. In contrast, the introduction of $\acute{\Theta}$ significantly amplifies its influence on S. This empirical evidence suggests that the models in this experiment, particularly those in the C-VAE and VABAM families, are engineered to regulate the orderliness of synthesis through ancillary variable inputs. Furthermore, with

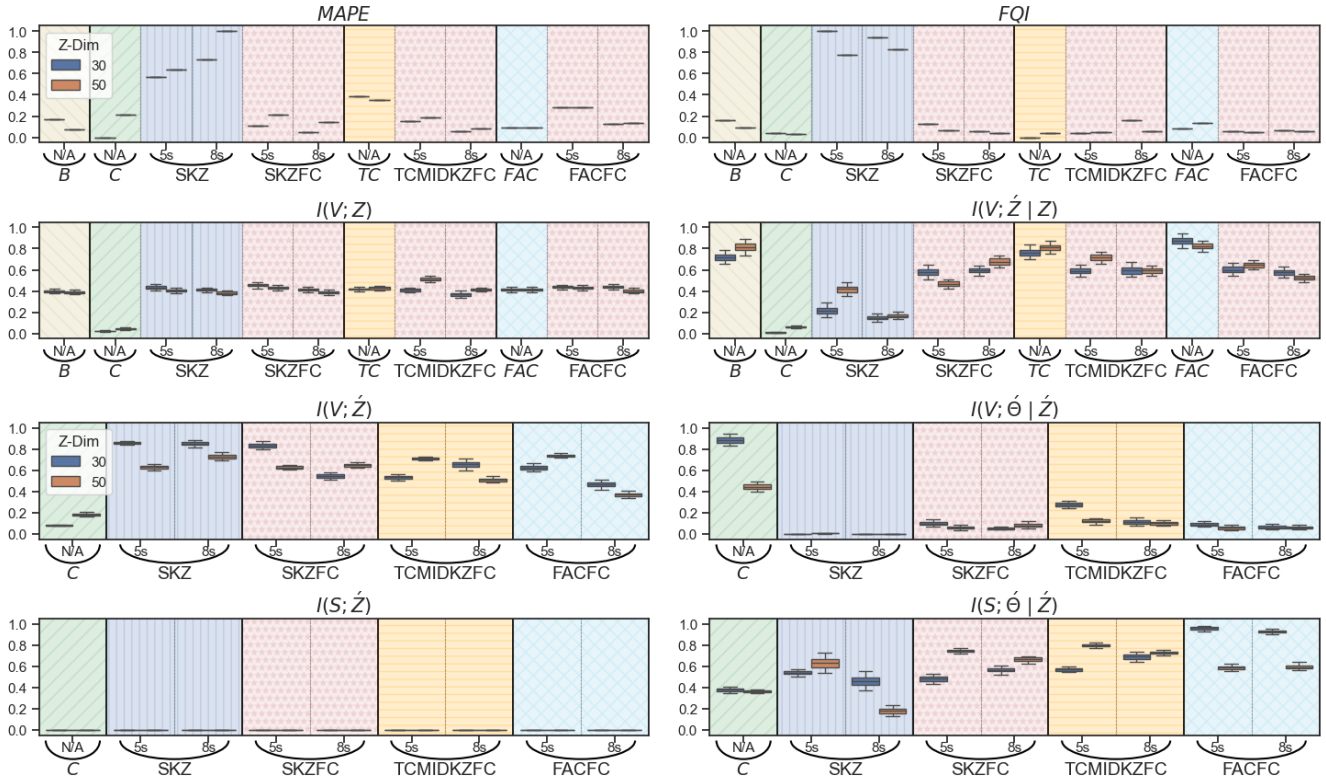

Fig. 5. Comparison of Model Performance on ECG Data: Metric Distributions; $MAPE$ and FQI are point estimates, whereas other metrics are interval estimates. $MAPE$ and FQI values are normalized between 0 and 1 for each respective metric, while other metrics are normalized across both baseline and key metrics.

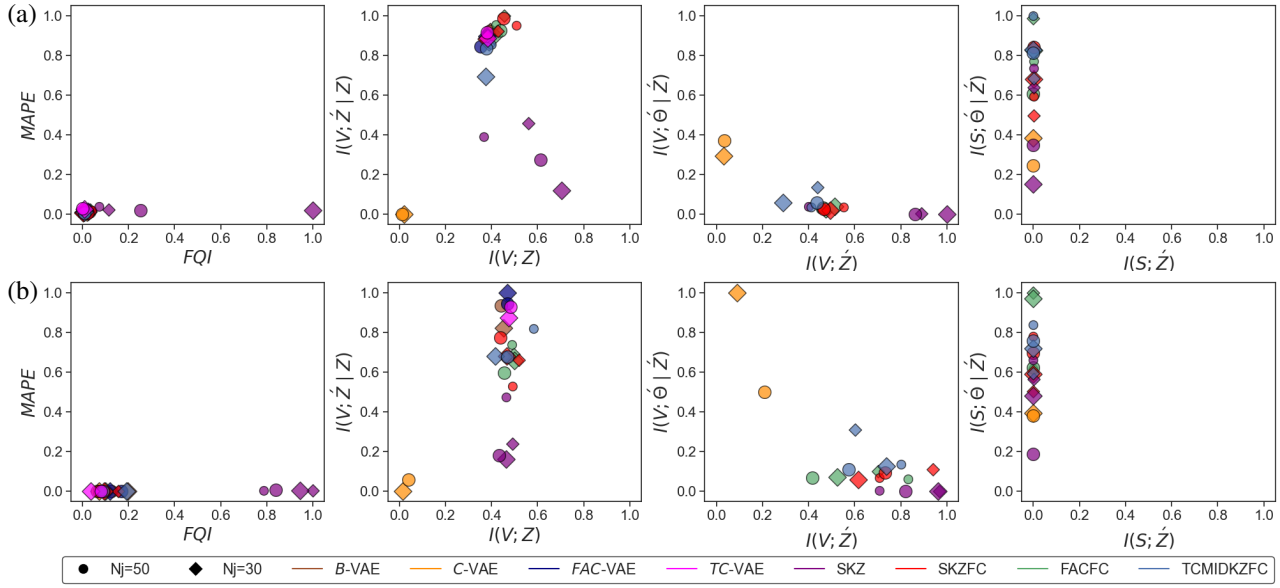

Fig. 6. Cross-Comparative Analysis of Baseline and Key Metrics; (a) Results from ABP. (b) Results from ECG. Values are normalized on both x and y axes.

the exception of SKZ, most configurations demonstrate that VABAM variants outperform C-VAE concerning $I(S;\acute{\Theta}|\acute{Z})$ values, indicating their superior synthesis controllability.

G. Evaluation by Isolated Dimension Size in Z and Θ Ranges

We examined the performance variation based on isolated dimension size N_j and the range of Θ, utilizing SKZFC with $C = 8s$ and $J = 50$ (Fig.7). We systematically varied the isolated dimension size across the set $1, 10, 20, 30, 40, 50$. Furthermore, we constrained the parameter Θ by setting its maximum values to $0.05, 0.1, 0.5$, and 1.0. This approach limited Θ to values from 0 up to these upper limits, while ensuring the same number of syntheses across all cases.

Across both signals, a consistent pattern was observed when varying the isolated dimension size N_j and Θ ranges, except for $I(S;\acute{\Theta}|\acute{Z})$. Notably, $I(V;\acute{Z}|Z)$ exhibited a gradual change in relation to N_j. In particular, at $N_j = 50$, $\acute{Z}$ becomes identical to Z, rendering $I(V;\acute{Z}|Z)$ to zero. In contrast, setting $N_j = 1$

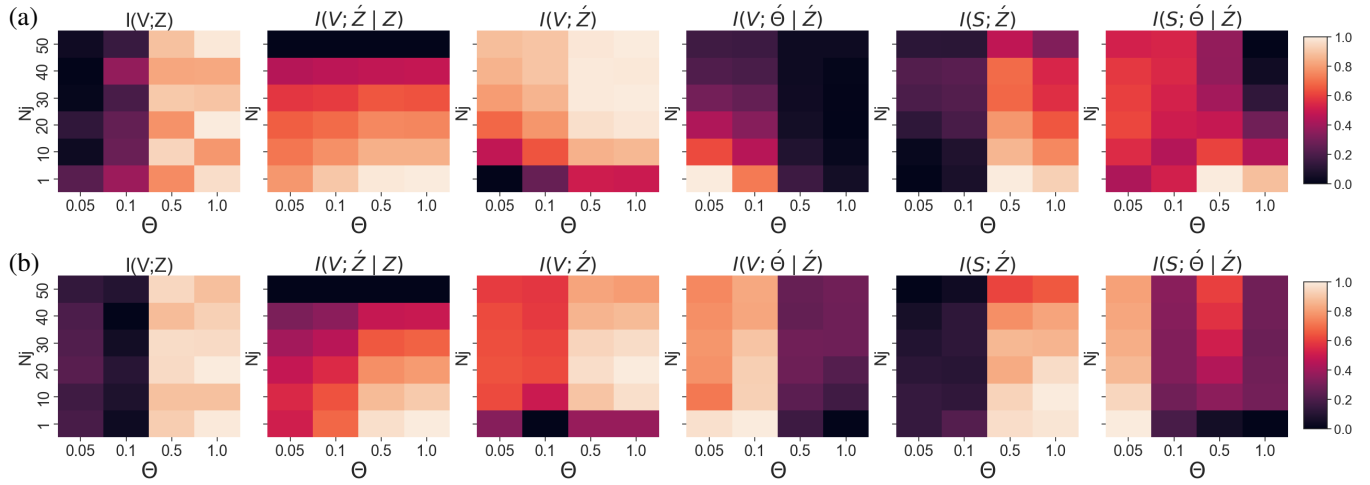

Fig. 7. Performance Variation by Isolated Dimension Size N_j and Θ Ranges; (a) ABP results. (b) ECG results. Values are normalized for each metric.

maximally isolates a single dimension of Z, resulting in the highest $I(V;\acute{Z}|Z)$. This suggests effective disentanglement, as the influence of an individual $\acute{Z}$ on V diverges from the collective effect of Z. Concurrently, $I(V;\acute{Z})$ assesses $\acute{Z}$'s unconditioned direct impact on V, gauging differences between the distributions of ground truth samples and those of batch samples (Eq.(32)). An increasing $I(V;\acute{Z})$ with greater N_j signifies enhanced diversity in the representation of the batch's morphology, suggesting effective disentanglement through the distinct contributions from each dimension to morphology.

Another notable result is the pronounced sensitivity of the metrics to variations in the maximum limit of Θ, particularly for $I(V;Z)$, $I(V;\acute{\Theta}|\acute{Z})$, and $I(S;\acute{Z})$. Among these, $I(V;Z)$ and $I(S;\acute{Z})$ exhibit a positive relationship with the Θ max limit. For $I(V;Z)$, this can be interpreted as the synthesis of a more diverse set of signals within the morphological identity of the ground truth, facilitated by the broader range of Θ. In the case of $I(S;\acute{Z})$, when the intervals of the pass filter effect are sufficiently spaced, it leads to the generation of morphologically distinct signals, thereby enhancing orderliness. However, if the intervals are insufficiently spaced, this may result in the generation of almost indistinguishable signals, potentially disrupting the orderliness of V.

Conversely, $I(V;\acute{\Theta}|\acute{Z})$ exhibits a distinct negative correlation with Θ's maximum limit. This suggests that a lower limit directs the pass filter effect of $\acute{\Theta}$ towards lower-frequencies, generating signals with strong amplitudes at these frequencies, which may differ from the original. However, setting a higher limit enables the synthesis of not only low-frequency enhanced signals but also those closely resembling the original.

A consistent pattern across the signals or a similar pattern with other metrics was not identified in $I(S;\acute{\Theta}|\acute{Z})$. Nonetheless, the results underscore the importance of finding optimal hyperparameter settings to enhance orderliness, which may require trade-offs with other metrics.

H. Qualitative Evaluation

Qualitative assessments were conducted on 500 synthesized samples per signal for each model (Fig. 8). Reflecting the unique goals of each model, configurations were tailored to enable the best possible signal synthesis. For B-, TC-, and FAC-VAEs, designed without ancillary inputs, synthesis variations were explored by adjusting Z. This process involved sampling 500 random numbers from a normal distribution $N(0,1)$, sorting them in ascending order, and then adding them to Z for synthesis. For C-VAE, a strategy regarding ancillary input, distinct from previous quantitative evaluations, was adopted to highlight its differences from other models, necessitating careful interpretation. This approach involved randomly selecting 500 normalized PSD values from ground truth signals, identifying the frequency index with the maximum value for each sample, and ordering the PSD values based on the maximum frequency indices. Such sorting accentuated the functional strengths of C-VAE in synthesis exploration, as the morphology tends to be predominantly determined by the frequency with the maximum PSD value [51].

The primary findings suggest that while TC-VAE, known for its proficiency in disentangling Z, manages to preserve morphological identity to some extent, it still exhibits accompanying horizontal or phase shifts. In contrast, C-VAE faces significant challenges in maintaining morphological identity, showing noticeable horizontal or phase shifts. This difficulty is likely attributable to the ancillary input strategy employed in this experiment. Given that this experimental setup was optimized to highlight C-VAE performance, it emphasizes the inherent difficulties in preserving morphological identity solely through the straightforward utilization of PSD arranged by maximum frequency indices.

Regarding the main models, SKZFC, TCMIDKZFC, and FACFC maintain the morphological identity without horizontal or phase shifts, demonstrating effective amplitude-based synthesis. The PSD degrees in Fig. 8 illustrate that, in these models, amplitudes are coherently manifested along the synthesis axis within a few frequencies, in sharp contrast to the irregular expression of amplitudes across a wide frequency spectrum in the benchmark models. This qualitative assessment aligns well with our introduced metric, PD-PSD (Eq.(26)), which quantifies structured amplitude-based modulation. Additionally, these main models tend to synthesize distinct signals, especially in their response to θ band, resulting in noticeable differences in

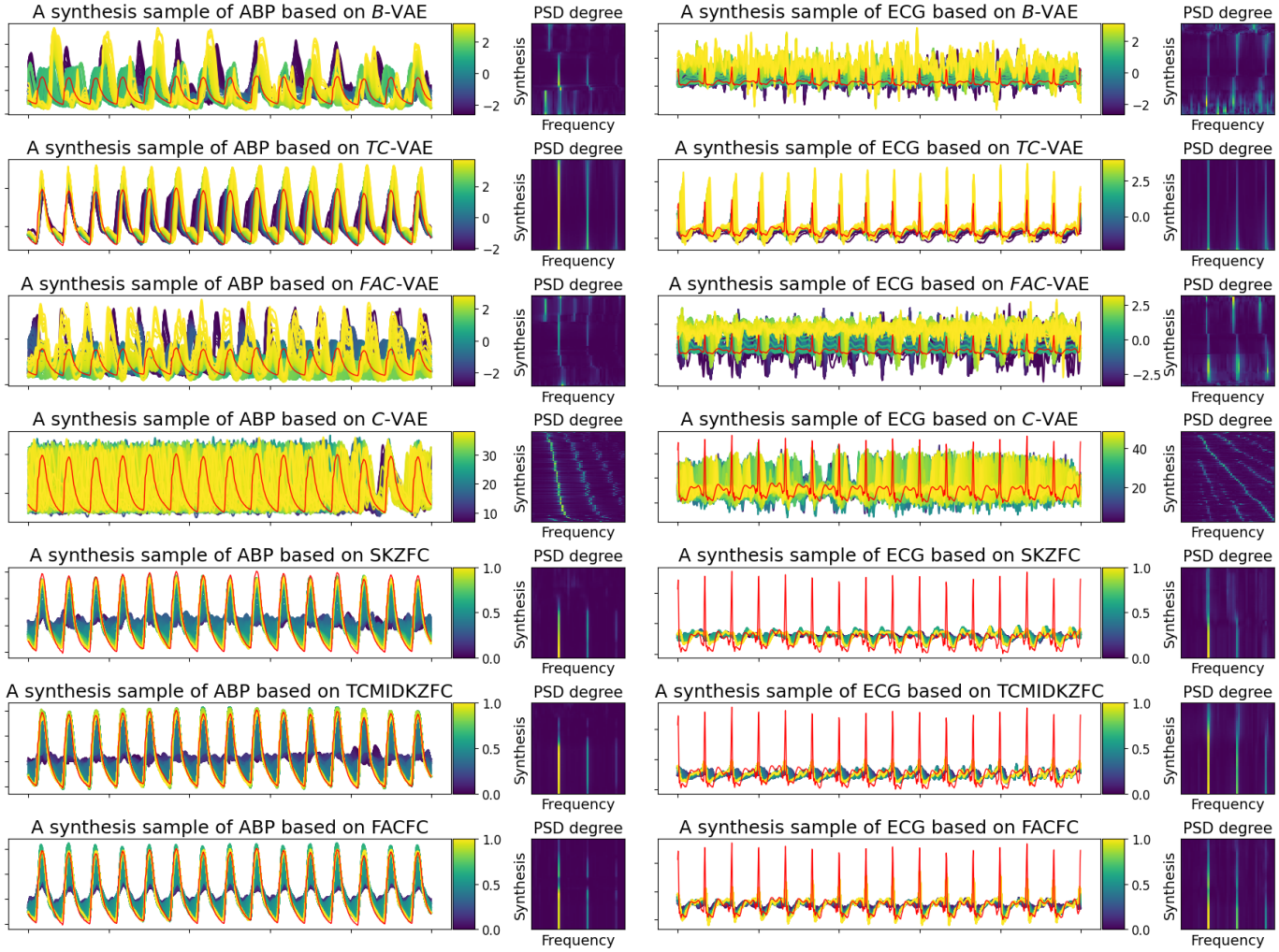

Fig. 8. Synthesized Samples for ABP and ECG Data; The colors of the synthesized signal correspond to the values of the synthesis modulation parameter. Color bars for B, TC, and FAC-VAE show mean Z values; for C-VAE, the maximum frequency index; and for SKZFC, TCMIDKZFC, and FACFC, the Θ values. The red lines represent the original signals. The PSD degree matrix depicts the intensity of the PSD across frequency and synthesis indices.

low-frequency synthesis outcomes. Considering the variety of methods for disentangling Z among these models, our findings encourage further investigation to refine the synergy between disentanglement of Z and the pass filter effects.

VI. DISCUSSION AND CONCLUSION

In this study, we introduced VABAM, a method designed to enable amplitude-based synthesis of pulsatile physiological signals while preserving their morphological identity. As a member of the VAE family, VABAM presents distinct outcomes compared to C-VAE and demonstrates strong compatibility with the disentanglement-centric models such as TC- and FAC-VAEs. This compatibility highlights the potential of VABAM for further advancement and broader application.

Moreover, we proposed three novel metrics grounded in JMI: 1) disentanglement of Z in signal morphology, 2) contribution of Θ to signal morphology, and 3) controllability of Θ-based synthesis. These metrics provide empirical evidence on the quantification of unique characteristics across various models. Furthermore, our efforts to quantify the morphological features of signals via PSD, alongside the innovative integration of permutation entropy with MI for controllability assessment, constitute a pioneering and theoretically robust methodology. These underline the innovative essence and academic rigor of our study, thereby establishing a new standard in this domain.

However, exercising caution in the use of JMI-based metrics is imperative. These metrics, formulated as KLD expressions, tend to yield higher values when the distributions of two random variables differ significantly [52]. Factors such as incomplete model training, resulting in inadequate reconstruction, or ineffective amplitude-based synthesis can contribute to these high values. Therefore, relying solely on a single metric for performance evaluation is inadequate. A comprehensive assessment incorporating various metrics is essential for a thorough evaluation of the performance of a model.

Moving forward with our research, areas for improvement have been identified. First, in terms of measuring controllability with $I(S;\acute{\Theta}|\acute{Z})$, synthesis that entails minimal changes between signals could yield high values if the PSD orderliness is maintained across synthesis samples. However, synthesis resulting in negligible differences between signals may not be advantageous from an application standpoint. Thus, methods that consider the PSD variance among synthesized signals along

the frequency indices must be developed, prioritizing samples that demonstrate both significant variance and orderliness.

Second, exploring a broader range of synthesis via model adjustments and optimization, coupled with an in-depth investigation of its physiological implications, promises substantial contributions to medical research, particularly in anesthesiology and physiology. Moreover, expanding VABAM to incorporate control over horizontal or phase shifts is expected to refine the synthesis process, potentially widening its application range.

Recently, the use of pulsatile physiological signals has grown rapidly, resulting in the development of various models for foundation, imputation, and prediction. However, efforts to integrate domain-specific knowledge have been limited. In this context, our research is positioned to serve as a benchmark for embedding domain theories within deep learning models. We envision this laying the groundwork for further research that more deeply integrates the theoretical aspects of physiological signals.

REFERENCES

[1] Z. Yue, M. Shi, and S. Ding, "Facial video-based remote physiological measurement via self-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[2] K. P. Birman, "Rule-based learning for more accurate ecg analysis," *IEEE transactions on pattern analysis and machine intelligence*, no. 4, pp. 369–380, 1982.

[3] R. Mukkamala, A. T. Reisner, H. M. Hojman, R. G. Mark, and R. J. Cohen, "Continuous cardiac output monitoring by peripheral blood pressure waveform analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 459–467, 2006.

[4] S. Kim, H. Kim, E. Yun, H. Lee, J. Lee, and J. Lee, "Probabilistic imputation for time-series classification with missing data," in *International Conference on Machine Learning*, pp. 16654–16667, PMLR, 2023.

[5] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng, "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 734–740, 2010.

[6] E. Hwang, Y.-S. Park, J.-Y. Kim, S.-H. Park, J. Kim, and S.-H. Kim, "Intraoperative hypotension prediction based on features automatically generated within an interpretable deep learning model," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[7] J. G. Webster, "Reducing motion artifacts and interference in biopotential recording," *IEEE transactions on biomedical engineering*, no. 12, pp. 823–826, 1984.

[8] A. Chatterjee and U. K. Roy, "Ppg based heart rate algorithm improvement with butterworth iir filter and savitzky-golay fir filter," in *2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, pp. 1–6, IEEE, 2018.

[9] A. de Cheveigné and I. Nelken, "Filters: when, why, and how (not) to use them," *Neuron*, vol. 102, no. 2, pp. 280–293, 2019.

[10] S. Iravanian and L. Tung, "A novel algorithm for cardiac biosignal filtering based on filtered residue method," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 11, pp. 1310–1317, 2002.

[11] M. Ahlstrom and W. Tompkins, "Digital filters for real-time ecg signal processing using microprocessors," *IEEE Transactions on Biomedical Engineering*, no. 9, pp. 708–713, 1985.

[12] I. Winkler, S. Debener, K.-R. Müller, and M. Tangermann, "On the influence of high-pass filtering on ica-based artifact reduction in eeg-erp," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4101–4105, IEEE, 2015.

[13] E. Hwang, H.-S. Park, H.-S. Kim, J.-Y. Kim, H. Jeong, J. Kim, and S.-H. Kim, "Development of a bispectral index score prediction model based on an interpretable deep learning algorithm," *Artificial Intelligence in Medicine*, p. 102569, 2023.

[14] M. T. Islam, I. Zabir, S. T. Ahamed, M. T. Yasar, C. Shahnaz, and S. A. Fattah, "A time-frequency domain approach of heart rate estimation from photoplethysmographic (ppg) signal," *Biomedical Signal Processing and Control*, vol. 36, pp. 146–154, 2017.

[15] G. McGibney, M. Smith, S. Nichols, and A. Crawley, "Quantitative evaluation of several partial fourier reconstruction algorithms used in mri," *Magnetic resonance in medicine*, vol. 30, no. 1, pp. 51–59, 1993.

[16] D. Gogolewski, "Influence of the edge effect on the wavelet analysis process," *Measurement*, vol. 152, p. 107314, 2020.

[17] A. T. Reisner, G. D. Clifford, and R. G. Mark, "The physiological basis of the electrocardiogram," *Advanced methods and tools for ECG data analysis*, vol. 1, p. 25, 2006.

[18] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data–a practical approach," *Journal of neuroscience methods*, vol. 250, pp. 34–46, 2015.

[19] A. Gharehbaghi and M. Lindén, "A deep machine learning method for classifying cyclic time series of biological signals using time-growing neural network," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 9, pp. 4102–4115, 2017.

[20] S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, "Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models," *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[21] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[22] J. H. McClellan, R. Schafer, and M. Yoder, *Dsp first*. Prentice Hall Press, 2015.

[23] S. Agarwal, A. Rani, V. Singh, and A. P. Mittal, "Eeg signal enhancement using cascaded s-golay filter," *Biomedical Signal Processing and Control*, vol. 36, pp. 194–204, 2017.

[24] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2507–2515, 2006.

[25] R. Sharma, R. B. Pachori, and P. Sircar, "Automated emotion recognition based on higher order statistics and deep learning algorithm," *Biomedical Signal Processing and Control*, vol. 58, p. 101867, 2020.

[26] X. Li, A. H. H. Ngu, and V. Metsis, "Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation," *arXiv preprint arXiv:2206.13676*, 2022.

[27] Y. Sang, M. Beetz, and V. Grau, "Generation of 12-lead electrocardiogram with subject-specific, image-derived characteristics using a conditional variational autoencoder," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2022.

[28] A. H. Afandizadeh Zargari, S. A. H. Aqajari, H. Khodabandeh, A. Rahmani, and F. Kurdahi, "An accurate non-accelerometer-based ppg motion artifact removal technique using cyclegan," *ACM Transactions on Computing for Healthcare*, vol. 4, no. 1, pp. 1–14, 2023.

[29] H. Li, G. Ditzler, J. Roveda, and A. Li, "Descod-ecg: Deep score-based diffusion model for ecg baseline wander and noise removal," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[30] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Generalized generative deep learning models for biosignal synthesis and modality transfer," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 968–979, 2022.

[31] V. Kuznetsov, V. Moskalenko, and N. Y. Zolotykh, "Electrocardiogram generation and feature extraction using a variational autoencoder," *arXiv preprint arXiv:2002.00254*, 2020.

[32] J. Lee, W. Kim, D. Gwak, and E. Choi, "Conditional generation of periodic signals with fourier-based decoder," *arXiv preprint arXiv:2110.12365*, 2021.

[33] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, "Variational autoencoder," in *Variational Methods for Machine Learning with Applications to Deep Networks*, pp. 111–149, Springer, 2021.

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

[35] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *International conference on learning representations*, 2016.

[36] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," *Advances in neural information processing systems*, vol. 28, 2015.

[37] R. T. Chen, X. Li, R. B. Grosse, and D. K. Duvenaud, "Isolating sources of disentanglement in variational autoencoders," *Advances in neural information processing systems*, vol. 31, 2018.

[38] H. Kim and A. Mnih, "Disentangling by factorising," in *International Conference on Machine Learning*, pp. 2649–2658, PMLR, 2018.

[39] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations

by mutual information estimation and maximization," *arXiv preprint arXiv:1808.06670*, 2018.
[40] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.
[41] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.
[42] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712*, 2016.
[43] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.
[44] P. Stoica, R. L. Moses, *et al.*, *Spectral analysis of signals*, vol. 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
[45] A. D. Wyner, "A definition of conditional mutual information for arbitrary ensembles," *Information and Control*, vol. 38, no. 1, pp. 51–59, 1978.
[46] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *science*, vol. 334, no. 6062, pp. 1518–1524, 2011.
[47] B. Moody, G. Moody, M. Villarroel, G. D. Clifford, and I. Silva, "Mimic-iii waveform database," 04 2020.
[48] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *Journal of Open Research Software*, vol. 7, no. 1, pp. 1–9, 2019.
[49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
[50] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-vae," *arXiv preprint arXiv:1804.03599*, 2018.
[51] R. De Boer, J. Karemaker, and J. Strackee, "Relationships between short-term blood-pressure fluctuations and heart-rate variability in resting subjects i: a spectral analysis approach," *Medical and biological engineering and computing*, vol. 23, pp. 352–358, 1985.
[52] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

Junetae Kim received the B.B.A. degree from Hanyang University, Seoul, South Korea, in 2013, and the Ph.D. degree in management information systems from the Korea Advanced Institute of Science and Technology (KAIST), Seoul, in 2018. He worked on both the Smart IT Team and the Strategy Marketing Team with the Device Solutions Division, Samsung Electronics, Hwaseong, South Korea. In 2019, he joined the Graduate School of Cancer Science and Policy (GCSP), National Cancer Center, Goyang, South Korea, where he is currently an Assistant Professor. His research fields include the development of artificial intelligence models with biosignals and genetic data.

Kyoungsuk Park received the master's and Ph.D. degree in mathematics with a specialization in combinatorics from Ajou University, Suwon, South Korea, in 2010 and 2016, respectively. From 2016 to 2021, she worked on algorithm development and AI research in the technology development department across several industries. Since 2022, she has been a researcher at the Graduate School of Cancer Science and Policy (GCSP), National Cancer Center, Goyang, South Korea. Her research interests include artificial intelligence and biosignal processing.

Kyunglim Kim received B.S. and M.S. degrees from Korea Advanced Institute of Science and Technology (KAIST). She received the Ph.D. degree in applied mathematics at University of California at Berkeley in 2017. During her Ph.D., she worked as a graduate student researcher at Los Alamos National Laboratory. She later worked at New York University (NYU) as a postdoctoral associate until 2021. She is now working as a data scientist in Data Intelligence team at Samsung Research. Her current research interest includes artificial intelligence in healthcare.

VABAM: Variational Autoencoder for Amplitude-based Biosignal Augmentation within Morphological Identities

Junetae Kim, Kyoungsuk Park, Kyunglim Kim

This additional material offers a step-by-step breakdown of the formulae introduced in the main article. Readers are encouraged to refer to the main document for a more integrated understanding of the context and concepts.

I. Details in Evidence Lower Bound Terms

A. Reconstruction of Raw Signals

$$\mathbb{E}_{q(x,z,\theta|y)}[\log p(y|x)] = \int_x \int_z \int_\theta \log p(y|x) q(x,z,\theta|y)\, d\theta\, dz\, dx \tag{1}$$

$$\approx \frac{1}{L}\sum_{l=1}^{L} \log p(y|x^{(l)}), \text{where } x \in \{x_{HH}, x_{HL}, x_{LH}, x_{LL}\}. \tag{2}$$

Empirically, the following equation, which accounts for a weighting factor denoted as ψ_y for the loss term and batch training, is applied:

$$J(\Phi_y; y) = \psi_y \left(\frac{1}{|B|} \sum_{b\in B} \left(\frac{1}{|D|} \sum_{d\in D} (y_{b,d} - g_y(x_{b,d}))^2 \right) \right), \tag{3}$$

where b and d represent the batch index and a sample index within the batch, respectively.

B. Reconstruction of Feature Signals

$$\mathbb{E}_{q(z,\theta|y)q(x|y,\theta)} \left[\log \frac{p(x|z,\theta)}{q(x|y,\theta)} \right] = \int_z \int_\theta \left(\int_x \log \frac{p(x|z,\theta)}{q(x|y,\theta)} q(x|y,\theta)\, dx \right) q(z,\theta|y)\, d\theta\, dz \tag{4}$$

$$= \int_z \int_\theta -\text{KLD}\Big(q(x|y,\theta) \,||\, p(x|z,\theta) \Big) q(z,\theta|y)\, d\theta\, dz \tag{5}$$

$$\approx -\frac{1}{L}\sum_{l=1}^{L} \text{KLD}\left(q(x|y,\theta^{(l)}) \,||\, p(x|z^{(l)},\theta^{(l)}) \right),$$

$$\text{where } z^{(l)} \sim \mathcal{N}(0,1),\ \theta^{(l)} \sim \text{Bernoulli}(0.5), \text{and } x \in \{x_{HH}, x_{HL}, x_{LH}, x_{LL}\} \tag{6}$$

In practice, the subsequent equation is utilized, considering a weight ψ_x for the loss term and batch training:

$$J(\Phi_x; x) = \psi_x \left(\frac{1}{|B|} \sum_{b\in B} \left(\frac{1}{|D|} \sum_{d\in D} (g_x(y_{b,d}, \theta_{b,d}) - g_{x'}(z_{b,d}, \theta_{b,d}))^2 \right) \right). \tag{7}$$

* First author: Junetae Kim (e-mail: lyjune0070@gmail.com)
† Corresponding author: Kyunglim Kim (e-mail: eugeniakkim@gmail.com)
Junetae Kim is with the Graduate School of Cancer Science and Policy and the Healthcare AI Team, National Cancer Center, Goyang, Korea
Kyoungsuk Park is with the Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Korea
Kyunglim Kim is with Samsung Research, Seoul, Korea
The source code the reproducibility of this work is available at https://github.com/JunetaeKim/VABAM

C. Variational Inference for $p(z)$

$$\mathbb{E}_{q(z|y)q(\theta|y)q(x|y,\theta)}\left[\log\frac{q(z|y)}{p(z)}\right] \approx \int_\theta\left(\int_z \log\frac{q(z|y)}{p(z)}q(z|y)\,dz\right)q(\theta|y)\,d\theta \tag{8}$$

$$= \mathrm{KLD}(q(z|y)\,||\,p(z))\int_\theta q(\theta|y)\,d\theta \tag{9}$$

$$= \mathrm{KLD}(q(z|y)\,||\,p(z)). \tag{10}$$

In our study, we employed the KLD method to regularize the latent variable z within the framework of the standard VAE [1], as shown below:

$$\mathrm{KLD}(q(z|y)\,||\,p(z)) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{2}\sum_{j=1}^{J}\left(\mu_{z_{n,j}}^2+\sigma_{z_{n,j}}^2-1-\log\sigma_{z_{n,j}}^2\right)\right), \tag{11}$$

where $q(z|y)$ represents the approximate posterior distribution of the latent variables, conjectured to be Gaussian, encapsulated by the expression $g_z(\cdot)\sim\mathcal{N}(0,1\,|\,y)$, with $g_z(\cdot)$ denoting the z-sampler in VABAM. Conversely, $p(z)$ denotes the prior distribution of these latent variables, presumed to be a standard normal distribution, given as $z\sim\mathcal{N}(0,1)$.

For batch-level implementation, the subsequent equation is employed:

$$J(\Phi_z;z) = \psi_z\left(\left|\frac{1}{|B|}\sum_{b\in B}\left(\frac{1}{2|D|}\sum_{d\in D}\sum_{j=1}^{J}\left(\mu_{z_{b,d,j}}^2+\sigma_{z_{b,d,j}}^2-1-\log\sigma_{z_{b,d,j}}^2\right)\right)-\xi_z\right|\right), \tag{12}$$

where ψ_z denotes the weight assigned to this loss term and ξ_z signifies the capacity term [2].

D. Variational Inference for $p(\theta)$

$$\mathbb{E}_{q(z|y)q(\theta|y)q(x|y,\theta)}\left[\log\frac{q(\theta|y)}{p(\theta)}\right] \approx \int_z\left(\int_\theta \log\frac{q(\theta|y)}{p(\theta)}q(\theta|y)\,d\theta\right)q(z|y)\,dz \tag{13}$$

$$= \mathrm{KLD}(q(\theta|y)\,||\,p(\theta)). \tag{14}$$

Here, $q(\theta|y)$ represents the approximate posterior distribution of θ, postulated to follow a uniform distribution, while $p(\theta)$ denotes the prior distribution of the parameter, also assumed to be uniformly distributed.

While θ inherently follows a uniform distribution, for implementing the reparameterization trick, this uniform distribution is approximated using a Bernoulli distribution with a mean of 0.5. [3], [4]. The KLD is articulated as

$$\mathrm{KLD}(q(\theta|y)\,||\,p(\theta)) = \mathrm{KLD}(q(\theta|y)||\mathrm{Uniform}(0,1.0)) \tag{15}$$

$$\approx \mathrm{KLD}(q(\theta|y)\,||\,\mathrm{Bern}(\mu_\theta=0.5)) \tag{16}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{c\in\{0,1\}}\left(\mu_{\theta_{n,k}}^c(1-\mu_{\theta_{n,k}})^{1-c}\ln\frac{\mu_{\theta_{n,k}}^c(1-\mu_{\theta_{n,k}})^{1-c}}{0.5^c0.5^{1-c}}\right)\right), \tag{17}$$

with K set to 6, based on the structure of VABAM.

Empirically, the following equation is implemented at the batch level:

$$J(\Phi_\theta;\theta) = \psi_\theta\left(\left|\frac{1}{|B|}\sum_{b\in B}\left(\frac{1}{K|D|}\sum_{d\in D}\sum_{k=1}^{K}\sum_{c\in\{0,1\}}\left(\mu_{\theta_{b,d,k}}^c(1-\mu_{\theta_{b,d,k}})^{1-c}\ln\frac{\mu_{\theta_{b,d,k}}^c(1-\mu_{\theta_{b,d,k}})^{1-c}}{0.5^c0.5^{1-c}}\right)\right)-\xi_\theta\right|\right), \tag{18}$$

where ψ_θ and ξ_θ represent a weight for this term and a capacity term, [2], respectively.

II. Synthesis Model Structures

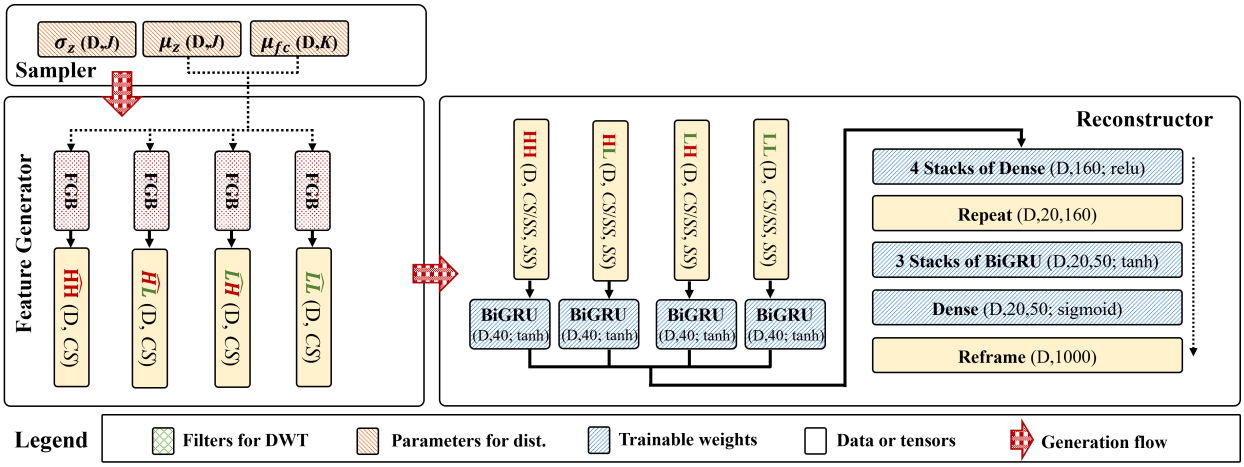

Fig. 1. Illustration of the Synthesis Model Structures: CS and SS denote the signal's compressed and sliding sizes, respectively. D, J, and K denote the data size in a batch, and the sizes of dimensions Z and θ, respectively.

III. Mutual Information-Based Novel Metric

In this section, we detail the derivations of six joint mutual information (JMI)-based metrics, offering both analytical and empirical forms. For programming purposes, it is advisable to refer to the empirical formulation. Moreover, we present a modified formulation designed for benchmarking models.

A. Assumptions

For estimating MI, we made some assumptions.

- **Uniform Distribution of Data Points:** Each test dataset point is uniquely indexed by an integer, associated with a uniformly distributed random variable [5]. Consequently, for a randomly selected instance y_n where $n \sim \mathcal{U}(1, N)$, we define its probability as $q(y_n) = q(n) = \frac{1}{N}$.
- **Uniform Distribution of Frequency Index:** Similarly, the probability of observing the random variable v, which denotes frequency index, is assumed to follow a uniform distribution, defined as $q(v) = \frac{1}{|V|}$.
- **Sampling via VABAM:** We adopt $z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1 \,|\, y)$ and $\theta \sim \mathcal{U}(0, 1 \,|\, y)$ to sample latent variables without replacement from the sampler in VABAM. This approach is rooted in the structural characteristics of VABAM, where y influences z and θ, and both z and θ subsequently influence v and s. This configuration leverages conditional independence, as the inclusion of the index y_n does not impact the outcomes of v and s once z and θ are known. This assumption allows the total number of Monte Carlo samples, L, to equal the size of the test dataset, N, which can be expressed by the equations $q(v|z, \theta) = q(v|z, \theta, y_n)$ or $q(s|z, \theta) = q(s|z, \theta, y_n)$.

Ultimately, these assumptions simplify computations and facilitate more stable and reliable inference, aligning with the foundational practices of deep learning and Bayesian inference.

B. $I(V; Z, \acute{Z})$

$$I(V; \acute{Z}, Z) = \underbrace{I(V; Z)}_{\text{(i)}} + \underbrace{I(V; \acute{Z}|Z)}_{\text{(ii)}}$$

(i) $\boldsymbol{I(V; Z)}$: The MI between the normalized PSD and Z.

a) Basic Analytical Derivation of $\boldsymbol{I(V; Z)}$

$$I(V;Z) = \mathbb{E}_{q(v,z)}\left[\log \frac{q(v,z)}{q(v)q(z)}\right] \tag{19}$$

$$= \mathbb{E}_{q(v|z)q(z)}\left[\log \frac{q(v|z)q(z)}{q(v)q(z)}\right] \tag{20}$$

$$= \mathbb{E}_{q(v|z)q(z)}\left[\log \frac{q(v|z)}{q(v)}\right] \tag{21}$$

$$= \mathbb{E}_{q(z)}\left[\sum_{v\in V} q(v|z)\log \frac{q(v|z)}{q(v)}\right]. \tag{22}$$

b) Approximation of $I(V;Z)$

$$I(V;Z) = \int_z q(z)\left(\sum_{v\in V} q(v|z)\log \frac{q(v|z)}{q(v)}\right)dz \tag{23}$$

$$\approx \frac{1}{L}\sum_{l=1}^{L} \mathrm{KLD}\left(q(v|z^{(l)}) \parallel q(v)\right) \tag{24}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(q(v|z^{(n)}, y_n) \parallel q(v)\right) \geq 0, \tag{25}$$

where $q(v|z^{(n)}, y_n)$ and $q(v)$ are approximated via Monte Carlo marginalization as follows:

$$q(v|z^{(n)}, y_n) = \int_\theta q(v|z^{(n)}, \theta^{(n)}, y_n)q(\theta)d\theta \;\approx\; \frac{1}{M}\sum_{m=1}^{M} q(v|z^{(n)}, \theta^{(n,m)}, y_n). \tag{26}$$

$$q(v) = \sum_{y_n\in Y} q(v|y_n)q(y_n) \;\approx\; \frac{1}{N}\sum_{n=1}^{N} q(v|y_n) \;\approx\; p(v). \tag{27}$$

c) Final Expression of $I(V;Z)$

$$I(V;Z) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(\frac{1}{M}\sum_{m=1}^{M} q(v|z^{(n)}, \theta^{(n,m)}, y_n) \parallel p(v)\right). \tag{28}$$

d) Empirical Formulation for $I(V;Z)$

$$I^*(V;Z) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D} \mathrm{KLD}\left(\frac{1}{M}\sum_{m=1}^{M} q(v|z^{(b,d)}, \theta^{(b,d,m)}, y_{b,d}) \parallel p(v)\right)\right). \tag{29}$$

e) Modification of $I(V;Z)$ for Benchmark Models without θ

$$I^z(V;Z) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D} \mathrm{KLD}\left(q(v|z^{(b,d)}, y_{b,d}) \parallel p(v)\right)\right). \tag{30}$$

f) Clarification of Prior Derivations

- V indicates the normalized PSD treated as a random variable.
- L, the number of Monte Carlo simulations, assumed to be equal to the total number of observations, N, in the test dataset utilized for calculating MI.
- $|B|$ is the total number of batches.

- $|D|$ is the number of samples in each batch, calculated as $\frac{N}{|B|}$.
- $z^{(b,d)}$ is a 4D tensor represented as $z^{(b,d)} \in \mathbb{R}^{B\times|D|\times M\times J}$, indexed by a batch index b, a sample index in the batch d, sub-repeat index m, and latent vector dimension index j. The values of z are randomly sampled at the dimensions b, d, and j, but remain constant across the dimension m, denoted as $z_{b,d,m,j} = z_{b,d,j}\ \forall m$.
- $\theta^{(b,d,m)}$ is a 4D tensor denoted as $\theta^{(b,d,m)} \in \mathbb{R}^{B\times|D|\times M\times K}$, with indexing based on a batch index b, a sample index in the batch d, sub-sampling index m, and latent vector dimension index k. The values of θ are sampled randomly at dimensions b, d, m, and k.
- $q(v)$ is a 4D tensor, expressed as $q(v) \in \mathbb{R}^{B\times|D|\times M\times|V|}$, with indices comprising a batch index b, a sample index in the batch d, sub-repeat index m, and a frequency index v treated as a random variable.
- $p(v)$ is a 1D distribution vector for the normalized PSD across the frequency index, v, in the test dataset.

ⓘⓘ $\boldsymbol{I(V;\acute{Z}|Z)}$: The MI between the normalized PSD and $\acute{Z}$ conditioned on Z. It quantifies how much information is gained about V by knowing $\acute{Z}$ in addition to Z, reflecting the additional information that $\acute{Z}$ provides over Z.

a) Basic Analytical Derivation of $\boldsymbol{I(V;\acute{Z}|Z)}$

$$I(V;\acute{Z}|Z) = \mathbb{E}_{q(v,\acute{z},z)}\left[\log\frac{q(v,\acute{z}|z)}{q(v|z)q(\acute{z}|z)}\right] \tag{31}$$

$$= \mathbb{E}_{q(v,\acute{z},z)}\left[\log\frac{(q(v,\acute{z},z)/q(z))}{q(v|z)q(\acute{z}|z)}\right] \tag{32}$$

$$= \mathbb{E}_{q(v,\acute{z},z)}\left[\log\frac{(q(v|\acute{z})q(\acute{z}|z)q(z)/q(z))}{q(v|z)q(\acute{z}|z)}\right] \tag{33}$$

$$= \mathbb{E}_{q(v|\acute{z})q(\acute{z},z)}\left[\log\frac{q(v|\acute{z})}{q(v|z)}\right] \tag{34}$$

$$= \mathbb{E}_{q(\acute{z},z)}\left[\sum_{v\in V} q(v|\acute{z})\log\frac{q(v|\acute{z})}{q(v|z)}\right]. \tag{35}$$

b) Approximation of $\boldsymbol{I(V;\acute{Z}|Z)}$

$$I(V;\acute{Z}|Z) = \int_{\acute{z}}\int_{z} q(\acute{z},z)\left(\sum_{v\in V} q(v|\acute{z})\log\frac{q(v|\acute{z})}{q(v|z)}\right)d\acute{z}\,dz \tag{36}$$

$$\approx \frac{1}{L}\sum_{l=1}^{L}\text{KLD}\left(q(v|\acute{z}^{(l)})\parallel q(v|z^{(l)})\right) \tag{37}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N}\text{KLD}\left(q(v|\acute{z}^{(n)},y_n)\parallel q(v|z^{(n)},y_n)\right) \geq 0, \tag{38}$$

where $q(v|\acute{z}^{(n)},y_n)$ and $q(v|z^{(n)},y_n)$ are approximated as follows:

$$q(v|\acute{z}^{(n)},y_n) \approx \frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)},\theta^{(n,m)},y_n). \tag{39}$$

$$q(v|z^{(n)},y_n) \approx \frac{1}{M}\sum_{m=1}^{M} q(v|z^{(n)},\theta^{(n,m)},y_n). \tag{40}$$

c) Final Expression of $\boldsymbol{I(V;\acute{Z}|Z)}$

$$I(V;\acute{Z}|Z) = \frac{1}{N}\sum_{n=1}^{N}\text{KLD}\left(\frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)},\theta^{(n,m)},y_n)\parallel\frac{1}{M}\sum_{m=1}^{M} q(v|z^{(n)},\theta^{(n,m)},y_n)\right)\geq 0. \tag{41}$$

d) Empirical Formulation for $I(V;\acute{Z}|Z)$

$$I^*(V;\acute{Z}|Z) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D}\text{KLD}\left(\frac{1}{M}\sum_{m=1}^{M}q(v|\acute{z}^{(b,d)},\theta^{(b,d,m)},y_{b,d}) \parallel \frac{1}{M}\sum_{m=1}^{M}q(v|z^{(b,d)},\theta^{(b,d,m)},y_{b,d})\right)\right). \tag{42}$$

e) Modification of $I(V;\acute{Z}|Z)$ for Benchmark Models without θ

$$I^z(V;\acute{Z}|Z) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D}\text{KLD}\left(q(v|\acute{z}^{(b,d)},y_{b,d}) \parallel q(v|z^{(b,d)},y_{b,d})\right)\right). \tag{43}$$

f) Clarification of Prior Derivations

- $\acute{z}^{(b,d)}$ is a 4D tensor represented as $\acute{z}\in\mathbb{R}^{B\times|D|\times M\times J}$, indexed by a batch index b, a sample index in the batch d, sub-repeat index m, and latent vector dimension index j. Initially, the values of z are randomly sampled at the dimensions b, d, and j, but remain constant across the dimension m, denoted as $z_{b,d,m,j}=z_{b,d,j}$ for all m. Then, all other elements, except for the isolated element $z_{j=j^*}$, which retains its original value from z_j, are replaced by zero.

C. $I(V;\acute{Z},\acute{\Theta})$

$$I(V;\acute{Z},\acute{\Theta}) = \underbrace{I(V;\acute{Z})}_{\text{(iii)}} + \underbrace{I(V;\acute{\Theta}|\acute{Z})}_{\text{(iv)}}$$

(iii) $I(V;\acute{Z})$: The MI between the normalized PSD and $\acute{Z}$.

a) Basic Analytical Derivation of $I(V;\acute{Z})$

$$I(V;\acute{Z}) = \mathbb{E}_{q(v,\acute{z})}\left[\log\frac{q(v,\acute{z})}{q(v)q(\acute{z})}\right] \tag{44}$$

$$= \mathbb{E}_{q(v|\acute{z})q(\acute{z})}\left[\log\frac{q(v|\acute{z})}{q(v)}\right] \tag{45}$$

$$= \mathbb{E}_{q(\acute{z})}\left[\sum_{v\in V}q(v|\acute{z})\log\frac{q(v|\acute{z})}{q(v)}\right]. \tag{46}$$

b) Approximation of $I(V;\acute{Z})$

$$I(V;\acute{Z}) = \int_{\acute{z}}q(\acute{z})\left(\sum_{v\in V}q(v|\acute{z})\log\frac{q(v|\acute{z})}{q(v)}\right)d\acute{z} \tag{47}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N}\text{KLD}\left(q(v|\acute{z}^{(n)},y_n) \parallel q(v)\right) \geq 0, \tag{48}$$

where $q(v|\acute{z}^{(n)}, y_n)$ and $q(v)$ are approximated via as follows:

$$q(v|\acute{z}^{(n)}, y_n) \approx \frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)}, \theta^{(n,m)}, y_n). \tag{49}$$

$$q(v) \approx \frac{1}{N}\sum_{n=1}^{N} q(v|y_n) \approx p(v). \tag{50}$$

c) Final Expression of $I(V;\acute{Z})$

$$I(V;\acute{Z}) = \frac{1}{N}\sum_{n=1}^{N} \text{KLD}\left(\frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)}, \theta^{(n,m)}, y_n) \parallel p(v)\right). \tag{51}$$

d) Empirical Formulation for $I(V;\acute{Z})$

$$I^*(V;\acute{Z}) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D}\text{KLD}\left(\frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(b,d)}, \theta^{(b,d,m)}, y_{b,d}) \parallel p(v)\right)\right). \tag{52}$$

ⓘⓥ $\boldsymbol{I(V;\acute{\Theta}|\acute{Z})}$: The MI between the normalized PSD and $\acute{\Theta}$ conditioned on $\acute{Z}$. It quantifies how much information about V is gained by knowing $\acute{\Theta}$ in addition to $\acute{Z}$, reflecting the additional information that $\acute{\Theta}$ provides over $\acute{Z}$.

a) Basic Analytical Derivation of $I(V;\acute{\Theta}|\acute{Z})$

$$I(V;\acute{\Theta}|\acute{Z}) = \mathbb{E}_{q(v,\acute{\theta},\acute{z})}\left[\log\frac{q(v,\acute{\theta}|\acute{z})}{q(v|\acute{z})q(\acute{\theta}|\acute{z})}\right] \tag{53}$$

$$= \mathbb{E}_{q(v,\acute{\theta},\acute{z})}\left[\log\frac{\Big(q(v,\acute{\theta},\acute{z})/q(\acute{z})\Big)}{q(v|\acute{z})q(\acute{\theta}|\acute{z})}\right] \tag{54}$$

$$= \mathbb{E}_{q(v,\acute{\theta},\acute{z})}\left[\log\frac{\Big(q(v|\acute{\theta},\acute{z})q(\acute{\theta}|\acute{z})q(\acute{z})/q(\acute{z})\Big)}{q(v|\acute{z})q(\acute{\theta}|\acute{z})}\right] \tag{55}$$

$$= \mathbb{E}_{q(v|\acute{\theta},\acute{z})q(\acute{\theta},\acute{z})}\left[\log\frac{q(v|\acute{\theta},\acute{z})}{q(v|\acute{z})}\right] \tag{56}$$

$$= \mathbb{E}_{q(\acute{\theta},\acute{z})}\left[\sum_{v\in V} q(v|\acute{\theta},\acute{z})\log\frac{q(v|\acute{\theta},\acute{z})}{q(v|\acute{z})}\right]. \tag{57}$$

b) Approximation of $I(V;\acute{\Theta}|\acute{Z})$

$$I(V;\acute{\Theta}|\acute{Z}) = \int_{\acute{z}}\int_{\acute{\theta}} q(\acute{\theta},\acute{z})\left(\sum_{v\in V} q(v|\acute{\theta},\acute{z})\log\frac{q(v|\acute{\theta},\acute{z})}{q(v|\acute{z})}\right)d\acute{\theta}d\acute{z} \tag{58}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N}\text{KLD}\left(q(v|\acute{\theta}^{(n)},\acute{z}^{(n)},y_n) \parallel q(v|\acute{z}^{(n)},y_n)\right) \geq 0, \tag{59}$$

where $q(v|\acute{z}^{(n)}, y_n)$ is approximated as follows:

$$q(v|\acute{z}^{(n)}, y_n) \approx \frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)}, \theta^{(n,m)}, y_n). \tag{60}$$

c) Final Expression of $I(V;\acute{\Theta}|\acute{Z})$

$$I(V;\acute{\Theta}|\acute{Z}) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(q(v|\acute{z}^{(n)},\acute{\theta}^{(n)},y_n) \parallel \frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(n)},\theta^{(n,m)},y_n)\right). \tag{61}$$

d) Empirical Formulation for $I(V;\acute{\Theta}|\acute{Z})$

$$I^*(V;\acute{\Theta}|\acute{Z}) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D} \mathrm{KLD}\left(q(v|\acute{z}^{(b,d)},\acute{\theta}^{(b,d)},y_{b,d}) \parallel \frac{1}{M}\sum_{m=1}^{M} q(v|\acute{z}^{(b,d)},\theta^{(b,d,m)},y_{b,d})\right)\right). \tag{62}$$

e) Clarification of Prior Derivations

- $\acute{\theta}^{(b,d)}$ is a 3D tensor denoted as $\acute{\theta}^{(b,d)} \in \mathbb{R}^{B\times|D|\times K}$, with indexing based on a batch index b, a sample index in the batch d, and latent vector dimension index k. Initially, the values of θ are sampled randomly at the dimensions b, d, and k, and then $\theta_{b,d,k}$ is sorted along the d-dimension to yield $\acute{\theta}_{b,\acute{d},k}$.

D. $I(S;\acute{Z},\acute{\Theta})$

$$I(S;\acute{Z},\acute{\Theta}) = \underbrace{I(S;\acute{Z})}_{\text{(v)}} + \underbrace{I(S;\acute{\Theta}|\acute{Z})}_{\text{(vi)}}$$

(v) $\boldsymbol{I(S;\acute{Z})}$: The MI between the permutation density of PSD (PD-PSD) and $\acute{Z}$.

a) Basic Analytical Derivation of $I(S;\acute{Z})$

$$I(S;\acute{Z}) = \mathbb{E}_{q(s|\underline{\acute{z}})q(\underline{\acute{z}})}\left[\log\frac{q(s,\underline{\acute{z}})}{q(s)q(\underline{\acute{z}})}\right] \tag{63}$$

$$= \mathbb{E}_{q(s|\underline{\acute{z}})q(\underline{\acute{z}})}\left[\log\frac{q(s|\underline{\acute{z}})}{q(s)}\right] \tag{64}$$

$$= \mathbb{E}_{q(\underline{\acute{z}})}\left[\sum_{s\in S} q(s|\underline{\acute{z}})\log\frac{q(s|\underline{\acute{z}})}{q(s)}\right]. \tag{65}$$

b) Approximation of $I(S;\acute{Z})$

$$I(S;\acute{Z}) = \int_{\underline{\acute{z}}} q(\underline{\acute{z}})\left(\sum_{s\in S} q(s|\underline{\acute{z}})\log\frac{q(s|\underline{\acute{z}})}{q(s)}\right)d\underline{\acute{z}} \tag{66}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(q(s|\underline{\acute{z}}^{(n)},y_n) \parallel q(s)\right) \geq 0, \tag{67}$$

where $q(s|\underline{\acute{z}}^{(n)},y_n)$ and $q(s)$ are approximated as follows:

$$q(s|\underline{\acute{z}}^{(n)},y_n) = \int_{\underline{\theta}}\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(n)},y_n,\underline{\theta}^{(n)})q(\underline{\theta}^{(n)})d\underline{\theta} \tag{68}$$

$$\approx \frac{1}{R}\sum_{r=1}^{R}\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(n)},\underline{\theta}^{(n,r)},y_n). \tag{69}$$

$$q(s) = \int_{\underline{z}} \int_{\underline{\theta}} \sum_{v \in V} q(s, v|\underline{z}, \underline{\theta}) q(\underline{\theta}) q(\underline{z}) d\theta dz \tag{70}$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} \sum_{v \in V} q(s, v|\underline{z}^{(n,r)}, \underline{\theta}^{(n,r)}, y_n). \tag{71}$$

c) Final Expression of $I(S; \acute{Z})$

$$I(S; \acute{Z}) = \frac{1}{N} \sum_{n=1}^{N} \text{KLD} \left(\frac{1}{R} \sum_{r=1}^{R} \sum_{v \in V} q(s, v|\underline{\acute{z}}^{(n)}, \underline{\theta}^{(n,r)}, y_n) \parallel \frac{1}{R} \sum_{r=1}^{R} \sum_{v \in V} q(s, v|\underline{z}^{(n,r)}, \underline{\theta}^{(n,r)}, y_n) \right). \tag{72}$$

d) Empirical Formulation for $I(S; \acute{Z})$

$$I^*(S; \acute{Z}) = \frac{1}{B} \sum_{b=1}^{B} \left(\frac{1}{|D|} \sum_{d \in D} \text{KLD} \left(\frac{1}{R} \sum_{r=1}^{R} q(s|\underline{\acute{z}}^{(b,d)}, \underline{\theta}^{(b,d,r)}, y_{b,d}) \parallel \frac{1}{R} \sum_{r=1}^{R} q(s|\underline{z}^{(b,d,r)}, \underline{\theta}^{(b,d,r)}, y_{b,d}) \right) \right). \tag{73}$$

e) Clarification of Prior Derivations

- R denotes the number of Monte Carlo simulations, distinct from N and M in level, serving as the number of data sub-partitions for computational efficiency.
- $\underline{z}^{(b,d,r)}$ is defined as $\underline{z}^{(b,d,r)} = \{z_{(b,d,r,1)}, z_{(b,d,r,2)}, \ldots, z_{(b,d,r,M)}\}$, where each $z_{(b,d,r,m)}$ is an element in the set and m is an index ranging from 1 to M. In practice, $\underline{z}$ is treated as a 5D tensor, indexed by a batch index b, a sample index within the batch d, a data-partition indicator r, a sub-repeat index m, and a latent vector dimension index j. The values of $\underline{z}$ are randomly sampled along the dimensions b, d, r, and j, and remain constant across the dimension m, denoted as $\underline{z}_{b,d,r,m,j} = \underline{z}_{b,d,r,j}\ \forall m$.
- $\underline{\acute{z}}^{(b,d)}$ is defined as $\underline{\acute{z}}^{(b,d)} = \{\acute{z}_{(b,d,r,1)}, \acute{z}_{(b,d,r,2)}, \ldots, \acute{z}_{(b,d,r,M)}\}$ and treated as a 5D tensor. It is indexed by a batch index b, a sample index in the batch d, data-partition indicator r, sub-repeat index m, and latent vector dimension index j. Initially, the values of $\underline{z}$ are randomly sampled across the dimensions b, d, and j, and remain constant across the dimensions r and m, denoted as $\underline{z}_{b,d,r,m,j} = \underline{z}_{b,d,j}\ \forall r, m$. Subsequently, except for the isolated element $\underline{z}_{j=j^*}$, which retains its original value from $\underline{z}_j$, all other elements are replaced by zero.
- $\underline{\theta}^{(b,d,r)}$ is defined as $\underline{\theta}^{(b,d,r)} = \{\theta_{(b,d,r,1)}, \theta_{(b,d,r,2)}, \ldots, \theta_{(b,d,r,M)}\}$ and operated as a 5D tensor, indexed based on a batch index b, a sample index in the batch d, data-partition indicator r, sub-sampling index m, and latent vector dimension index k. The values of $\underline{\theta}$ are randomly sampled along the dimensions b, d, r, m and k.
- $q(s, v)$ is defined as a 5D tensor, $q(s, v) \in \mathbb{R}^{B \times |D| \times R \times |V| \times |S|}$, with indices comprising a batch index b, a sample index in the batch d, data-partition indicator r, a random variable frequency index v, and a random variable permutation index s.

(vi) $\boldsymbol{I(S; \acute{\Theta} | \acute{Z})}$: MI between the PD-PSD and a sorted sequence of M pieces of ancillary information, denoted as $\underline{\acute{\theta}}$, conditioned on $\acute{Z}$. It quantifies how much information is gained about S by knowing $\acute{\Theta}$ in addition to $\acute{Z}$, thereby reflecting the additional information that $\acute{\Theta}$ provides over $\acute{Z}$.

a) Basic Analytical Derivation of $I(S; \acute{\Theta} | \acute{Z})$

$$I(S; \acute{\Theta}|\acute{Z}) = \mathbb{E}_{q(s|\underline{\acute{\theta}}, \underline{\acute{z}}) q(\underline{\acute{\theta}}, \underline{\acute{z}})} \left[\log \frac{q(s, \underline{\acute{\theta}}|\underline{\acute{z}})}{q(s|\underline{\acute{z}}) q(\underline{\acute{\theta}}|\underline{\acute{z}})} \right] \tag{74}$$

$$= \mathbb{E}_{q(\underline{\acute{\theta}}, \underline{\acute{z}})} \left[\sum_{s \in S} q(s|\underline{\acute{\theta}}, \underline{\acute{z}}) \log \frac{q(s|\underline{\acute{\theta}}, \underline{\acute{z}}) q(\underline{\acute{\theta}}, \underline{\acute{z}}) / q(\underline{\acute{z}})}{q(s|\underline{\acute{z}}) q(\underline{\acute{\theta}}|\underline{\acute{z}})} \right] \tag{75}$$

$$= \mathbb{E}_{q(\underline{\acute{\theta}}, \underline{\acute{z}})} \left[\sum_{s \in S} q(s|\underline{\acute{\theta}}, \underline{\acute{z}}) \log \frac{q(s|\underline{\acute{\theta}}, \underline{\acute{z}})}{q(s|\underline{\acute{z}})} \right]. \tag{76}$$

b) Approximation of $I(S;\acute{\Theta}|\acute{Z})$

$$I(S;\acute{\Theta}|\acute{Z}) = \int_{\underline{\acute{z}}}\int_{\underline{\acute{\theta}}} q(\underline{\acute{z}})q(\underline{\acute{\theta}}) \left(\sum_{s\in S} q(s|\underline{\acute{\theta}},\underline{\acute{z}}) \log \frac{q(s|\underline{\acute{\theta}},\underline{\acute{z}})}{q(s|\underline{\acute{z}})} \right) d\underline{\acute{\theta}} d\underline{\acute{z}} \tag{77}$$

$$\approx \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(q(s|\underline{\acute{\theta}}^{(n)},\underline{\acute{z}}^{(n)},y_n) \parallel q(s|\underline{\acute{z}}^{(n)},y_n) \right) \geq 0, \tag{78}$$

where $q(s|\underline{\acute{z}}^{(n)},y_n)$ is approximated as follows:

$$q(s|\underline{\acute{z}}^{(n)},y_n) \approx \frac{1}{R}\sum_{r=1}^{R}\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(n)},\underline{\theta}^{(n,r)},y_n). \tag{79}$$

c) Final Expression of $I(S;\acute{\Theta}|\acute{Z})$

$$I(S;\acute{\Theta}|\acute{Z}) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{KLD}\left(\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(n)},\underline{\acute{\theta}}^{(n)},y_n) \parallel \frac{1}{R}\sum_{r=1}^{R}\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(n)},\underline{\theta}^{(n,r)},y_n) \right). \tag{80}$$

d) Empirical Formulation for $I(S;\acute{\Theta}|\acute{Z})$

$$I^*(S;\acute{\Theta}|\acute{Z}) = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{|D|}\sum_{d\in D} \mathrm{KLD}\left(\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(b,d)},\underline{\acute{\theta}}^{(b,d)},y_{b,d}) \parallel \frac{1}{R}\sum_{r=1}^{R}\sum_{v\in V} q(s,v|\underline{\acute{z}}^{(b,d)},\underline{\theta}^{(b,d,r)},y_{b,d}) \right) \right). \tag{81}$$

e) Clarification of Prior Derivations

- $\underline{\acute{z}}^{(b,d)}$ is defined as $\underline{\acute{z}}^{(b,d)} = \{\acute{z}_{(b,d,1)},\acute{z}_{(b,d,2)},\ldots,\acute{z}_{(b,d,M)}\}$ and treated as a 4D tensor. It is indexed by a batch index b, a sample index in the batch d, sub-repeat index m, and latent vector dimension index j. Initially, the values of $\underline{z}$ are randomly sampled across the dimensions b, d, and j, but remain constant across the dimension m, denoted as $\underline{z}_{b,d,m,j} = \underline{z}_{b,d,j}\ \forall m$. Subsequently, except for the isolated element $\underline{z}_{j=j^*}$, which retains its original value from $\underline{z}_j$, all other elements are replaced by zero.
- $\underline{\acute{\theta}}^{(b,d,m)}$ is defined as $\underline{\acute{\theta}}^{(b,d,m)} = \{\acute{\theta}_{(b,d,1)},\acute{\theta}_{(b,d,2)},\ldots,\acute{\theta}_{(b,d,M)}\}$ and treated as a 4D tensor. It is indexed based on a batch index b, a sample index in the batch d, sub-sampling index m, and latent vector dimension index k. The values of $\underline{\acute{\theta}}$ are randomly sampled along the dimensions b, d, m, and k, and then sorted in ascending order over the index m.

IV. Descriptions of Supplemental Animated Materials

For enhanced understanding, additional animated materials are provided, with each file accessible at https://github.com/JunetaeKim/VABAM/tree/main/Figures.

Anim. 1 VABAM (Our Model) Synthesis Results displays the outcomes of 100 synthesis experiments with VABAM, showcasing the ability to maintain the original morphology of signals by preventing phase changes and horizontal movements along the time axis.

Anim. 2 C-VAE Synthesis Results illustrates the results of 100 synthesis experiments using C-VAE, showing the difficulty of preserving morphological identities when incorporating PSD values as conditional inputs.

Anim. 3 Pass Filter Convolution Operation Animation depicts how a pass filter functions through the convolution between a signal and the filter, emphasizing the resultant loss of information due to edge effects observable in the right part of the animation.

Anim. 4 Trade-off between Cleansing Efficacy and Morphological Alteration elucidates the trade-off relationship between the effectiveness of signal cleansing and the morphological changes resulting from the repeated application of a pass filter.

References

[1] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, "Variational autoencoder," in *Variational Methods for Machine Learning with Applications to Deep Networks*, pp. 111–149, Springer, 2021.

[2] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in $beta$-vae," *arXiv preprint arXiv:1804.03599*, 2018.

[3] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712*, 2016.

[4] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[5] R. T. Chen, X. Li, R. B. Grosse, and D. K. Duvenaud, "Isolating sources of disentanglement in variational autoencoders," *Advances in neural information processing systems*, vol. 31, 2018.