NAVIGATING SLEEP: NEURAL AUDIO SYNTHESIS AS COMPOSITIONAL MEDIUM FOR EEG DATA SONIFICATION

Stefanos Vasilakis

Intelligent Instruments Lab / ACUTE Research Lab
University of Iceland
stv@hi.is

Areti Andreopoulou *Thanos Polymeneas Liontiris*

Laboratory of Music Acoustics and Technology (LabMAT)
National and Kapodistrian University of Athens, Greece
a.andreopoulou@music.uoa.gr
athpoli@gmail.com

ABSTRACT

This paper presents a practice-based artistic research on data-driven composition that uses EEG bandpower to navigate the latent spaces of RAVE models, constituting what Grond and Berger refer to as 'musification'. Through this practice-based exploration, the paper examines how RAVE's latent space architecture shapes the transformation of EEG bandpower into extended musical form, and what creative possibilities and structural constraints emerge from this process. This study intersects human-machine co-musicianship and artistic data sonification in the form of data-driven composition. RAVE-based data navigation presents challenges for traditional sonification taxonomies, suggesting that sonification discourse may need to accommodate this class of algorithmically opaque, practice-dependent approaches that blur the boundary between parameter mapping and compositional material exploration.

1. INTRODUCTION

Sleep represents a liminal state characterized by unconscious processes we can measure but not fully comprehend. Sleep occupies one third of the lifetime of an individual and is a cognitive state in which brain activity manifests in an extraordinary manner, as in dreams or peculiar examples of sleep disorders unified under the umbrella term of "parasomnias" [2]. Even in sleep tests like polysomnography, we capture electrical traces of brain activity without grasping the subjective experience of sleep itself.

The choice of neural audio synthesis as the primary sound generation method originated from conceptual resonances between the phenomenon under study and the technology employed. The poetics emerging from the metaphorical association between neural oscillations of brain function and neural audio synthesis is hard to overlook. The aspect of unexplored brain function resonates with the indeterminate function of neural networks: RAVE models operate as complex systems whose latent representations we can navigate without fully understanding the learned features that structure their sonic output. According to this metaphor, the latent representations of the models used map onto the sleep data as latent representations of the phenomenon of sleep.

This inquiry is guided by three research questions. **RQ1:** How do RAVE models transform EEG bandpower data into sonic material, and what creative possibilities and limitations emerge from this process? **RQ2:** How do the three "actors" (sleep data structure, RAVE model topology, and compositional intervention) negotiate agency in shaping the musical outcome? **RQ3:** What pipeline design decisions proved critical when adapting physiological time-series data for neural synthesis, and what considerations from this process might inform other data-driven composition practices?

2. BACKGROUND

EEG analyzes macroscopic electromagnetic brain activity through the measurement of signals that reflect summed electrical potentials from large populations of cortical neurons [3]. The primary brain waves Delta (0.5–4 Hz), Theta (4–7 Hz), Alpha (8–12 Hz), Sigma (12–16 Hz), and Beta (13–30 Hz) are present simultaneously in various mental and cognitive states and can be measured in different proportions and locations. Sleep staging is the process of analyzing and classifying monitored sleep into discrete stages based on these brain wave patterns. Sleep is divided into non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep; NREM is further subdivided into N1 (light sleep), N2, and N3 (deep or slow-wave sleep). Under normal conditions, sleep cycles consist of recurring sequences of these stages in 90–120 minute intervals across the night [3].

RAVE is a Variational Auto-Encoder (VAE) algorithm, i.e. a deep learning generative model for unsupervised learning, modelling and generation of audio, developed by the Artificial Creative Intelligence and Data Science (ACIDS) group at the IRCAM laboratory [4]. Generative models aim to represent a given dataset by modelling its underlying distribution in the form of latent variables $(z_0, z_1, z_2, \ldots, z_n)$ that constitute the sonic identity of the dataset. The spatial distribution of these data-points contained in the model across the axes of each learned feature is what forms the "latent space." What differentiates the interaction with VAEs and makes them promising as musical instruments is their potential for unconditional audio generation, that is, generation without any audio input to the encoder. By feeding numeric values directly to the exposed dimensions of the latent space, one can navigate the space of learned sonic properties and provide the model with control signals to synthesize sound by modulating the latent representation variables. Intuitively, this way of using RAVE is equivalent to using a synthesizer, where the input data acts as the control signal directly targeting a position in the model's learned sound manifold [4].

3. LITERATURE REVIEW

The literature on sonification of brain function data is extensive, both for utilitarian and artistic purposes. The most notorious earliest example of artistic work exploring the use of brain function data for musical purposes is Alvin Lucier's *Music for Solo Performer*, in which sound is controlled by the brain signals of the performer. Alongside Lucier's experimentations, pioneering composers such as David Rosenboom and Richard Teitelbaum extensively incorporated biosignals, especially EEGs, as control signals for various parameters of synthesizers [5]. Sonification of EEG data during sleep has also been examined: Fernandes et al. [6] converted sleep EEG data into music with tempo varying by sleep stages, while Wu et al. [7] experimented with the mapping of EEG features like power, amplitude and frequency to intensity, pitch and duration of notes in order to differentiate between deep sleep and REM stages.

In recent years, many artists, musicians and researchers have leveraged the potential of neural networks for sound generation and timbral exploration [8]. The affordances of these new technologies has redefined the traditional notion of creativity, intention and authorship, extended the boundaries of musical expression and control, and raised discussion around the agency of actors in the co-creative process. As described in practice-based research with neural synthesis, the main three components of composing with neural synthesis are: entanglement, arbitrariness and continuity [9]. Entanglement refers to the interdependency of the sonic features of the latent space, where a characteristic of the learned sound cannot be separately controlled without affecting the others. Arbitrariness refers to the lack of control on the formation of sound features during the training process, implying that the knowledge acquired through interaction with a specific model is not transferable to interaction with another one. Continuity refers to the structure of latent spaces which "represent continuous dimensions and cannot be selectively deactivated" [9]. These characteristics impose that interaction with neural audio synthesis is primarily explorative since "the music changes, moves and transforms in an uncertainty" [8].

4. SONIFICATION PIPELINE

The data used were acquired from the Bitbrain Open Access Sleep (BOAS) dataset [10], containing PSG recordings from 108 healthy adults. Subject selection employed a two-stage process: quantitative ranking by sleep stage balance using a sum-of-squared-deviations metric from ideal 20% distribution across all five stages, and qualitative assessment by visual inspection of the hypnogram structure to ensure physiologically representative sleep architecture with at least 3–4 REM and Deep Sleep transitions. Subject sub-19 was selected for exhibiting both balanced stage distribution (Wake: 18.8%, N1: 3.3%, N2: 43.8%, N3: 15.9%, REM: 18.3%) and naturalistic sleep progression beginning with wakefulness.

A one-to-one data-to-sound transformation would result in an equal duration sonic result which would barely convey the trajectory of sleep stages in a perceivable manner, since the transitions would blur through extended audition and subsequent attention fatigue. A 30:1 time compression ratio was applied by extracting bandpower features from 2.56-second epochs and transmitting them at intervals of 0.0853 seconds, enabling audition of a full-night session in approximately 14 minutes. The RAVE models' 2048-sample block size at 48 kHz establishes a frame rate of approximately 23.44 Hz, the optimal frequency for updating latent vector controls to achieve smooth synthesis. Initial experiments using standard 30-second epochs with 1 Hz transmission rate produced stuttering due to insufficient update frequency relative to RAVE's 23.44 Hz frame rate: the model expects high variation of numerical values to move organically in the latent space, instead of the sequential, slow incrementing between distant values. By switching to 2.56 seconds, the stuttering effect disappeared.

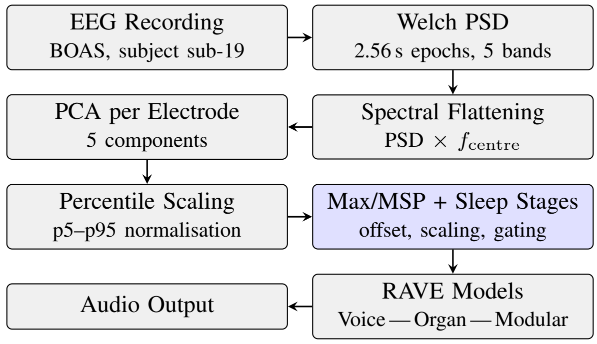

Figure 1: Sonification pipeline from EEG recording to audio output. Sleep stage labels in the Max/MSP patch serve as symbolic contextual signifiers [11] that gate which electrode–model combination is active at each moment.

Bandpower was calculated per epoch using Welch's periodogram to estimate Power Spectral Density, yielding five absolute values corresponding to Delta, Theta, Alpha, Sigma, and Beta bands. The 1/f aperiodic component of neural oscillations means that Delta waves naturally dominate the power spectrum. Without correction, this Delta dominance would sonically mask the contributions of higher-frequency bands like Beta and Sigma. These bands are critical for differentiating wake, REM, and N2 sleep stages in the composition. Spectral flattening was applied by multiplying each band's PSD by its center frequency, equalizing representational weight across bands. This approach trades representational accuracy for compositional range: where scientific sonification demands fidelity to data relationships as they exist, musification permits a transformation that serves aesthetic and experiential goals.

PCA was then applied independently to each electrode, extracting five orthogonal principal components ordered by variance. This decorrelation prevents any single frequency band from dominating the mapping while creating a variational hierarchy where PC1 captures the dominant spectral pattern (slow-wave vs. fast-wave balance) and PC2 the next pattern (sleep spindle activity), mirroring RAVE's own variance-ordered latent dimensions. PC1 (66–76% variance) captured sleep depth through slow-wave vs. fast-wave balance, creating a clear compositional hierarchy where PC1 controlling the first latent dimension becomes the primary narrator of sleep progression. Data transformation through PCA enabled mapping data to sound in order of salience [1].

For the data-driven composition, the whole night recording of EEG data was used, keeping the inherent temporal evolution of the dataset. The data from each electrode were kept separate, and the manually referenced sleep staging dataset was used alongside

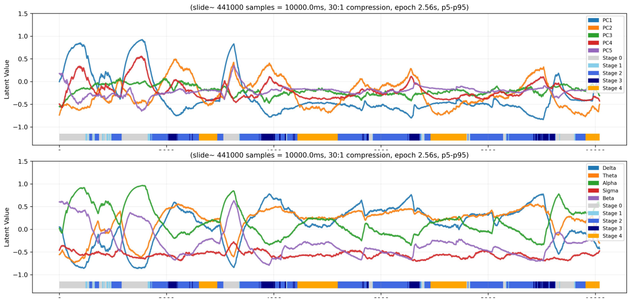

Figure 2: Subject sub-19, electrode C4 (30:1 time compression, 2.56 s epochs). Top: principal components of relative bandpower across the hypnogram. Bottom: relative bandpower across the hypnogram. The colour bar indicates sleep stages.

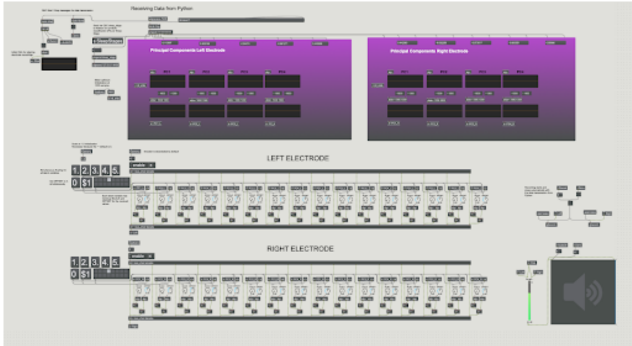

Figure 3: Max/MSP patch for real-time EEG-to-RAVE latent space navigation. The upper panels display incoming PCA components per electrode; the lower sections provide offset and scaling controls for each latent dimension of the left and right electrode RAVE models.

the neural activity data, providing knowledge of which sleep stage was associated with the data as a function of time. All sonic material was produced by unconditional audio generation (i.e. without feeding audio input to the encoder) by navigating the latent space with the processed data, while the sleep stages were used to signify which electrode-model combinations would sonify each sleep stage.

This distinction of roles in the sonification design draws from the concept of signal and symbol high-order structure of data transformation as a framework for data display design formulated in [11]. The concept of signal pertains to structural transformations that encode spatial and temporal relations of the dataset; the concept of symbol represents contextual transformations that allow for comparative insights across the dataset [11]. Following this structure, the EEG bandpower data across time constitutes the signal and the sleep stages as categorical distinctions constitute the symbol providing the contextual meaning. Three pre-trained RAVE models were used although several were tested: a voice model (RAVE v3, 16 latent dimensions), an organ model (modified RAVE v1, 16 latent dimensions), and a model trained on modular synthesizer recordings (RAVE v3, 8 latent dimensions), all trained at the Intelligent Instruments Lab, University of Iceland.

5. FINDINGS

Models trained on homogeneous, smooth corpora produced artifacts when latent values exceeded approximately ± 1.5, while models trained on more diverse material remained stable across the full ± 3 range. This reveals that the useful navigable regions of a model's latent space is determined not by the model architecture itself, but by the diversity of the training corpus. This finding shaped compositional decisions about which models to assign to which electrodes based on their expected dynamic range.

For the Modular model, trained on diverse content with heterogeneous sonic textures, encoder activation proved essential: without it, discontinuities emerged at block boundaries as a stuttering effect when adjacent blocks contained very different material. The Organ and Voice models, trained on homogeneous sustained sounds, worked better with the encoder deactivated, producing pad-like textures. In other words, homogeneous training data creates similar adjacent blocks requiring no temporal context, while diverse corpora demand hidden state continuity to prevent audible incoherent jumps in the latent space. This technical constraint became a compositional parameter: encoder activation shaped whether models responded with immediate, static textures or with the emergence of micro-structures hidden in the training dataset.

The interaction between model and data were the main actors in defining the sonic material of each electrode. A significant constraint arose from data with negative values (e.g. the PC1 in the NREM stages) when modulating latent variables corresponding to loudness: the output of the decoder was silent. Such type of data agency in the form of signal decreases directly constrained aesthetic decisions. The offset and scaling mechanism had more of a taming role, an influence on what seemed to be external to compositional intention. For that reason, sleep stages were treated as symbolic contextual signifiers that directed which sonic material would be audible [11], effectively muting portions of each soundscape deemed problematic while keeping the temporal structure intact.

Another constraint that shaped the practice was the dynamic character of neural audio synthesis. It would be useless to manually navigate the data points in an order other than their order of appearance and tune the system according to compositional intention: specific decisions for each sleep stage could not be made in a static manner because the result would be different when the actual navigation of the data points occurred. The only efficient solution was to tune the system as the data was flowing in the form of a closed interaction loop involving real-time control and auditory feedback. In that sense, the sonic material was emerging in the making. The musical outcome was dictated indivisibly by the entanglement between data, models and aesthetic intentions: data trajectory determined temporal evolution, model topology constrained sonic possibilities, and compositional intervention navigated between these constraints.

6. THEORETICAL POSITIONING

RAVE-based data sonification presents challenges for traditional sonification taxonomies. Unlike Parameter Mapping Sonification (PMSon), where explicit rules connect data features to sonic parameters, RAVE navigation operates through implicit, learned relationships within a high-dimensional latent space. The mapping is indirect: data values position the decoder within a sonic manifold whose structure was shaped by the training corpus, not by designer intention. In traditional parameter mapping, a designer specifies "PC1 $\rightarrow$ filter cutoff" or "bandpower $\rightarrow$ amplitude," establishing explicit, interpretable control relationships. In RAVE navigation, "PC1 $\rightarrow z_0$" is specified, but it can not be predetermined that z_0 in a specific model would control e.g. both amplitude and pitch, or that negative values would produce silence. Examples of model specific behaviors, such as that the voice model's z_{15} contained content from the recording's space reverb were discoveries rather than design choices. The mapping exists since data modulates latent dimensions which affect sonic characteristics but the sonic meaning of each dimension is arbitrary, formed by algorithmic processes during training rather than human specification.

RAVE navigation shares significant characteristics with Model-Based Sonification (MBSon) that distinguish both from traditional PMSon. Like MBSon, RAVE uses a dynamic model to mediate between data and sound, introducing a model space where temporal evolution generates acoustic output [12]. MBSon requires user interaction to repeatedly excite a dissipative system that returns to silence without continued excitation: users "actively query the object, which answers with sounds" [12, p. 402]. In MBSon designers have complete knowledge of model behavior. RAVE navigation, by contrast, involves double opacity: both the model's sonic behavior and the data patterns are discovered simultaneously. This positions the composer not as tool-designer but as co-explorer alongside the algorithmic system, navigating learned representations whose internal logic remains inscrutable.

This suggests that sonification taxonomies developed before machine learning methods may need expansion. Current frameworks distinguish sonification techniques by how mappings are specified and data is queried [12]. RAVE navigation introduces a third category: methods where mappings are learned from training data and discovered through practice, creating implicit rather than explicit control relationships. As neural synthesis tools become more prevalent in artistic practice, sonification discourse may need to accommodate this class of algorithmically opaque, practice-dependent approaches that blur the boundary between parameter mapping and compositional material exploration.

7. CONCLUSIONS

The metaphorical association between sleep's neural opacity and RAVE's computational opacity proved productive throughout the creative process. Both sleep and neural synthesis operate as measurable yet opaque systems. Methodologically, this research provides a documented pipeline for adapting physiological time-series data to neural audio synthesis, addressing challenges such as PCA for decorrelation and structural alignment, percentile scaling for outlier management, and epoch length optimization for frame rate compatibility. Theoretically, the work positions RAVE latent space navigation as a hybrid sonification technique that resists traditional taxonomy.

Several directions could extend this research. Using multiple subjects' data could enable comparative studies or multi-movement compositions exploring individual differences in sleep architecture. Training custom RAVE models on sleep-related sounds such as breathing and ambient night recordings might create stronger conceptual coherence. This research suggests that neural audio synthesis, despite (or perhaps because of) its opacity and unpredictability, offers creative potential for data-driven composition. The challenge lies not in controlling these tools precisely but in learning to navigate their learned spaces purposefully, embracing uncertainty as creative constraint rather than technical limitation.

8. REFERENCES

[1] F. Grond and J. Berger, "Parameter mapping sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Publishing House, 2011.

[2] J. A. Fleetham and J. A. E. Fleming, "Parasomnias," *CMAJ*, vol. 186, no. 8, pp. E273–E280, 2014. https://doi.org/10.1503/cmaj.120808

[3] A. S. BaHammam, *Clinical Atlas of Polysomnography*. CRC Press, 2018.

[4] A. Caillon and P. Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv:2111.05011*, Dec. 2021.

[5] M. A. Ortiz, N. Coghlan, J. Jaimovich, and R. B. Knapp, "Biosignal-driven art: Beyond biofeedback," in *Proc. NIME 2011*, 2011.

[6] C. M. Fernandes, D. Migotina, and A. C. Rosa, "Brain's Night Symphony (BraiNSy): A methodology for EEG sonification," *IEEE Trans. Affective Computing*, vol. 12, no. 1, pp. 103–112, 2021. https://doi.org/10.1109/TAFFC.2018.2850008

[7] D. Wu, C.-Y. Li, and D. Yao, "Scale-free music of the brain," *PLoS ONE*, vol. 4, no. 2, e5915, Feb. 2009. https://doi.org/10.1371/journal.pone.0005915

[8] K. Tahiroğlu, M. Kastemaa, and O. Koli, "AI-terity 2.0: An autonomous NIME featuring GANSpaceSynth deep learning model," in *Proc. NIME 2021*, 2021. https://doi.org/10.21428/92fbeb44.3d0e9e12

[9] N. Privato, G. Lepri, T. Magnusson, and E. T. Einarsson, "Sketching magnetic interactions for neural synthesis," in *Proc. Int. Conf. Technologies for Music Notation and Representation*, pp. 89–97, 2024.

[10] E. López-Larraz *et al.*, "Bitbrain Open Access Sleep Dataset," OpenNeuro, 2025. https://doi.org/10.18112/OPENNEURO.DS005555.V1.1.0

[11] M. A. Crozzoli, "Dark sonification: An entangled post-interaction multimodal data display system," in *Doctoral Consortium, Sixth Decennial Aarhus Conference*, Aug. 2025.

[12] T. Hermann, A. Hunt, and J. Neuhoff, *The Sonification Handbook*, 1st ed. Logos Publishing House, 2011.