NEBULA: A PCA-Based Method to Explore RAVE-Encoded Audio Representations

Moisés Horta Valenzuela (`moises.valenzuela@kunstuni-linz.at`)[1] and
Enrique Tomás (`enrique.tomas@kunstuni-linz.at`)[1]

[1]*Tangible Music Lab*, **Kunstuniversität Linz**, Hauptplatz 6, 4020 Austria

ABSTRACT

The challenges of exploring the latent spaces of deep-learning audio models often lead artists to rely on chance, randomness, and combinatorial approaches, making it difficult to steer these models toward musically meaningful outcomes. In this paper, we explore how Principal Component Analysis (PCA) applied to pre-encoded RAVE (Realtime Audio Variational Autoencoder) latent representations can provide a more controlled and curated approach to navigating these high-dimensional spaces. By restricting exploration to selected regions of the latent space, musicians gain clearer pathways to achieving specific sonic goals. Although t-SNE and UMAP effectively preserve intricate local structures, we show how the linearity, computational efficiency, and interpretability of PCA offer distinct advantages for real-time applications. In addition, we introduce a graphical user interface (GUI) and a sensor system for manipulating 'timbral vectors' derived from PCA components, providing an intuitive tool for identifying, refining, and shaping sonic transformations. To evaluate the effectiveness of PCA, we systematically compare its performance with t-SNE and UMAP, highlighting the trade-offs among these methods.

1. INTRODUCTION

Neural synthesis methods have significantly advanced toward real-time generative audio capabilities and are increasingly being adopted by artists for creative applications [1–4]. Early approaches utilized deep autoregressive models such as WaveNet [5], SampleRNN [6], and WaveRNN [7], as well as Fourier-based models such as Tacotron [8] and GANSynth [9].

Modern approaches based on Variational Autoencoders (VAEs) have gained popularity because they allow for fast, high-quality audio synthesis, and direct control over generation by exposing latent variables. VAEs require a time-intensive training phase using large datasets of audio. However, once trained, models like RAVE [10] enable real-time high-quality sound synthesis at relatively low latency. Recent projects developed with RAVE include *La-*

tent Terrain by Shuoyang Jasper Zheng [1] (2024), *Mouja* by Nicola Privato [2] (2024), *semilla.AI* by Moisés Horta Valenzuela (a.k.a. hexorcismos, first author) [3] (2023).

To comprehend the methods of sound generation within these artistic practices, it is necessary to first examine a few key technical concepts. The training model process produces high-dimensional latent spaces. They are a multidimensional compressed representation of data points informing about the model structure, and the dataset's learned audio features. Latent spaces only preserve essential features that will inform input data structures to generate the output space, in our case sonic results. Each dimension of a latent space corresponds to a latent variable learned from the original data. Latent variables are underlying characteristics that inform the way data are distributed, but they are usually entangled, not observable, and difficult to navigate in a linear way.

In audio synthesis projects using RAVE, artists can directly access the signals that feed the encoder and the decoder of the system (Figure 1 top schematic). In this aesthetic exploration of the variables of the latent space, we have identified a number of methods that are usually applied in this representational realm:

- Seed Interpolation: two or more latent vectors are selected, and intermediate states are generated, through linear or spherical interpolation. This approach typically results in smooth transitions between timbres or musical motifs. However, while visually appealing in dimensionality-reduced projections, these interpolations can be unpredictable, especially when the latent space is not well understood.

- Unconditional Exploration: this method refers to randomly sampling the latent space, a method frequently used in generative adversarial networks (GANs) and variational autoencoders (VAEs) for sonic discovery. Artists often embrace chance-based approaches to produce new or unexpected sounds. While this method can yield serendipitous results, it provides little control over musical direction or timbre consistency, making it difficult to achieve artistically coherent material.

- Timbre Transfer: the goal of this method is to apply

---

[1] `https://github.com/jasper-zheng/nn_terrain`
[2] `https://nicolaprivato.com/mouja`
[3] `https://semilla.ai/`

the spectral qualities of one sound onto another. It has gained popularity for cross-domain transformations, such as making a voice sound like a violin or transforming an environmental recording into a synthetic pad. Models like RAVE allow for low-latency transformations with high fidelity. However, aligning timbre nuances from different domains is not always intuitive and can involve large latent jumps that are difficult to navigate without a structured framework like dimensionality reduction.

Despite the creative potential of these approaches, they are largely based on trial and error. The key challenge is how to effectively explore the black-box nature of learned latent variables. Due to the high dimensionality of latent representations, direct parameter manipulation often results in unpredictable or unintended sounds. In models like RAVE, where latent spaces typically range from 4 to 32 dimensions, navigating them can feel like a process of random exploration and chance.

The training process yields high-dimensional latent spaces—compressed representations whose coordinates describe the audio features the model has learned. In RAVE, the first few latent dimensions (often < 4) already show partial disentanglement: they correlate consistently with broad timbral cues such as spectral centroid, overall loudness, or spectral spread. However, as dimensionality increases, the correspondence between any single coordinate and a perceptual factor rapidly degrades; later dimensions tend to encode mixtures of several lower-level attributes and become harder to interpret directly. This uneven interpretability motivates post-hoc structuring strategies, such as the PCA approach proposed here.

One way to address this complexity is by predefining control values to explicitly condition the generation process [11, 12]. Their strategy is to model the distribution of high-dimensional output space as a generative model conditioned on the input observation. Another approach, proposed by Vigliensoni and Fiebrink [13], applies Interactive Machine Learning through regression techniques. In this method, users iteratively provide training sets that pair locations in the human-performance space with corresponding locations in the model's latent space. Then, a regression algorithm learns to map between the two, enabling users to explore intermediate points. However, this approach requires a large number of training pairs. Without sufficient data, the model generates mappings that fail to accurately reflect the intended relationships.

This paper proposes an alternative approach: leveraging Principal Component Analysis (PCA) on RAVE encoded audio data to make latent spaces more interpretable and musically guided. By applying PCA, pre-encoded data points are automatically clustered based on their key characteristics, forming data clouds directly connected to the sonic properties of the encoding audio materials. By structuring access to the latent space in this way, we aim to bridge the gap between exploratory navigation and purposeful control, offering artists a clearer and more intuitive way to shape sound.
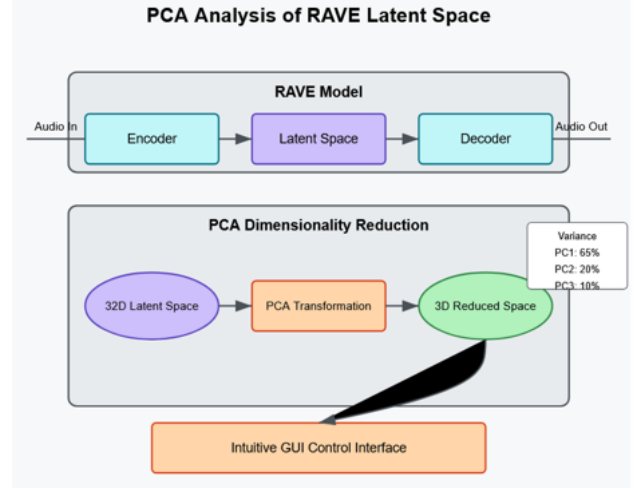

Figure 1. PCA Analysis of RAVE Latent Space (top) and schematic of PCA Dimensionality Reduction (bottom)

2. A PCA-BASED APPROACH TO EXPLORE ENCODED REPRESENTATIONS

2.1 Theoretical Foundations and Benefits of PCA in RAVE Models

Principal Component Analysis (PCA) provides a complementary lens through which to view RAVE's latent representations. While the variational autoencoder (VAE) already compresses the audio into 4–32 coordinates and partially separates dominant factors in the earliest axes, the later coordinates remain decidedly entangled and opaque to the user. PCA therefore does not seek to create disentanglement from scratch; instead, it re-orders, re-scales, and groups the existing latent directions so that the most perceptually coherent variations—whether they originate in dimension 0 or dimension 13—are surfaced first, giving musicians a clear set of "timbral sliders" to manipulate.

At its core, PCA involves the eigendecomposition of the data covariance matrix, yielding eigenvectors that define orthogonal directions of maximum variance. When applied to RAVE encodings, these eigenvectors can be conceptualized as "timbral vectors" capturing significant sonic variations across the dataset, with corresponding eigenvalues quantifying the variance explained by each principal component [14].

Several distinct advantages make PCA particularly well-suited for audio latent space exploration. Its linearity and interpretable nature ensure that movements in a particular direction yields predictable sonic transformations, helping artists build intuitive control over parameter adjustments [15]. PCA's computational efficiency makes it ideal for real-time applications, as the core transformation matrix can be pre-computed offline and applied rapidly during performance, even on resource-constrained systems like single-board computers used in RAVE deployment [14].

The parameter stability of PCA contrasts sharply with methods like t-SNE and UMAP, which depend on stochastic processes and complex hyperparameter tuning. PCA

yields deterministic results based solely on the input data, providing consistency, an essential element for musical performance and composition. Furthermore, PCA ranks dimensions by their variance contribution, allowing users to focus first on the most significant latent dimensions, providing a structured approach to exploring broad timbral shifts before addressing more subtle variations.

2.2 Methodological Implementation for Audio Analysis

2.2.1 Data Preparation and Encoding with RAVE

Our implementation of a PCA-based method for RAVE latent space exploration must effectively bridge computational analysis with artistic objectives. The process begins with assembling a curated audio corpus which encompasses the desired sonic palette, ensuring careful consideration of audio quality, diversity, and representativeness. Audio samples are then encoded using the pretrained RAVE encoder to produce a dataset of latent vectors. During encoding, audio files are converted into fixed-length frames, typically with the RAVE compression ratio being 2048 audio samples per one latent embedding of n-dimensional values (2048:1). The size of the VAE latent space dimension can vary depending on the model's training hyperparameters.

2.2.2 Feature Extraction and Analysis

Once the audio dataset is encoded into its latent representation, PCA decomposition reveals several types of information that guide artistic exploration. The eigenvalue spectrum indicates how variance is distributed across dimensions, offering an insight into the intrinsic dimensionality of the sonic material.

In well-trained RAVE models, the first three to five principal components (PCs) typically capture 70–85% of the corpus variance [14], and these high-energy PCs align neatly with salient perceptual descriptors. For instance, PC 1 often tracks spectral centroid, PC 2 follows amplitude-envelope shape, and PC 3 reflects harmonic richness [15]. Beyond this point, the variance explained by each additional PC drops sharply, and the corresponding directions increasingly blend several low-level features at once—mirroring the growing entanglement already present in the later raw RAVE coordinates. The PCA spectrum thus offers a quantitative map of where the model is already partially disentangled and where user guidance is still required.

2.2.3 Real-time Implementation Considerations

Implementing PCA-based navigation for real-time performance introduces several technical considerations. Linear interpolation in PCA space generally produces more predictable sonic results compared to direct interpolation in the original latent space, though non-linear interpolation curves may be used for expressive control.

Typically, the most significant principal components are assigned to primary controllers (e.g., accelerometer, x/y pads), while less significant components are mapped to secondary controllers or automated via envelopes [14]. Establishing soft or hard boundaries within the PCA space helps prevent excursions into unstable or undesirable sound regions, particularly in areas where the RAVE model may exhibit latent instabilities [15]. Although PCA is not a clustering algorithm, projecting the data onto the principal components often reveals natural groupings corresponding to playing techniques, articulation types, or instrument families, clarifying the topology of the latent space [16].

Additionally, tracking the temporal evolution of sounds in the PCA-reduced space can uncover characteristic trajectories—such as those representing the attack-sustain-release envelope—which can be reproduced, modified, or combined to generate new, predictably controlled gestures.

2.3 Comparison with Other Dimensionality Reduction Techniques

2.3.1 t-SNE (t-Distributed Stochastic Neighbor Embedding)

Systematic comparisons between PCA and prominent non-linear dimensionality reduction methods reveal important distinctions for audio applications. t-SNE (t-Distributed Stochastic Neighbor Embedding) is widely adopted for visualization due to its capacity to preserve local structure, making it effective at revealing cluster relationships in high-dimensional data [17]. Its strengths lie in preserving local neighborhoods exceptionally well, revealing subtle timbre relationships that might be lost in linear projections. The intuitive visual groupings of perceptually similar sounds are advantageous for exploratory analysis. However, t-SNE's computational expense (with $O(n^2)$ complexity) and its non-linear, stochastic nature make it challenging for real-time musical control. Its dependence on the perplexity parameter and tendency to distort global relationships limit its utility for predictable transitions between sonic states.

2.3.2 UMAP (Uniform Manifold Approximation and Projection)

UMAP (Uniform Manifold Approximation and Projection) offers a more recent alternative that preserves both local and global structures while being computationally more efficient than t-SNE [18]. UMAP demonstrates superior preservation of both local and some global structures compared to t-SNE and produces clearer separations between different playing techniques and articulations. Nonetheless, UMAP's inherent non-linearities complicate intuitive navigation. Parameters like n-neighbors require careful tuning, and different initialization seeds can lead to inconsistent embeddings, making it less reliable for real-time performance systems.

2.3.3 PCA Revisited: Linearity as an Advantage

Evaluations consistently demonstrate that while non-linear methods may produce sonically impressive results that highlight complex relationships, PCA offers more reliable and intuitive control for real-time musical applications.

3. IMPLEMENTATION

The implementation code and examples of use can be accessed from `https://github.com/tamlablinz/RAVE_PCA`

3.1 Dataset Composition and Experimental Framework

Our research leverages pre-trained open RAVE models from ACIDS-IRCAM[4] : 'wheel.ts' and 'darbouka.ts', trained on human speech and percussion audio datasets, respectively.

For comparative analysis, we encoded a 10-minute audio comprising two contrasting sonic categories: harsh noise improvisation (complex spectral content, minimal harmonic structure) and melodic Ondes Martenot compositions (clear harmonic content, sustained tones). This juxtaposition tests whether PCA can effectively separate and organize distinct timbral categories within a shared latent space. The resulting latent vectors (dimension=4) served as input for comparing dimensionality reduction techniques, enabling evaluation of how effectively each method could visualize latent organization and provide intuitive pathways for creative exploration. Figures 2-4 present a comparative visualization of PCA, t-SNE, and UMAP analyses using the 'wheel.ts' model, while Figures 5-7 show the same three techniques applied to the 'darbouka.ts' model. Figure 8 provides an additional PCA analysis of the 'darbouka.ts' model using a different dataset focused on percussion recordings, demonstrating the technique's consistency across varied audio sources.
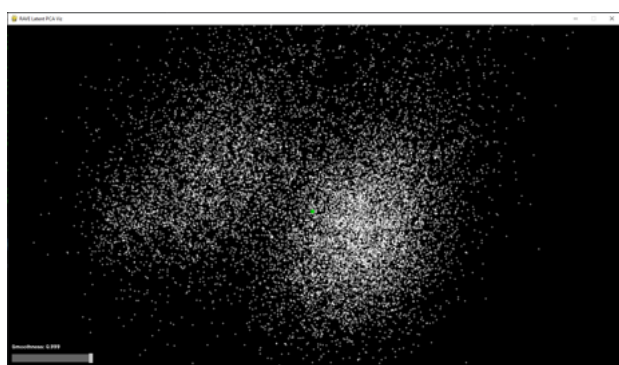

Figure 2. PCA Analysis with model 'wheel.ts'

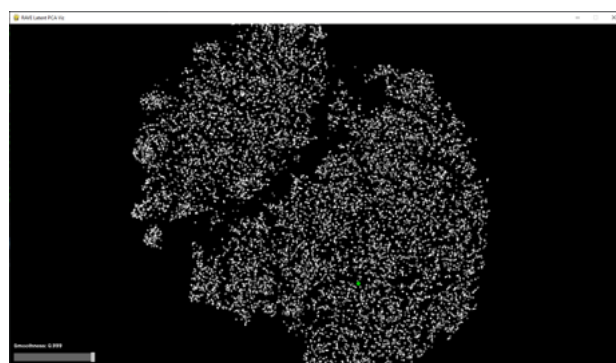Figure 6. t-SNE Analysis with model 'darbouka.ts'

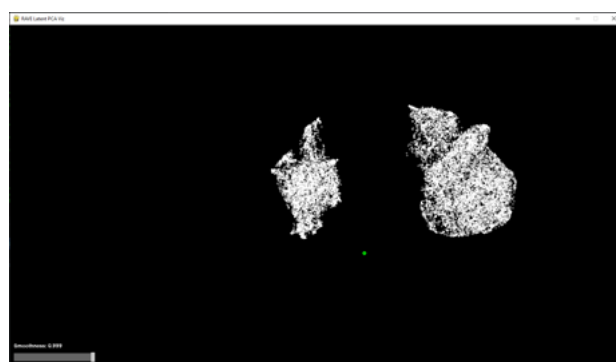Figure 3. t-SNE Analysis with model 'wheel.ts'
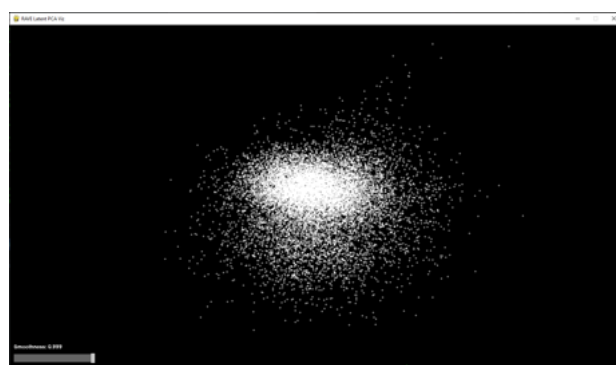

Figure 4. UMAP Analysis with the model 'wheel.ts'

Figure 5. PCA Analysis with model 'darbouka.ts'

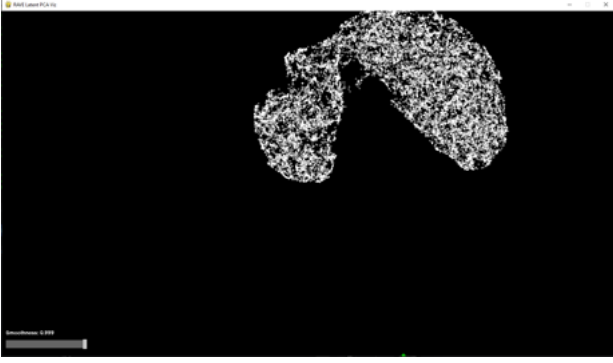[4] `https://acids-ircam.github.io/rave_models_download` accessed 3.3.25

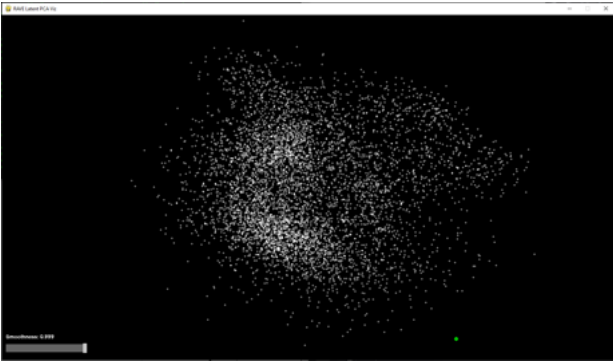Figure 7. UMAP Analysis with the model 'darbouka.ts'

Figure 8. PCA Analysis with the model 'darbouka.ts' of a different dataset (recording of percussion)
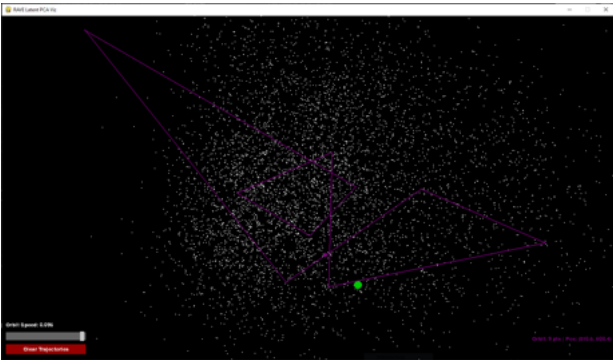

Figure 9. GUI with trajectory editor and player

3.2 Software Architecture and GUI Development

We developed a comprehensive framework integrating real-time audio synthesis with interactive visualization using a modular architecture: an Analysis Module (Python with scikit-learn), a Synthesis Engine (Pure Data, PyTorch with CUDA optimizations), and an Interactive Interface (Python with Dash/Plotly). This design efficiently bridges computational techniques with creative applications while maintaining sub-10ms response times even on modest hardware, ensuring suitability for live performance and sound design.

The graphical interface (figure 9) emphasizes intuitive interaction through key elements: an interactive projection visualization displaying latent vectors on principal components with immediate audio feedback, trajectory design

tools for creating and manipulating paths through the latent space, and flexible export capabilities for audio, control data, and visualization. Users can capture real-time movements, edit waypoints, control interpolation speed, and develop complex sonic patterns, effectively leveraging the organizational power of PCA within a creative workflow.

To operationalize these insights, a purpose-built graphical user interface leverages PCA-reduced representations for intuitive exploration of RAVE latent spaces. A central 2D or 3D display projects the encoded audio corpus onto the principal components, providing visual orientation within the sonic landscape. Users can define key points in the latent space by selecting exemplary sounds or manually setting coordinates, enabling smooth navigation between these points with adjustable interpolation curves [19].

The interface includes a path recorder system to capture routes through the latent space—via real-time controller input or algorithmic generation—allowing users to capture, loop, and combine trajectories for complex timbral gestures. A flexible parameter mapping matrix permits the assignment of principal components to various control inputs (MIDI controllers, OSC messages, internal LFOs/envelopes) with adjustable scaling and response curves. Real-time visual feedback displays spectral changes resulting from latent space navigation, helping users intuitively correlate visual positions with sonic outcomes [20].

This interface design prioritizes musical usability over technical complexity, providing artists with clear, predictable control over the sonic possibilities offered by the RAVE model.

3.3 Hardware

For musical experimentation with the system, we designed and built two digital musical instruments (see Figure 10) based on single-board computers that autonomously run the visualization system, audio synthesis engine, and user interface. At the core of our instruments is a Raspberry Pi 5 with 8GB RAM and active cooling. We connected it to a 4-inch round LCD display, which also functions as a multitouch interface for exploring the graphical visualization. On the display, users can select data points as well as pan, tilt, and zoom in or out of the graphical visualization.

To enable gestural interaction, we integrated a MPU6050 accelerometer and gyroscope, along with custom-made switches directly wired to the Raspberry Pi's GPIO. A simple Python script reads the General Purpose Input/Output (GPIO) inputs and transmits this data via the Open Sound Control (OSC) protocol to a sound engine programmed in Pure Data (Pd). This hardware setup allows users to select data points and timbral trajectories via the multitouch display while controlling sound expressively through the inclination of the instrument.

For sound generation, we utilized RAVE's compiled *nn-tilde* object for Pure Data on the Raspberry Pi, enabling real-time performance with custom-trained RAVE models. These models were trained on GPU-equipped computers using a configuration optimized for smaller model sizes. In practice, we did not perceive any noticeable increase in

latency when synthesizing sound on the Raspberry Pi compared to using *nn-tilde* on our laptops. The gestural sensor data was mapped to trigger sound envelopes, creating the sensation of playing discrete notes. Additionally, inclination along the two horizontal axes controlled the central frequency of a band-pass filter and the instrument's overall volume.

The single-board computers and sensor systems were housed in custom 3D-printed enclosures (figure 11). On one hand, we explored the possibility of generating physical objects derived from the PCA-generated clusters of data points. Given the complexity of these data clouds, we had to simplify them to create a printable structure. On the other hand, we also designed a more neutral enclosure optimized for gestural interaction with the instrument. In the interaction section, we discuss both approaches in detail.
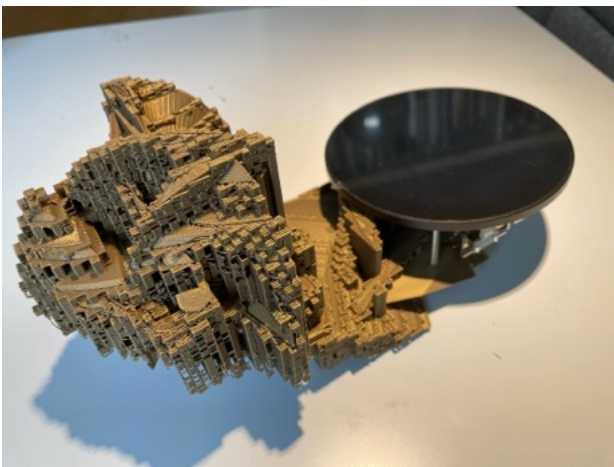

Figure 10. Instrument on a PCA-based 3D-printed representation (RaspberryPi 5, Waveshare multitouch LCD screen and MPU6050 sensor)

4. RESULTS

4.1 Comparative Analysis of Dimensionality Reduction Approaches

We systematically compared PCA, t-SNE, and UMAP using consistent hyperparameter optimization across the same corpus of encoded latent vectors. Our PCA implementation retained three principal components explaining 79.4% of variance, with clear correlations to audio features: PC1 with spectral centroid (r = 0.78), PC2 with temporal envelope characteristics, and PC3 with harmonic content. The t-SNE implementation used perplexity = 50 with 1000 iterations after testing multiple configurations, while UMAP parameters (n-neighbors = 15, min-dist = 0.1) balanced local and global structure preservation.

Quantitative evaluation metrics included neighborhood preservation, trustworthiness, continuity, and feature correlation. Results showed t-SNE excelled in neighborhood preservation (76.3%) but performed poorly on continuity (0.71). UMAP achieved better balance (68.9% preservation, 0.83 continuity), while PCA offered the most consistent global organization despite lower neighborhood preservation (53.2%). Crucially, PCA demonstrated the strongest correlation between its dimensions and perceptually relevant audio features, providing more interpretable navigation axes for creative applications.

4.2 Timbre Space Musical Exploration

The practical implementation of the system helped us to test the hypothesis that interacting with pre-encoded audio representations offers a more intuitive approach to exploring latent space in real-time sound synthesis.

Through a series of short musical improvisation sessions with volunteers from our department, we observed that PCA visualization effectively enabled participants to associate each cluster with a specific sonic quality and navigate the timbral map. In particular, we observed how participants observed that the central regions of each cluster exhibited minimal timbral variation, making them highly predictable, while the outer boundaries contained more distinguishable content, grouping sonically similar audio segments (e.g., based on pitch or noisiness) into distinct regions.

As participants gained familiarity with the timbral clusters, they were able to structure musical improvisations around their recognition of different sonic regions. The graphical interface played a crucial role in enhancing participants' understanding of how RAVE organizes timbral variations. It allowed for more deliberate navigation toward desired sonic outcomes, such as bright harmonic textures or subdued, textural sounds—a level of control that was less apparent in t-SNE or UMAP projections, where latent spaces appeared more entangled.

In conclusion, PCA-based navigation yielded interactions that participants described as more predictable and musically coherent than scanning the full latent space at random. We stress, however, that a base level of predictability already exists in the un-rotated RAVE axes—particularly the first two or three— which are only loosely correlated in our PCA but remain partially disentangled in their original form. What PCA chiefly adds is an ordering and weighting of those axes, highlighting the "clean" ones while compressing or combining the more entangled directions, thereby streamlining real-time control without discarding the expressive potential of the native latent space.

4.3 Interaction

Interacting with a multidimensional data space in real time is one of the most challenging tasks for a musician. The PCA approach helps mitigate this complexity by reducing the number of dimensions to explore. However, even with dimensionality reduction, designing interfaces that balance the inherent complexity of the data with intuitive control remains difficult.

In our case, the rounded multitouch LCD interface allowed for direct selection of data points within the interactive space. Without the need for intermediary controls, we could quickly choose timbres and create expressive trajectories between data points. The ability to loop sonic trajectories, a technique often used in musical video games, freed us from manually selecting data points in real time, enabling a more fluid interaction with the gestural system.

Figure 11. Performing two NEBULA musical instruments

Despite these advantages, physical interaction with the screen proved ergonomically limiting, restricting a more embodied exploration of the sonic content linked to the data points. Recognizing this limitation, we integrated a gestural sensor system to enhance the depth of embodied interaction.

To further investigate the musical potential of this system, the authors conducted three improvisation sessions with the instruments (see Figure 11). Video excerpts from our improvisations are available online[5] As experienced digital musicians already familiar with RAVE models, we found that the PCA-based method creatively constrained our musicking, shaping how we engaged with the pre-encoded audio material. The necessity of pre-encoding sound prior to performance structured our improvisations, leading us to develop a repertoire of PCA-generated data clouds that could be interactively loaded. These functioned as a form of graphic score, guiding our sonic explorations in new directions. Consequently, the PCA approach not only structured our performances but also inspired us to design a series of PCA-based sound maps for live improvisation.

5. DISCUSSION

The implementation of our Principal Component Analysis (PCA)-based method warrants further discussion. We have identified key advantages and limitations that inform its effectiveness in the context of audio navigation and sound design.

Advantages: The proposed PCA-based approach offers a fast, stable, and transparent means of reducing dimensionality, making it particularly well-suited for both live and studio applications. The principal components serve as intuitive, interpretable control parameters, effectively functioning as multidimensional "sliders" that allow artists to explore sonic transformations with deliberate intent. This structured navigation provides an accessible yet powerful framework for interacting with complex timbral spaces.

Limitations:

Despite its advantages, the linear nature of PCA presents certain constraints. Specifically, it may not effectively cap-

---

[5] https://github.com/tamlablinz/RAVE_PCA.

ture fine-grained variations in sound, particularly those associated with highly nuanced timbral characteristics. In contrast, nonlinear dimensionality reduction techniques (e.g., t-SNE or UMAP) can offer a more refined separation of features, which may be preferable for applications requiring highly specific timbral control. This limitation highlights a potential trade-off between computational efficiency and perceptual accuracy in sound exploration. Additionally, although PCA helps reorganise the latent space, it does not create disentanglement; RAVE's encoder already handles some of that work in its low-index dimensions, leaving only the higher-index coordinates densely entangled and hard to label. Consequently, our linear projection can still miss subtle timbral cues that live in those later, mixed dimensions.

To address these limitations and enhance the method's applicability, we propose the following refinements:

- Hybrid Dimensionality Reduction: A combined approach leveraging PCA for initial global structuring, followed by UMAP or t-SNE for localized refinements, could facilitate more context-sensitive sound design. This hybrid method would retain PCA's efficiency while integrating nonlinear adaptability where finer control is needed.

- Temporal Analysis of Latent Representations: Investigating the time-dependent evolution of embeddings within RAVE's latent space could provide new strategies for gesture-based transformations. This could enable dynamic sound-shaping methods informed by temporal patterns and expressive performance gestures.

- User-Labeled Axes for Enhanced Interpretability: Integrating semantic descriptors (e.g., "warm," "metallic," "grainy") alongside PCA-derived principal components could reinforce interpretability and afford users greater direct control over sonic exploration. By aligning computational features with perceptually meaningful attributes, this approach could bridge the gap between data-driven and artist-driven sound manipulation.

These developments aim to refine and extend the PCA-based method, ensuring greater expressivity, precision, and adaptability in creative audio applications.

6. CONCLUSIONS

This paper has introduced a PCA-based methodology for navigating and visualizing the latent space of a RAVE autoencoder, addressing a fundamental challenge in deep generative sound synthesis—the difficulty of directing models toward specific musical or timbral outcomes. By pre-encoding audio data and subsequently projecting the resulting latent vectors using Principal Component Analysis (PCA), this approach enables a more intuitive selection, refinement, and manipulation of sound.

A comparative analysis with t-SNE and UMAP underscores the trade-offs between interpretability and granularity in latent space representations. While PCA provides

transparent and efficient dimensionality reduction, nonlinear techniques such as t-SNE and UMAP offer finer detail at the cost of computational efficiency and real-time applicability.

Furthermore, the graphical user interface (GUI) developed as part of this study demonstrates how interactive visual and auditory feedback loops enhance artistic curation and musical exploration of the latent space. The findings suggest that PCA serves as an effective, lightweight, and interpretable tool for both sound design and performance-oriented applications. This research lays the groundwork for future exploration into hybrid approaches, combining PCA with nonlinear dimensionality reduction techniques, and for the development of user-guided strategies that further refine the controllability and expressivity of generative sound synthesis systems.

7. REFERENCES

[1] B. Caramiaux and M. Donnarumma, "Artificial intelligence in music and performance: a subjective art-research inquiry." in *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, 2021, pp. 75–95.

[2] B. Caramiaux and S. F. Alaoui, "Explorers of unknown planets. practices and politics of artificial intelligence in visual arts," in *Proceedings of the ACM on Human-Computer Interaction 6, CSCW2*, 2022, pp. 1–24.

[3] P. Esling and N. Devis, "Creativity in the era of artificial intelli- gence," in *arXiv preprint arXiv:2008.05959*, 2020.

[4] J. Kim, R. Bittner, A. Kumar, and J. Bello, "Neural music synthesis for flexible timbre control," 2019. [Online]. Available: https://doi:10.1109/ICASSP.2019.8683596

[5] A. van den Oord, S. Dieleman *et al.*, "Wavenet: A generative model for raw audio." in *arXiv preprint arXiv:1609.03499*, 2016.

[6] N. Kalchbrenner, E. Elsen *et al.*, "Efficient neural audio synthesis," in *arXiv preprint arXiv:1802.08435*, 2018.

[7] S. Mehri, K. Kumar *et al.*, "Samplernn: An unconditional end-to-end neural audio generation model," in *arXiv preprint arXiv:1612.07837*, 2016.

[8] Y. Wang, R. Skerry-Ryan *et al.*, "Tacotron: Towards end-to-end speech synthesis," in *In INTERSPEECH, 2017*, 2017.

[9] J. Engel, K. K. Agrawal *et al.*, "Gansynth: Adversarial neural audio synthesis," in *In International Conference on Learn- ing Representations*, 2019.

[10] N. Holighaus *et al.*, "Rave: A realtime audio variational autoencoder," in *Proceedings of the Sound and Music Computing Conference*, 2020.

[11] N. Devis, N. Demerlé, S. Nabi, D. Genova, and P. Esling, "Continuous descriptor-based control for deep audio synthesis," in *In ICASSP 2023- 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

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

[13] G. Vigliensoni and R. Fiebrink, "Steering latent audio models through interactive machine learning," in *14th International Conference on Computational Creativity (ICCC'23), Waterloo, ON (CA), 19-23 June, 2023*, 2023.

[14] I. T. Jolliffe, "Principal component analysis." Springer, 2002. [Online]. Available: https://www.springer.com/gp/book/9780387954424

[15] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: Data mining, inference, and prediction." Springer, 2009. [Online]. Available: https://web.stanford.edu/~hastie/ElemStatLearn/

[16] M. Müller, "Fundamentals of music processing: Audio, analysis, algorithms, applications." Springer, 2015. [Online]. Available: https://link.springer.com/book/10.1007/978-3-319-19121-9

[17] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," vol. 9, 2008, pp. 2579–2605. [Online]. Available: http://www.jmlr.org/papers/volume9/vandermaaten08a.html

[18] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2018. [Online]. Available: https://arxiv.org/abs/1802.03426

[19] J. Nielsen, "Usability engineering." Morgan Kaufmann, 1994. [Online]. Available: https://www.elsevier.com/books/usability-engineering/nielsen/978-1-55860-506-2

[20] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, "Designing the user interface: Strategies for effective human-computer interaction." Pearson, 2016.