Exploring High-Dimensional Spaces: Evaluation of a Node-Based Interface for Neural Audio Synthesis

Leonardo Auri
Audio Communication Group
Technische Universität Berlin
Berlin, Germany
auri@campus.tu-berlin.de

Tom Beckmann
Hasso Plattner Institute
University of Potsdam
Potsdam, Germany
tom.beckmann@hpi.uni-potsdam.de

Henrik von Coler
School of Music
Georgia Institute of Technology, USA
hvc@gatech.edu

Robert Hirschfeld
Hasso Plattner Institute
University of Potsdam
Potsdam, Germany
robert.hirschfeld@hpi.uni-potsdam.de

Abstract

Surveying and making use of high-dimensional representations is a relevant challenge in science, engineering, and, increasingly, artificial intelligence. We present a novel node-based interface for exploring high-dimensional spaces by traversing random axes, exploring points' neighbourhoods, and targeting underexplored areas. The interface supports users in mapping a meaningful manifold without requiring knowledge of or biasing the user toward a data distribution. We apply the interface to a 128-dimensional space controlling a real-time neural audio decoder. In a user study, participants created mappings for a musical instrument using our interface and a baseline (slider controls per dimension). Users' subspace exploration differed structurally with our interface, mapping dense, multi-dimensional vectors vs. vectors largely restricted to coordinate subspaces in the baseline; furthermore users created more mappings and indicated higher creativity support index scores in our interface. Our interface thus presents an effective method for mapping high-dimensional spaces without techniques requiring prior knowledge such as dimensionality reduction.

1 INTRODUCTION

Interfacing with high-dimensional spaces is a problem with increasing practical importance for many fields. Neural audio synthesizers based on variational autoencoders, for example, encode a training distribution as a continuous latent space from which perceptually varied sounds can be decoded in real time (Caillon and Esling, 2021). Working with such a system as a musical instrument requires the performer or composer to have knowledge of that space: to know of sonically interesting regions, to have intuitions of their topology, in order to ultimately map[1] points onto playable controls. We argue that, without appropriate tooling, exploration of such a space proceeds haphazardly and less effectively.

Miranda and Wanderley (2006) proposed a digital musical instrument (DMI) model that decouples the gestural control layer of a musical instrument from its sound-production layer. Under this model, sculpting a timbral palette and deciding how to actuate it are distinct creative activities: "the notion of a fixed causality between an action and a sound in a DMI does not necessarily apply. The same gesture (cause) can lead to the production of completely different sounds" (Miranda and Wanderley, 2006). Realising this separation in practice for a neural synthesizer means that the first activity — finding the sounds — must itself be supported by dedicated tooling before the second activity can begin.

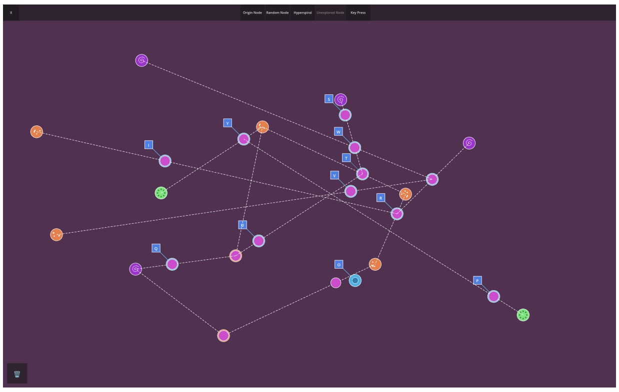

Figure 1: Final interface state of one participant using our proposed node-based interface for high-dimensional space exploration.

In this context we define "exploration" in line with the Cambridge Dictionary (2026): "the activity of searching in a place in order to find a particular thing, usually something valuable". In the example above, the "place" being explored would be the space of high-dimensional representations, while points of aesthetic interest or creative utility make up the valuable "thing" being searched for. Tooling which is better for the purpose of exploring such a high-dimensional space will be recognisable by the number

[1] For this paper we explicitly borrow Hunt and Wanderley's definition of *mapping* – "the designed link between an instrument's playing interface and its sound source" (Hunt and Wanderley, 2002). This should be understood in contrast to a cartographic definition, in which a map, for example, could be expected to retain distance relations up to scaling and projection.

of points deemed to be of creative value by a practitioner surfaced by the tool, ultimately leading to a “better map” – one which makes more valuable mappings available for creative work.

The standard approach to making the exploration of a high-dimensional space tractable is dimensionality reduction: PCA, t-SNE, or UMAP project a high-dimensional latent space onto a lower-dimensional manifold that can be mapped directly to a controller surface (Caillon and Esling, 2021; Bacon et al., 2025; Scurto and Postel, 2023). This expedites exploration, but at the cost of the variability available to the user: dimensionality reduction inevitably discards structure, and the regions reachable through the reduced representation are a strict subset of those present in the original space. PCA in particular has been shown to distort complex manifolds in ways that produce systematically misleading conclusions (Elhaik, 2022). For a creative endeavour, a reduced palette risks constraining the very aesthetic possibilities the practitioner set out to discover. Furthermore, dimensionality reduction requires prior knowledge of encodings of a representative dataset in the high-dimensional space, which may not be given.

An additional consideration comes from the topology of learnt latent representations themselves. Information in a trained neural network is not aligned to coordinate axes; the “privileged basis” assumption — that individual dimensions carry interpretable, disentangled meaning — fails in practice, with most neurons exhibiting polysemanticity across multiple unrelated network states (Marshall and Kirchner, 2024; Elhage et al., 2023). Any system of exploration which tends to induce axis-alignment will therefore systematically under-sample regions where much meaning actually resides. We contend that an effective exploration strategy rests on sampling along many arbitrary trajectories through the space, not only its coordinate axes or the principal axes of some prior distribution.

To address these two challenges, we propose an interface (depicted in fig. 1) that supports practitioners in exploring a latent space. Rather than reducing the dimensionality of the target space, the interface offers a set of complementary tools for constructing and traversing arbitrary axes: random anchor points, coverage-maximising sampling of underexplored regions, linear interpolation and extrapolation between any two saved points, and dense quasi-random trajectories (hyperspirals) around a chosen centre. Together these tools aid the user in surveying the high-dimensional space without requiring prior knowledge of the data distribution and without sacrificing the structure available in the full-dimensional space.[2]

In order to assess its capabilities in real use, we apply this interface to a 128-dimensional neural audio decoder and evaluate it against a per-dimension slider baseline in a controlled within-subjects experiment. Our analysis focuses on the following questions:

RQ1 Does the node-based interface lead users to explore a structurally richer subspace of the latent space compared with a slider-based baseline?

RQ2 Does the node-based interface provide higher creativity support for the task of mapping a playable instrument from a neural audio synthesizer?

RQ3 Does the node-based interface lead participants to commit to more latent space mappings?

Section 1.1 gives an overview of other work related to utilisation of high-dimensional representations.

Section 2 outlines our methodological approach, including detailed descriptions of our proposed interface (section 2.2) and our baseline comparison (section 2.3).

Section 3 summarises our experimental observations across creativity support (section 3.1) and behavioural (section 3.2) metrics as well as a structural analysis of the explored subspaces (section 3.3).

A synthesis of our results (section 4) offers insights into the mechanisms underpinning the significant differences in outcomes observed through the use of our proposed interface.

1.1 Related Work

The proliferation of sonic possibilities in electronic music has been noted as a defining challenge since the advent of electro-acoustic technologies (Emmerson, 1986, p. 18; Demers, 2010, p. 21-22), a trend that accelerated sharply with computational tools (Chadabe, 1997, ch. 5; Wanderley, 2023). As synthesizers and audio software exposed ever-larger parameter spaces to practitioners, a dedicated research tradition coalesced around the problem of *mapping*: constructing tractable, performable control relationships between gestural input and sonic output (see e.g. Hunt and Kirk, 2000; Miranda and Wanderley, 2006; Wanderley, 2023). The high-dimensional continuous latent spaces produced by variational autoencoders constitute a modern instance of the same challenge.

The predominant strategy for making such spaces navigable is dimensionality reduction. PCA projection is built directly into the RAVE architecture (Caillon and Esling, 2021) and is adopted in numerous musical applications of RAVE-based synthesis (Visi, 2024; Shepardson and Magnusson, 2023; Reus, 2024). Alternative non-linear mappings via variational autoencoders (Limberg and Zhang, 2024) or other learned methods (Fried et al., 2014) project the high-dimensional space onto a 2D control surface, supporting both navigation and visual inspection of a sound dataset. Liu et al. (2019) introduce *latent space cartography*, a framework that makes the topology of a learned space inspectable; the approach itself, however, operates on an already dimensionality-reduced representation. Embodied exploration approaches invite practitioners to move through a reduced latent space using physical controllers or body motion, via an embodiment framing of compositional tasks (Bacon et al., 2025) or the soundwalk paradigm of Scurto and Postel (2023), in which participants navigate a physical space mapped onto a latent space. All of these strategies trade off coverage of the full-dimensional space in exchange for tractability, and many presuppose some prior knowledge of that space’s topology.

A complementary body of work seeks to either imbue the structure of learnt spaces with predefined characteristics, or to find mappings to already known spaces. Paek et al. (2025) demonstrate that perceptually meaningful acoustic features correspond to near-linear axes in learned audio representations, enabling direct manipulation of timbral properties along interpretable directions. Esling et al. (2018) enforce such structure at training time, using perceptual ratings from timbre studies as a regularisation objective.

[2] A live demo of the proposed interface, as well as the baseline interface, is available at https://appendix.leoauri.com/ehds/demo/

Analogous approaches in the visual domain regularise via geometric priors (Zeng et al., 2025) or exploit functional maps between disparate latent spaces (Fumero et al., 2024). Text–audio contrastive pre-training provides an alternative access channel by aligning audio latent spaces with a jointly trained language representation (Huang et al., 2022; Wu et al., 2023). Other strategies sidestep high-dimensionality by disentangling interpretable factors at the model level (Limberg et al., 2025; Nitzan et al., 2020), or by recasting generation as a language-modelling problem over discrete codebooks (Borsos et al., 2023; Kumar et al., 2023).

Visual interface design offers a final set of precedents. Research in live coding (TOPLAP, 2005; Rein et al., 2018) investigates ad-hoc, exploratory creation of music and visuals for live audiences and correspondingly values real-time discovery. While most live coding environments are text-based, visual programming extensions for Snap*!* (Romagosa et al., 2025; Ball et al., 2019) offer node-based interfaces with a comparable visual vocabulary; their focus, however, remains on programming logic rather than direct manipulation of a continuous parameter space.

2 METHOD

We developed a node-based interface to support high-dimensional space exploration as well as a baseline slider-based interface for comparison. Complete code of the interface, backend, and analysis implementation is available from our repository.[3]

2.1 Concept and Design Process

We designed the proposed interface through multiple iterations with the authors themselves attempting to explore various 128-dimensional latent spaces. Through the design iterations, we identified a workflow where users sample random points in the latent space and then explore their neighbourhood, for example by interpolating between two interesting points, or by sampling the vicinity of a known point. We further refined the user interactions of the proposed interface, as well as those of the baseline comparison interface, through pilot testing with users that were involved neither in the design process nor in the subsequent user study. Thereby we brought both interfaces to a state where usability issues did not impede smooth use, such that the interfaces could be fairly compared on the basis of their differing conceptual underpinning.

Conceptually, the node-based interface exposes a high-dimensional space through a visual canvas of interconnected nodes (fig. 1). A node on the canvas represents a single N-dimensional vector that prescribes a point in the target space. To give users feedback, the selected node's corresponding vector is continuously streamed to produce audio output as further described below in subsection 2.4.

Importantly, the canvas is only a workspace for the exploration itself, not a rendering of the high-dimensional space the nodes represent. Users place nodes freely, arranging the points they have found or generated and laying out the trajectories along which they are exploring, without affecting the vectors the nodes represent. The interface serves as a "map" in the sense laid out in section 1: a designed link between interface and sound source, and, importantly, not in a "cartographic" sense: the interface does not impose a distance or direction correspondence between canvas and target space. Any approach that did preserve such correspondence would amount to a linear projection of the space onto two dimensions, reintroducing exactly the loss of reachable structure this work sets out to avoid.

Our aim is therefore not to render a high-dimensional space spatially legible, but to give practitioners the means to explore it in the service of creative work. What the free layout does afford is an intuition for the structure of the user's own exploration: which points were reached from which, which axes have been traversed and how far, and the opportunity to arrange points to reflect the user's own emerging spatial understanding of the manifold they have exposed (spatial representations are shown to underlie even non-spatial cognition, see Bellmund et al., 2018; Behrens et al., 2018, including for sound-frequency tasks, see Aronov et al., 2017).

In the following, we describe our prototype in more detail.

2.2 Node-based interface (proposed)

Four node types each represent a fixed point in the target N-dimensional space; we refer to these collectively as *anchor nodes*. When a user clicks on (selects) an anchor node, the interface begins to output the vector (point in N-dimensional space) represented by that node. Users can also assign a key to a selected node to quickly switch the output vector. The four anchor nodes are:

Origin nodes represent the zero vector $\mathbf{0}^N$. This allows users to access the origin of the high-dimensional space, which may carry special significance (due to training regularization, it happens to encode silence in some learnt latent spaces), as well as interpolating towards it, or sampling its vicinity.

Random nodes allow a user to sample randomly from the high-dimensional space. Random nodes hold a point sampled uniformly from $\{-1, 1\}^N$ at creation time; pressing the space bar while a random node is selected resamples it, substituting the currently held vector for the newly sampled one.

Unexplored nodes provide a mechanism to identify and sample from regions of the high-dimensional space which are most sparsely represented by existing nodes in the workspace. Unexplored nodes apply a coverage-maximising algorithm (described below) to propose a point in a region not yet represented by any existing node on the canvas.

Pin nodes allow the user to save points from explorations between or around nodes. Below we detail the mechanisms by which users can interpolate between, extrapolate from, and explore around existing nodes. Any point along these explorative trajectories can be "pinned" by clicking a pin icon, which saves that point to a new pin node. The pin node can then be used as an anchor for further exploration.

Interpolation and extrapolation. The anchor nodes outlined above provide fixed points; the *Interpolation* node described here turns any two of them into a direction of travel. Two points in the space jointly define an axis: the line running through both and onwards in either direction.

[3] https://github.com/leoauri/fulldim-interface/tree/publication

Such an axis is not aligned with the space's coordinate axes in general, and constructing it demands no prior knowledge of the space's structure — it is determined solely by points the user has retained. The stretch of the axis lying between the two points passes through their intermediate mixtures, while continuing past an endpoint carries on in the same direction, potentially pushing further into the features distinguishing one point from another. This is the interface's means of constructing arbitrary axes as motivated by section 1: it hands the user a one-dimensional, continuously playable handle on the full-dimensional space, oriented by material the user opts to explore rather than by the model's basis.

Concretely, dragging a connection between any two anchor nodes creates an Interpolation node together with a track joining the two parents. The Interpolation node is a type of *connection node*: unlike an anchor node it holds no fixed point of its own, but rides on the track between its parents and, when selected, outputs whichever point of the trajectory its current position designates. Sliding the Interpolation node along the track outputs the linearly interpolated vector between the two endpoints; moving the node beyond either endpoint extrapolates along the same axis. Hitting the pin icon on a connection node captures its current output into a new Pin node, which can in turn be selected, interpolated, or assigned to a key.

Unexplored sampling algorithm. This algorithm selects a new Unexplored node's vector, providing a means to sample from relatively unexplored regions of the N-dimensional space. Given the set $\mathcal{V}$ of all vectors currently saved on the canvas, the algorithm estimates the empirical mean $\boldsymbol{\mu}$ and isotropic variance σ^2 of $\mathcal{V}$, then draws 500 candidates from a uniform distribution scaled to the same spread. Each candidate $\mathbf{c}$ is scored as

$$s(\mathbf{c}) = \exp\left(-\frac{\beta}{2}\left(\frac{\|\mathbf{c}-\boldsymbol{\mu}\|}{\sigma\sqrt{N}}\right)^2\right) \cdot \min_{\mathbf{v}\in\mathcal{V}} \|\mathbf{c}-\mathbf{v}\|. \quad (1)$$

The score rewards large nearest-neighbour distance from existing points while penalising candidates far from the empirical centroid, preventing spurious tail samples. $\beta = 2/N$ scales inversely with dimensionality, with the numerator chosen empirically to balance the tail penalty with the nearest-neighbour reward. The candidate with the highest score is adopted as the node's vector.

Hyperspiral traversal. An interpolation axis is restricted to the single axis between two points. Frequently, though, the user holds a single point of interest and wants to hear what surrounds it, in order to ascertain variations that are possible via small deviations in the latent space. A hyperspiral winds outward from a chosen centre, turning continuously as it travels, such that a single control sweeps through many directions while its distance from the centre grows steadily (see fig. 2). Positions early in the trajectory are slight perturbations of the centre, those later in it depart ever more sharply, and each arrives from a direction the previous ones did not take. The trajectory is fixed by coefficients drawn at random, so a second hyperspiral around the same centre sweeps an independent family of directions, letting the user survey a point's surroundings repeatedly without retracing the same path.

Concretely, a *Hyperspiral* node encodes the set of coefficients required to trace a hyperspiral trajectory (see the hyperspiral algorithm below) around a centre point. A Hyperspiral node does not anchor a specific point in the N-dimensional space. Linking a Hyperspiral node to an anchor node generates a dedicated track carrying a connection node of its own. This node looks and behaves exactly as an Interpolation node does — it is slid along its track, pinned, and assigned to keys in the same way — but the positions it outputs follow the hyperspiral around the anchor rather than a straight-line path.

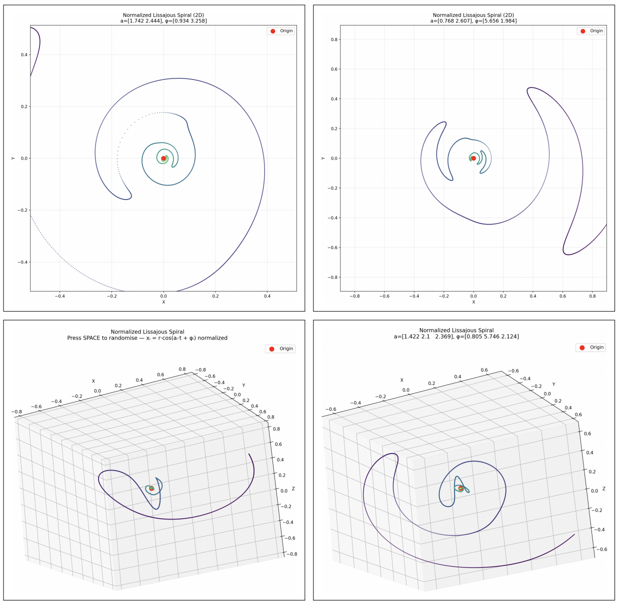

Figure 2: Example hyperspirals shown to participants in 2 dimensions (top row) and 3 dimensions (bottom row).

At initialisation a Hyperspiral node draws N coefficients $a_1, \ldots, a_N$ uniformly from $[0, 25]$. Because they are drawn continuously, they are rationally independent with probability one, ensuring the trajectory is dense on the hypersphere and never self-repeats. Given a centre vector $\mathbf{v}_0$, the output is $\mathbf{v}(t) = \mathbf{v}_0 + \mathbf{r}(t)$, where the trajectory parameter t is given by the position of the connection node along its track: $t = 0$ at the anchor node holding the centre and $t = 1$ at the Hyperspiral node itself, with positions beyond the latter continuing the trajectory outward. The offset $\mathbf{r}(t)$ is computed as follows. An unnormalised Lissajous point is formed with $x_i(t) = \cos(a_i t)$ for each dimension, then projected onto a hypersphere of radius $r = t\sqrt{2N/3}$:

$$\mathbf{r}(t) = \frac{r}{\|\mathbf{x}(t)\|}\,\mathbf{x}(t). \quad (2)$$

The radius scaling ensures that at $t = 1$ the offset equals $\sqrt{2N/3}$, the expected Euclidean distance between two points drawn uniformly from $[-1, 1]^N$, providing a dimensionality-independent sense of scale.

Keyboard mapping. As mentioned before, nodes can be bound and selected via key presses for rapid output vector switching. A *Keypress* node binds a keyboard key to an existing anchor node, activating that node when the assigned key is pressed. As a shortcut, pressing an unassigned key while any unmapped anchor node is selected automatically creates the corresponding Keypress node. When invoked on a connection node, this shortcut additionally creates a Pin node capturing the node's current output, so that a single keypress recalls a specific position along the interpolation axis or hyperspiral.

2.3 Slider-based interface (baseline)

To analyse the effect of our proposed interface, we designed a baseline interface that exposes the latent space directly, without any additional tools for exploration such as sampling a point's neighbourhood. Notably, all other exploration tools for high-dimensional spaces known to us from the literature rely on dimensionality reduction techniques to make an exploration manageable, which this paper specifically aims to avoid. As such, the baseline interface reflects the most common approach to high-dimensional parameter control for audio equipment: a dedicated continuous control per dimension. Unlike parameter control for audio equipment, we are specifically dealing with an entirely unknown high-dimensional space. As such, we deliberately avoided any analysis-driven arrangement of the controls or labelling. Instead, sliders are placed at random positions at initialisation, while the endpoints of each slider can be repositioned freely by the user (see fig. 3). This allows participants to, for one, organise

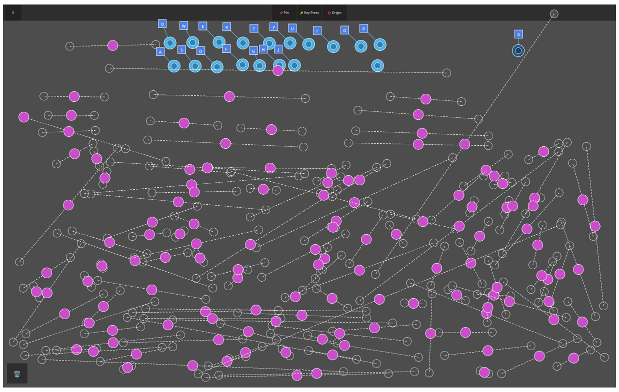

Figure 3: One participant's completed slider interface preset.

the layout according to their own emerging understanding of the space, as well as deciding between precision (more distant slider endpoints) and increased range (closer slider endpoints) per dimension as required.

Each slider maps linearly to a single dimension of the target N-dimensional vector. Slider endpoints represent values from -1 to 1. Sliders can be moved beyond the endpoints, so that the output values are not constrained to the range defined by the endpoints.

To support recall of discovered sounds, the interface shares a subset of the node-based interface's vocabulary: *Pin*, *Origin*, and *Keypress* nodes are available via a top drawer. A Pin node captures the current output vector and can be dragged onto the canvas; an Origin node holds the zero vector $\mathbf{0}^N$. Keypress assignment follows the same shortcut as the node-based interface: pressing an unassigned key while an unmapped Pin or Origin node is selected binds that key to the node's saved vector.

2.4 Neural audio backend

Both interfaces stream their N-dimensional output vectors in real time via the Open Sound Control (OSC) protocol (Wright and Freed, 1997). An audio backend running in Max[4] receives these vectors and routes them to a trained RAVE (Caillon and Esling, 2021) neural audio decoder running via the *anira~* external[5] for real-time audio inference. Continuous, low-latency audio feedback is essential to the exploratory nature of the task: immediate response of the audio output as movement occurs supports participants' navigation of the latent space.

To give participants access to the full 128-dimensional latent space, the RAVE models were exported without the PCA reduction layer described in Caillon and Esling (2021). Only the decoder components are used; the encoder is discarded for the experiment.

Participants may select between two models trained on qualitatively distinct datasets. The first was trained on GTSinger (Zhang et al., 2024), a large multilingual dataset of singing voices spanning a range of vocal techniques and styles. The second was trained on a synthetic dataset of summed sine tones with randomised frequencies and amplitudes[6], producing a timbral space of inharmonic, spectrally varied sounds. Offering a choice of model was motivated by pilot observations in which participant responses to a single model varied considerably; providing two contrasting timbral worlds reduces the risk that a dispreferred sound aesthetic distracts from the interface evaluation.

2.5 Participants

We recruited participants from a list of individuals maintained by the principal investigator who have previously expressed interest in study participation.

$N = 16$ individuals enrolled in our study. Participants reported their genders as follows: 9 female (56.2%), 6 male (37.5%), 1 chose not to say (6.2%). Of the 15 who provided demographic information, participants had a mean age of 34.93 ($SD = 4.82$).

We asked participants to provide scores for the Goldsmiths Musical Sophistication Index (Gold-MSI), a survey instrument designed to assess "musical skills and behaviours on multiple dimensions in the general population" (Müllensiefen et al., 2014). We used the short-scale version of the survey available from GMSI Team (2013). Figure 4 summarises the Gold-MSI scores across our participants. Compared to the reference sample of the general population, our participants are up to 1 point higher on the Gold-MSI subscales on average (7-point scales), while still showing spread on all subscales other than "Emotions".

2.6 Task

Our primary interest is in revealing how effectively each interface supports exploration of a high-dimensional space — specifically, whether participants can surface a structurally diverse set of points that they judge to be sonically interesting or useful to them. To obtain a concrete, measurable endpoint, we asked participants to construct a playable

[4] https://cycling74.com/products/max

[5] *anira~* (https://github.com/anira-project/anira_tilde) wraps the anira library (Ackva and Schulz, 2024), a real-time safe neural network inference library, for processing in Max.

[6] The summed sine dataset is available from https://appendix.leoauri.com/ehds/#dataset or https://doi.org/10.5281/zenodo.22012306 and code to generate synthetic datasets with similar characteristics from https://github.com/leoauri/tone-datagen/tree/publish-EHDS.

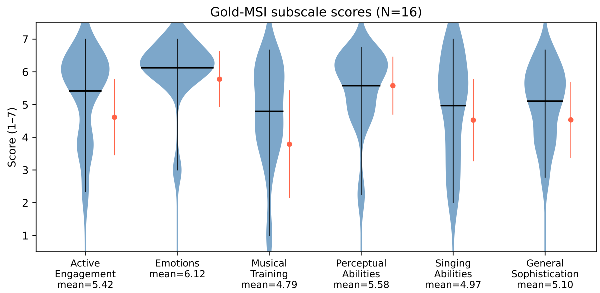

Figure 4: Averaged scores over the Gold-MSI subset applied. Each scale shows mean and range of participants. In red the reported population mean and standard deviations from Müllensiefen et al. (2014).

instrument by assigning discovered sounds to keys of the computer keyboard. This framing is advantageous because it requires participants to commit to specific points in the latent space, giving us a well-defined, quantifiable artefact of exploration. Furthermore it can be specifically motivated by our asking participants to perform a live demonstration of their finished instrument mapping using only the keyboard interface in a second step.

During the study overview participants were given the following task description alongside background material introducing the concept of a high-dimensional space and the specific concepts relevant to the task (see repository). The task description was also shown during each interface condition for reference.

Task description:

1. Explore the diversity of possible sounds available in the sound space.
2. Create an instrument mapping for the keyboard by assigning sounds you discover to keyboard keys. Aim to create an instrument that demonstrates the breadth of sounds available in the sound space.
3. Record a demonstration — once your mapping is complete, we will create a short recording in which you demonstrate the range of sounds you have discovered using the mapped keyboard keys.

Referring again to our definitions of “exploration” and “mapping” from the introduction, our task is directly aimed at eliciting a search of a high-dimensional space for creatively valuable material. The points surfaced should be used to create an instrument mapping, which both motivates the search as well as giving us a tangible outcome we can analyse. The participant’s own valuation of the material is intrinsic to the design goal of the tool. Our choice of a broad task which leaves room for participants’ own aesthetic opinions is therefore deliberate. A more narrowly defined task might allow for a more concrete analysis of outcomes, but would no longer compare the interfaces in terms of our intent. Our interest is in developing a more useful interface for surfacing creatively valuable material in an unknown high-dimensional space.

2.7 Experimental procedure

Each session proceeded through the following steps, with interface condition order counterbalanced across participants:

1. Study overview.
2. Tutorial for the first interface condition.
3. Model selection: participants chose between the two audio-producing models. In order to make participants aware of the models’ respective sonic characteristics, random samples were generated by sampling from $[-1, 1]^N$ at regular intervals.
4. First interface condition.
5. Demonstration recording for first condition.
6. Creativity Support Index (CSI) agreement statements (see section 3.1) for first condition.
7. Optional break.
8. Tutorial for the second interface condition.
9. Second interface condition.
10. Demonstration recording for second condition.
11. CSI agreement statements for second condition.
12. CSI paired-factor comparisons.
13. Gold-MSI background questionnaire.

The complete experimental flow was implemented as a single automated sequence (see footnote 3).

Tutorial. Each interface condition was preceded by a tutorial during which each element of the respective interface is introduced and explained. The participant was asked to demonstrate use of the interface elements. In order not to precede familiarity with the space of sounds generated by the neural audio synthesizer, during the tutorial the interface drives an 8 parameter basic synthesizer patch (oscillator + biquad filter) (see footnote 3).

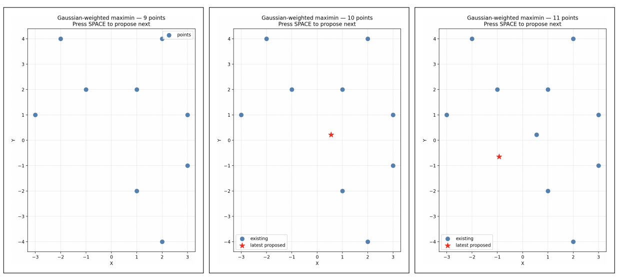

Figure 5: Unexplored node algorithm visualisations shown to participants.

In order to scaffold concepts important to the node-based interface, visualisations depicting hyperspiral trajectories (see fig. 2) and the operation of the unexplored sampling algorithm (see fig. 5 and section 2.2) are shown during the respective tutorial.

Demonstration recording. No time limit was imposed on participants’ task completion. Once participants indicated they were finished with the task, the experimenter prompted them to make sure that all desired sounds had been mapped to keys, as they were only to use the keyboard interface for the demonstration recording.

3 RESULTS

We compared our proposed interface with the baseline interface in a controlled within-subject experiment. We collected participant usage data including final interface presets, timing data, and demonstration data, as well as questionnaire responses. The raw data is available from our supplementary materials[7].

Here we outline the collected metrics and our analyses. Analysing both users' objective interface use as well as their reported subjective experience allows us to finally synthesise stronger conclusions with regards to the effectiveness of the proposed interface for making a high-dimensional space tractable.

3.1 Creativity Support Index

The Creativity Support Index (CSI) is "a psychometric survey designed for evaluating the ability of a creativity support tool to assist a user engaged in creative work" (Cherry and Latulipe, 2014). The instrument comprises two parts: a set of twelve agreement statements and a paired-factor comparison section. The agreement statements produce subscores along 6 dimensions of creativity support for a given creativity support tool. The paired-factor comparisons are administered per-task and produce weightings for the 6 factors, allowing the calculation of a participant's total CSI score out of 100. This weighting makes the index "more sensitive to the factors that are the most applicable in a given situation" (Cherry and Latulipe, 2014). Since we compare a single task over two interfaces, the agreement statements were administered twice, directly after each condition, and the paired-factor comparisons once in conclusion.

Figure 6 shows participant ratings on CSI subscales per condition. Figure 7 summarises participants' individual paired-factor comparison counts. Figure 8a shows CSI scores computed from participants' subscale scores and paired-factor comparisons.

Table 1, fig. 9, and fig. 8b show the per condition effect on CSI scores: significant effects for Expressiveness, Enjoyment, Exploration, and Results Worth Effort subscales, as well as CSI total.

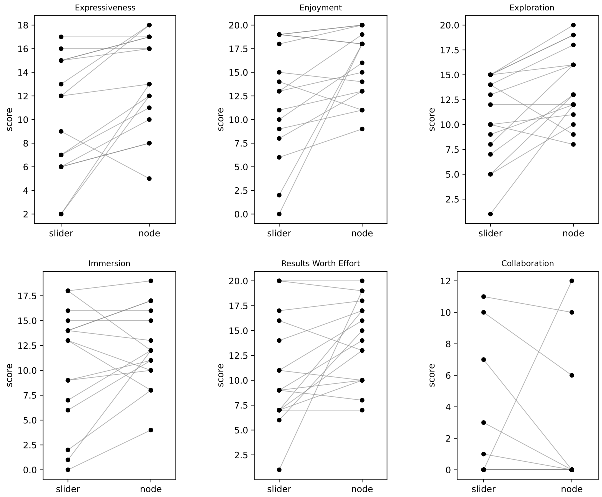

Figure 6: Paired CSI subscale scores by condition (node vs. slider) for all six subscales. Each line represents one participant.

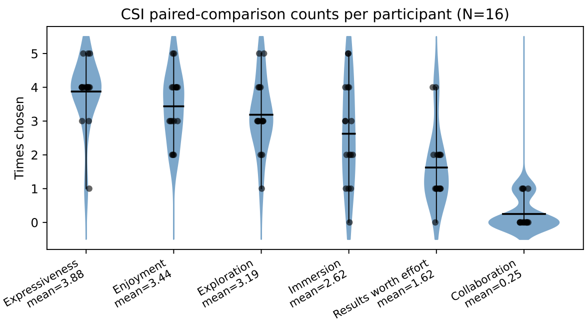

Figure 7: Distribution of CSI paired-factor comparison counts per participant. Each point is the number of times a participant chose a given factor over the others; the violin shows the density of those counts and the horizontal line marks the mean.

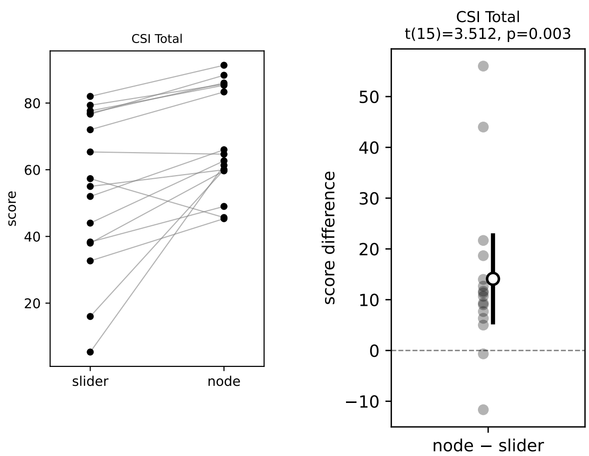

Figure 8: Computed total CSI scores: a) *(left)* per condition; b) *(right)* difference between conditions per participant, mean, and 95% CI.

3.2 Behavioural outcomes

We recorded the duration participants spent on task as well as the number of nodes with keyboard mappings assigned. Table 2 and fig. 10 show differences in behaviour between conditions: a significant effect on number of keyboard mappings was observed.

3.3 Structural analysis of explored subspace

We collate the complete set of vectors with keyboard mappings assigned across all participants for each condition and analyse the geometric structure of the explored subspace. To characterise how far participant explorations departed from the coordinate axes, we apply two complementary sparsity measures to each mapped vector.

[7] All raw data collected from our experiments, including completed presets, preset screenshots, timing data, experiment metadata and questionnaire responses is available from Zenodo at 10.5281/zenodo.22012307.

Table 1: Condition effects (node vs. slider) on CSI scales. * indicates $p < 0.05$.

Scale	Mean (slider)	Mean (node)	Mean diff	$t(15)$	p	d
Expressiveness	10.000	13.188	3.188	3.404	0.004*	0.851
Enjoyment	12.188	15.812	3.625	2.497	0.025*	0.624
Exploration	10.500	14.000	3.500	3.530	0.003*	0.882
Immersion	10.562	12.188	1.625	1.516	0.150	0.379
Results Worth Effort	10.688	14.125	3.438	2.556	0.022*	0.639
Collaboration	2.000	1.750	-0.250	-0.262	0.797	-0.065
CSI Total	54.292	68.396	14.104	3.512	0.003*	0.878

Table 2: Condition effects (node vs. slider) on behavioural outcomes. * indicates $p < 0.05$.

Metric	Mean (slider)	Mean (node)	Mean diff	$t(15)$	p	d
Duration	808.413	900.436	92.024	0.641	0.531	0.160
Key Mappings	9.500	13.375	3.875	3.139	0.007*	0.785

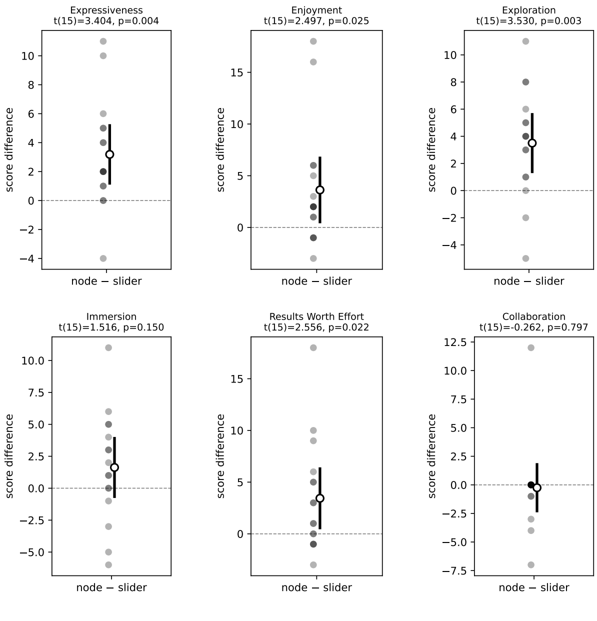

Figure 9: CSI subscale score differences (node minus slider) for all six subscales, mean, and 95% CI.

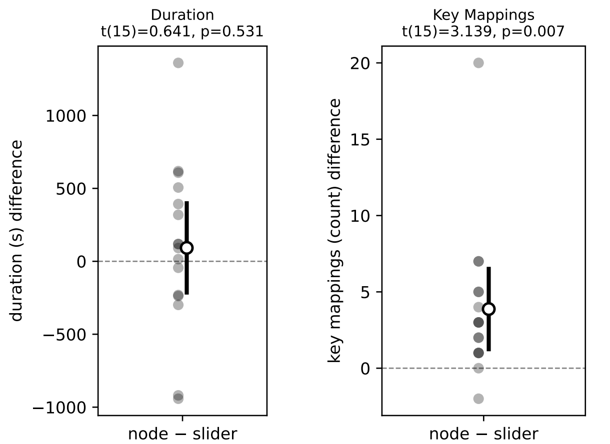

Figure 10: Behavioural differences between conditions per participant, mean, and 95% CI: a) *(left)* time spent on task; b) *(right)* total keyboard mappings at task completion.

Hoyer sparsity (HS), introduced in Hoyer (2004), is derived from the ℓ_1/ℓ_2 ratio:

$$\mathrm{HS}(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1/\|\mathbf{x}\|_2}{\sqrt{n} - 1} \tag{3}$$

HS equals 1 for a one-hot (coordinate-axis) vector and 0 for a vector with equal energy in all components.

Participation ratio (PR) (Bell and Dean, 1970; Gao et al., 2017) quantifies the effective number of active dimensions:

$$\mathrm{PR}(\mathbf{x}) = \frac{\|\mathbf{x}\|_2^4}{\|\mathbf{x}\|_4^4} \tag{4}$$

PR equals 1 for a one-hot vector and n for a vector with perfectly uniform energy. Because it weights dimensions by their squared contribution, it is less sensitive than HS to components with small magnitude. Figure 11 visualises both sparsity metrics on 2-dimensional values.

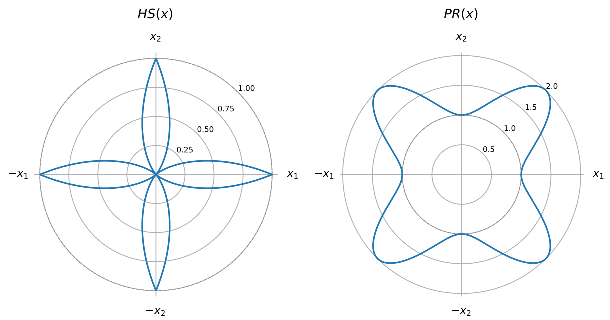

Figure 11: 2D polar plots comparing Hoyer sparsity (HS) and participation ratio (PR).

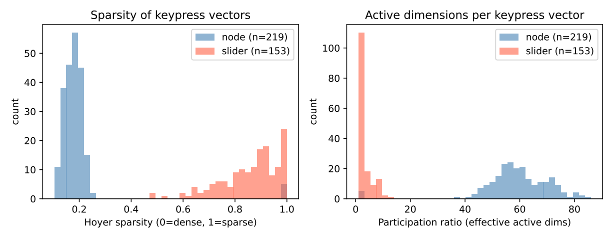

Figure 12: Sparsity and effective participation ratio of latent vectors produced by node and slider exploration.

Figure 12 shows the distribution of both measures by condition. Slider-mapped vectors had a median HS of 0.887

Table 3: Ordering effects (first vs. second condition). * indicates $p < 0.05$.

Scale/Metric	Mean (first)	Mean (second)	Mean diff	$t(15)$	p	d
Expressiveness	12.500	10.688	-1.812	-1.569	0.138	-0.392
Enjoyment	15.250	12.750	-2.500	-1.560	0.139	-0.390
Exploration	12.750	11.750	-1.000	-0.760	0.459	-0.190
Immersion	11.438	11.312	-0.125	-0.109	0.915	-0.027
Results Worth Effort	13.188	11.625	-1.562	-1.002	0.332	-0.250
Collaboration	2.375	1.375	-1.000	-1.085	0.295	-0.271
CSI Total	64.438	58.250	-6.188	-1.194	0.251	-0.299
Duration	1042.929	665.920	-377.009	-3.484	0.003*	-0.871
Key Mappings	12.812	10.062	-2.750	-1.934	0.072	-0.484

and a median PR of 2.18, indicating energy concentrated in very few dimensions. Node-mapped vectors had a median HS of 0.177 and a median PR of 58.91, indicating energy spread across a substantial fraction of the 128 dimensions.

4 DISCUSSION

Our experimental results lend strong evidence to our proposed interface offering an effective tool for exploration of high-dimensional spaces for creative purposes. Participants rated the interface significantly higher on the CSI, addressing RQ2. Participants mapped more points in the latent space using the proposed interface, adressing RQ3. While the behavioural measure alone might be explained by arbitrary differences in interface design, taken together with the significantly higher CSI ratings, it suggests that our interface better surfaced points participants deemed interesting or useful, thus offering greater creativity support.

Aside from time spent in condition, no strong ordering effects were observed (see table 3). No effect of model choice on CSI difference (see fig. 13) was observed. No significant correlation of Gold-MSI subscales with CSI per condition effect was observed (see fig. 14).

Our structural analysis reveals qualitative differences in how participants explored the latent space across the two conditions, directly addressing RQ1. The 2D PCA projection (fig. 15) makes this difference visually apparent: vectors mapped under the node condition form a compact, distributed cloud, while vectors from the slider condition cluster along narrow corridors or extend to distant extremes.

The distribution of vector norms (fig. 16) illuminates the predominant strategy in the slider condition: participants drive individual sliders to extreme values, producing vectors far from the origin, while the region close to the origin — where combinations of many dimensions might have produced interesting sounds — remain comparatively unexplored. This reflects a natural tendency when working with independent per-dimension controls: one slider is turned up to find a sound, then another, without the interface encouraging combinations across many dimensions simultaneously.

The sparsity measures (fig. 12) clearly quantify the underlying structural difference. Slider-mapped vectors evidence high sparsity and participation ratios indicating energy tending to be concentrated in fewer than three dimensions — vectors lying very close to (on) coordinate axes. Node-mapped vectors were comparatively less sparse and showed a participation ratio consistent with energy distributed across a substantial fraction of the 128-dimensional space.

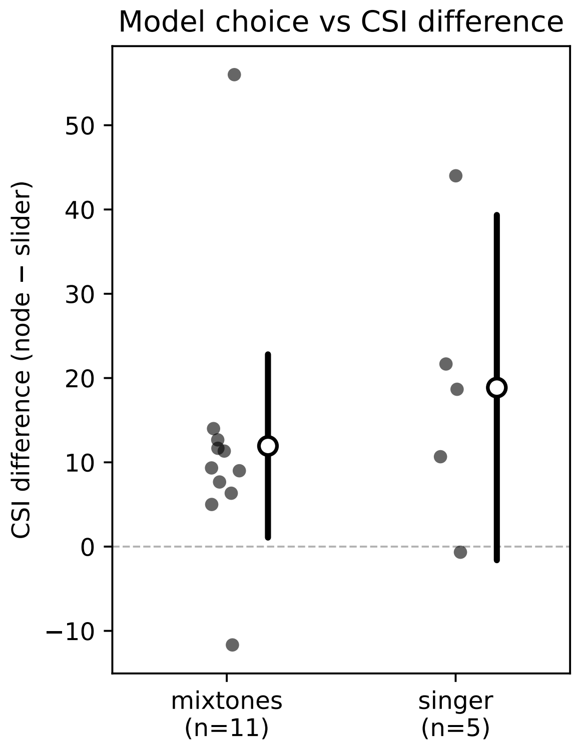

Figure 13: CSI difference by model choice subgroup. Confidence interval of each subgroup encompasses the mean of the other: no evidence for an effect on CSI difference by model choice.

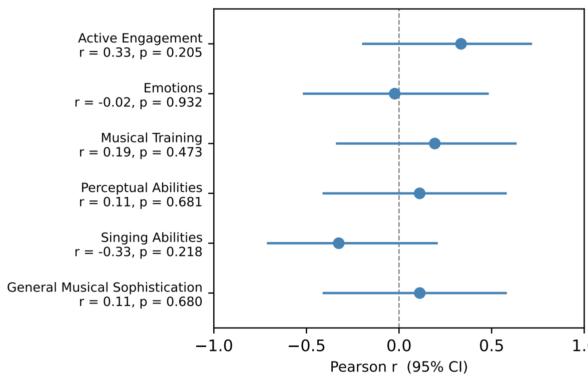

Figure 14: Correlations of Gold-MSI scales with CSI per condition effects, 95% CIs. No significant effects observed.

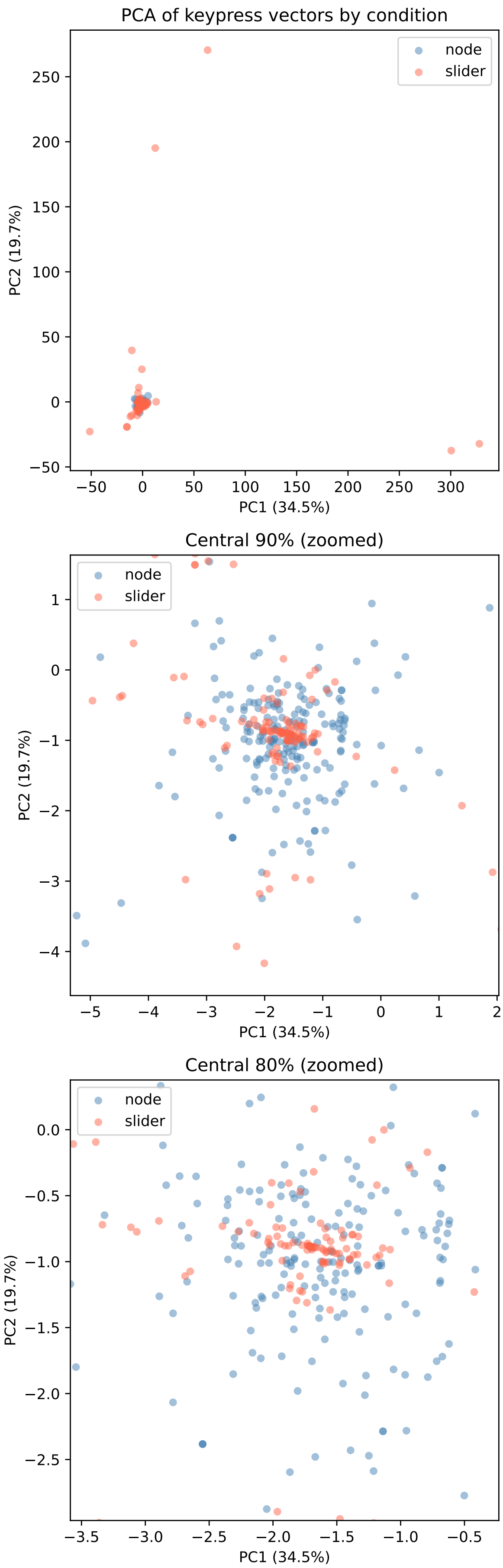

Figure 15: PCA of keypress vectors by condition.

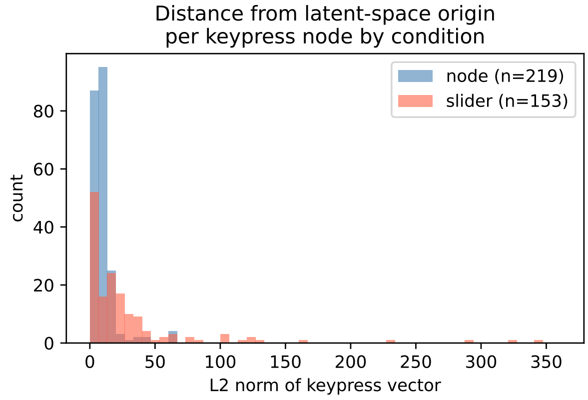

Figure 16: L2 norms of mapped vectors. Participants consistently map vectors far from the origin in the baseline interface.

Node-mapped vectors resemble typical draws from a multivariate distribution, while slider-mapped vectors are largely restricted to low-dimensional coordinate subspaces.

These results speak directly to the concern raised in section 1: the slider interface does not merely lead users to explore a different region of the latent space, but rather a fundamentally different *structure* of it. Because information in trained neural representations is not aligned to coordinate axes (Marshall and Kirchner, 2024; Elhage et al., 2023), axis-aligned exploration will systematically miss much of the perceptually meaningful variation available in the full space. The node interface, by encouraging interpolation along arbitrary directions between saved points, steers participants naturally toward the denser, more combinatorially rich regions of the latent space. Taken together with the significant CSI advantage observed in section 3.1, these findings suggest that the structural richness of explored subspace and perceived creativity support are mutually reinforcing: delving more deeply into the space led to explorative experiences which participants found more expressive, enjoyable, and worth the effort.

4.1 Limitations

We tested nine outcomes for statistical significance at $p < 0.05$, of which six were significant. Under the Benjamini–Hochberg framework (Benjamini and Hochberg, 1995), the estimated false discovery rate among these six results is $9 \times 0.05/6 \approx 7.5\%$, meaning fewer than one of the six reported effects is expected to be a false positive.

Our sample size of $n = 16$ limits the possibility to detect more subtle effects and analyse subgroups. Further, giving participants more time and situating the tool within a more realistic setting for their creative work might also produce different results. Given the short time, the novelty of the interface may have had an effect on participants' perception.

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed a user interface for exploring and mapping high-dimensional spaces. By providing users with a set of tools in a novel node-based interface, we enabled their exploration to proceed beyond the space's coordinate axes. In a user study, we compared our interface to a baseline interface that simply provides sliders controlling each of the 128 parameters of a real-time neural audio decoder. We found that users in our interface were more likely to deviate from the coordinate axes, mapped more points in the space they deemed interesting, and indicated a comparatively high creativity support index.

In the future, it would be interesting to repeat our experiment with different models—both other audio models as well as models for text or image generation. A systematic ablation of the provided tools should also be undertaken in order to uncover the effect of individual exploration strategies. Our interface already provided various ways for users to interactively "move" through the space using direct manipulation. Beyond this, it would be interesting to investigate further ways for users to interactively navigate high-dimensional spaces, such as through simulated forces and/or external musical instrument interfaces. This would allow us to extend our analyses beyond mere static points to encompass the possibilities of dynamic trajectories moving through high-dimensional space.

AUTHOR DECLARATIONS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The first author's doctoral research is supported by the Erhard Höpfner Stiftung. This study received no specific grant funding.

The study was approved by TU Berlin Kommission für Forschungsethik der Fakultät I. The study was conducted in accordance with the local legislation and institutional requirements. The participants provided their informed consent to participate in the study.

AUTHOR CONTRIBUTIONS

Contributor roles are given according to the CRediT taxonomy.[8] LA - Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. TB - Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft, Writing – review & editing. HvC - Supervision. RH - Funding acquisition, Resources, Supervision.

ACKNOWLEDGEMENTS

We thank the participants of our user study for their time. Model training was carried out on the HPI Data Center at the Hasso Plattner Institute, Potsdam, made available through a guest-researcher arrangement.

All research content, scientific ideas, analyses, and conclusions in the paper are human-generated. Draft formulations from initial notes were generated by Claude Opus 4.8. LaTeX code was edited by Claude Opus 5. Claude Code was used for development of all software components.

REFERENCES

Ackva, V. and Schulz, F. (2024). Anira: An architecture for neural network inference in real-time audio applications. In *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, pages 1–10. IEEE.

Aronov, D., Nevers, R., and Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal/entorhinal circuit. *Nature*, 543(7647):719–722.

Bacon, B. K., Auri, L., and McKee, H. A. (2025). Approaches to notation for embodied engagement with a novel neural network-based musical instrument. *Frontiers in Computer Science*, 7.

Ball, M., Mönig, J., Romagosa, B., and Harvey, B. (2019). Snap! A Look at 5 Years, 250, 000 Users and 2 Million Projects. In Hawthorne, E. K., Pérez-Quiñones, M. A., Heckman, S., and Zhang, J., editors, *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE 2019, Minneapolis, MN, USA, February 27 - March 02, 2019*, page 1279. ACM.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, 100(2):490–509.

Bell, R. J. and Dean, P. (1970). Atomic vibrations in vitreous silica. *Discussions of the Faraday Society*, 50:55.

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., and Doeller, C. F. A. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):1–13.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., Roblek, D., Teboul, O., Grangier, D., Tagliasacchi, M., and Zeghidour, N. (2023). AudioLM: a Language Modeling Approach to Audio Generation. arXiv:2209.03143.

Caillon, A. and Esling, P. (2021). RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. arXiv:2111.05011.

Cambridge Dictionary (2026). Exploration. https://dictionary.cambridge.org/dictionary/english/exploration. Accessed: 2026-08-08.

Chadabe, J. (1997). *Electric sound: the past and promise of electronic music*. Prentice Hall, Upper Saddle River, N.J. OCLC: 35138370.

Cherry, E. and Latulipe, C. (2014). Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction*, 21(4):1–25.

Demers, J. (2010). *Listening through the noise: the aesthetics of experimental electronic music*. Oxford university press, Oxford New York.

Elhage, N., Lasenby, R., and Olah, C. (2023). Privileged bases in the transformer residual stream. *Transformer Circuits Thread*, 24.

Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, 12(1):14683.

[8] https://credit.niso.org/

Emmerson, S., editor (1986). *The Language of Electroacoustic Music*. Palgrave Macmillan UK, London.

Esling, P., Chemla-Romeu-Santos, A., and Bitton, A. (2018). Generative timbre spaces: regularizing variational auto-encoders with perceptual metrics. arXiv:1805.08501.

Fried, O., Jin, Z., Finkelstein, A., and Oda, R. (2014). AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 281–286. International Conference on New Interfaces for Musical Expression.

Fumero, M., Pegoraro, M., Maiorca, V., Locatello, F., and Rodolà, E. (2024). Latent Functional Maps: a spectral framework for representation alignment. *Advances in Neural Information Processing Systems*, 37:66178–66203.

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., and Ganguli, S. (2017). A theory of multineuronal dimensionality, dynamics and measurement.

GMSI Team (2013). The Goldsmiths musical sophistication index (Gold-MSI). `https://shiny.gold-msi.org/gmsi_toplevel/`. Accessed: 2026-05-02.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. arXiv:cs/0408058.

Huang, Q., Jansen, A., Lee, J., Ganti, R., Li, J. Y., and Ellis, D. P. W. (2022). MuLan: A Joint Embedding of Music Audio and Natural Language. arXiv:2208.12415.

Hunt, A. and Kirk, R. (2000). Mapping strategies for musical performance. *Trends in gestural control of music*, 21(2000):231–258.

Hunt, A. and Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound*, 7(2):97–108.

Kumar, R., Seetharaman, P., Luebs, A., Kumar, I., and Kumar, K. (2023). High-Fidelity Audio Compression with Improved RVQGAN. arXiv:2306.06546.

Limberg, C., Schulz, F., Zhang, Z., and Weinzierl, S. (2025). Pitch-Conditioned Instrument Sound Synthesis From an Interactive Timbre Latent Space. arXiv:2510.04339.

Limberg, C. and Zhang, Z. (2024). Mapping the Audio Landscape for Innovative Music Sample Generation. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, ICMR '24, pages 1207–1213, New York, NY, USA. Association for Computing Machinery.

Liu, Y., Jun, E., Li, Q., and Heer, J. (2019). Latent Space Cartography: Visual Analysis of Vector Space Embeddings. *Computer Graphics Forum*, 38(3):67–78.

Marshall, S. C. and Kirchner, J. H. (2024). Understanding polysemanticity in neural networks through coding theory. arXiv:2401.17975.

Miranda, E. R. and Wanderley, M. (2006). *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series)*. A-R Editions, Inc., USA.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE*, 9(2):e89642.

Nitzan, Y., Bermano, A., Li, Y., and Cohen-Or, D. (2020). Face Identity Disentanglement via Latent Space Mapping. arXiv:2005.07728.

Paek, N., Zang, Y., Yang, Q., and Leistikow, R. (2025). Learning Interpretable Features in Audio Latent Spaces via Sparse Autoencoders. arXiv:2510.23802.

Rein, P., Ramson, S., Lincke, J., Hirschfeld, R., and Pape, T. (2018). Exploratory and Live, Programming and Coding. *The Art, Science, and Engineering of Programming*, 3(1):1:1–1:33.

Reus, J. (2024). Tungnaá: a hyper-realistic voice synthesis instrument for real-time exploration of extended vocal expressions. In *NIME*.

Romagosa, B., Mor, E., and Maloney, J. (2025). Adding live coding music capabilities to the microblocks educational programming language.

Scurto, H. and Postel, L. (2023). Soundwalking Deep Latent Spaces. In *23rd International Conference on New Interfaces for Musical Expression, NIME 2023, Mexico City, Mexico, May 31 - June 3, 2023*. nime.org.

Shepardson, V. and Magnusson, T. (2023). The living looper: Rethinking the musical loop as a machine action-perception loop. In *NIME*.

TOPLAP (2005). Manifestodraft. `http://toplap.org/wiki/ManifestoDraft`.

Visi, F. (2024). The Sophtar: a networkable feedback string instrument with embedded machine learning. In *International conference on new interfaces for musical expression (NIME), utrecht, netherlands, september 4-6, 2024*, pages 142–148. International Conference on New Interfaces for Musical Expression.

Wanderley, M. (2023). Prehistoric NIME: Revisiting Research on New Musical Interfaces in the Computer Music Community before NIME. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 60–69.

Wright, M. and Freed, A. (1997). Open SoundControl: A New Protocol for Communicating with Sound Synthesizers. In *International Computer Music Conference (ICMC)*.

Wu, Y., Chen, K., Zhang, T., Hui, Y., Berg-Kirkpatrick, T., and Dubnov, S. (2023). Large-scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation. arXiv:2211.06687.

Zeng, W., Chang, X., Su, J., Gu, X., Sun, J., and Xu, Z. (2025). GMapLatent: Geometric Mapping in Latent Space. arXiv:2503.23407.

Zhang, Y., Pan, C., Guo, W., Li, R., Zhu, Z., Wang, J., Xu, W., Lu, J., Hong, Z., Wang, C., Zhang, L., He, J., Jiang, Z., Chen, Y., Yang, C., Zhou, J., Cheng, X., and Zhao, Z. (2024). GTSinger: A Global Multi-Technique Singing Corpus with Realistic Music Scores for All Singing Tasks. In *Advances in Neural Information Processing Systems 37*, pages 1117–1140. arXiv:2409.13832.