Organised Sound (2025), **30**, 355–366
doi:10.1017/S1355771825100915

CAMBRIDGE UNIVERSITY PRESS

Article

Expanding the machine: Notating generative synthesis with a state-based representation and a navigable timbre space

Vincenzo Madaghiele[1], Leonard Lund[2], Derek Holzer[3], Tejaswinee Kelkar[4], Kıvanç Tatar[5] and Andre Holzapfel[6]

[1]Department of Musicology, University of Oslo, Oslo, Norway; [2]KTH Royal Institute of Technology, Stockholm, Sweden; [3]Division of Media Technology and Interaction Design (MID), KTH Royal Institute of Technology, Stockholm, Sweden; [4]Department of Musicology, University of Oslo, Oslo, Norway; [5]Data Science and AI Division, Computer Science and Engineering Department, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden and [6]Division of Media Technology and Interaction Design (MID), KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

Notating electroacoustic music can be challenging due to the uniqueness of the instruments employed. Electronic instruments can include generative components that can manipulate sound at different time levels, in which parameter variations can correlate non-linearly to changes in the instrument's timbre. The way compositions for electronic instruments are notated depends on their interfaces and the parameter controls available to performers, which determine the state of their sound-generating system. In this article, we propose a notation system for generative synthesis based on a projection from its parameter space to a timbre space, allowing to organise synthesiser states based on their timbral characteristics. To investigate this approach, we introduce the Meta-Benjolin, a state-based notation system for chaotic sound synthesis employing a three-dimensional, navigable timbre space and a composition timeline. The Meta-Benjolin was developed as a control structure for the Benjolin, a chaotic synthesiser. Framing chaotic synthesis as a specific instance of generative synthesis, we discuss the advantages and drawbacks of the state- and timbre-based representation we designed based on the thematic analysis of an interview study with 19 musicians, who composed a piece using the Meta-Benjolin notational interface.

(Received 25 February 2025; revised 18 July 2025; accepted 4 September 2025; First Published online 29 January 2026)

1. Introduction

The development of generative music systems has been a major focus of electroacoustic musicians throughout the years. Electronic instruments are often composed of modular, interchangeable components, which can act at different time scales with some degree of automation. Electroacoustic compositions are usually associated with custom instrument systems, often played by composers themselves in performance. These systems increasingly employ components capable of autonomously generating materials and creating complex variations at different levels of musical structure. These components can be described as generative synthesisers, as they can autonomously control medium- and long-term aspects of musical structures (Boden and Edmonds 2009). When composing and performing with generative instruments, musicians engage with continuous semi-autonomous processes, modifying their characteristics through changes in the systems' parameters (Di Scipio and Sanfilippo 2019).

The integration of generative components into digital instruments creates novel possibilities for composition. At the same time, it creates the need for new approaches to notation that consider the specific characteristics of generative music systems. Historically, notation has had the function of aiding memory for the execution of a piece, communicating the piece's characteristics to collaborators and creating possibilities for interpretation (Magnusson 2019). More generally, notation systems also have the function of aiding abstract thought using symbols, enabling representation of complex concepts and ideas. Ultimately, as suggested by Hope (2017), the function of notation is to represent the abstract idea of a work such that the piece is identifiable regardless of who performs it and the choices they make within their execution.

Several challenges impede attempts to notate generative music systems. One of them is that the systems evolve as their versions progress. Often, fixed compositions can only be related to a specific version of a generative music system (Birnbaum et al. 2017). Second, compositions may not be feasible to play on other instruments, blurring the boundary between control interfaces and compositional interfaces (Magnusson 2010). This makes it difficult to create a notation system that allows interpretation across instruments, because different instruments allow to control different characteristics of sound, and they have different time scales of operation. Third, generative and/or stochastic components make it impractical to notate short-term features such as pitch and duration, given that the short-term aspects of the sound evolution can be controlled by the algorithmic process, and performers are often in charge of parameters governing higher-level characteristics of the sounds generated by the algorithm (Magnusson et al. 2022; Bown et al. 2009).

Corresponding author: Vincenzo Madaghiele; Email: vincenzo.madaghiele@imv.uio.no
Vincenzo Madaghiele and Leonard Lund contributed equally as first authors of this article.

Cite this article: Madaghiele V, Lund L, Holzer D, Kelkar T, Tatar K, Holzapfel A. (2025). Expanding the machine: Notating generative synthesis with a state-based representation and a navigable timbre space. *Organised Sound* **30**: 355–366, https://doi.org/10.1017/S1355771825100915

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

Music systems with generative components require appropriate modes of notation, which must consider the mutable characteristics of these systems and the non-linear timbral effects of their control parameters. In this article, we propose a notation system for the Benjolin (Hordijk 2009), a chaotic synthesiser based on feedback and logic operations. We consider chaotic synthesis a subclass of music generative system, thus making the Benjolin a generative instrument.

Generative instruments are accessed through interfaces that describe the meta-level of music the instrument acts upon, a representation of the system's state often indicated by parameter values. Therefore, a composition with a generative system can be represented as a sequence of the states the instrument visits over time. In this article, we propose a notation method in which the system's state is represented by its position in the three-dimensional timbre space of the instrument relative to the other possible states, rather than a set of parameter values. Consequently, a composition in our system is represented as a three-dimensional path in the timbre space, describing how the timbre of the instrument evolves in time throughout the sequence of states with respect to its full range of possibilities. To notate pieces composed with the Benjolin, we introduce the Meta-Benjolin, a state-based interactive notation system based on a mapping between the parameter space and the timbre space of the synthesiser, and an interactive timeline.

We present the Meta-Benjolin as both a notation system and a meta-instrument, since it enables control of the Benjolin in novel ways. After describing the design of the Meta-Benjolin, we discuss the results of a qualitative user study in which musicians used the instrument to compose a one-minute piece. Considering notation as a conceptual representation able to capture the basic idea of a piece, we developed a thematic analysis of the interviews, specifically focusing on the following research questions: (1) How can we appropriately represent graphically a piece composed with a generative synthesiser, describing the temporal evolution of its timbral characteristics? (2) How did musicians employ the proposed interactive notation system to create expressive musical structures? (3) What are the advantages and shortcomings of the proposed notation system in providing an identifiable abstract representation of musical ideas?[1]

2. Background

2.1. Generative synthesis and process-based music

Generative music and generative art have a long history, with many examples of systems and machines within musical and artistic works or appearing within the processes of making those artworks. A recent definition by Brian Eno describes generative music as 'the idea that it's possible to think of a system or a set of rules which once set in motion will create music for you' (D'Errico 2022). The assumption behind the creation of music as a result of a process is that music can be generated by applying a combination of algorithmic techniques based on music theory. This topic has been approached by serialist and minimalist composers such as Michael Nyman and Steve Reich, who have explored rule-based transformation processes with notated musical works (Schwarz 1980). Musicians interested in cybernetics have employed feedback networks as a way to construct semi-autonomous systems (Sanfilippo and Valle 2013), and several other techniques have been used in custom algorithmic systems over the years in contexts of composed and improvised music in many styles, encompassing ambient electronic works such as 65daysofstatic's *Wreckage Systems* (65daysofstatic 2022) and interactive improvisation systems like George Lewis' *Voyager* (Steinbeck 2018).

In this article, we are specifically interested in generative synthesis. We define generative synthesis models as sound-generating systems that can organise sound patterns and their spectromorphology from the *micro* up to the *meso* time scale of musical perception (Smalley 1986). The *meso* time scale is described by Roads (2003) as a local time scale with respect to a global music composition, whose duration is measured in seconds. At the *meso* time scale, sound objects are organised into melodic, harmonic, rhythmical and contrapuntal relations. In electronic music, this is the time scale at which the evolution of sound masses is perceived, with variations in amplitude, tempo, density, harmonicity and spectrum (Roads 2001). Events at this time scale are described by Wishart (1994) as *sequences*, while Smalley identifies this scale as the one in which *spectromorphological* evolution occurs.

Several algorithmic techniques have been used to create such generative synthesis models. Some examples are modular synthesis patches (Teboul and Kitzmann 2024), stochastic sequencers (Luke and Carnovalini 2024), rule-based systems (Spangler 1999), statistical sequence modelling (Van Der Merwe and Schulze 2010), genetic algorithms (Alfonseca et al. 2007), cellular automata (Agostini et al. 2014), feedback systems (Sanfilippo and Valle 2013) and chaotic algorithms, the latter of which is explored in this article. The recent development of deep learning has introduced a large amount of novel deep generative models (Tomczak 2022) such as autoregressive models, autoencoders and generative adversarial networks. These algorithms can be used to generate control data for sound synthesis (Barkan et al. 2019), symbolic notation (Herremans et al. 2017; Huang et al. 2018; Mittal et al. 2021) or to generate audio (Caillon and Esling 2021; Mehri et al. 2022; Tatar et al. 2023).

When performing with generative synthesis, musicians have control over a set of sound-generating parameters, whose effect is typically non-linearly correlated with changes in behaviour of the system. A change in control parameters affects the system's state, which determines a change in its time behaviour up to the *meso* time scale. This control approach reconfigures the relationship between the musician and the instrument, transforming it into a dialogue in which agency is shared, a collaboration between the system and the composer (Di Scipio and Sanfilippo 2019; Erdem et al. 2022; Magnusson et al. 2022).

In this article, we specifically investigate control and notation strategies for the Benjolin, a generative synthesiser whose sound qualities depend on a finite set of eight continuous synthesis parameters, with no external sound input. The Benjolin is a chaotic synthesis model (Slater 1998) which, as we argue in this article, represents a specific instance of generative synthesis technique.

2.2. Timbre-based control of synthesis

The possibilities given by electronic synthesis and elaboration of samples have shifted the focus of electroacoustic composers from categories like pitch, rhythm and harmony to the evolution of spectral properties. In this context, timbre has been used to refer to sonic properties that cannot be described by traditional categories. However, timbre is a debated concept whose definition varies widely across disciplines and musical practices, so the representation (and therefore notation) of timbre is an open problem in contemporary research. A wide range of disciplines have discussed

[1]The Meta-Benjolin is available online at https://meta-benjolin.com. The code for the Meta-Benjolin is released as open-source software at https://github.com/vincenzomadaghiele/Meta-benjolin.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

the concept of timbre, including but not limited to music psychology, music informatics and (ethno)musicology. Timbre, as argued by ethnomusicologist Cornelia Fales (2005), exists only in the mind of the listener, and musicologist Nina Sun Eidsheim (2019, p.10) even disputes that timbre is a knowable entity. In contrast, in music psychology, timbre spaces have been proposed as a means to map the similarity of sounds as perceived by experiment participants to a Euclidean space (Grey 1975).

A timbre space is a representation of the relations between a set of sounds, such that neighbouring points are similar, while distant points sound very different from each other (Wessel 1979). Movements in the timbre space should map linearly to a perceived change in the quality of the sound. Initially, timbre spaces have been constructed using multidimensional scaling on a dataset of pairwise similarity scores obtained from surveys (McAdams 2019). Recent approaches are based on computing perceptually relevant audio features like spectral shape, loudness and mel-frequency cepstrum coefficients (MFCCs) on a dataset, reducing the dimensionality of that dataset using dimensionality reduction techniques like principal component analysis (PCA) or uniform manifold approximation projection (UMAP) (Garber et al. 2020; Moore and Brazeau 2023), or deep learning approaches like variational autoencoders (VAEs) (Tatar et al. 2021; Esling et al. 2021; Caillon and Esling 2021).

Timbre spaces have been applied to explore the sonic output of complex synthesis models to develop timbre-based control strategies (Fasciani and Wyse 2012). The timbre space of a synthesis model can be constructed by evenly sampling its control parameters and encoding the corresponding synthesis sounds based on their timbral properties. Controlling parameters based on their timbre representation can be challenging, especially for synthesis models whose relationship between changes in parameters and changes in timbre is non-linear (Hayes et al. 2025).

Several software to compute and navigate timbre spaces of sound libraries and synthesis models have been released in recent years. Examples are Audiostellar (Garber et al. 2020), Timbre Space Analyzer & Mapper (TSAM) (Fasciani 2016), FlowSynth (Esling et al. 2021), Flucoma (Tremblay et al. 2022) and Latent Timbre Synthesis (Tatar et al. 2021). Audiostellar visualises a user-supplied audio corpus as an interactive scatter plot in two dimensions by extracting timbral features and then reducing the dimensionality using PCA, t-SNE or UMAP (Garber et al. 2020); users interact with the timbre space with the mouse, and hovering over a point triggers the sample to play (Moore and Brazeau 2023). An alternative option to sample playback is mapping synthesis controls to timbre space positions. An example of this is the TSAM by Fasciani (2016), a tool which analyses the possible timbres of a given synthesiser, finds a two- or three-dimensional representation using Isomap and then maps the synthesiser parameters to the timbre-based control space. Another approach for synthesiser control through a latent space is FlowSynth, which combines VAE and normalising flows (NF), using the VAE to learn a compact representation of the audio and NF to map this to the synthesis parameters (Esling et al. 2021).

These methods mainly employ timbre similarities computed over mathematical descriptors based on psychoacoustic perceptual studies. The limitations of this approach have been dissected by Morrison (2024) as going from the 'thick event' (Eidsheim 2019: 5) of perceiving sound in socioculturally situated interactions between listeners and sound producers to a single point in a Euclidean timbre space. To further complicate things, research in music informatics has developed with a 'curious divide' (Siedenburg et al. 2016) to music psychology, for instance, by using a larger number of signal descriptors to estimate the similarity between sounds.

In this article, we employ timbre descriptors as a mathematical way of representing audio where traditional categories of pitch, harmony and rhythm are not suited for representation of complex spectral evolutions. Our study entails the computation of timbre descriptors to arrive at a timbre space that provides meaningful interpretations in the 'thick event' of notating generative music, bridging the described extreme positions. We acknowledge that we 'flatten' (Morrison 2024) a thick event in this process, but we will return to a culturally situated depth through our interviews with composers using the system.

2.3. Notation of gesture, sound or process

The design of an interface (physical or screen-based) for a digital sound model is strictly related to how music for that interface can be notated. Usually, notation is based on some abstraction that is necessary for being able to translate musical instruction to gesture, process or sound form. Historically, notation embeds instructions relative to musical actions, characteristics of sound and/or ways in which musicians (and instruments) should behave. These aspects are embedded in different ways in various kinds of notation. For example, a live coding script describes a sequence of operations that may be followed to create similar musical profiles using different scripting environments. A graphical score such as Cage's *Aria* (1958) asks the performer to interpret melodic contour lines using thoughtful choices, steering the monodic instrument to represent changes in sound. Digital insctruments such as the T-stick may also be notated at the gesture level, instructing the performer how exactly to move the instrument (Moriceau et al. 2024).

The way instruments are notated is inseparable from the way they are accessed and controlled gesturally, retaining links to ideas that are central to the music theory(ies) informing the performance practice. For our case with the Benjolin, Figure 1 represents an example of graphically notated pieces for Benjolin created by Pete Gomes, where the notation directly instructs the player how to interact gesturally with the synthesis parameters describing the piece.

Several approaches to notation of generative processes have been proposed throughout the years. Pre-algorithmic generativity is discussed by Rochais (2024) with several examples of generative and score-like formulations. A common strategy for multidimensional notational representations is to use graphical notation. Schaeffer's propositions for acousmatic music describe an early abstraction based on a multimodal understanding of the interaction between shape and sound (Couprie 2018). Drawing as a strategy for representing electroacoustic music (Thiebaut et al. 2008) is found in works not only dealing with notation but also to study perception and musical representation of the acousmatic (Godøy 2006; Jensenius 2013; Nymoen et al. 2013; Kelkar et al. 2018). Xenakis' ideas (Nelson 1997) for UPIC graphical scores represent a graphical notation directly linked to the spectromorphology of sound, while contemporary cluster-based methods revolve around corpus-based, information-heavy styles of notating music (Scordato 2017; Garber et al. 2020; Bell, 2023; Roma et al. 2019; Esling et al. 2021; Coduys and Ferry 2004). Vickery (2014) proposes screen-based score paradigms for representing segmented scores in their articles, with an overview of interactions between temporality and screen display for score mappings. Bell, 2023) raises the question of the extent to which timbre space

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

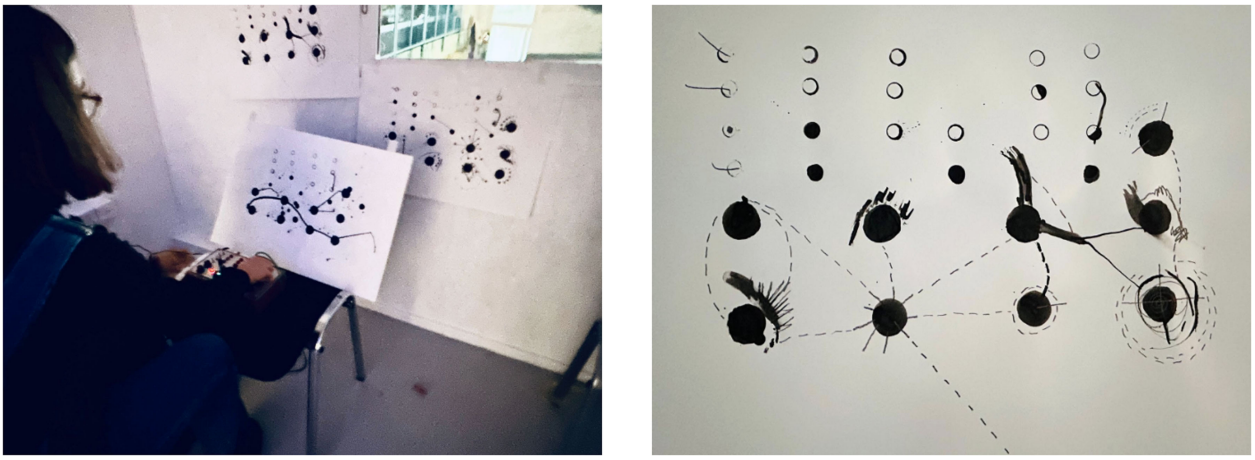


Figure 1. Notation for Benjolin pieces. Images by Pete Gomes, used with permission. This notation is based on the hardware Benjolin instrument, representing the eight knobs and the modular patch. The graphic signs around each knob portray different behaviours with which each knob governing synthesis parameters should be approached while performing.

representations (Couprie 2018) present a typology of graphical scores for composing and transcribing electroacoustic music. However, even the multidimensional potentials of graphical scores can be challenged when generative and chaotic instruments need to be represented due to the unpredictability and wide variations of their sonic outputs.

Magnusson (2014) presents several case studies for interpreting algorithmic scores, establishing the idea of computer codes being literally interpretable as musical scores. In their work on agential instruments, Armitage and Magnusson (2023) explore dynamic scores further with these ideas of agential notation. Animated notation and evolving graphical notation also represent paradigms lending themselves to multidimensional, time-varying representations. Newer representations of graphic notations integrated with machine learning have led to explorations of novel temporal, sonic and spatial interactions.

The dimensions of interface, control and music theory have many intersecting concepts when it comes to notation (Hope 2017). Many interactive notation systems can be considered at the same time as control interfaces and musical representations, because in these cases the notation system actively serves as a control structure for a sound model as well. UPICS, cluster-based methods, live coding languages and agential scores are clear examples of this duality.

3. The Meta-Benjolin

Generative synthesis models produce time-varying sound processes based on a set of control parameters. Generative processes control sonic structures up to the *meso* time scale, so even when the control parameters are static, the timbral and rhythmic characteristics of the generated sound can exhibit wide variations. Composing with generative systems requires being able to work with fixed states of the system for specific time durations, because allowing a state's sonic evolution to unfold is an essential part of its aesthetics. For this reason, we propose a notation system that combines the timbral representation of system *states* with a composition timeline, allowing users to select system states based on their timbral characteristics, manipulate their duration and introduce transitions between them.

In this section, we introduce the Meta-Benjolin as a notation system and a meta-instrument developed as a control structure for the Benjolin synthesiser. The Meta-Benjolin allows access to the possible timbres that can be produced by the Benjolin in the form of a three-dimensional, navigable point cloud. Each point in the cloud corresponds to a state of the Benjolin sound model, identified by a unique set of parameters, which users can select and organise in time.

The Meta-Benjolin controls the sound model of the Benjolin directly by changing its parameters as users interact with a screen-based interface. Besides organising states in time in the composition timeline, users can introduce transitions between them by navigating the shortest path between two states in the parameter space – a crossfade transition – or in the timbre space – a meander transition. This means that overall, the possible transitions between two consecutive states are three: a crossfade, a meander and a jump cut, which is a direct change of the parameters of the Benjolin from one state to the other. Theoretically, these transitions include all control options for the Benjolin, including single-parameter sweeps. The introduced transitions create new possibilities for the generative system of the Benjolin, because they feed different signals to its feedback component, affecting its resulting behaviour. This aspect makes the Meta-Benjolin a meta-instrument (Fiebrink 2017), because it provides control possibilities beyond the parametric controls of the original Benjolin based on a layer of abstraction[2].

In the Meta-Benjolin, a composition is graphically represented by the sequence of consecutive states, their duration, the transitions between them and their location in the timbre space. The notational representation of a piece in the Meta-Benjolin consists of the whole interface, including the composition timeline and the piece's representation as a path in the three-dimensional point cloud. In the following sections, we introduce the Benjolin, describe the construction of the timbre space and discuss the design decisions behind the specific implementation of the Meta-Benjolin.

3.1. The Benjolin

The Benjolin is a hardware synthesiser designed by Rob Hordijk in 2009, making features of his commercial Blippoo Box (Hordijk

[2]Video material *Video_example_1.m4v* of this article shows a video demo of the Meta-Benjolin functions.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press



Figure 2. A hardware Benjolin as a standalone synthesiser. Macumbista Instruments, 2025.

2009) more accessible as a do-it-yourself project for the analogue synthesist community. The Benjolin, shown in Figure 2, is made up of two voltage-controlled oscillators with frequency control parameters, Oscillator 1 (O1 FRQ) and Oscillator 2 (O2 FRQ), and one voltage-controlled low-pass filter with frequency (FIL FRQ) and resonance (FIL RES) controls. The Benjolin's audio output is the result of a comparator operation of the triangle waves from O1 and O2 fed into the low-pass filter (see Figure 3).

The defining characteristic of both the Benjolin and the Blippoo is a chaotic control structure called the Rungler. The Rungler utilises a shift-register-based pseudo-random number generator to generate a 3-bit, stepped analogue control voltage signal. The square wave of O1 provides a digital data source for the pseudo-random generator, while the square wave of O2 acts as a digital clock controlling the generator's rate of change. The resulting Rungler control voltage can then be applied as the variable modulation source to the frequency parameters of the oscillators and the filter. In total, there are eight user-controllable parameters in the instrument:

O1 FRQ: variable frequency of Oscillator 1;

O1 RUN: variable amount of Rungler stepped wave control over Oscillator 1 frequency;

O2 FRQ: variable frequency of Oscillator 2;

O2 RUN: variable amount of Rungler stepped wave control over Oscillator 2 frequency;

FIL FRQ: variable cutoff frequency of the filter;

FIL RES: variable resonance of the low-pass filter;

FIL RUN: variable amount of Rungler stepped wave control over filter cutoff frequency;

FIL SWP: variable amount of Oscillator 2 triangle wave control over filter cutoff frequency;

Because the output of the Rungler can modify the frequency of both the clock and data oscillators, the overall arrangement fulfils the basic criteria of a chaotic synthesis system: it displays a high sensitivity to initial conditions, feedback is present within the system and the system contains an element of non-linearity (Slater 1998). Hordijk notes that the interaction of the Rungler and its corresponding oscillators represents a chaotic attractor, which seeks a new balanced state from the turbulence created whenever its control parameters are disturbed (Hordijk 2009).

For the purposes of this project, we created a digital emulation of the Benjolin sound model using the Pure Data (PD) environment. The emulation facilitates network connections to these parameters from other programming environments, such as Python, via the Open Sound Control (OSC) protocol. Our PD Benjolin[3] emphasises the functional aspects of the original instrument, with less priority given to a completely accurate reproduction of its output sound. In this article, we refer to a state of the Benjolin as the sonic output of the Benjolin produced by a unique combination of its eight sound-generating parameters.

3.2. Latent space

The Meta-Benjolin is based on a latent representation of the timbre of the Benjolin. Each state in the Benjolin corresponds to a sequence pattern of changes in spectromorphology and rhythm whose period is determined by the clock. The rhythmic and spectromorphological aspects of the sound in each state are strongly interrelated and can be difficult to control separately. Therefore, a notation system for this instrument needs to consider this unique aspect. For this reason, we decided to use a latent timbre representation in which the complex multidimensional spectral characteristics of each state are reduced to three dimensions. This representation has been constructed using a VAE (Kingma and Welling 2013), an unsupervised machine learning model that learns a compressed representation of a dataset

[3]The PD Benjolin is available as open-source software at https://github.com/macumbista/benjolin.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

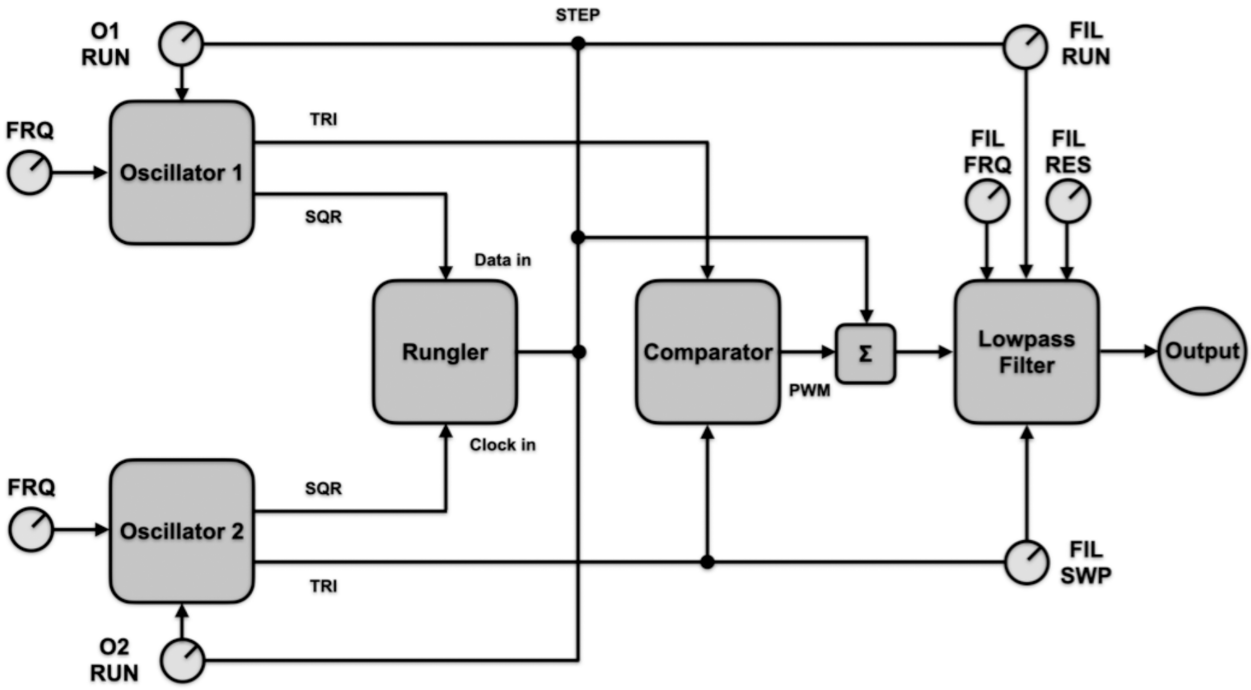


Figure 3. Diagram of signal flow and operations in the Benjolin synthesiser.

by projecting the distribution of training datasets to a lower-dimensional space.

3.2.1. Dataset

Creating a timbre space representation requires the generation of a dataset describing the states in which the Benjolin can be. To do so, it is necessary to sample the continuous values of each parameter such that the range of possibilities of the instrument is reasonably portrayed. The Benjolin is a chaotic system, meaning that its output can change widely with small parameter variations. However, we can count on some predictable behaviour at the timbral level, which means that a Benjolin state has similar timbral characteristics, regardless of previous states, if a state runs for a long enough time for it to reach an equilibrium. This time has been experimentally established as two seconds. To produce the dataset, we sampled two-second-long recordings of 85536 unique parameter settings of the PD Benjolin emulation at a sample rate of 44.1 kHz and a bit depth of 16. The recordings were done by sampling parameter settings at 20%, 40%, 60% and 80% for each parameter, leading to 4^8 = 65536 recordings to ensure an even spread of the parameter space. An additional 20000 unique settings were randomly sampled to complete the dataset.

Audio features were extracted to capture timbral information from each recording. We used torchaudio of PyTorch for MFCC calculation, and the remaining features were calculated with custom implementations in CuPy so that all computations were on GPU. These features comprised 13 MFCCs, Spectral Centroid, Spectral Flatness, Spectral Flux, Zero-Crossing Rate and Mean Square Amplitude, resulting in 18 feature dimensions. Features were calculated for each audio frame using a window size of 1056 samples and a hop size of 64 samples. Each audio recording was encoded using a bag-of-frames approach (Aucouturier et al. 2007). The mean and the standard deviation of each feature were calculated across all the frames within the recording, resulting in a 36-dimensional feature vector for each state. This approach allows for a focus on timbral features, treating features as a static distribution which ignores temporal relations between frames. This is helpful as the influence of the order of audio events – which two seconds of a longer pattern are sampled – is decreased, a clear advantage over working directly with spectrogram representations or audio. Finally, the dataset was standardised such that each feature would have a mean of 0 and a standard deviation of 1.

3.2.2. VAE training

A VAE neural network (Pinheiro Cinelli et al. 2021) was used to obtain a latent representation of the timbre space. The VAE model consists of an encoder network, a sampling layer and a decoder network. We employed a VAE where encoder and decoder networks each consist of three hidden fully connected layers. The encoder receives a 36-dimensional input vector, x, reducing it to 16 latent dimensions.

The loss function consists of the reconstruction error (mean-squared error) and the fl-weighted Kullback–Leibler divergence with fl = 0.01. Training ran for 200 epochs using an Adam optimiser with a learning rate of 0.001 and a learning rate scheduler (patience = 5, factor = 0.5). Mixed precision training was used to speed up training and reduce computational resources. After training was completed, the entire dataset was fed through the encoder network to compute the mean vector of each data point, which was used as the latent space projection for each data point. PCA was used to further reduce the latent space to three principal components, retaining 96.6% of the variance of feature vectors. Lastly, only the three-dimensional coordinates of each data point in the final latent space are used to map from coordinates to the corresponding synthesis parameters in a lookup table, as shown in Figure 4.

3.3. Screen-based notation interface

The screen-based interface of the Meta-Benjolin allows the user to navigate the virtual three-dimensional timbre space as a point

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

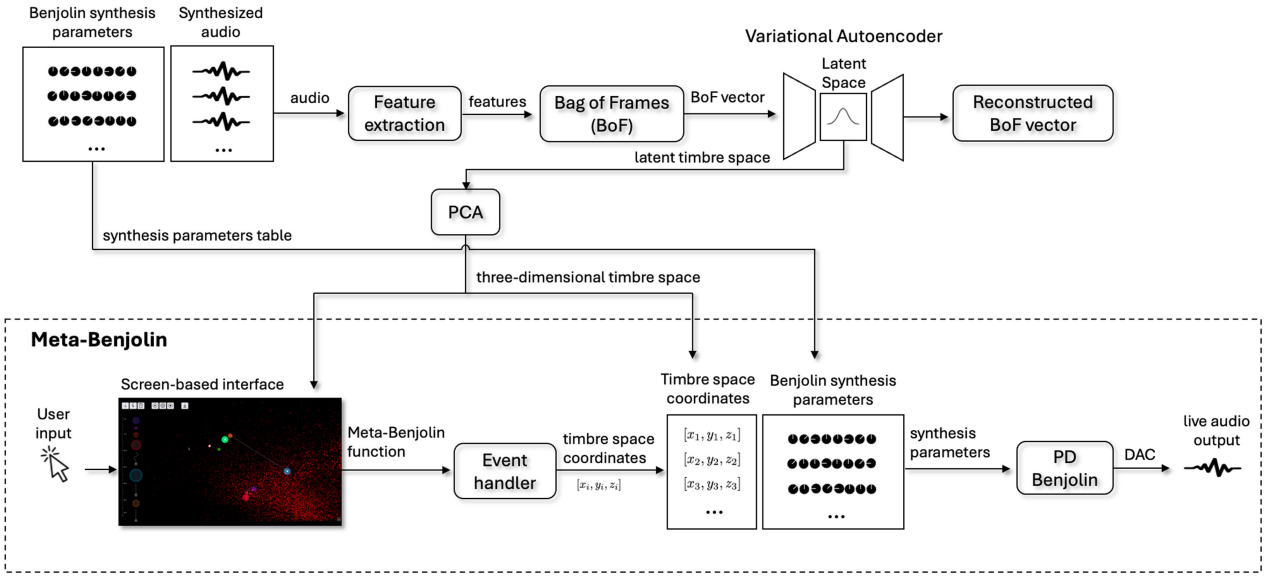


Figure 4. Data flow diagram of the Meta-Benjolin structure. The latent representation based on timbre learnt by the VAE is used in the screen-based interface to map three-dimensional coordinates to the corresponding synthesis parameters.

cloud. The design of the interface aims to enable visualisation and access to the full range of possibilities of the instrument, interacting with the sound process created by the Benjolin based on its timbral qualities, without any knowledge of the parameters of the original synthesis model. A diagram of the interface is shown in Figure 5.

The graphic user interface of the Meta-Benjolin has been developed in JavaScript, and it is accessed from a browser. The sounds are generated using the PD Benjolin running in the background. Whenever synthesis parameters change in response to a user action, the interface accesses a table mapping the three-dimensional coordinates of each point in the space to the corresponding eight-dimensional parameters of the synthesiser, whose values are sent to the PD Benjolin through OSC, causing a change in sound. The user interface is composed of three main components: a three-dimensional point cloud, a vertical composition timeline and a menu.

Point cloud. The three-dimensional point cloud occupies the entire screen. A three-dimensional representation has been chosen over a two-dimensional space to encourage nuance and variations. The user can navigate the space where the cloud is located using the mouse, zooming in and out to specific regions. Each point in the cloud corresponds to one state of the Benjolin from the training corpus of the VAE. When the user hovers the mouse pointer over a point in the cloud, the sound of the corresponding state of the Benjolin is played; if the mouse pointer is not hovering on any point, the instrument is silent. This interaction was designed as a 'scrubbing' of the corpus (Wessel and Wright 2002). When a user clicks on a point, it changes in colour, and a circle of the same colour is created in the composition timeline, indicating that the state has been chosen by the user to be part of the piece. The colours are chosen randomly.

Composition timeline. The composition timeline is located on the left side of the interface. The composition flows vertically from the top to the bottom as a sequence of states and transitions between them. States are represented as circles and transitions as arrows. The states selected by the user are represented as circles of different colours, stacked vertically. The vertical order of the states in the composition timeline represents the time evolution of the piece. The size of each circle in the composition timeline represents the time duration of the corresponding state, which can be increased or decreased by dragging the circle's border. The user can change the order of the states in time by dragging elements in the composition bar in between each other. When two states are subsequent in the timeline, a line is drawn between them in the point cloud, indicating that there is a transition between those two states. In this way, the whole composition is represented as a path in the three-dimensional space, superimposed on the point cloud.

Menu. The menu is located at the top left of the screen. It is composed of seven buttons in three groups. The buttons in the menu allow to introduce crossfade and meander transitions between states, delete states from the composition timeline, play back, stop and record the whole composition, and download the coordinates of the states as a text file.

4. Evaluation methodology

The Meta-Benjolin was evaluated through a qualitative user study. The aim of the study was to understand how musicians engage with the functions of the Meta-Benjolin, focusing on how they use the three-dimensional representation of sound we propose. The study methodology combined think-aloud (Charters 2003) and semi-structured interviews (Adams 2015) facilitated by stimulated recall (Dempsey 2010). The transcript of the interviews, together with the sound material and the screen recordings of the participants interacting with the Meta-Benjolin, has been analysed using thematic analysis (Braun and Clarke 2024). The study was carried out at KTH Royal Institute of Technology in Stockholm (Sweden). It involved 19 participants with various degrees of experience in the composition of electroacoustic music. The participants were gathered through an open call (posters, mailing lists and social media posts) to students of media technology at KTH, students of electroacoustic composition at the KMH Royal College of Music (Stockholm, Sweden) and electroacoustic composers involved with Fylkingen and Elektronmusikstudion EMS. In the following discussion, participants are referred to with a letter and a number (e.g., A5, D3) to anonymise their identity.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

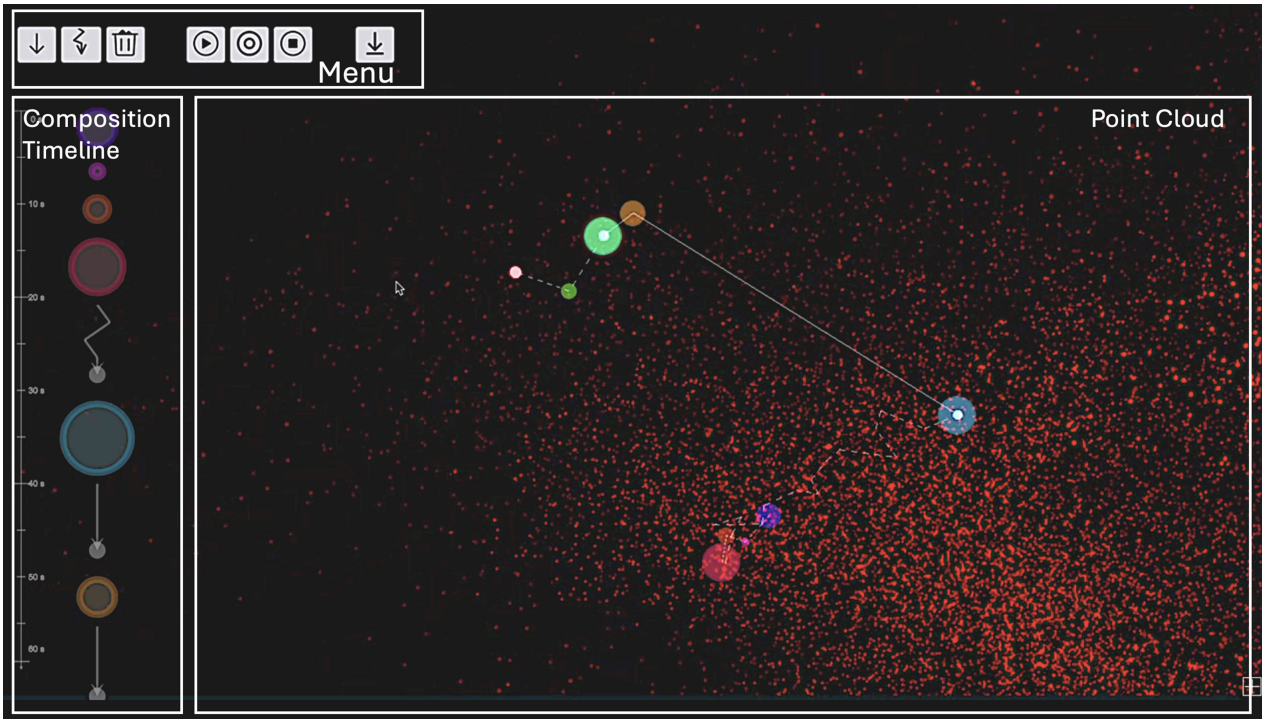


Figure 5. The Meta-Benjolin interface is composed of a three-dimensional navigable point cloud, a vertical composition timeline (left) and a menu (top-left).

At the start of the study, participants were given a short introduction to the Benjolin instrument and to the idea behind the Meta-Benjolin. Participants did not play the physical Benjolin instrument. The study was structured in two parts: in the first section, participants were asked to explore the point cloud for six minutes, with the task of finding five sounds that are as different from each other as possible, so that the participants get familiar with the navigation. In the second section, participants were tasked with composing a short piece of one-minute duration starting from scratch, using the Meta-Benjolin, in 15 minutes. During composing, participants were encouraged to express their thoughts out loud. Afterwards, participants were asked to listen back to their one-minute composition and communicate their thoughts about it within a semi-structured interview. The aim of this section was articulating the compositional strategies employed by them when creating their piece and how the functions of the Meta-Benjolin were used[4].

5. Results

We developed our thematic analysis with the objective of understanding the musical possibilities afforded by the three-dimensional representation and the effectiveness of the state-based notation at representing the Benjolin sounds. Consequently, we explored two main themes: the use of three-dimensional space to create musical structures and the capability of the notational system to express visually the unique identity of a piece.

5.1. Composing in three-dimensional timbre space

When discussing their composition process, several users referred to the point cloud as a sonic palette from which they could choose the source material for their piece. When composing, users usually started by looking for an area with interesting sonic materials. After selecting a few states from the cloud, they then sorted them in the composition timeline based on the states' sonic relationship to each other, changing their duration and connecting them with transitions. The composition process usually involved several phases of exploration of the cloud, finding a set of states and sorting them in time.

'When I chose one sound, I tried to think of what would naturally make sense for a next sound, what should it progress to? So I guess in that way, I was looking around for things like, oh, maybe I would want something louder now, something more squiggly now'. (A5)

The way musicians chose to order the states in time was often based on temporal variation of specific aspects, identified by musicians themselves by listening and comparing the sounds of states according to their proximity. Some musicians focused on the variation of pitch and dynamic aspects of the Benjolin, while others combined descriptions of *micro* aspects such as pitch, rhythm and dynamics with more abstract *meso*-level descriptions:

'Well, it starts from the darkness and the void and it comes closer. It's livelier here, it's almost biological in a way, from the cosmic background. It's something more melodic, periodic, which results in chaos and noise [. . .], from the dark noise to the lighter noise'. (B3)

Participants described the state's *meso*-level characteristics, referring to the *complexity* and *simplicity* of sounds and their *softness*, *darkness* or *liveness*. When sorting the chosen states on the timeline, participants often created narrative, tension and evolution in the structure of their piece by varying these aspects.

'What I wanted to have done is something starting quite simple, adding complexity, and then back to simplicity'. (B4)

Distances in the three-dimensional space were used to explore variations. Consecutive states belonging to the same neighbourhood in the cloud were used to create nuanced variations of timbre, often within the same section of the piece, while large distances were used

[4]The study procedure, participant demographics and the 19 short compositions are available at https://doi.org/10.5281/zenodo.15781184, as sound and visual representation in the Meta-Benjolin.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

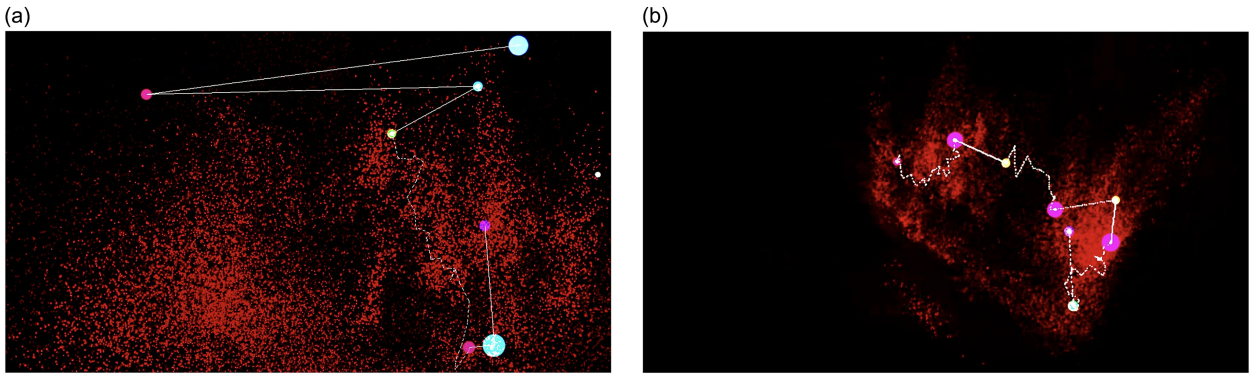


Figure 6. Examples of the use of transitions to navigate long distances. D2 used a meander transition in the middle of the piece to connect two sections; within a section, neighbouring states are connected using crossfades. A5 used a crossfade and a meander transition to navigate between two neighbourhoods in the cloud, each corresponding to a section in their piece. (a) Composition by D2 (detail), Sound_example_4.m4a in the sound material. (b) Composition by A5 (detail), Sound_example_5.m4a in the sound material.

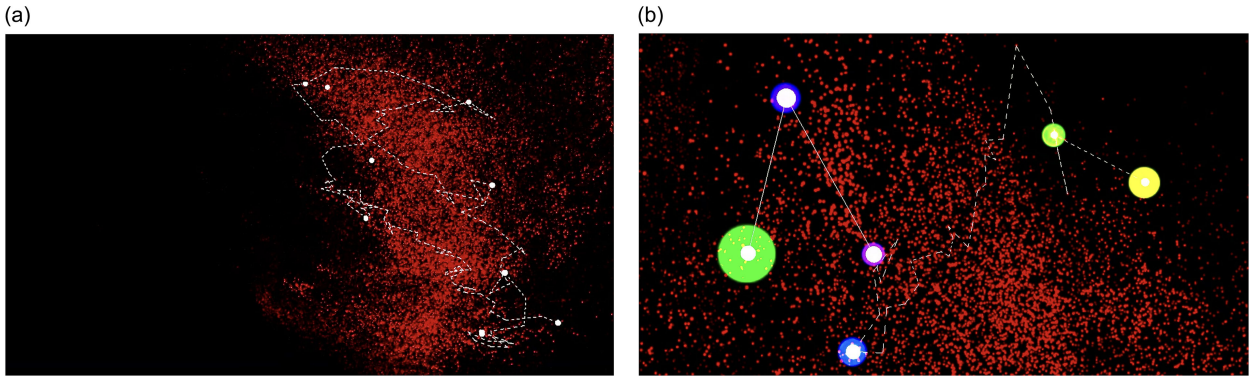


Figure 7. The space distribution of these two compositions gives information about how the sound evolves in time. While composer D3 created a gradual and constant evolution by navigating the whole point cloud using the meander transition, B3 was interested in exploring local variations and nuances. This difference can be seen by the fact that the viewpoint is zoomed far out in (a), while it is much closer to the cloud in (b). (a) Composition by D3 (detail), *Sound_example_1.m4a* in the sound material. (b) Composition by B3 (detail), *Sound_example_2.m4a* in the sound material.

to create contrast between sections, often employing a meander or crossfade transition between two distant states, as in Figure 6.

An alternative approach was using transitions regularly to create a continuous evolution, as in the case of the piece of participant D3, shown in Figure 7a, who created a loop exploring the whole cloud in small steps between selected points using the meander transition. Other participants remained in the same neighbourhood for the whole piece, travelling short distances to develop gradual, minute alterations. An example of this is the piece composed by B3, portrayed in Figure 7b, which can be divided into two short sections, both starting with noisy rhythmical elements evolving to pitched material.

5.2. The Meta-Benjolin as a notation system

Notation can be used for communication between musicians, as an aid to memory for a performer, as a compositional language to facilitate abstract thoughts, and ultimately, to represent the conceptual nature of a piece, its identity. In the notation system of the Meta-Benjolin, the point cloud representation and the timeline have a complementary function. This is confirmed by participants' perceptions, as pointed out by user A2:

'I thought about the timeline more in terms of what song I was creating than the cloud. The cloud was for me the sound, the material I could use. This was like, this was the material, and this was what I was working on'. (A2)

A recurring theme in the interviews was that the Meta-Benjolin does not include any *semantic* information about the quality of the sound. In the interface, the distances between timbres are represented as relative to each other in the point cloud, without any absolute referents. This is a design feature, based on our assumption that a non-semantic approach would be appropriate for an instrument such as the Benjolin, in which the spectral and rhythmic properties of each state are necessarily intertwined, and it is not possible to describe one aspect separately from the other.

The lack of semantic descriptions had some specific advantages and disadvantages. On the one hand, this aspect promoted discovery, serendipity and exploration; not having access to a qualitative description, users composed by comparing the states in terms of their visual distance and their sonic qualities. As participant C3 pointed out, this meant that the choice of states had to be based mainly on listening:

'You really rely a lot on what you're hearing in a very direct way. There's not a lot of control, which is good and bad, I guess'. (C3)

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

A clear drawback of non-semantic representation is that it can make navigating the cloud confusing, since the three axes have no explicit meaning. This made it sometimes frustrating for users to realise their more specific musical ideas. As expressed by participant C2:

'I tried to go in different directions, up and down, left and right, but then since I didn't understand the relation of the position and the sound, I didn't really pay attention too much to this, the most useful thing was to listen to the sounds repeatedly'. (C2)

Part of the confusion regarding the meaning of axes might be due to the lack of experience with the cloud. With more experience, users might be able to identify similar-sounding areas in the cloud intuitively and more securely. At the same time, the large number of points and their high density made it difficult to precisely delimit specific areas. As several users pointed out, the large number of points made it difficult to meaningfully think about small-scale variations because, without semantic axes, the sonic effect of small-scale steps is uncertain.

Overall, this system was seen as particularly effective in representing the differences between timbres and the amount of variation from one section to another:

'Whenever I think of making music, one thing that I'm really overwhelmed with is how things seem to move. There's just a lot of things moving at once and I'm like, okay wait, how are things moving? So in that way, I found it really cool that I was able to understand how things were moving'. (A5)

Despite the ambiguity related to the lack of semantic descriptions, users were able to express musical structures and construct meaningful musical ideas by employing notions from their specific styles, and the corresponding graphical representations showcase this wide variation of possibilities.

6. Discussion

The Meta-Benjolin shares with similar interactive notations the duality of being a notation method and a control system for a synthesis model. As for live coding, UPICS and agential scores, the notation representation cannot be completely separated from the control system, and we embrace this characteristic as a fundamental feature of digital systems. We argue that the Meta-Benjolin, besides being a control interface, can be used as a notation system because, by complementing the timbre space with a composition timeline, it abstracts the behaviour of the Benjolin's dynamic process – therefore allowing access to it for planning and execution of musical ideas, bypassing the technical meanings of the system parameters by which the instrument (as most synthesisers) is traditionally accessed.

As discussed in Section 2.3, notation can have several functions, among which are communication of musical ideas, facilitating interpretation across instruments and aiding abstract thought. We designed the Meta-Benjolin specifically to facilitate abstract thought through symbolic representations. Participants's tendency to refer to the cloud as a 'sound palette' for the organised states in the timeline confirms that our symbolic representation was intuitively understood, and participants were able to navigate it to express musical ideas with varying degrees of satisfaction. Participants did not have access to the eight sound-generating parameters while using the Meta-Benjolin; therefore, it was not possible to evaluate whether they would have preferred to compose using them with respect to using our representation. The interview study suggests that the main drawback of this system is the lack of semantic sound descriptions. In the future, this aspect could be introduced by colour-coding states according to specific spectral or rhythmic properties, or by clustering the cloud into similar-sounding neighbourhoods to facilitate semantic reasoning.

7. Conclusions

In this article, we have proposed a state- and timbre-based interactive notation system for generative synthesis. To explore its potential, we have introduced the Meta-Benjolin, an interactive notation system for the Benjolin synthesiser. Overall, our study suggests that the Meta-Benjolin has the potential to notate and visually represent complex compositional ideas. Despite a degree of ambiguity due to its non-semantic representation, composers were able to express themselves within their own styles and produce a wide variety of musical works. The visualisations express the unique identity of each piece, representing its temporal evolution through relative distances, transitions and relative point sizes.

To further evaluate the potential of our representation as a notation system, we encourage its application to other generative synthesis models and its expansion. For example, our user study does not show if this notation can be used for interpretation across instruments or how users would employ the Meta-Benjolin together with the physical Benjolin. We do not argue that this representation method is better or worse than possible alternatives; a comparison of the state and timbre representation we propose with respect to alternative notation methods for the Benjolin, such as the parameter-based notation in Figure 1, while interesting, is out of the scope of this article.

We propose that the structure of the Meta-Benjolin could be applied to other state-based generative synthesisers, not necessarily chaotic, and it could be used as a tool to represent electroacoustic compositions employing generative systems, while at the same time expanding the possibilities for control of these instruments. As the generative instruments are emerging faster than ever, our work is an initial step into investigating notation approaches to address the complexity of highly autonomous systems. We think that this research direction has potential in impacting a broader variety of actors, such as musicians, composers, sound designers, and creative practitioners in the film, content, or game industry who use musical instruments with machine learning or AI approaches in their practices.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S1355771825100915

Acknowledgements. This work is partially funded by the Department of Musicology at the University of Oslo and by the Swedish Research Council (2019-03694). This work is partially funded by the Wallenberg AI, Autonomous Systems and Software Program-Humanity and Society (WASP-HS), funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

The authors would like to thank Stefano Fasciani, Emil Kraugerud, Balint Laczko and the musicians who participated in the user study for their time, creativity and valuable reflections.

Conflicts of interest. The authors report no conflicts of interest.

Ethics Statement. The computations required for the machine learning models employed in this article were comparable to a daily computer usage of an

individual. Thus, the machine learning component of this project was not a significant part of its environmental impact. The study did not require an ethics approval inline with the Swedish regulations, and it follows the ethics guidelines of KTH Royal Institute of Technology.

References

Adams, W. C. 2015. Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation*, Hoboken, NJ: John Wiley & Sons., 864: 492–505.

Agostini, A., Daubresse, E. and Ghisi, D. 2014. Cage: A High-Level Library for Real-Time Computer-Aided Composition. In *ICMC*.

Alfonseca, M., Cebrian, M. and Ortega, A. 2007. A Simple Genetic Algorithm for Music Generation by Means of Algorithmic Information Theory. In *2007 IEEE Congress on Evolutionary Computation*, 3035–42. IEEE.

Armitage, J. and Magnusson, T. 2023. Agential Scores: Exploring Emergent, Self-Organising and Entangled Music Notation. In *Proceedings of the 8th International Conference on Technologies for Music Notation and Representation (Northeastern University, Boston, Massachussetts, USA, 2023)*.

Aucouturier, J. -J., Defreville, B. and Pachet, F. 2007. The Bag-of-Frames Approach to Audio Pattern Recognition: A Sufficient Model for Urban Soundscapes but not for Polyphonic Music. *The Journal of the Acoustical Society of America* **122**(2): 881–91.

Barkan, O., Tsiris, D., Katz, O. and Koenigstein, N. 2019. Inversynth: Deep Estimation of Synthesizer Parameter Configurations from Audio Signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **27**(12): 2385–96.

Bell, J. 2023. Maps as Scores:" Timbre Space" Representations in Corpus-based Concatenative Synthesis. In *International Conference on Technologies for Music Notation and Representation (TENOR)*.

Birnbaum, D., Fiebrink, R., Malloch, J. and Wanderley, M. M. 2017. 2005: Towards a Dimension Space for Musical Devices. In *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression, Current Research in Systematic Musicology*, Springer International Publishing, **3**: 211–22.

Boden, M. A. and Edmonds, E. A. 2009. What is Generative Art? *Digital Creativity* **20**(1–2): 21–46. https://doi.org/10.1080/14626260902867915.

Bown, O., Eldridge, A. and McCormack, J. 2009. Understanding Interaction in Contemporary Digital Music: From Instruments to Behavioural Objects. *Organised Sound* **14**(2): 188–96.

Braun, V. and Clarke, V. 2024. Thematic analysis. In *Encyclopedia of quality of life and well-being research*, Springer, 7187–93.

Cage, J. 1958. Aria. Peters Edition EP 6701.

Caillon, A. and Esling, P. 2021. Rave: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis. arXiv preprint arXiv:2111.05011.

Charters, E. 2003. The Use of Think-Aloud Methods in Qualitative Research an Introduction to Think-Aloud Methods. *Brock Education Journal* **12**(2).

Coduys, T. and Ferry, G. 2004. Iannix Aesthetical/Symbolic Visualisations for Hypermedia Composition. *Journees d'informatique musicale*.

Couprie, P. 2018. Methods and Tools for Transcribing Electroacoustic Music. In *Technologies for Music Notation and Representation (TENOR)*.

D'Errico, M. 2022. Worlds of Sound. In *Push: Software Design and the Cultural Politics of Music Production*. Oxford University Press.

Dempsey, N. P. 2010. Stimulated Recall Interviews in Ethnography. *Qualitative Sociology*, **33**: 349–67.

Di Scipio, A. and Sanfilippo, D. 2019. Defining Ecosystemic Agency in Live Performance: The Machine Milieu Project as Practice-based Research. 2019: array. *The Journal of the International Computer Music Association*, 28–43.

Eidsheim, N. S. 2019. *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Duke University Press, 288.

Erdem, Ç., Wallace, B., Glette, K. and Jensenius, A. R. 2022. Tool or Actor? Expert Improvisers' Evaluation of a Musical ai "toddler". *Computer Music Journal* **46**(4): 26–42.

Esling, P., Masuda, N. and Chemla-Romeu-Santos, A. 2021. Flowsynth: Simplifying Complex Audio Generation through Explorable Latent Spaces with Normalizing Flows. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 5273–5.

Fales, C. 2005. Short-Circuiting Perceptual Systems: Timbre in Ambient and Techno Music. Wired for Sound: Engineering and Technologies in Sonic Cultures, 156–80.

Fasciani, S. 2016. Tsam: A Tool for Analyzing, Modeling, and Mapping the Timbre of Sound Synthesizers. In *Sound and Music Computing (SMC)*.

Fasciani, S. and Wyse, L. 2012. Adapting General Purpose Interfaces to Synthesis Engines using Unsupervised Dimensionality Reduction Techniques and Inverse Mapping from Features to Parameters. In *ICMC 2012*.

Fiebrink, R. 2017. Machine Learning as Meta-Instrument: Human-Machine Partnerships Shaping Expressive Instrumental Creation. *Musical Instruments in the 21st Century: Identities, Configurations, Practices*. Berlin: Springer, **10**: 137–51.

Garber, L., Ciccola, T. and Amusategui, J. C. 2020. Audiostellar, an open source corpus-based musical instrument for latent sound structure discovery and sonic experimentation. In *Proceedings of ICMC*.

Godøy, R. I. 2006. Gestural-Sonorous Objects: Embodied Extensions of Schaeffer's Conceptual Apparatus. *Organised sound* **11**(2): 149–57.

Grey, J. M. 1975. An Exploration of Musical Timbre using Computer based tchniques: Analysis, synthsis and peceptual scling. [PhD dissertation]. Stanford University.

Hayes, B., Saitis, C. and Fazekas, G. 2025. Audio Synthesizer Inversion in Symmetric Parameter Spaces with Approximately Equivariant Flow Matching. Proceedings of of the 26th Int. Society for Music Information Retrieval Conf., Daejeon, South Korea, 2025.

Herremans, D., Chuan, C.-H. and Chew, E. 2017. A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys (CSUR)* **50**(5): 1–30.

Hope, C. 2017. Electronic Scores for Music: The Possibilities of Animated Notation. *Computer Music Journal* **41**(3): 21–35.

Hordijk, R. 2009. The Blippoo Box: A Chaotic Electronic Music Instrument, Bent by Design. *Leonardo Music Journal* **19**: 35–43.

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Simon, I., Hawthorne, C., Shazeer, N., Dai, A. M., Hoffman, M. D., Dinculescu, M. and Eck, D. 2018. Music Transformer: Generating Music with Long-term Structure. In *International Conference on Learning Representations*.

Jensenius, A. R. 2013. An Action-Sound Approach to Teaching Interactive Music. *Organised Sound* **18**(2): 178–89.

Kelkar, T., Roy, U. and Jensenius, A. R. 2018. Evaluating a Collection of Sound-Tracing Data of Melodic Phrases. *Institut de Recherche et Coordination Acoustique/Musique*.

Kingma, D. P. and Welling, M. 2013. Auto-Encoding Variational Bayes.

Luke, S. and Carnovalini, F. 2024. A Hierarchical, Modular Music Sequencer. In *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*, 429–38.

Magnusson, T. 2010. Designing Constraints: Composing and Performing with Digital Musical Systems. *Computer Music Journal* **34**(4): 62–73.

Magnusson, T. 2014. Scoring with Code: Composing with Algorithmic Notation. *Organised Sound* **19**(3):268–275.

Magnusson, T. 2019. *Sonic Writing*. Bloomsbury Publishing.

Magnusson, T., Kiefer, C. and Ulfarsson, H. 2022. Reflexions Upon FEEDBACK. In *NIME* 2022, The University of Auckland, New Zealand.

McAdams, S. 2019. The Perceptual Representation of Timbre. In *Timbre: Acoustics, Perception, and Cognition*, 23–57. Cham: Springer International Publishing.

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A. and Bengio, Y. 2022. Sample-rnn: An Unconditional End-to-End Neural Audio Generation Model. In *International Conference on Learning Representations*.

Mittal, G., Engel, J., Hawthorne, C. and Simon, I. 2021. Symbolic Music Generation with Diffusion Models. Proceedings of of the 22nd Int. Society for Music Information Retrieval Conf., Online, 2021.

Moore, T. and Brazeau, J. 2023. Serge Modular Archive Instrument (SMAI): Bridging Skeuomorphic & Machine Learning Enabled Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Moriceau, G., Yan, Y., Thibault, D. and Wanderley, M. M. 2024. The Obstacle Course of dmi Performance: Two Case Studies with T-stick and Karlax. In

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press

Proceedings of the International Conference on New Interfaces for Musical Expression.

Morrison, L. 2024. Timbre Space: On the Flat History of a Multidimensional Metaphor. *Music & Science*, **7**: 20592043241268720.

Nelson, P. 1997. The UPIC System as an Instrument of Learning. *Organised Sound* **2**(1): 35–42.

Nymoen, K., Godøy, R. I., Jensenius, A. R. and Torresen, J. 2013. Analyzing Correspondence between Sound Objects and Body Motion. *ACM Transactions on Applied Perception (TAP)* **10**(2): 1–22.

Pinheiro Cinelli, L., Araujo Marins, M., Barros da Silva, E. A. and Lima Netto, S. 2021. Variational autoencoder. In *Variational Methods for Machine Learning with Applications to Deep Networks*. Switzerland: Springer Nature, 111–49.

Roads, C. 2001. *Microsound*. Cambridge, MA: MIT Press.

Roads, C. 2003. The Perception of Microsound and its Musical Implications. *Annals of the New York Academy of Sciences*, **999**(1): 272–81.

Rochais, G. 2024. Randomness, Rule, Program: How the Issues of Generative art take Shape in Prealgorithmic Music. In C., Soddu and E., Colabella (eds.) *Proceedings of th XXVIIth Generative Art Conference – GA2024*, Venise, Italy, Vol. 1: 212–28.

Roma, G., Green, O. and Tremblay, P. A. 2019. Adaptive Mapping of Sound Collections for Data-Driven Musical Interfaces. In *NIME*, 313–18.

Sanfilippo, D. and Valle, A. 2013. Feedback Systems: An Analytical Framework. *Computer Music Journal* **37**(2): 12–27.

Schwarz, K. R. 1980. Steve Reich: Music as a Gradual Process: Part I. *Perspectives of New Music* **19**(1/2): 373–92.

Scordato, J. 2017. Composing with Iannix. In *Proceedings of the Fifth Conference on Computation, Communication, Aesthetics and X*, Vol. 5: 389.

Siedenburg, K., Fujinaga, I. and McAdams, S. 2016. A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research* **45**(1): 27–41.

Slater, D. 1998. Chaotic Sound Synthesis. *Computer Music Journal* **22**(2): 12.

Smalley, D. 1986. Spectro-Morphology and Structuring Processes. In *The Language of Electroacoustic Music*, Springer, 61–93.

Spangler, R. R. 1999. *Rule-Based Analysis and Generation of Music*. California Institute of Technology.

Steinbeck, P. 2018. George Lewis's voyager. In *The Routledge companion to Jazz studies*, Routledge, 261–70.

Tatar, K., Bisig, D. and Pasquier, P. 2021. Latent Timbre Synthesis: Audio-based Variational Auto-Encoders for Music Composition and Sound Design Applications. *Neural Computing and Applications* **33**: 67–84.

Tatar, K., Cotton, K. and Bisig, D. 2023. Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures. Proceedings of the Sound and Music Computing conference 2023.

Teboul, E. J. and Kitzmann, A. 2024. *Modular Synthesis: Patching Machines and People* (1st ed.). Focal Press.

Thiebaut, J. -B., Healey, P. G. and Bryan-Kinns, N. 2008. Drawing Electroacoustic Music. In *ICMC*. Citeseer.

Tomczak, J. M. 2022. *Deep Generative Modeling*. Springer International Publishing, Cham.

Tremblay, P. A., Green, O., Roma, G., Bradbury, J., Moore, T., Hart, J. and Harker, A. 2022. *Fluid corpus manipulation toolbox*.

Van Der Merwe, A. and Schulze, W. 2010. Music Generation with Markov Models. *IEEE multimedia* **18**(3): 78–85.

Vickery, L. 2014. The Limitations of Representing Sound and Notation on Screen. *Organised Sound* **19**(3): 215–27.

Wessel, D. and Wright, M. 2002. Problems and Prospects for Intimate Musical Control of computers. *Computer Music Journal* **26**(3): 11–22.

Wessel, D. L. 1979. Timbre Space as a Musical Control Structure. *Computer Music Journal* **3**(2): 45.

Wishart, T. 1994. *Audible Design*. York (United Kingdom): Orpheus the Pantomime.

65daysofstatic. *Wreckage systems*. https://www.youtube.com/65propaganda/live.

https://doi.org/10.1017/S1355771825100915 Published online by Cambridge University Press