# Deckblatt einer

# wissenschaftlichen Bachelorarbeit

**Familienname:**

**Vorname:**

**Matrikelnummer:**

**Studienkennzahl:**

**Thema der Arbeit:**

**Angefertigt in der Lehrveranstaltung:**

Name der Lehrveranstaltung

**Vorgelegt am:**

Datum

**Beurteilt durch:**

Leiter_in der Lehrveranstaltung

| Familienname, Vorname | Matrikelnummer |
|---|---|
| Lekavski, Dominik | 11928650 |

# Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich die darin enthaltenen Bestimmungen eingehalten habe. Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, den  20.06.2023
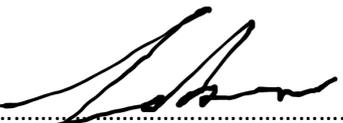
.......................................................................
Unterschrift der Verfasserin/des Verfassers

# Noisebox: Augmenting the Expressivity of Matter through Audio Mosaicing

Dominik Lekavski
Matriculation Number: 11928650

BACHELOR THESIS

Institute of Electronic Music and Acoustics
University of Music and Performing Arts, Graz

Bachelor's Programme Computer Music and Sound Art
UV 033 104

Supervisor: Mag.art. Mag.phil. Mag.rer.nat. Daniel Mayer

JUNE 2023

# Abstract

In this thesis, the Noisebox, a custom-made acoustic instrument, is examined as a new interface for musical expression, with a particular emphasis on electroacoustic improvisation. The live performance system leverages audio mosaicing via the Fluid Corpus Manipulation (FluCoMa) extension in SuperCollider. Building upon the principles of Electroacoustic Percussion Boards (EPB), digital musical instruments (DMI), and material computation, the research explores the Noisebox from historical, conceptual, and technological standpoints. Through a practice-based methodology, this work encourages an iterative refinement process of both the Noisebox and its controlling algorithms. The core objective of the study is a multi-faceted exploration of this hyperinstrument, assessing its potential as an expressive and embodied interface for musical expression.

# Table of contents

# 1. Introduction

## 1.1 Motivations

> 'Physical information pervades the world and it is through its continuous production that matter may be said to express itself.' (DeLanda 2006).

The transition from acoustic to electronic, and ultimately to digital musical instruments (DMI), has fundamentally reshaped the musical landscape. This evolution represents more than a shift in technology; it signifies a shift in our conceptualization of timbre, interaction and musical expression itself.

Every musical instrument, by virtue of its unique characteristics, profoundly influences the music one can create. The sounds of a xylophone, for instance, differ considerably from those of a violin, and both deviate significantly from the boundless potential of a computer. However, it is not just about the sounds these instruments produce. Their physical attributes and affordances also contribute to specific playing styles and repertoires.

With the advent of the digital age, we've been released from these physical limitations. Designing digital musical instruments is no longer restricted by the constraints of acoustics and physical materials. Instead, it has evolved into a complex dance of sensors, algorithms, and computational power, all working together to create new idioms of expression, novel timbres and unique ways of interaction and performance (Fiebrink 2017).

However, while digital tools open up a new world of possibilities, they often struggle to replicate the inherent expressiveness of acoustic instruments. The expressiveness and tactile sensations that come from the interaction between a musician and an acoustic instrument are challenging to achieve in the digital domain. Indeed, much recent research is being drawn towards designing elaborate new interfaces for musical expression (NIME). The term originates from the community and yearly conference known as NIME (NIME 2023).

Haptic feedback is crucial for embodying musical intentions, and diverse musical expression is achieved through multimodal feedback between the performer and instrument. The exploration of instrument resistance, whether physical or virtual, can enhance performance, and haptic interaction can foster shared participation between the performer, instrument, and performance space (Hayes 2012).

Unfortunately, digital musical instruments often lack the resistance, physical properties, and forces of acoustic instruments, leading to a loss of expressive potential. What's needed are innovative interfaces that emphasize the reintroduction of tangible forces and offer an alternative or expansion to traditional strategies. The use of *machine learning* and *music information retrieval (MIR)* tools has proven to be a successful strategy.

In this project, the use of machine learning and music information retrieval (MIR) tools stems from the unique possibilities these technologies offer. We can harness specific acoustic properties and translate these into the digital realm, both mirroring and expanding upon the expressiveness of traditional instruments. By analysing and processing data from an acoustic instrument, machine learning algorithms can adapt and respond to changes in real time, creating a more fluid and interactive experience during live performance. They can also perform an algorithmic selection of sound based on similarity, leading to a more organic and coherent integration of diverse sonic elements. Focusing specifically on sample-based content, the concept of creating a sonic mosaic, where smaller samples combine to form a larger piece or musical gesture, is analogous to this process. Concatenative synthesis, specifically the sub-set technique of mosaicing, offers means to achieve this.

## 1.2 Objectives

Initially, this thesis was inspired by a custom-made acoustic instrument—the *Noisebox*—which is constructed from a wooden box with an array of playable objects affixed to its frame and a contact microphone that picks up the sound. Following the exploration of concatenative synthesis, it seemed like a logical

evolution to expand and enrich the possibilities of the Noisebox with more recent technology.

The goal of this research is to investigate this hybrid instrument from historical, conceptual, and technological perspectives, specifically within the context of electroacoustic music, and further explore and assess its potential as an expressive and embodied new interface for musical expression. Guided by principles of Electroacoustic Percussion Boards (EPB), digital musical instruments (DMI) and material computation, the possibilities for gestural articulation and interaction are discussed, as well as inherent errors and uncertainties of this setup. This analysis paves the way for the creation of a system specifically designed for live performance and improvisation.

## 1.3 Software

While incorporating this technology into open-source computer music coding environments has been technically demanding, current toolkits, such as the Fluid Corpus Manipulation project (FluCoMa) created by the University of Huddersfield's Department of Music and Music Technology, make their creative and personalized use more accessible. The project will utilize the FluCoMa extension package for SuperCollider as its critical technological infrastructure (McCartney 2002; Tremblay et al. 2022).

Just as mastering instruments demands specific physical skills and musical knowledge, the creation and use of digital musical instruments require an understanding of the technology. In Chapter 3, we'll delve into the underlying technology behind the live performance system. We'll demonstrate how to create sonic gestures from a database of sound snippets, effectively blending the expressiveness of the materiality of the instrument with the flexibility offered by digital manipulation.

## 2. From Micromontage to Concatenative Sound Synthesis

The realm of *concatenative synthesis*, an innovative technique of sound synthesis originating from the late 20th century, has made pivotal contributions to the evolution of electronic music and audio technology. In contrast to more traditional synthesis methods, such as subtractive synthesis, concatenative synthesis does not rely on oscillators or filters to produce sound. Literature even references it as an 'adaptive digital audio effect' (Verfaille and Arfib, 2001). Instead, it revolves around the idea of leveraging a database, or *corpus*, of sound *units*, which are then pieced together or 'concatenated' to generate new sonic sequences (Schwarz, 2007).

## 2.1 Origins of Concatenative Sound Synthesis (CSS)

The conceptual origins of concatenative synthesis, as described by Schwarz (2006), can be traced back to the development of *Musique Concrète* in the 1940s. Championed by composer Pierre Schaeffer at *Groupe de Recherche Musicale (GRM)*, *Musique Concrète* introduced the concept of using found sounds or recorded audio to create music. In the seminal work *'Traité des Objets Musicaux,'* Pierre Schaeffer introduced the concept of the *'sound object,'* a principle bearing similarities to the *'unit'* in concatenative synthesis. A sound object, as defined by Schaeffer, is a clearly 'delimited segment of a source recording', serving as the foundational unit for composition. This aligns with the *unit* in concatenative synthesis, which is an isolated sound segment that is used as the building block for synthesis. The unit's role in concatenative synthesis thus echoes the function of Schaeffer's sound object as a constituent element (Schwarz, 2006). Building upon this principle, it is worth noting that, in contrast to methods like granular synthesis, which rely on much smaller fragments, concatenative synthesis frequently employs larger sections of a sound in order to preserve the distinctive qualities of the original source material.

These developments laid the groundwork for composers like John Cage and Iannis Xenakis, who started to compose with the tiniest segments of sounds, a technique termed 'micromontage'. However, the specific term 'micromontage'

wasn't popularized until later, primarily through the work of composers and theorists like Curtis Roads. His work in the latter part of the 20th century helped define and explain the concepts of micromontage and microsound. The terms refer to the manipulation of sound on a microscopic time scale. The idea was to treat sound as a malleable material that could be sliced, rearranged, and sculpted to create entirely new auditory experiences (Roads 2001).

An early example is John Cage's composition *Williams Mix*. For this composition, Cage and his team meticulously arranged and spliced approximately 600 small pieces of magnetic audiotape, a labor-intensive process that spanned over nine months (Sturm, 2006). Following suit in the late 1950s, Iannis Xenakis utilized similar micromontage techniques, assembling hundreds of short magnetic tape segments to create *Concret PH* and *Analogique B* (Roads, 2001). In 1967, Bernard Parmegiani employed the same methodology to construct parts of *Capture Ephémère* (Sturm, 2006).

The advancement of digital technologies has allowed us to control tiny sounds, or microsounds, with greater precision. As these technologies improved and automated some parts of the process, granular synthesis has proven to be the most impactful (Roads 2001).

*Granular synthesis* is a technique in which audio is fragmented into small windowed pieces referred to as 'grains'. The term *'grains'*, in the context of sound synthesis, didn't come into wide usage until the advent of the technique, which was introduced and developed throughout the 1970s by pioneers, namely Iannis Xenakis and Curtis Roads. The practical implementation of granular synthesis in personal computer systems, however, wasn't fully realized until the late 1980s. (Roads 2001, p. 110) While these grains can vary in length, even down to individual cycles of a waveform, they typically range from 1 to 100 milliseconds (Roads 2001, p.87). Each grain can then be individually manipulated and recombined in numerous ways to create new sounds. The subsequent section will explore granular synthesis in more depth, discussing its parameters, particularly in its role within the concatenative synthesis algorithm of the thesis.

To recap, the transition from micromontage to granular synthesis was driven by technological advancements that allowed for more precise, versatile, and automated sound manipulation. At the same time, the term 'grains' came to be adopted widely to reflect the new understanding of how these minute pieces of sound could be synthesized to create entirely new sonic phenomena. This shift represented a major development in our ability to manipulate sound on a microscopic level, leading to countless new possibilities and compositions. This established the foundational elements for the development of concatenative synthesis.

The early implementations of concatenative synthesis for industrial purposes appeared in *Concatenative Text-to-Speech (CTTS)* systems in the 1980s (Sturm 2006). These pioneering systems constructed realistic-sounding spoken language by stitching together pre-recorded speech segments. The method demonstrated significant potential for producing *high-quality synthetic speech*. The CTTS systems work by combining distinct units of phonemes, diphones, and syllables, among others. Sourced from an expansive *database* of recorded speech waveforms, these units are chosen for their optimal match to the utterance being synthesized, as noted by Schwarz in 2000. This unit selection algorithm, a method of database retrieval, exemplifies a *data-driven* approach to text-to-speech synthesis (Schwarz 2000). With the onset of the 21st century, inspired by advancements in speech synthesis, researchers began exploring the application of concatenative synthesis in music. Notably, Diemo Schwarz became a pivotal contributor in this emerging field.

## 2.2 Corpus-Based Concatenative Synthesis (CBCS)

The concept of *'corpus-based concatenative synthesis'* was first coined to distinguish its novel approach and specific application in music from its counterpart in speech synthesis (Schwarz 2007). It builds on the foundation of basic concatenative synthesis, but diverges by adopting a more *exploratory* approach to the selection and combination of sound units, instead of pursuing realistic modelling. The process begins with the creation of a rich, multifaceted *corpus* of *sound units*. This corpus is then analyzed and described in terms of various *sound features* or *audio descriptors* like *pitch, timbre, loudness, and duration*. These features are used to guide the *unit selection process*, *matching target sound properties* with those in the corpus. This facilitates the creation of new sound sequences that bear specific desired sonic characteristics (Schwarz 2007). An early example of corpus-based concatenative synthesis software is CataRT, developed by Schwarz from 2005 onwards at the IRCAM institute in Paris. See Schwarz (2006) for a comprehensive comparison of concatenative synthesis work in chronological order.

Figure 2.1 provides an overview of the 'data flow model of a concatenative synthesis system' (Schwarz, 2007). It includes three components: (a) 'general structure,' (b) 'analysis component,' and (c) 'synthesis component.' 'Rounded boxes in the figure represent data, rectangular boxes represent components, and arrows represent the flow of data' (Schwarz, 2007). The figure outlines the tasks that are executed in any concatenative synthesis system, although the details can vary (Schwarz, 2006).
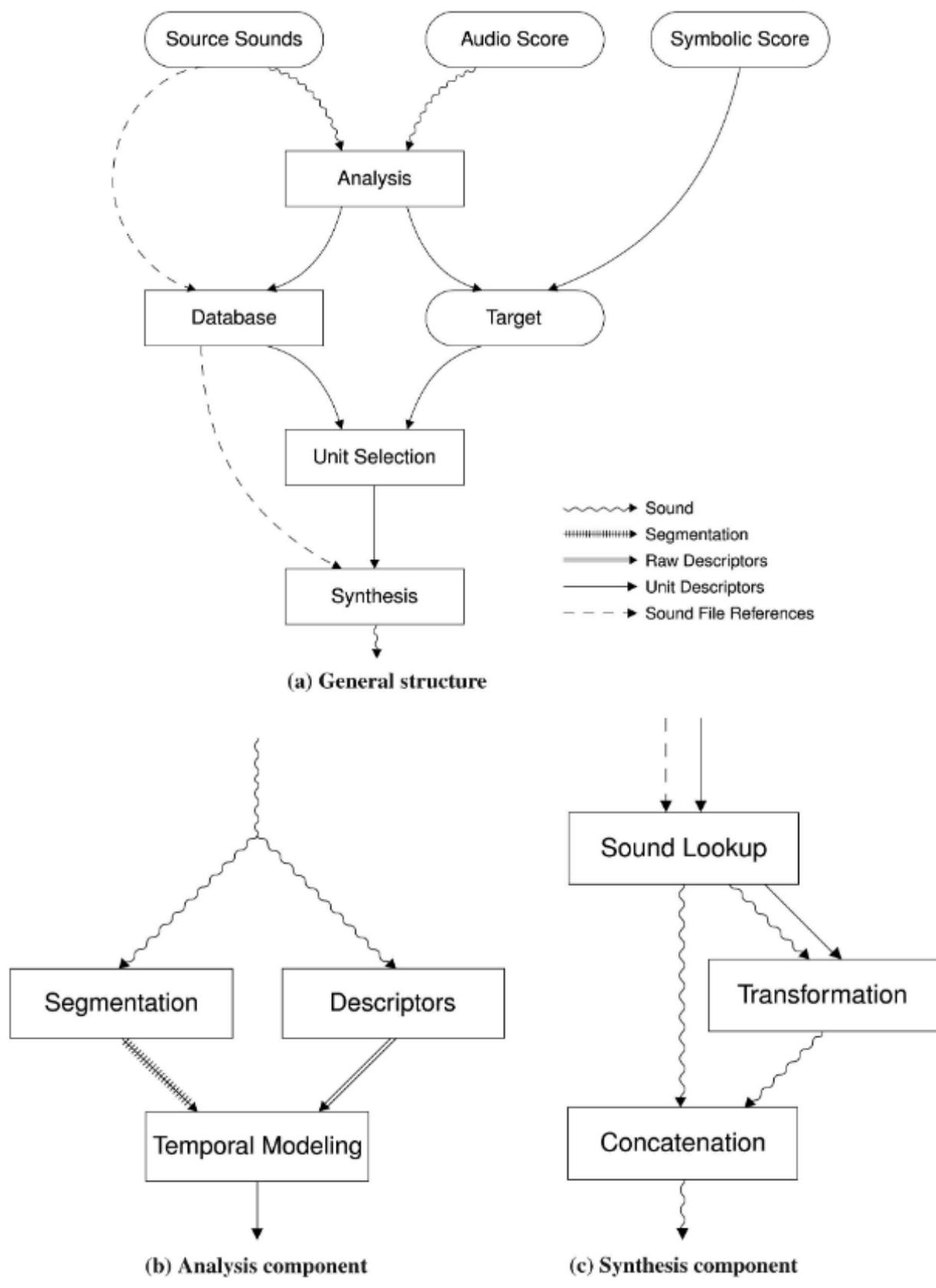
Figure 1: Data flow model of a concatenative synthesis system (Schwarz 2006, 6).

In the figure:

**(a) General structure**

- Source Sounds
- Audio Score
- Symbolic Score
- Analysis
- Database
- Target
- Unit Selection
- Synthesis

Legend:
- Sound
- Segmentation
- Raw Descriptors
- Unit Descriptors
- Sound File References

**(b) Analysis component**

- Segmentation
- Descriptors
- Temporal Modeling

**(c) Synthesis component**

- Sound Lookup
- Transformation
- Concatenation

The live *audio mosaicing* technique deployed in the performance system discussed in this thesis represents a sub-method of corpus-based concatenative synthesis. Here, the concatenation is determined through real-time descriptor analysis of the live audio input. This offers a very direct and interactive control of the spectrotemporal characteristics of the resulting sound and can be categorized as a type of resynthesis (Schwarz 2007). In the forthcoming chapter, we will delve into the intricacies of mosaicing, unpacking both the technical and creative aspects of this process. However, this chapter will not offer an in-depth explanation of the algorithms that fuel the process of audio mosaicing, as such a thorough exploration extends beyond the intended scope of the thesis. The aim will be to illustrate how these techniques are employed, specifically within the context of the live performance system. Thus, the focus will bridge the gap between theoretical knowledge and practical application, enhancing understanding of its implementation in the hybrid setup.

## 3. Overview of Audio Mosaicing

The live performance system integrates audio mosaicing by leveraging the *FluCoMa (Fluid Corpus Manipulation) extension* in *SuperCollider* (Tremblay et al. 2022; McCartney 2002). FluCoMa offers a wide range of tools for creative applications, including, but not limited to, *machine learning*, *music information retrieval* (MIR) and corpus-based concatenative synthesis. Through this chapter, we will clarify the practicalities of this extension in the process of audio mosaicing, focusing on the segmentation, database construction, analysis, selection, and normalization processes.

### 3.1 The Corpus and its Content

The corpus, a dataset of the broader database, represents the collection of audio materials that are sampled and concatenated in the process of mosaicing. In this system, the corpus contains pre-recorded audio files that are organized and segmented based on their distinct characteristics. During the development, I experimented with an eclectic mix of samples, ranging from textural field recordings, percussive sounds to samples of acoustic instruments altered

through prepared techniques. This wide-ranging selection ensured a rich palette of textures and timbres for experimentation with mosaicing. A variation in the length of these samples was tested, further contributing to the heterogeneity of the corpus. It proved beneficial to have a diverse range of sounds with distinct timbres. Additionally, introducing variations of the same sound sources with differing dynamics and slight shifts in tone enhanced the playability and expressivity of the corpus. The choice of audio samples represents a balance between the desire for varied and interesting sonic materials and the need for coherence within the live performance system. With the overarching improvisation in mind, the corpus is organized. The selection of sounds for the corpus, therefore, can be viewed as an integral part of the artistic process.

Another strategy involves using multiple mosaicing instances, each employing a separate corpus. This approach significantly enhances creative control over the system's output. By utilizing distinct corpora, each with its specific purpose, users can achieve greater flexibility and customization in generating desired outcomes. For instance, one corpus may primarily comprise short transient sounds, while another corpus may feature textural or tonal sounds. They can be combined, gradually blended over time, or used separately, depending on the desired behaviour and artistic intent. Thus, a higher degree of expressivity and flexibility is possible.

The audio files are in mono Waveform Audio File Format (.wav), given their compatibility with the analysis process. To facilitate the analysis process, described in section 3.2, and to streamline retrieval, all the audio files for one corpus are arranged into a single buffer. Each unit, determined by the onset of a distinct sonic event, is assigned an index, which will later be associated with its corresponding analysis data. This information is stored in a dataset, ensuring lookup of the units during the mosaicing process.

## 3.2 Analysis: Audio Feature Extraction

At the core of the mosaicing process is *audio feature extraction*, a practice rooted in *Music Information Retrieval (MIR)*. This procedure transforms audio data into data values (or streams) that articulate specific information about the signal. This includes the detection of possible onsets and a comparison of the timbre of the input signal, or target, to the timbre of segments within the corpus.

Mosaicing would not be possible without this *numerical representation* of sound. Like constructing a mosaic without distinguishing between the tiles' colours, sizes, or shapes would lead to chaos, an auditory mosaic without numerical representation would be similarly unfeasible. In audio mosaicing, the numerical representation of sound allows us to differentiate between various audio elements based on their inherent features.

Both the audio input from the Noisebox (target) and the audio material of the corpus require analysis via the same procedures, specifically *onset slicing* and *Mel-frequency cepstral coefficient (MFCC)* analysis. The primary distinction lies in the timing of these analyses; while the live audio input necessitates *real-time analysis*, the corpus analysis is an *offline process*.

### 3.2.1 Onset Slicing

The segmentation of the corpus is accomplished through a process known as onset slicing. This process results in separate units for analysis and concatenation. The corresponding object, FluidBufOnsetSlice, in FluCoMa implements several curated metrics, each possessing unique characteristics, to determine slice points in a sound based on the difference between spectral frames. It generates a time-series representing changes from one moment to the next. If a change exceeds the threshold parameter, a slice point corresponding to that event is produced (Tremblay et al. 2022). The *Rectified Complex Phase Deviation* metric was selected, given its optimal performance for sounds that fluctuate between stable and unstable sonic states. An illustration of such a sound

would be one that unpredictably oscillates between tonal and noisy states. (Tremblay et al. 2022).

As mentioned before, each slice in the corpus is assigned an index during the onset slicing process, which is subsequently linked and saved with the corresponding descriptor data in a dataset. The process applied to the live audio input from the Noisebox largely mirrors that of the corpus. However, because of the real-time nature of the live audio input, we employ the real-time version of the FluCoMa object, known as *FluidOnsetSlice*. This object generates a trigger when it detects an onset. The generated trigger message is pivotal for executing the lookup of the corpus unit that most closely aligns with the descriptor data from our live input.

## 3.2.2: Audio Descriptor: Mel-Frequency Cepstral Coefficient

While there is no universally agreed-upon definition of timbre, there are certain central aspects that help describe it. Indeed, timbre extends beyond the basic auditory attributes of sound events, such as pitch, loudness, duration, and spatial position. It is seen in the realm of psychoacoustic research as a *multidimensional* entity, embodied in the concept of *'timbre spaces.'* In order to make sense of these spaces, it is important to derive reliable acoustical parameters from the audio signal that can serve as potential physical correlates, also known as *audio descriptors*. These descriptors can help map out the multiple dimensions of timbre, contributing to our understanding and analysis of it (Peeters et al. 2011).

*MFCC*, an acronym for *Mel-Frequency Cepstral Coefficient* (pronounced 'kepstral'), is an analysis technique commonly used for characterizing and comparing timbre. It condenses the entire spectrum into fewer coefficients, which collectively represent the broad shape of the spectrum (Tremblay et al. 2022). The quantity of coefficients, identified as *'numCoeffs'* in the related FluCoMa objects *FluidBufMFCC* and its real-time counterpart *FluidMFCC*, usually varies between 13 and 20. In my experience, using 20 coefficients proved to be the most effective for the mosaicing system. Both objects need to use the same amount of coefficients.

I have chosen to use MFCC analysis, given its proven effectiveness in similar projects and prior research. As one example, a study by Terasawa, Berger, & Makino (2012) highlighted its reliability as a perceptual metric for timbre. Consequently, it offers a suitable method for comparing the timbral similarity between two sound sources. In the practical application within my system, this method successfully retrieves sounds from the corpus that bear perceptual similarities to the target.

## 3.2.3: Statistical Analysis

In most cases, a buffer contains time-series data represented as 'frames.' These could be audio samples, such as in an audio buffer for playback; alternatively, they could be a sequence of audio descriptors generated through an analysis process, as seen in FluidBufMFCC buffer-based analyses. If diverse time-series (for instance, sound slices) vary in their frame count, making comparisons might become challenging. Utilizing the *statistical summary* offered by the object *BufStats,* can aid in comparing these time series (Tremblay et al. 2022). For this purpose, I selected the *'mean'* option. The mean represents the *average* value of the data. It is calculated by summing all values and dividing by their total count. The mean can be a useful measure to show the *central tendency* within a set of descriptor values (Tremblay et al. 2022).

## 2.3.4: Scaling: Normalization

*Scaling* plays an important role in processing the *multi-dimensional data* derived from the MFCC analysis, where each *dimension* refers to a *coefficient* from this analysis. Particularly, for the optimal functioning of the nearest neighbour search (NN) algorithm, which is performed by the KDTree object that the mosaicing system employs for the lookup of units. KDTree, as used in the FluCoMa toolkit, relies on distance metrics (*Euclidean distance*) to find the nearest neighbours in the feature space. In this context, the 'distance' between points in this multidimensional space is often considered a measure of their 'similarity', meaning that points closer in the feature space are likely more similar. Normalization ensures that all dimensions of the feature space contribute equally

to the distance computation. Without this step, dimensions with larger ranges might overshadow those with smaller ranges, leading to biased results (Tremblay et al. 2022).

In the case of our mosaicing system, we are considering two distinct sources: the corpus, which contains a variety of sound sources, and the Noisebox, which provides a more specific audio input. Given the variety in the corpus and the unique attributes of the Noisebox, it is highly likely that the ranges of MFCC data can be quite diverse.

By using the *Normalize* object in FluCoMa, we transform each dimension of our dataset into a 'normal' range between 0 and 1. This ensures that the distance calculation within the KDTree algorithm is based on a balanced consideration of all dimensions, and not disproportionately influenced by any single dimension due to its range. In other words, normalization enables a more objective measure of 'similarity' between the live input from the Noisebox and the diverse sounds in the corpus (Tremblay et al. 2022).

Therefore, the role of scaling is crucial for our mosaicing system's successful operation. This will ensure that the multidimensional data, irrespective of its source, is processed in a way that leads to unbiased and meaningful results in the lookup of the units. This approach further ensures an expanded range of playable units, thereby enhancing the system's expressivity and potential for sonic diversity.

## 2.3.6: Lookup: K-Dimensional Tree & Nearest Neighbour Search

The lookup process refers to the selection of audio units from the corpus database based on their similarity to the target. This process uses the data generated in the earlier steps to select units from the corpus. As discussed in the previous section, lookup is facilitated by *a k-dimensional tree (KDTree)* using the *nearest neighbour search (NN)* algorithm. This method allows for efficient nearest neighbour searches of multi-dimensional data, enabling a quick selection of relevant audio units based on their distance in the feature space. In order to

perform queries on a *KDTree*, it is necessary to first fit the tree to the corpus dataset that contains the units it will later search for (Tremblay et al. 2022).

A trigger message generated by the *FluidOnsetSlice* object starts each individual unit-lookup. This trigger message starts a nearest neighbour search in the k-dimensional tree. Meaning at the moment of an onset of the live audio input, the real-time descriptor data is sampled and compared to those of the units in the corpus. The relevant unit will be queried and played back in real-time using granular synthesis as the playback method.

## 3.3 Granular Synthesis

In this system, the *granular synthesis* technique is implemented using the *TGrains3 buffer granulation* object from SuperCollider. This particular choice is informed by TGrains3's potential for the definition of a custom envelope and manipulation of the rise/decay shape of the grains. This feature gives the performer more creative flexibility in altering the sound produced by the mosaicing system.

The granulator is triggered by the onsets from the Noisebox. This trigger prompts the retrieval of a grain from the corpus buffer, which is essentially the selected unit from the corpus as identified by the previously discussed lookup process. The integration of granular synthesis with the lookup process thus results in the production of grains that mirror the attributes of the real-time input.

To further enhance performance capabilities, two switchable playback modes have been incorporated into the granulator. The first mode is onset-triggered, in which grains are played exclusively when there is an onset from the live audio input. This mode allows a granular response that is closer to the input's spectrotemporal characteristics. While the grain's duration aligns with that of the analysed unit, I have introduced a feature permitting the specification of the duration percentage to be played back. This provides further control options, enhancing the capacity to fine-tune the sound output.

The second mode triggers grains at a constant, definable rate. This mode allows for continuous sound production, irrespective of the live input. However, every time an onset is detected in the live audio input, the grain selection jumps to the respective unit position in the buffer, corresponding to the characteristics of the detected onset. This makes possible a 'freezing' of the sound, as the same grain can be replayed and manipulated continuously, facilitating creative exploration and control over the produced sound. To prevent the output sound from being too static, I have implemented a random deviation in the buffer position and the possibility to randomly pan each grain in the stereo field.

The combination of these modes and the flexibility of the TGrains3 object in SuperCollider offers a wide range of expressive possibilities for improvisation. The performer can additionally shape and control the produced sound, resulting in a more dynamic performance.

# 4. The Instrument 'Noisebox"

## 4.1 Overview

The sound source and instrument used in the system is a self-constructed *'Noisebox.'* Attached to its wooden housing are several objects, including three springs, two metal rods, and two guitar strings. These are then excited by various tools such as friction mallets, a comb, a violin bow, a metal sponge, and even the hand. A miniature condenser vibration pickup captures the sound.
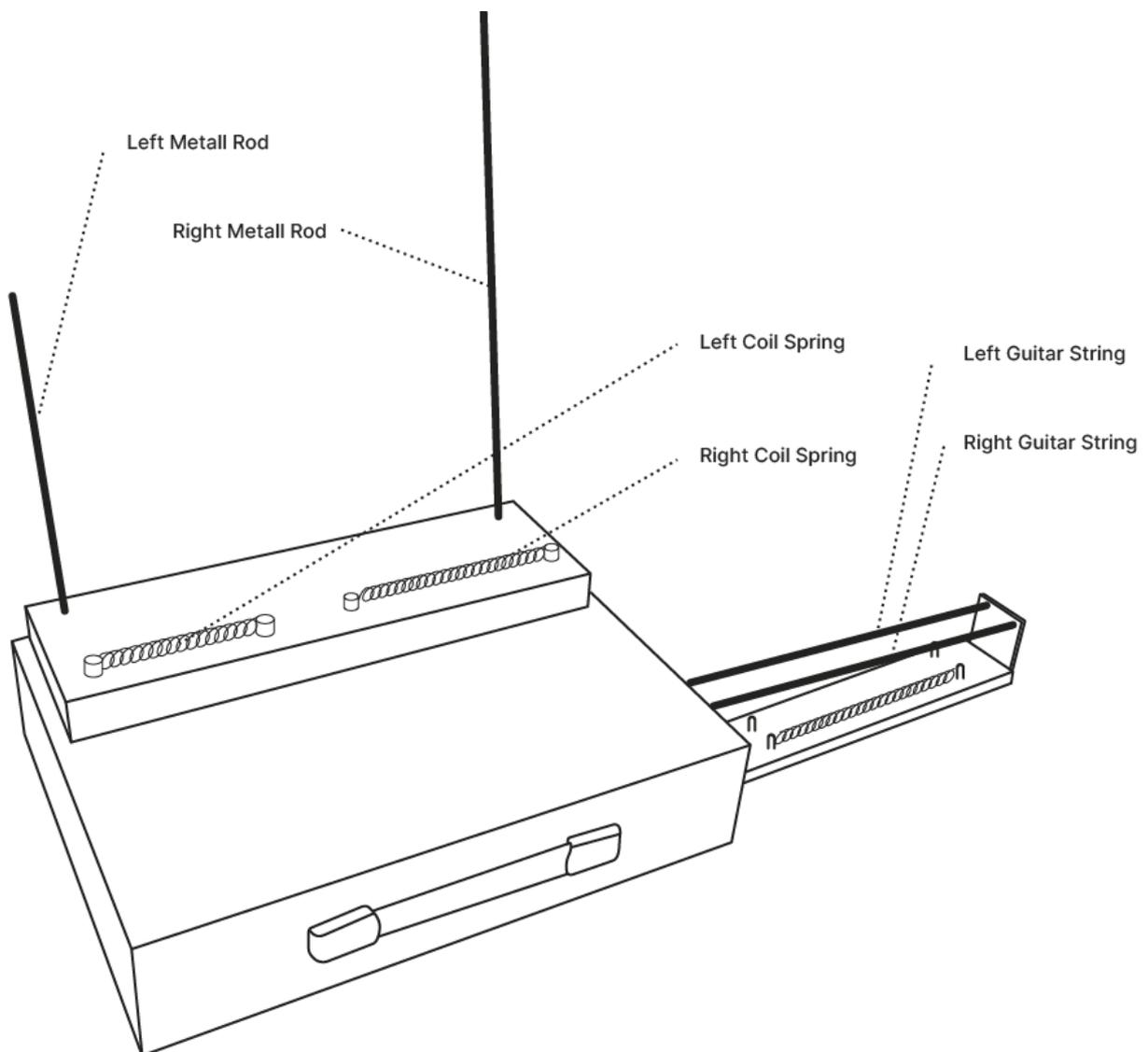
## 4.2 Specifications

The instrument's main body is a wooden housing:

- o Dimensions of the wooden housing:
  - ▪ Length: 36 cm
  - ▪ Width: 27 cm
  - ▪ Height: 8 cm

Assembly for springs and metal rods:

- o Dimensions of the assembly:
  - ▪ Length: 33.5 cm
  - ▪ Width: 10 cm
  - ▪ Height: 3 cm

Coil springs clamped on two metal screws each:

- o The left spring is stretched to a length of 10 cm, with 20 windings.
- o The right spring is stretched to a length of 14 cm, with 48 windings.
- o The third spring is attached to the neck, stretched to 16 cm with 44 windings.

Two guitar strings are strung on the right side of the box:

- o Both are tuned by a mechanism inside of the box.
- o Each string should have a length of 36 cm.

## 4.3 Signal Flow



loudspeaker array
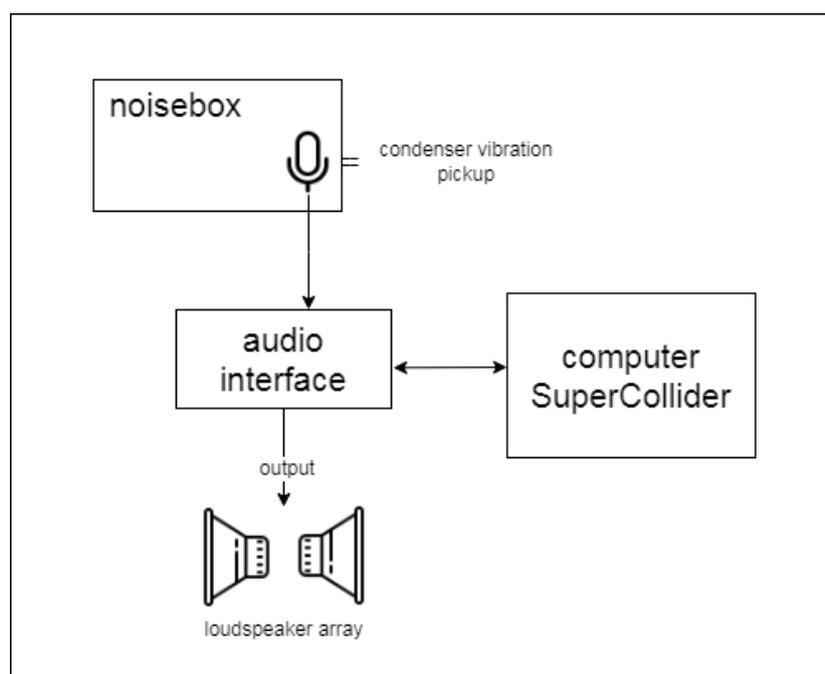
## 4.4 Origins of the Noisebox

The notion of a noise box is far from novel. The underpinning idea of objects generating noise, encapsulating considerable artistic potential, can be traced back to the pioneering experiments of the *Italian Futurists* in the early 20th century. In *'L'Arte dei Rumori'* (The Art of Noises), Luigi Russolo proposed an innovative musical theory embracing the noise of the industrial age as an under-explored artistic resource. He advocated for expanding the sonic palette of music, challenging traditional musical aesthetics, and classifying noise into six categories to be manipulated through specially designed noise-generating acoustic or mechanical instruments, or *'intonarumori'*. Drawing from his Futurist background, he urged music to evolve and reflect the dynamic, technological realities of contemporary society. His manifesto reshaped the understanding of music, promoting the integration of all sounds, including urban and industrial noises, into musical composition (Russolo, 1986).
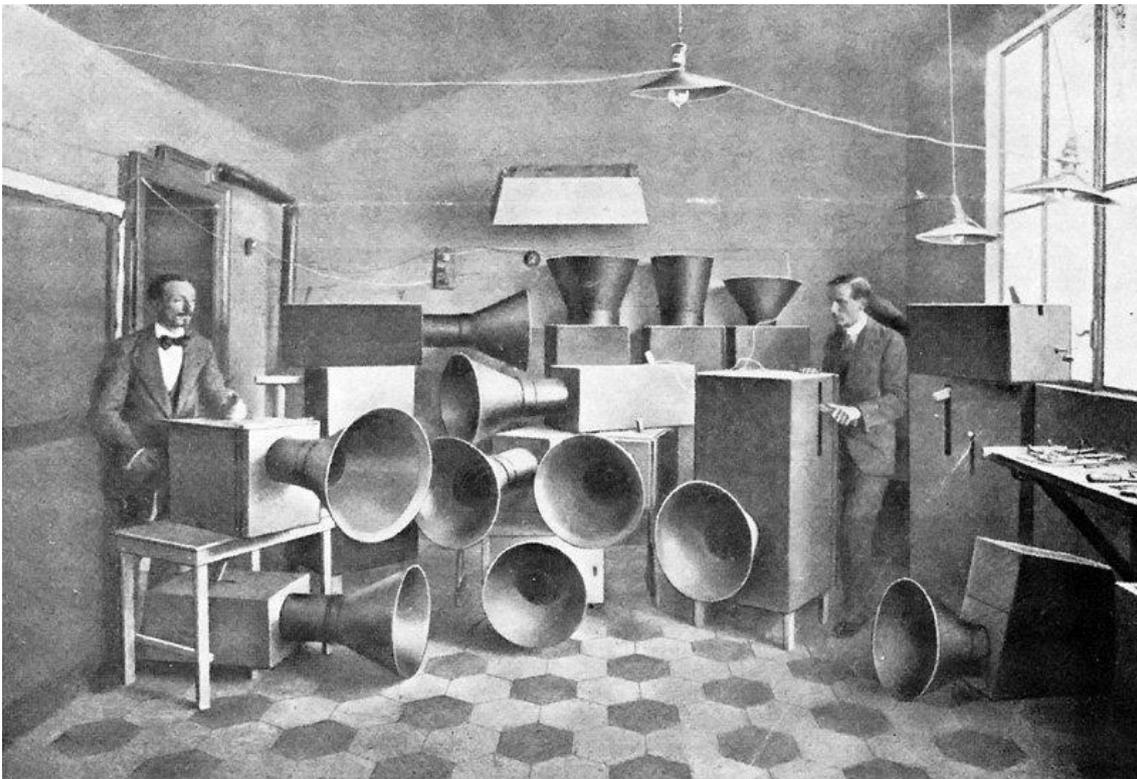


Figure *2:* Luigi Russolo (left) and his assistant Ugo Piatti with their Intonarumori.

The futurists' works served as a watershed moment in music history, signifying a shift in compositional possibilities and practice. They adopted noises as compositional timbres, thus liberating music from the traditional confinement of pitch-based tonality. This approach paved the way for an unlimited palette of sonic colours (Russolo, 1986).

Numerous composers, influenced by the Futurist movement, incorporated the sounds of industrialization into their orchestral works. A prominent example is George Antheil's *'Ballet Mecanique'* (1924), known for its unique use of an airplane engine. Similarly, Mosolov's *'Iron Foundry'* (1928) is recognized for its consistent use of a resonating thunder sheet, while Varèse's *'Ionisation'* (1929-1931) stands out for its inclusion of sirens, anvils, and a sound referred to as the 'lion's roar' (Burtner 2005).

Additionally, traditional instruments were used in non-standard ways, exemplifying the practice of what is referred to as *extended instrumental techniques*. For the composers of the European and American avant-garde during the 20th century, the expanding use of these techniques became a significant trend, as they continuously explored and incorporated new sound possibilities into their compositions. John Cage's *'prepared piano'* works serve as emblematic examples of this trend (Burtner 2005).

Transitioning to the 1970s, artists like Keith Rowe and Hugh Davies, prominent figures in the free improvisation scene, started employing everyday objects as musical instruments. This approach, inherent to their improvisational performances, effectively broadened the understanding of what could be considered a musical instrument and called into question traditional performance norms.

Presently, the Noisebox, while building upon the lineage of these historical progressions, aligns most closely with the electroacoustic percussion boards (EPBs), a term invented by Tom Nunn, a lesser-known figure from the same period. The term EPBs refers to his inventions, which are sheets of plywood cut into particular shapes, with various hardware devices attached. These sculptural

instruments can be played in various ways, such as striking, scraping, and plucking, with their sounds amplified by a contact microphone (Nunn 1988). Nunn devoted much of his life to the creation of these Electroacoustic Percussion Boards (EPBs), developing hundreds of unique instruments (Tewari 2017).
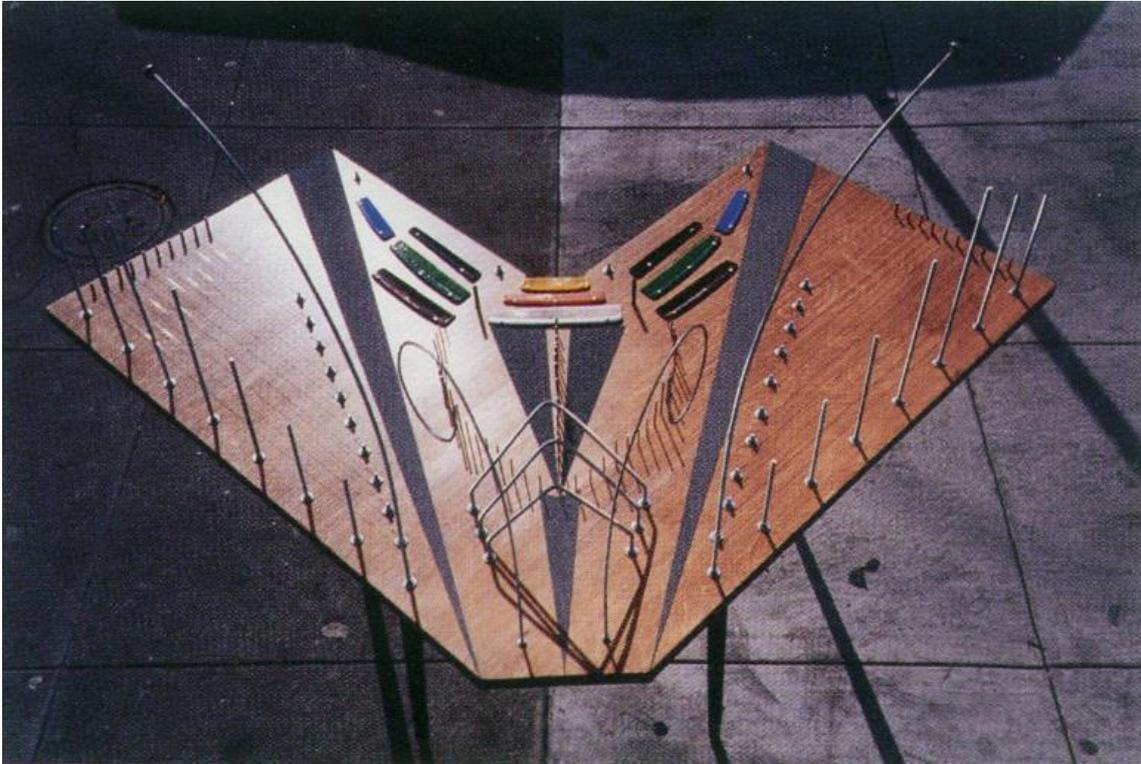


Figure 3: Tom Nunn, Earwarg II, plywood, steel, plastic, rubber, 48 x 67 x 55 in, 1987. (Photo: James R. Russell) (Nunn 1988).

## 4.5 Playing Techniques

In his work from 1988, Nunn describes the multifaceted techniques available for playing Electroacoustic Percussion Boards (EPBs). As part of my personal exploration, I have experimented with several of these techniques while playing the Noisebox and am committed to delving into additional methods that Nunn outlines in the future. According to him, a technique comprises three integral parts: 'the implement (the object held in the hand, or the hand itself), the action (striking, plucking, scraping, etc.) and the device (the thing attached to the board-combs, nails, etc.)'.

He outlines a variety of implements used in this context, including different combs, metal and wood sticks, small rubber mallets, threaded rods, guitar picks, homemade bows, and even the players' own fingers and fingernails.

Nunn categorizes over 50 foundational techniques, including 'mallets striking threaded rods, sticks scraping combs, fingernails plucking nails, bows bowing threaded rods, sticks striking strings, combs scraping surfaces, etc.' He underscores the significant variety found within a single technique, elaborating on this point with the example of sticks scraping combs:

> 'where the point of the stick touches the teeth of the comb (from the base to the tip) determines pitch; the pressure of the stick against the comb affects the dynamics, or loudness; how fast the scrape is made along the comb creates another pitch; the angle of the stick as well as its length, diameter and material affects timbre; and so forth.' (Nunn 1988).

Therefore, he concludes EPBs can yield to an impressive range of sound nuances and expressions.

According to Nunn, another important attribute of these devices is their inherent capacity for improvisation. The EPBs are 'specifically designed for improvisation' and they incorporate elements of 'non-linearity, randomness, ambiguity and unpredictability.'

The key ideas that Nunn embodies here align with what inspired me to explore the concept of the 'Noisebox'. But as someone raised in the digital age, I was naturally drawn to combining the advantages of the physical and digital worlds, which culminated in the idea of combining the Noisebox with corpus-based concatenative synthesis.

## 5. Hyperinstruments, Gesture and Interaction

In 2010, Tremblay and Schwarz explored the concept of utilizing a physical instrument, specifically an electric bass, as a 'subtle, expressive, and intuitive interface' for corpus-based concatenative synthesis or, in this case, *live audio mosaicing*. According to the authors, this was the first attempt at implementing this approach within a performance-oriented new interface, with the primary focus on expressivity. They proposed a method to directly map numerous descriptors of the audio stream, thereby enabling interaction with a corpus. This was achieved using a modified version of *CataRT* (Tremblay & Schwarz 2010).

Such instruments can be categorized under the umbrella term *digital musical instrument (DMI)*. According to Miranda and Wanderley (2006), DMIs are systems that create 'computer-generated sound' and consist of two crucial elements: a control surface and a sound generation unit. The control surface, also known as a gestural or performance controller, acts as an interface for the musician's tangible interaction with the instrument.

The authors describe the sound generation unit as a synthesis algorithm with its own parameters, which essentially creates the sound. The linkage between the output of the control surface and the input controls of the sound generation unit is facilitated by a 'mapping layer', a concept that refers to the strategies used for this connection (Miranda and Wanderley, 2006).

In the above application, the mapped 'parameters' are identified as audio descriptors, which direct the selection of sound units in the corpus. The sound generation unit is embodied by the electric bass. Miranda and Wanderley (2006) further categorize these DMIs as *'augmented'* or *'hyperinstruments'* based on their features. Interactions with these sophisticated instruments are guided by recorded actions, often termed as 'gestures.' Within the music domain, 'gesture' can imply various meanings. Here, we consider two: the 'musical gesture,' referring to the imaginary movements induced by sound forms, and the 'instrumental gesture,' which pertains to the performer's movements (Miranda and Wanderley, 2006).

The *'instrumental gesture'* is captured via a method known as *'indirect acquisition.'* This term, as defined by Miranda and Wanderley (2006), describes the process of obtaining information about performer actions from the evolution of the sound's structural properties (musical gesture) produced by an instrument. In this context, a microphone is employed as the sensor to capture this data.

The classification and definitions in the context of the Noisebox not only provide a clear placement for this device within the broad spectrum of Digital Musical Instruments (DMIs), but they also establish a theoretical foundation for subsequent exploration and future improvements. In addition, such understanding helps to identify the unique features and performance capabilities inherent in such a specific *hyperinstrument*.

The notion of '*material computation*', as articulated by Navab, Van Nort, and Wei (2014), provides a revealing lens through which to view this matter.

# 6. Material Computation & Expressivity

Navab, Van Nort, and Wei (2014) introduce a concept to designing musical instruments that mirrors the approach I have chosen to design the Noisebox. The authors argue that purely digital instruments are often limited by their interfaces. Instead, they propose a *'material computation'* design approach that perceives physical matter as a means to encode gestures. To further illustrate their point, Navab et al. (2014) elaborate:

> 'By this we do not mean simply the physical properties of sensing technologies, but the encoded gesturally and computationally modulatable potential of physical matter itself - the spatial and temporal encoding of gestural potential - and how this is coupled with environment, human interaction and sonic output in continuous and connected fashion."

In this framework proposed by Navab et al. (2014), the materiality of the instrument is not simply considered but is used as a *computational substrate*. This plays a pivotal role in determining the instrument's sonic quality, control structure, and mapping design. Navab et al. (2014) argue that, when dealing with instruments that utilize contact microphones and physical engagement with objects, the texture and resonance of the material itself take on a significant role. Additionally, the physical and spatial structure of the object conditions the kinesthetic gestural interactions, making them an essential part of the equation.

They emphasize the integration of physical matter and sound synthesis/processing techniques to achieve a rich computational response to an array of non-schematized gestures. The authors also discuss the concept of audio mosaicing, which they suggest is not solely for imitating target gestures but also for preserving the continuous morphologies of the target phrase. Lastly, they advocate for the idea of matter as a computational substrate to bridge conventional dichotomies in design metaphors, such as performer/performed and digital/analog.

Navab et al. (2014) present a comprehensive viewpoint on audio mosaicing, underscoring its distinctive characteristics within the realm of Digital Musical Instruments (DMIs). They emphasize the advantages of leveraging the expressive capabilities of physical materials, which serve as the '*computational substrate"* in this setup.

## 7. Improvisation

The Noisebox, with its inherent complexity and unpredictability, creates a rich environment for improvisation. The performer, while understanding the underlying system and potential outcomes, is continually presented with elements of surprise.

Central to the performance is the idea of physical resistance and interaction with the material affordances of the Noisebox. Thus, when improvising with it, the overall form often assumes a secondary role. Instead, the focus gravitates towards gestural moments, dynamics, textures and the spectromorphological trajectories of these sounds. This approach resonates with Stockhausen's concept of '*moment form'*, which emphasizes the autonomy and individuality of moments over traditional formal development (Stockhausen 1963).

Performing with this hyperinstrument is about striking a balance—retaining control while embracing the unpredictable. This balance translates into a struggle of sorts, a tussle between the performer's intent and the system's response. Yet, it is within this struggle that the improviser finds space for creativity and expression. There is, however, a degree of control that the performer maintains over the outcome, including the ability to shape and guide the direction of the overall timbral qualities by interacting with the granulator and different corpora.

In sum, improvisation with the Noisebox is a dynamic process, one that thrives on unpredictability while valuing control. It explores the tension between the intended musical expression and the surprises that emerge during performance.

# 8. Conclusion

The primary aim of this research was to develop a comprehensive understanding of the Noisebox as a new interface for musical expression (NIME), particularly in the context of improvisation within electroacoustic music. The instrument was examined through historical, conceptual, and technological lenses, while also investigating related work in the field.

Using a practice-based methodology, the instrument and its governing algorithms underwent iterative refinement. This approach facilitated extensive experimentation with diverse corpus materials, algorithms, and playing techniques. Due to the inherent complexity and uncertainties of the machine learning and machine listening algorithms, there are often deviations from the expected responses, leading to surprising outputs. However, these unforeseen outcomes introduce an element of unpredictability into the improvisation, augmenting the interaction with the Noisebox.

At present, the instrument exists as a prototype, representing the initiation of a personal artistic practice. The research aimed to establish an understanding of technical considerations and their musical implications, with a focus on the live audio mosaicing system and the underlying algorithms.

It should shed light on the immense expressive potential that the materiality of such Electroacoustic Percussion Boards holds, which can be harnessed as a captivating computational substrate.

The findings and analysis presented offer valuable insights into the challenges and opportunities for further exploration in this field. While the application of the Noisebox in a compositional context was not explored, it holds significant promise and warrants separate research.

In conclusion, this research has opened up fresh avenues for my artistic expression and compositional activity, resulting in a deeper understanding of the possibilities offered by this setup.

# References

## Bibliography

Burtner, Matthew. 2005. "Making Noise: Extended Techniques after Experimentalism." NewMusicBox. Accessed June 3, 2023. https://newmusicusa.org/nmbx/making-noise-extended-techniques-after-experimentalism/.

DeLanda, Manuel. 2006. "Material Expressivity." Domus 893: 122-23. Accessed June 7, 2023. https://lebbeuswoods.wordpress.com/2009/01/05/manuel-delanda-matters-4/.

Fiebrink, Rebecca. 2017. "Machine Learning as Meta-Instrument: Human-Machine Partnerships Shaping Expressive Instrumental Creation." In Musical Instruments in the 21st Century, 137-151, edited by Bovermann, T., de Campo, A., Egermann, H., Hardjowirogo, SI., Weinzierl, S. Singapore: Springer. https://doi.org/10.1007/978-981-10-2951-6_10.

Hayes, Lauren. 2012. "Performing Articulation and Expression through a Haptic Interface." Proceedings of the International Computer Music Conference, 400–403.

McCartney, James. 2002. "Rethinking the Computer Music Language: SuperCollider." Computer Music Journal 26, no. 4: 61-68. doi: 10.1162/014892602320991383.

Miranda, Eduardo R., and Marcelo M. Wanderley. 2006. New Digital Instruments: Control and Interaction Beyond the Keyboard, 1-30. Middleton, WI: A-R Editions.

NIME. 2023. "New Interfaces for Musical Expression." Accessed June 20, 2023. https://www.nime.org/.

Navab, Navid, Doug Van Nort, and Sha Xin Wei. 2014. "A Material Computation Perspective on Audio Mosaicing and Gestural Conditioning." Proceedings of the International Conference on New Interfaces for Musical Expression, 387–390. https://doi.org/10.5281/zenodo.1178893

Nunn, Tom. 1988. "Electroacoustic Percussion Boards: Sculptured Musical Instruments for Improvisation." Leonardo 21, no. 3: 261-265.

Peeters, Geoffroy, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. 2011. "The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals." Journal of the Acoustical Society of America 130, no. 5: 2902-2916. DOI: 10.1121/1.3642604.

Roads, Curtis. 2001. Microsound. Cambridge, MA: MIT Press.

Russolo, Luigi. 1986. The Art of Noises. Translated by Barclay Brown. New York: Pendragon Press.

Schwarz, Diemo. 2007. "Corpus-Based Concatenative Synthesis." IEEE Signal Processing Magazine 24: 92-104.

Schwarz, Diemo. 2006. "Concatenative Sound Synthesis: The Early Years." Journal of New Music Research 35: 3-22.

Schwarz, Diemo. 2000. "A System for Data-Driven Concatenative Sound Synthesis." In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-00), December 7-9. Verona, Italy.

Stockhausen, Karlheinz. 1963. "Momentform: Neue Zusammenhänge zwischen Aufführungsdauer, Werkdauer und Moment." In Texte zur elektronischen und instrumentalen Musik. Aufsätze 1952-1962 zur Theorie des Komponierens, edited by D. Schnebel. Cologne: Texte 1, 189-210.

Sturm, Bob L. 2006. "Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Composition." Computer Music Journal 30, no. 4: 46-66.

Terasawa, Hiroko, Jonathan Berger, and Shoji Makino. 2012. "In Search of a Perceptual Metric for Timbre: Dissimilarity Judgments among Synthetic Sounds with MFCC-Derived Spectral Envelopes." Journal of the Audio Engineering Society 60, no. 9: 674-685.

Tewari, Sudhanshu Ganesh. 2017. "Confluence: Tom Nunn's Sculptured Musical Instruments For Improvisation." PhD diss., University of California, Santa Cruz.

Tremblay, Pierre Alexandre, O. Green, G. Roma, J. Bradbury, T. Moore, J. Hart, and A. Harker. 2022. "The Fluid Corpus Manipulation Toolbox (v.1)." Zenodo. https://doi.org/10.5281/zenodo.6834643.

Tremblay, Pierre Alexandre, and Diemo Schwarz. 2010. "Surfing the Waves: Live Audio Mosaicing of an Electric Bass Performance as a Corpus Browsing Interface." In New Interfaces for Musical Expression (NIME), 447-450. Sydney, Australia.

Verfaille, Vincent, and D. Arfib. 2001. "A-DAFx: Adaptive Digital Audio Effects." In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-01), 10-14. Limerick: University of Limerick.

## List of Figures

Figure 1. Data flow model of a concatenative synthesis system (Schwarz 2006, 6).

Figure 2. Luigi Russolo (left) and his assistant Ugo Piatti with their Intonarumori. "Intonarumori." ThereminVox, accessed June 17, 2023, https://www.thereminvox.com/stories/instruments/intonarumori/.

Figure 3. Tom Nunn, Earwarg II, plywood, steel, plastic, rubber, 48 x 67 x 55 in, 1987. Photo by James R. Russell. Nunn 1988.