


# DISCRIMINATING MUSIC SEQUENCES METHOD FOR MUSIC THERAPY - DIMUSE

Emil A. Canciu <sup>1</sup> , Florin Munteanu <sup>2</sup> , Valentin Muntean <sup>3</sup>  and Dorin-Mircea Popovici <sup>4,1</sup> \*

<sup>1</sup> Faculty of Electrical Engineering and Computer Science, Transilvania University of Brasov / Politehnicii 1, Brasov, 500024, Romania; e-mail@f-iesc@unitbv.ro

<sup>2</sup> Academy of Romanian Scientist, Brasov Organization / Brasov, Ilfov nr. 3, sector 5, Bucuresti, Romania; e-mail@secretariat@aosr.ro

<sup>3</sup> Faculty of Music, Transilvania University of Brasov / Andrei Saguna 2, Brasov, 500123, Romania; e-mail@f-muzica@unitbv.ro

<sup>4</sup> Faculty of Mathematics and Informatics, Ovidius University of Constanta / Mamaia Blvd. 124, Constanta, 900552, Romania; e-mail@mateinfo@univ-ovidius.ro

\* Correspondence: e-mail@emil.canciu@unitbv.ro

## Abstract

The purpose of this research was to investigate whether music empirically associated with therapeutic effects contains intrinsic informational structures that differentiate it from other sound sequences. Drawing on ontology, phenomenology, nonlinear dynamics, and complex systems theory, we hypothesize that therapeutic relevance may be linked to persistent structural patterns embedded in musical signals rather than to stylistic or genre-related attributes. The paper introduces the Discriminating Music Sequences (DiMuSe) method, an unsupervised, structure-oriented analytical framework designed to detect such patterns. The method applies 24 scalar evaluators derived from statistics, fractal geometry, nonlinear physics, and complex systems, transforming sound sequences into multidimensional vectors that characterize their global temporal organization. Principal Component Analysis (PCA) reduces this feature space to three dominant components (PC1–PC3), enabling visualization and comparison in a reduced informational space. Unsupervised k-Means clustering is subsequently applied in the PCA space to identify groups of structurally similar sound sequences, with cluster quality evaluated using Silhouette and Davies–Bouldin indices. Beyond clustering, DiMuSe implements ranking procedures based on relative positions in the PCA space, including distance to cluster centroids, inter-item proximity, and stability across clustering configurations, allowing melodies to be ordered according to their structural proximity to the therapeutic cluster. The method was first validated using synthetically generated nonlinear signals with known properties, confirming its capacity to discriminate structured time series. It was then applied to a dataset of 39 music and sound sequences spanning therapeutic, classical, folk, religious, vocal, natural, and noise categories. The results show that therapeutic music consistently forms a compact and well-separated cluster and ranks highly in structural proximity measures, suggesting shared informational characteristics. Notably, pink noise and ocean sounds also cluster near therapeutic music, aligning with independent evidence of their regulatory and relaxation effects. DiMuSe-derived rankings were consistent with two independent studies that identified the same musical pieces as highly therapeutic. The present research remains at a theoretical stage. The method has not yet been tested in clinical or experimental therapeutic settings and does not account for individual preference, cultural background, or personal music history, all of which strongly influence therapeutic outcomes. Consequently, DiMuSe does not claim to predict individual efficacy but rather to identify structural potential at the signal level. Future work will focus on clinical validation, integration of biometric feedback, and

Received:

Revised:

Accepted:

Published:

**Citation:** Canciu, E. A.; Munteanu, F.; Muntean, V.; Popovici, D. M. Dimuse method. *Appl. Sci.* **2025**, *1*, 0. <https://doi.org/>

**Copyright:** © 2026 by the authors. Submitted to *Appl. Sci.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the development of personalized extensions that combine intrinsic informational structure with listener-specific response data.

**Keywords:** DiMuSe method; Music Therapy; Therapeutic Music; Informational Structure; Nonlinear Time Series; Principal Component Analysis (PCA); k-Means Clustering; Structural Ranking of Music; Unsupervised Analysis of Sound Sequences

## 1. Introduction

Does music contain an informational structure that is interpreted uniquely by each individual? If such a structure exists, could it be analyzed from ontological perspectives to reveal natural or cultural characteristics? Understanding whether such informational structures exist is particularly relevant in the context of music therapy, where therapeutic effects are observed but not yet fully explained. Despite extensive empirical evidence supporting its benefits, the mechanisms underlying the therapeutic efficacy of music remained insufficiently characterized, particularly from a signal-structural and informational perspective. Most existing studies focus on subjective assessments, genre classifications, or emotional descriptors, leaving open the question of whether therapeutic music possesses objective, detectable structural properties.

Looking for empirical evidence on the therapeutic effect of music, researchers found positive outcomes for improving the well-being of patients suffering from dementia, autism spectrum disorder[1], stress and related illnesses[2]. Some reviews described the therapeutic effect ([3] or [4]). Interesting empirical evidence has been found regarding melodies that had increased therapeutic effect, compared to others. Among these findings, some melodies appear especially potent. One such example was the Mozart Sonata for Two Pianos in D major, K448[5].

To explain why certain melodies exert such effects, researchers have investigated the neural processes of music perception. Existing research explained this as a very complex process, that involves brain areas responsible for interpreting sound (superior olivary complex and inferior colliculus) and also motor and pre-motor regions[6][7][8]. When listening to music, the auditory cortex decodes the sound streams on a "segment-by-segment" and not on a "moment-by-moment" basis. There was also a possibility that the neural circuits for memory operate using firing-rate codes. The neurons in these neural circuits have the ability to synchronize with the fundamental frequency (carrier frequency) or the temporal structure (envelope) of a sound[9]. Pitch processing occurs in a hierarchical manner. The auditory stream is perceived as a melody over a period of time[10]. The following idea resulted, suggesting that therapeutic effects may be tied to specific informational codes embedded in musical structure:

*The frequency and temporal structure of the auditory information appears to contain a code that the auditory cortex has the ability to decode. The sound stream length is important for the brain to process its pitch and decode it as a melody.*

If perception depends on how the brain decodes information, then therapeutic outcomes may also depend on individual differences. Investigating further the mechanism by which music exerts therapeutic effect, research papers were identified on the topic of personalized music playlist. The personalized playlists were able to enhance the efficacy of music based interventions in clinical settings[6]. Listening to pleasurable music activated brain regions associated with pleasure and reward-seeking behaviour. As such, this paper acknowledged that the type of music used in therapy should align with the therapeutic context, specific method, disorder or disease, and, crucially, the musical preferences of

the listeners. Music preferences can be influenced by many factors like demographics[11], personality traits[12] or social influences.

In parallel with clinical and neurobiological research, recent advances in computational music analysis have increasingly focused on quantifying musical features in order to predict measurable physiological or behavioural outcomes [13]. Within the field of Music Information Retrieval (MIR), music is commonly represented through predefined acoustic descriptors capturing rhythm, spectral content, and temporal organization. Considering, machine-learning and neural-network-based approaches have been successfully employed to map musical features to outcomes such as movement synchronization, arousal, or stress reduction. In these supervised frameworks, therapeutic effect is operationally defined through measurable responses, and model performance is evaluated using predictive metrics such as explained variance. While such approaches demonstrate strong predictive power within controlled experimental settings, they are inherently outcome-driven and depend on large labelled datasets. Moreover, the learned feature representations are often difficult to interpret, limiting insight into the intrinsic properties of the music itself [14].

While personalized music playlist approaches have demonstrated benefits in enhancing therapeutic engagement and efficacy, they rely predominantly on subjective criteria such as self-reported preferences, demographic correlations, genre labels, or coarse emotional tags. These strategies provide limited insight into the intrinsic properties of the music itself and offer little capacity for systematic screening or comparison across musical pieces regarding their possible therapeutic impact. In particular, current personalization frameworks do not address whether certain musical sequences share objective structural characteristics that may contribute to their therapeutic potential, independent of individual taste or cultural context. This gap highlighted the need for analytical tools capable of identifying and comparing music based on its underlying informational structure, thereby complementing preference-driven approaches with an objective, structure-based perspective.

Recent advances in signal processing and non-linear analysis suggest that complex auditory stimuli can be characterized beyond conventional spectral descriptors, revealing persistent patterns related to structure, coherence, and organization. In this context, music may be approached not merely as a cultural artefact or aesthetic experience, but as a structured temporal signal whose internal organization can be quantitatively analysed. This perspective opens the possibility that music associated with therapeutic effects may share common informational features that are independent of style, instrumentation, or cultural origin.

Several studies have hinted at this direction by reporting correlations between therapeutic outcomes and specific signal characteristics, such as  $1/f$  noise behaviour [15], fractal properties [16], or reduced complexity relative to highly dynamic musical forms. However, a systematic and reproducible method for identifying and discriminating such structures across diverse musical corpora is still lacking. Addressing this gap requires a methodology capable of capturing global structural similarities rather than relying on local or surface-level features.

The present work adopted a complementary perspective, rather than predicting therapeutic outcomes directly, it investigated whether music empirically associated with therapeutic effects exhibits identifiable intrinsic informational structures, independent of listener-specific or experimental conditions. This distinction motivates the development of an unsupervised, structure-oriented analytical method. This work assumed that therapeutic music is not defined primarily by genre or preference, but by persistent informational structures detectable in non-linear signal space.

In this work, we introduce the Discriminating Music Sequences Method (DiMuSe), a signal-based analytical framework designed to detect and compare structural similarities between music sequences in a reduced, non-linear feature space. The method combines robust signal descriptors with dimensionality reduction techniques, enabling the identification of persistent clustering patterns across large and heterogeneous datasets. Importantly, DiMuSe does not presuppose any specific musical genre, therapeutic label, or emotional annotation, relying instead on the intrinsic informational structure of the audio signals.

The central working hypothesis of this study is that music commonly associated with therapeutic effects exhibits stable structural regularities that can be operationally detected using appropriate signal representations. From this perspective, therapeutic efficacy is not attributed solely to subjective preference or contextual factors, but is hypothesized to correlate with persistent informational configurations embedded in the temporal organization of the signal. Ontological considerations are therefore introduced not as metaphysical claims, but as an operational assumption: if form carries information, then recurring therapeutic effects should correspond to detectable and reproducible signal structures.

To evaluate this hypothesis, DiMuSe is applied to multiple datasets, including curated collections of music labelled as therapeutic, control datasets of non-therapeutic music, and reference signals such as natural sounds and synthetic noise. The method's robustness is assessed through repeated clustering analyses, external validation using independent studies, and comparative evaluation against known signal classes. The results demonstrate consistent separation between therapeutic and non-therapeutic music, as well as notable proximity between therapeutic music and natural sound patterns, suggesting the presence of shared structural characteristics.

The paper proposed the following research question:

*RQ: Can objective, quantifiable 'fingerprints' related to therapeutic efficacy be extracted from the informational structure of music?*

This study does not claim to establish causal mechanisms between signal structure and therapeutic outcomes. Rather, it provides a methodological contribution: a reproducible framework for identifying and comparing informational structures in music signals, offering new avenues for research in music therapy, auditory neuroscience, and complex systems analysis. By shifting the focus from genre-based classification to structural coherence, DiMuSe lays the groundwork for future investigations into how organized temporal patterns may support cognitive and physiological regulation.

## 2. Structured information in non-linear time series like music

This section introduces the concept of the music's *informational structure*, exploring it through the ontological perspective of information. It further examines the distinction between *linear and non-linear systems*, offering a rationale for selecting *Principal Component Analysis (PCA)* as the core research method.

### 2.1. Ontology, Phenomenology, and the Informational Structure of Music

The present study adopted an informational perspective on music signals, treating them as structured temporal processes rather than solely as aesthetic or cultural artefacts. In this framework, information was not reduced to semantic content or symbolic meaning, but was understood operationally as the organization and persistence of patterns within a signal across time and scale. Such an approach was well established in signal processing, complex systems theory, and non-linear dynamics, where structure is inferred from statistical regularities, correlations, and reduced-dimensional representations.

Music signals exhibit multiple layers of organization, ranging from local spectral features to long-range temporal correlations. While conventional analyses often emphasize

short-time descriptors, recent work [17] has shown that global structural properties—such as fractal scaling, coherence, and reduced complexity [18]—may play a significant role in how auditory stimuli are perceived and processed. These properties are particularly relevant in contexts where music is associated with regulation, relaxation, or therapeutic effects, suggesting that structure, rather than stylistic elements alone, may be a key factor.

From an informational standpoint, form is treated as a carrier of organization that can be detected independently of semantic interpretation. This does not imply any intrinsic therapeutic quality of specific musical forms, but rather motivates the hypothesis that signals associated with similar functional effects may share common structural configurations. The role of analysis, therefore, is not to assign meaning to music, but to identify reproducible patterns that distinguish one class of signals from another in a data-driven manner.

In this work, informational structure is operationalized through a set of signal descriptors designed to capture both variability and organization in the temporal domain. Dimensionality reduction techniques are employed to reveal dominant modes of variation and to facilitate comparison across heterogeneous datasets. Importantly, these methods are used as exploratory operators: they do not impose predefined categories, but allow clusters and proximities to emerge from the data itself.

Ontology refers to the study of existence and the underlying structure of reality. These philosophical perspectives are important for understanding music as information, not just as sound. This subsection investigates the nature of information from a structural phenomenological standpoint, focusing on how information manifests and is experienced within conscious awareness. It does not aim to provide a comprehensive philosophical analysis of music, but rather to establish a conceptual basis for treating music as an information-carrying structure that can be analysed using computational methods.

Ontological considerations are introduced here in a minimal and pragmatic sense. The assumption that form carries information is not treated as a metaphysical claim, but as a working hypothesis guiding the analysis. If therapeutic effects are recurrent and robust across individuals and contexts, it is reasonable to expect that the corresponding signals exhibit detectable regularities. The aim of this section is therefore to clarify the conceptual basis for treating music as an informational object amenable to structural analysis, providing the foundation for the DiMuSe methodology presented in the following sections.

## 2.2. Linear vs Non Linear system

Linear systems are fundamental in mathematics, engineering and physics because of their predictable behaviour and simplicity in analysing them. A linear system respects the principles of *superposition* and *homogeneity*[19]. This means that the outcome is proportional with the input causing factors. The direct implication is that very small influences can be neglected as input factors as their effect is very small.

From an application perspective, this means that the systems can be decomposed in parts, where each part can be investigated individually and summing the effect of the parts will result in the effect of the entire system.

The *Superposition principle* states that if a system responds to multiple inputs, the total response is equal to the sum of the individual responses to each input applied separately. Mathematically, if a system produces an output  $y_1(t)$  for an input  $x_1(t)$  and an output  $y_2(t)$  for an input  $x_2(t)$ , then the response to a combined input  $x_1(t) + x_2(t)$  is given by:

$$y(t) = y_1(t) + y_2(t) \quad (1)$$

This property ensures that linear systems exhibit *additivity*, meaning the system's response to multiple stimuli is simply the sum of the responses to each stimulus individually.

The *Principle of Homogeneity* states that if the input to a linear system is scaled by a constant factor, the output will be scaled by the same factor. If an input  $x(t)$  produces an output  $y(t)$ , then for a scaled input  $a \cdot x(t)$ , where  $a$  is a constant, the system will produce:

$$y'(t) = a \cdot y(t) \quad (2)$$

This principle ensures that linear systems exhibit *scalability*, meaning that multiplying the input by a factor results in a proportionally scaled output.

A system is considered linear if it satisfies both the superposition and homogeneity principles. Therefore, for any inputs  $x_1(t)$  and  $x_2(t)$ , and any scalar constants  $a$  and  $b$ , a linear system obeys:

$$y(t) = a \cdot y_1(t) + b \cdot y_2(t) \quad (3)$$

These properties are fundamental in analysing mechanical, predictable systems, such as electrical circuits, vibrations in structures, and signal processing. In contrast, *non-linear systems*, including biological organisms or chaotic phenomena, do not obey these principles, leading to complex and unpredictable behaviours. As a consequence, the response of a non-linear system can not be analysed as a sum of the responses of its components [20]. The relationship between its components together with the environment where the system is functioning has to be considered for such an analysis.

A non-linear system is generally considered a *black box*, where no clear linear relationship can be established between the system's input parameters and its output response. Unlike linear systems, which obey principles such as superposition and homogeneity, non-linear systems exhibit complex behaviours, including bifurcations, chaos, and emergent properties.

Since exact analytical solutions are often unavailable, *data-driven identification methods* are commonly used to model such systems. As music can be considered a non-linear system, such methods are directly applicable to analysing musical sequences. One effective approach would be the Principal Component Analysis (PCA) method for detecting patterns in the system's behaviour.

Non-linear system identification is typically performed through experimental analysis, where the system's response to a set of known inputs is recorded. A reliable methodology consists of the following steps:

- Introduce input signals with known properties into the system.
- Measure and record the system's output responses.
- Observe changes and correlations in system behaviour.

Machine learning techniques provide powerful tools for identifying non-linear system properties. Unsupervised neural networks offer a way to extract structure from unknown systems without requiring predefined output labels. These networks analyse patterns within the observed data and detect recurring features.

PCA is a mathematical technique used to reduce the dimensionality of data while retaining significant variance in the data set. It helps identifying the most relevant features of a non-linear system by:

- Transforming correlated variables into a set of linearly uncorrelated components.
- Extracting dominant modes of behavior from system response data.
- Improving computational efficiency by reducing complexity.

The mathematical representation of PCA is given by:

$$Z = XW \quad (4)$$

where  $X$  represents the original dataset,  $W$  is the transformation matrix composed of eigenvectors, and  $Z$  denotes the new transformed feature space.

To analyze unknown input data, the following approach could be taken:

1. Compare previously recorded responses from empirical tests.
2. Use PCA to extract dominant patterns and features.
3. Use unsupervised neural networks like unsupervised clustering techniques to identify clusters of similar items.
4. Compare results with existing models to estimate system characteristics.

Such a technique has been used in the current paper and has been described in the following subsections.

### 2.3. Existing methods to discriminate non-linear time series like music

At the time of writing the paper, extracting information from music tracks was an established field known as *Music Information Retrieval (MIR)*, which has been widely used for various applications, including *context-based music retrieval (CB-MIR)*, artist identification, genre classification, query by humming, emotion recognition, instrument recognition, and music annotation[21]. One major challenge in MIR has been the classification of musical genres, where defining genre boundaries was an open problem[22].

Closer to the current date advances have shifted MIR applications toward *machine learning (ML)* approaches, where predefined datasets are used for training models in automatic classification and retrieval tasks. One example would be the *MediaEval Database for Emotional Analysis of Music (DEAM)*, which has been utilized for training deep learning networks aimed at mapping emotional content in music[6].

Beyond conventional MIR techniques, *complex systems analysis* provided additional insights into the discrimination of nonlinear time series, such as musical structures. Advances in *chaos theory* have facilitated the identification of intricate patterns in dynamic systems, including applications in geophysical signal analysis. A similar approach has been successfully applied in geophysics, where non-linear signal processing tools — including PCA — were used to cluster precursor signals of earthquakes. This geophysical application directly inspired the present research: just as geophysical data contain hidden precursors of major events, therapeutic music may contain informational structures that precede and trigger therapeutic effects. Building on this precedent, our study adapts these nonlinear analysis methods to music, aiming to uncover intrinsic patterns that differentiate therapeutic melodies from other sequences. For instance, nonlinear signal processing tools have been employed to discriminate signals generated by *geo dynamic structures*, aiding in the identification of earthquake precursor signals[23]. One such approach involved *Principal Component Analysis (PCA)* for extracting dominant patterns in highly complex data.

Given the complexity of non-linear systems, where input-output relationships cannot be decomposed into simple linear components, traditional statistical techniques struggled to reveal underlying patterns. To address this challenge, this paper leveraged the *Principal Component Analysis (PCA)*, a technique capable of identifying dominant structural variations within nonlinear datasets. By reducing dimensionality while preserving relevant features, PCA provided a systematic approach for detecting hidden patterns within music sequences.

The paper proposed the *DiMuSe method*, based on PCA, for identifying informational structures in musical sequences. By applying non-linear time series analysis techniques to music data, the aim was to develop a framework capable of detecting intrinsic patterns that may differentiate melodies with therapeutic properties from other sequences.

The researchers considered the proposed DiMuSe method a different epistemological tool, explaining in the Table 1 the difference between the Neural-network approach (used in [13]) and the proposed method.

Dimension	Neural-network approach	DiMuSe approach
Philosophy	Predictive, data-driven, outcome-oriented	Structural, exploratory, information-oriented
Primary goal	Predict physiological/behavioural outcomes	Detect intrinsic informational structure
Feature handling	Learned latent features (black box)	Explicit scalar descriptors (transparent)
Data requirements	Large labelled datasets, supervised learning	Small datasets, unsupervised

**Table 1.** Neural-network approach vs. DiMuSe

### 3. DiMuSe Method

Building on the rationale from Section 2, where non-linear analysis and Principal Component Analysis (PCA) were identified as suitable tools for uncovering hidden informational structures in music, this section introduces the *Discriminating Music Sequences* method (DiMuSe). While existing approaches such as Music Information Retrieval (MIR) and machine learning-based genre detection have advanced the field, they often depend on predefined labels or handcrafted features. Such reliance makes it difficult to reveal the deeper, intrinsic patterns that may underlie therapeutic effects in music. To address this limitation, the DiMuSe method was developed. It applies scalar evaluators drawn from diverse scientific domains, offering an unsupervised approach to analyzing musical sequences and enabling clusters to emerge naturally from their informational structure rather than from externally imposed categories.

#### 3.1. Methodological Overview

The DiMuSe (Discriminating Music Sequences) method is designed as a data-driven framework for identifying structural similarities between music signals based on their informational organization. Rather than relying on genre labels, emotional annotations, or subjective assessments, the method operates directly on signal-derived descriptors, allowing patterns to emerge from the intrinsic properties of the data.

The methodological pipeline consists of three main stages: signal representation, dimensionality reduction, and structural comparison. Each stage is chosen to progressively reduce complexity while preserving information relevant to global organization and temporal structure.

In the first stage, audio signals are transformed into a set of numerical descriptors capturing both spectral and temporal characteristics. These descriptors are selected to balance sensitivity to variability with robustness to local fluctuations, enabling meaningful comparison across heterogeneous musical samples. All signals are processed using identical preprocessing steps to ensure consistency and reproducibility.

In the second stage, dimensionality reduction is applied to the resulting feature space in order to reveal dominant modes of variation. Principal Component Analysis (PCA) is employed as an exploratory operator, allowing correlated descriptors to be projected onto a reduced set of orthogonal components. Importantly, PCA is not used here for classification, but as a means to expose latent structural relationships and facilitate visualization and clustering in a lower-dimensional space. The third stage involves the identification and analysis of clustering patterns within the reduced space. By examining the relative positions

and groupings of different signal classes, DiMuSe enables the detection of persistent proximities and separations between musical corpora. Stability is assessed through repeated analyses and comparisons across datasets, ensuring that observed patterns are not artefacts of specific selections or parameter choices.

Throughout the method, no a priori assumptions are made regarding the therapeutic value of specific musical genres or styles. Instead, therapeutic relevance is evaluated post hoc by examining whether signals labelled as therapeutic exhibit consistent structural grouping distinct from control datasets. This approach allows the method to remain neutral with respect to interpretation while providing quantitative evidence for structural regularities.

The DiMuSe framework is thus intended as an exploratory and comparative tool rather than a predictive or diagnostic model. Its primary contribution lies in offering a reproducible methodology for mapping informational structures in music signals, creating a foundation for subsequent investigations into the relationship between signal organization and functional effects in music therapy and related domains.

### 3.2. Constructing the Principal Components Vector for DiMuSe

Evaluating structured sound sequences such as music—characterized by beat, harmony, timbre, and melody—using only conventional statistical evaluators proved insufficient, as these approaches do not preserve the sequential order of signals. Since harmony is established through ordered sequences, losing this structural information would compromise the analysis of musical features.

*The novelty of the proposed method lied in the selection of the scalar evaluators.*

The scalar evaluators employed in the DiMuSe method were not designed as music-specific descriptors, but as general-purpose measures for characterizing nonlinear time-series signals. Their selection was motivated by the objective of identifying intrinsic informational and dynamical properties of sound sequences, rather than extracting stylistic or semantic musical features tied to genre, instrumentation, or timbral identity.

From this perspective, musical signals were treated as instances of nonlinear temporal processes, similar in nature to other complex signals encountered in physics, biology, or geophysics. Consequently, the evaluators used in this study originate from statistical analysis, fractal geometry, nonlinear physics, and complex systems theory, and are applicable beyond the musical domain to a wide range of non-musical or non-auditory signals. This domain-agnostic design would allow DiMuSe to function as a structural screening tool, capable of revealing similarities among signals based on their underlying informational organization rather than on music-specific acoustic conventions.

Rather than relying solely on statistical measures, the paper analysed sound sequences through multiple scientific perspectives. To classify music sequences objectively, the researchers explored scalar evaluators available for PCA across various scientific domains. The primary validation criterion was that these evaluators should be *sensitive to temporal variations in time series data*.

In contrast to commonly used music information retrieval features—such as Mel-frequency cepstral coefficients, spectral centroid, or chroma vectors—which are primarily designed as short-term, perceptually motivated summaries and therefore exhibit limited sensitivity to nonlinear dynamical properties, the present study intentionally adopted scalar descriptors selected for their responsiveness to nonlinear signal behaviour. Specifically, the DiMuSe evaluators were chosen based on their ability to respond to non linear time-series dynamics as a first condition, and also capture long-range temporal dependencies, multiscale organization, and history-dependent structure. This design choice reflects the aim of investigating whether the data sequences exhibits distinctive informational patterns

at a structural level that precedes sound related features like perceptual categorization, cultural interpretation, or genre-based description.

However, the researchers intend in the future to select a list of evaluators that are specific for sound sequences and add those as scalar evaluators to the DiMuSe evaluators to study their impact.

The scientific fields from which the suitable evaluators were sourced included:

- *Statistics* – Standard deviation, autocorrelation, entropy measures.
- *Fractal Geometry* – Measures of self-similarity and complexity.
- *Nonlinear Physics* – Chaos theory and dynamical system indicators.
- *Complex Systems* – Network-based metrics and emergent pattern analysis.

The evaluators and their interpretation the context of sound sequences were listed in Table 2.

Category	Evaluator	Structural property captured	Interpretation in musical / sound context
Statistics	Mean ( $\mu$ )	Central tendency	Average signal level; relates to overall energy balance
Statistics	Standard deviation ( $\sigma$ )	Variability	Dynamic range and amplitude fluctuations
Statistics	Skewness (M3)	Distribution asymmetry	Bias toward peaks or troughs; dynamic imbalance
Statistics	Kurtosis (M4)	Tail heaviness	Presence of transients or extreme events
Fractal geometry	dnfft	Multiscale roughness	Degree of structural irregularity across time scales
Fractal geometry	dnpfft	Extended-domain fractality	Persistence of structure across wider frequency ranges
Fractal geometry	dipfft	Energy-weighted fractality	Coupling between complexity and signal energy
Fractal geometry	dntimp	Time-domain fractal dimension	Temporal self-similarity and long-range organization
Complex systems	hhcor (Hh)	Correlation roughness	Texture smoothness versus jaggedness
Complex systems	Hurst exponent (H)	Long-range dependence	Predictability and persistence of temporal patterns
Nonlinear physics	Power spectrum entropy	Spectral disorder	Degree of randomness versus tonal organization
Nonlinear physics	sum_q	Spectral energy distribution	Global spectral content and balance

**Table 2.** Scalar evaluators used in the DiMuSe method and their associated structural properties and interpretations.

By integrating scalar evaluators from these diverse fields, the DiMuSe method aimed to detect intrinsic patterns in musical sequences, thereby enabling the clustering of melodies based on their informational structure.

### 3.2.1. Statistical Evaluators

#### Function *med*

The function *med* computes the arithmetic mean of the input signal and serves as a first-order statistical descriptor of its central tendency.

*Theoretical background.* Given a discrete signal  $x = \{x_1, x_2, \dots, x_N\}$ , the mean value is defined as:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i. \quad (5)$$

The mean characterizes the average level of the signal and is commonly used to describe its global offset.

*DiMuSe implementation.* In the MATLAB implementation, the mean is computed directly using the built-in function `mean(A)`, applied to the original input signal prior to normalization. The returned scalar value corresponds to the estimator *med*.

#### Function *sigma*

The function *sigma* computes the sample standard deviation of the input signal and quantifies the dispersion of values around the mean.

*Theoretical background.* For a discrete signal  $x = \{x_1, x_2, \dots, x_N\}$  with mean  $\mu$ , the sample standard deviation is defined as:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2}. \quad (6)$$

This formulation includes Bessel's correction and provides an unbiased estimate of variance for finite samples.

*DiMuSe implementation.* In the MATLAB code, the standard deviation is computed using the default `std(A)` function, which implements the sample standard deviation with  $N - 1$  normalization. The resulting scalar value is returned as the estimator *sigma*.

#### Functions *M3* and *M4*

The functions *M3* and *M4* compute root-transformed central moment descriptors that characterize the global shape of the amplitude distribution of a signal. These evaluators are derived from the third- and fourth-order central moments but are modified to provide scale-consistent scalar measures suitable for multivariate analysis.

*Theoretical background.* Given a discrete signal  $x = \{x_1, x_2, \dots, x_N\}$  with mean  $\mu$ , the third- and fourth-order central moments are defined as:

$$\mu_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3, \quad (7)$$

$$\mu_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4. \quad (8)$$

These moments capture asymmetry and tail heaviness of the distribution, respectively, but their magnitudes scale nonlinearly with signal amplitude.

*DiMuSe implementation.* In the MATLAB implementation, the central moments are first accumulated explicitly and normalized by the signal length. To obtain dimensionally comparable scalar descriptors, their absolute values are taken and root-transformed:

$$M3 = \left| \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3 \right|^{1/3}, \quad (9)$$

$$M4 = \left| \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4 \right|^{1/4}. \quad (10)$$

These transformations ensure that both descriptors have the same physical dimension as the original signal amplitude and reduce sensitivity to extreme outliers. The resulting scalars quantify the magnitude of distributional asymmetry (M3) and tail heaviness (M4), without encoding their sign.

### 3.2.2. Fractal Evaluators

The following fractal evaluators[24] provided a quantitative characterization of fractal dimensions in signal analysis.

#### Function *dnfft*

The function *dnfft* implements a fractal estimator related to the *Smoothing Dimension*, intended to quantify multiscale roughness in nonlinear signals. The description below distinguishes the general principle of the estimator from the exact computational steps used in the DiMuSe MATLAB implementation.

*Theoretical background.* Let  $s(t)$  denote the original signal and let  $x(t) = \frac{ds(t)}{dt}$  be its first derivative. The smoothing-dimension principle evaluates how the signal energy changes as the effective smoothing scale increases. In frequency-domain form, this can be expressed by considering the energy of the derivative after progressively restricting the bandwidth (equivalently, varying an effective cutoff frequency). Under fractal scaling assumptions, the resulting energy measure follows a power law with respect to the cutoff scale, and the slope in log–log coordinates defines a smoothing exponent.

*DiMuSe implementation.* The code computes the derivative and then uses the cumulative sum of the derivative power spectrum to implicitly represent a family of increasing cutoff frequencies: (i) the signal is truncated (*trunc*), (ii) differentiated numerically (*diff*), (iii) mean-centered, (iv) transformed by FFT, and the one-sided power spectrum is formed as  $|X(f)|^2$ . The cumulative energy curve is then computed as:

$$W(k) = \sum_{i=1}^k |X(i)|^2 \quad (11)$$

and the analyzed quantity is  $\sqrt{W(k)}$ , which corresponds to the Euclidean norm of the derivative content up to an implicit cutoff determined by bin index  $k$ .

Both the normalized frequency index and the cumulative energy are scaled to  $[0, 1]$ , and a linear regression is performed in log–log space over a fixed index range that excludes the lowest-frequency bins:

$$k \in \left[ \left\lceil \frac{l}{20} \right\rceil + 2, \left\lfloor \frac{l}{2} \right\rfloor - 1 \right] \quad (12)$$

where  $l$  is the signal length after differentiation. The regression slope is taken as the smoothing exponent, and the corresponding fractal dimension estimate is computed as  $df = 1 + dn$ .

The MATLAB implementation returns three scalar values (packed into a vector):

- *dnfft1*: the estimated smoothing exponent (slope of the log–log regression),
- *dnfft2*: the regression fit error returned by the function *reg*,
- *dnfft3*: the corresponding fractal dimension estimate computed as  $df = dnfft1 + 1$ .

#### Function *dnpfft*

The function *dnpfft* estimates the *smoothing dimension with extended frequency domain*. This evaluator is conceptually related to *dnfft*, but instead of explicitly applying a bank of low-pass filters with selectable cutoff frequencies, it derives an equivalent multiscale relationship directly from the cumulative spectrum of the signal derivative.

*Theoretical principle.* After computing the first derivative of the input signal, the method analyzes how the accumulated spectral energy grows with increasing normalized frequency. Under fractal scaling assumptions, this growth follows a power law in log–log coordinates, whose slope is used as a smoothing exponent.

*DiMuSe implementation (as used in this study).* Given a signal  $x$ , the code performs: (i) truncation (function `trunc`), (ii) mean removal, (iii) numerical differentiation (`diff`), followed by an FFT. The squared magnitude spectrum of the derivative is accumulated using a cumulative sum, producing a monotone energy curve. The regression is then performed in log–log space on the normalized pairs  $\{(f, \sqrt{\text{cumsum}(|X(f)|^2)})\}$ , where frequency and cumulative energy are normalized to  $[0, 1]$ .

Importantly, the “extended domain” aspect is implemented by fitting the regression over a broad range of FFT bins, excluding only the very low-frequency and boundary bins. Specifically, the regression uses indices from  $l/20 + 2$  to  $l/2 - 1$ , where  $l$  is the signal length.

The function returns three scalar outputs packed into a vector:

- `dnppfft1`: the estimated slope (smoothing exponent) from log–log regression,
- `dnppfft2`: the regression fit error returned by `reg` (implementation-dependent),
- `dnppfft3`: the corresponding fractal dimension estimate computed as  $df = 1 + \text{dnppfft1}$ .

#### Function `dipfft`

The function `dipfft` estimates a related fractal exponent that incorporates the *energy accumulation* of the derivative spectrum. In contrast to `dnppfft`, which accumulates squared spectral magnitude and then takes a square root, `dipfft` accumulates the magnitude spectrum directly, yielding a different scaling sensitivity.

*Theoretical principle.* The method assumes that the cumulative spectral magnitude of the signal derivative follows a power-law scaling with normalized frequency. The slope of the log–log regression provides an exponent  $di'$ , which can be mapped to a fractal dimension estimate via  $df = di' + 0.5$ .

*DiMuSe implementation.* The effective parameters are fixed internally: `min` is set to 20 and `max` is set to  $\lfloor l/3 \rfloor$ , where  $l$  is the signal length (the input arguments are overwritten). The processing steps are: truncation, mean removal, numerical differentiation, FFT of the derivative, magnitude extraction on the positive-frequency half-spectrum, cumulative sum, and normalization to  $[0, 1]$ . A linear regression is then performed in log–log space over indices 20 to  $l/3$ .

The function returns three scalar outputs:

- `dipfft1`: the estimated slope  $di'$  from log–log regression,
- `dipfft2`: the regression fit error returned by `reg`,
- `dipfft3`: the corresponding fractal dimension estimate computed as  $df = \text{dipfft1} + 0.5$ .

#### Function `dntimp`

The function `dntimp` provides a time-domain variant of the smoothing-dimension concept. Instead of forming cumulative quantities from the FFT spectrum directly, it constructs a sequence of progressively smoothed signals by truncating the derivative spectrum and measuring the resulting time-domain energy. The scaling of this energy across smoothing levels yields a fractal exponent.

*Theoretical principle.* If a signal exhibits multiscale (fractal-like) structure, then its energy after progressive low-pass smoothing follows a power-law scaling with the effective

bandwidth. Estimating the slope of this relationship in log–log space provides a smoothing exponent that can be mapped to a fractal dimension.

*DiMuSe implementation.* Given a signal  $x$ , the code performs truncation, numerical differentiation, mean removal, and FFT of the derivative. It then generates `nrpoints` progressively smoothed versions of the derivative by retaining only the lowest `nmftj` FFT bins (and mirroring them to preserve a real-valued inverse FFT), where the initial bandwidth is set by `nmin = 32` and the bandwidth increases geometrically across `nrpoints` steps. For each step, the inverse FFT is computed and the time-domain energy is measured as  $\sqrt{\sum s^2}$ . The resulting pairs (`nmftj`, energy) are normalized to  $[0, 1]$  and fitted by linear regression in log–log space.

The function returns three scalar outputs:

- `dntimp1`: the estimated slope from log–log regression,
- `dntimp2`: the regression fit error returned by `reg`,
- `dntimp3`: the corresponding fractal dimension estimate computed as  $df = 1 + dntimp1$ .

### 3.2.3. Complex Systems Evaluators

#### Function `hhcor`

The function `hhcor` implements the *height–height correlation* estimator, which quantifies scale-dependent roughness by analyzing how the root-mean-square (RMS) of signal differences grows with the time lag. The description below distinguishes the theoretical definition from the exact procedure used in the MATLAB implementation adopted in DiMuSe.

*Theoretical background.* Given a signal  $s(t)$ , the height–height correlation function is defined through the scaling relation

$$\sqrt{(s(t) - s(t + \tau))^2} \sim \tau^{Hh}, \quad (13)$$

where  $\tau$  is the lag and  $Hh$  is the roughness exponent. Under standard fractal assumptions, the exponent is related to the fractal dimension by

$$D_f = 2 - Hh. \quad (14)$$

*DiMuSe implementation.* In the MATLAB code (`hhcor(a, pas)`), the lag is implemented as an integer sample shift  $n$ . For each lag  $n$ , the code forms two aligned subsequences  $a_1 = a(1 : N - n + 1)$  and  $a_2 = a(n : N)$ , computes the mean squared difference, and then takes the RMS:

$$R(n) = \sqrt{\frac{1}{N - n + 1} \sum_{i=1}^{N-n+1} (a(i) - a(i + n - 1))^2}. \quad (15)$$

The regression is performed in log–log space using the pairs  $(\log(n), \log(R(n)))$ . In particular, the implementation evaluates lags over

$$n = 10 : \text{pas} : \left\lfloor \frac{N}{4} \right\rfloor,$$

and estimates  $Hh$  as the slope of the linear regression returned by `reg`.

The function returns three scalar values (packed into a vector):

- `hhcor1`: the estimated exponent  $Hh$  (slope of the log–log regression),
- `hhcor2`: the regression fit error returned by `reg`,
- `hhcor3`: the derived fractal dimension  $D_f = 2 - hhcor1$ .

### Function *hurst*

The function *hurst* estimates the *Hurst exponent* using the classical rescaled range (R/S) method, which quantifies long-range dependence in a time series. The description below distinguishes the theoretical definition from the exact MATLAB implementation used in DiMuSe.

*Theoretical background.* For a discrete signal  $s(t)$ , the R/S method evaluates, for each window length  $\tau$ , the rescaled range

$$\frac{R(\tau)}{S(\tau)}, \quad (16)$$

where  $R(\tau)$  is the range of the cumulative deviations from the mean within the window and  $S(\tau)$  is the standard deviation within the same window. For many self-affine processes, the expected scaling follows:

$$\mathbb{E}\left[\frac{R(\tau)}{S(\tau)}\right] \sim \tau^H, \quad (17)$$

and the Hurst exponent  $H$  is obtained as the slope of a linear fit in log–log space.

*DiMuSe implementation.* In the MATLAB code (`hurst(a, nnn)`), the analysis is applied to the first difference of the input signal (`a = diff(a)`). The signal is then partitioned into non-overlapping segments of length  $i$ , and for each segment the rescaled range  $R/S$  is computed. For each window length  $i$ , the implementation averages  $R$  and  $S$  across all segments ( $\lfloor N/i \rfloor$  segments), and stores the mean ratio  $R(i)/S(i)$ . The window length is swept as:

$$i = 100 : nnn : \left\lfloor \frac{N}{3} \right\rfloor,$$

where  $N$  is the length of the differenced signal. A linear regression in log–log coordinates is then performed by `reg1` on the pairs  $(\log(i), \log(R(i)/S(i)))$ , and the slope is taken as the estimated Hurst exponent.

The MATLAB function returns three scalar values (packed into a vector):

- *hurst\_1*: the estimated Hurst exponent  $H$  (slope of the log–log regression),
- *hurst\_2*: the regression fit quality returned by `reg1` (implementation-dependent),
- *hurst\_3*: the derived fractal dimension estimate computed in the code as  $D_f = 2 - \text{hurst}_1$ .

### 3.2.4. Physics Evaluators

#### Function *tft*

The function *tft* computes the (*normalized*) *power spectrum entropy*, which summarizes how uniformly the signal's spectral power is distributed across frequency bins. The description below distinguishes the general definition from the exact computation used in the DiMuSe MATLAB implementation.

*Theoretical background.* Given a discrete power spectrum that is normalized to form a probability mass function  $p_i$  (i.e.,  $p_i \geq 0$  and  $\sum_i p_i = 1$ ), the Shannon entropy is:

$$H = - \sum_i p_i \log p_i. \quad (18)$$

Low entropy indicates that spectral energy is concentrated in a small number of bins (e.g., near-periodic signals), while high entropy indicates a more uniform distribution of spectral energy (e.g., broadband noise).

*DiMuSe implementation.* In the MATLAB code, the normalized spectral distribution is provided as the vector `qq`. The entropy is computed as:

$$H = - \sum_{i: p_i > 0} p_i \log p_i, \quad (19)$$

where terms with  $p_i = 0$  are skipped (equivalently,  $0 \log 0 = 0$ ). The implementation then returns a *normalized entropy* by dividing by  $\log(nn)$ , where `nn` is the number of frequency bins used in the distribution:

$$H_{\text{norm}} = \frac{H}{\log(nn)}. \quad (20)$$

With this normalization,  $H_{\text{norm}} \in [0, 1]$ , where 0 corresponds to power concentrated in a single bin and 1 corresponds to a uniform distribution across `nn` bins. The MATLAB output `ttt` corresponds to  $H_{\text{norm}}$ .

Function `sum_q`

The function `sum_q` returns a scalar spectral magnitude summary computed from the discrete Fourier transform (DFT) of the normalized signal. The description below distinguishes the general spectral interpretation from the specific MATLAB implementation used in DiMuSe.

*Theoretical background.* A frequency-domain representation of a discrete signal can be obtained via the DFT, and global scalar summaries of the spectrum can be used to characterize the overall spectral magnitude level. Such summaries reflect how much spectral content is present in the analyzed frequency range, but they are distinct from physical energy measures unless power ( $|X(f)|^2$ ) and Parseval-consistent normalization are used.

*DiMuSe implementation.* In the MATLAB code, the input signal  $A$  is first centered and normalized to unit standard deviation. The one-sided magnitude spectrum is then computed as:

$$q(k) = \frac{|\text{FFT}(A)(k)|}{nn^2}, \quad k = 1, \dots, nn, \quad (21)$$

where  $nn = \lfloor n/2 \rfloor$  and  $n$  is the signal length. The scalar returned by this evaluator is:

$$\text{sum}_q = \sum_{k=1}^{nn} q(k), \quad (22)$$

i.e., the sum of the scaled one-sided FFT magnitudes. In addition, the normalized spectral distribution used for entropy estimation is computed as  $qq(k) = q(k) / \sum_k q(k)$ , ensuring  $\sum_k qq(k) = 1$ .

### 3.3. Validation of the DiMuSe Method Using Generated Signals

Before applying the method to real-world music, it was essential to test its reliability in controlled conditions. To achieve this, Matlab-generated arrays were created using rules that produced persistent, anti-persistent, fractal, and cumulative signals. This validation step ensured that the method could detect meaningful clusters even when applied to synthetic data, providing confidence that the same analytical approach could later reveal hidden patterns in actual therapeutic melodies.

#### 3.3.1. Array Generation Rules

The first step involved generating a natural (Gaussian) random array consisting of 8192 scalars, referred to as *originalArray*. This array served as the foundation for creating additional signals, except for fractal signals.

### Persistent Array Generation Function 647

A 10-iteration loop was implemented, where each iteration produced a new array based on the previous one, following the transformation rule: 648  
649

$$nextArray(elemLoop) = \frac{prevArray(elemLoop) + prevArray(elemLoop + 1)}{2} \quad (23)$$

The resulting arrays were labeled as *pers1*, *pers2*, ..., *pers10*. 650

### Anti-Persistent Array Generation Function 651

An initial 10-element array of percentages, *persArray* = [10%, 20%, ..., 100%], was defined. 652  
653

From the original generated array, a 10-iteration loop was conducted, where each iteration generated a new array using the following transformation rule: 654  
655

$$nextArray(elemLoop) = origArray(elemLoop) - origArray(elemLoop + 1) \cdot persArray(iterationLoop) \quad (24)$$

Each newly generated array progressively lost a certain amount of information from the initial Gaussian array due to the applied transformation. 656  
657

The resulting arrays were labeled as *antipers1*, *antipers2*, ..., *antipers10*. 658

### Fractal Array Generation Function 659

To generate fractal sequences, a 10-element array of predefined fractal dimensions, *fractalDimArray* = [1.1, 1.2, ..., 2.0], was initially constructed. 660  
661

Without directly utilizing the original Gaussian array as input, 10 additional arrays were generated corresponding to the predefined fractal dimensions in *fractalDimArray* using the *Takayasu method*[25]. These transformations ensured the generated signals exhibited fractal-like characteristics. The resulting arrays were labeled as *df1.1*, *df1.2*, ..., *df2.0*. 662  
663  
664  
665

### Cumulative Sum Array Generation Function 666

A single array was generated, representing the cumulative sum over all elements in the initial generated array. This process is equivalent to computing the integral of the original array, providing insight into its cumulative behavior. The resulting array was labeled *cumsum*. 667  
668  
669  
670

### Differential Array Generation Function 671

To capture local variations in the signal, an array was generated containing the difference between each pair of consecutive elements from the initial Gaussian array. This transformation effectively extracts high-frequency components, highlighting short-term fluctuations within the dataset. The resulting array was labeled *diff*. 672  
673  
674  
675

### 3.3.2. Running DiMuSe on the Generated Arrays 676

Applying the DiMuSe method in Matlab on the 32 previously generated arrays produced a .csv file containing 32 vectors, with each vector consisting of 24 scalar values. 677  
678

This file was used to compute the three Principal Component (PC) scalars for each vector, utilizing different combinations of generated array types. The results were subsequently visualized in scatter plots on the most representative PCA axis, PC1, PC2. The clustering was created using the k-Means Orange Data Mining Widget with 10 reruns and 679  
680  
681  
682

300 maximum iterations. The optimal number of clusters was assessed using 3 indicators [26]:

- k number of clusters for the k-Means unsupervised clustering method;
- *Silhouette score* (SIL), where values vary between -1 and +1, best score being between 0 and 1;
- *Davies–Bouldin Index* (DBI), where values vary between 0 and  $\infty$ , best score being closer to 0.

The *Silhouette Score* (SIL)[27] quantifies how similar a point is to its own cluster compared to other clusters, the interpretation of the values is:

- +1 means Excellent cluster assignment
- 0 means Overlapping clusters
- < 0 means Misclassified points

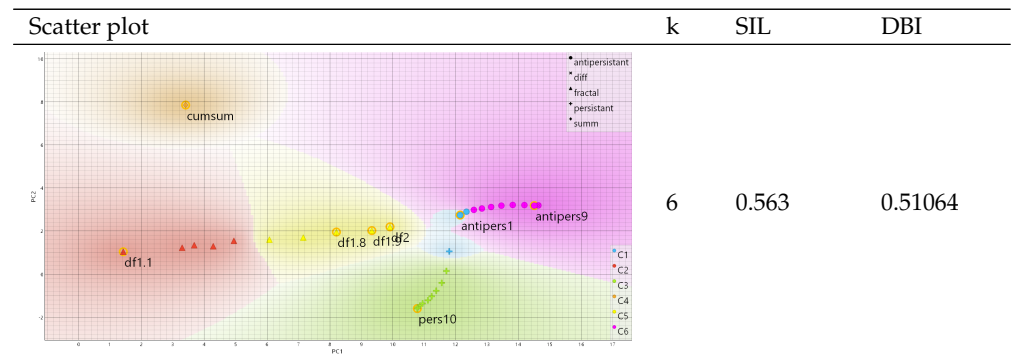
The *Davies–Bouldin Index* (DBI)[28] measures how compact clusters are relative to how far apart they are. The interpretation of the values is:

- Low ( $\rightarrow$  0) means Well-separated, compact clusters
- High means Overlapping clusters

The results were presented in Table 3.

**Table 3.** Generated signals PCA (PC1, PC2) plot, k-Means unsupervised clustering (Orange Data Mining v3.40 Scatter Plot).

Scatter plot	k	SIL	DBI
	3	0.569	0.53209
	4	0.601	0.38908
	5	0.644	0.38416



The best clustering option we identified was  $k=5$ , that matches the number of classes. Clusters were clearly divided in the plot, and the SIL and DBI indicators were well in the optimal values ranges.

*These results validate that the DiMuSe method successfully clusters time-series signals based on the selected estimators applied in the PCA analysis.*

### 3.4. Compiling a List of Sound Sequences for Clustering with the DiMuSe Method

Having confirmed that DiMuSe could distinguish between artificially generated signals, the next step was to apply it to actual music and sound recordings. To test its scope, a diverse dataset was assembled, covering multiple genres (classical, religious, therapeutic, folk, acapella, noise, and natural recordings). This diversity was intentional: by contrasting empirically validated therapeutic music with everyday or culturally significant sounds, the analysis could determine whether therapeutic pieces express unique informational structures that differentiate them from other categories.

A diverse dataset of sound sequences and melodies spanning various musical genres was created. The following genres were selected:

- Classical music
- Religious music
- Therapeutic music
- Traditional folk music
- Vocal acapella music
- Anthropogenic and natural sounds
- Computer-generated colored noise

For musical pieces, approximately one minute of each selected song was recorded from the online platform YouTube. The noises and environmental sounds were sourced from the royalty-free Pixabay platform.

This process resulted in the creation of 32 MP3 files, organized into six folders corresponding to the defined genres.

#### 3.4.1. Selected sound sequences

##### Noise (Colored)

The Table 4 contains colored noise sound sequences that are generated using specific algorithms, classified as Noise in the scatter plot.

##### Noise (Analog)

The Table 5 contains recorded noise sound sequences from antropic or natural sources, classified as Noise in the scatter plot.

Name	Description
noise_white	White Noise(1 min), a mix of all the frequencies that the human ear can hear, from 20 Hz to 20.000 Hz, all frequencies having the same intensity simultaneously. It's main characteristic is that it can annihilate any other sound with similar intensity [29]
noise_pink	Pink Noise(1 min), very close to the white noise, but with more energy towards one of it's intensity spectrum ends. It's frequency is progressing logarithmically and it's spectral density (power/frequency) is varying by 3db per octave or by 10 db per decade (proportional with 1/f) [29].
noise_gray	Gray Noise(1 min), it is white noise with an adjusted loudness curve over a given range of frequencies, giving the perception that is equally loud on all frequencies[29].
noise_violet	For Violet Noise(1 min), the power density increases 6.02 dB per octave, and it is the differentiation of a white noise signal[29].
noise_brown	Brown Noise(1 min) is also called Brownian, its power density decreases 6.02 dB per octave with increasing frequency, and it is generated with a temporal integration of a white noise[29].

**Table 4.** Colored noise

Name	Description
noise_tvstatic	1 min. sound of a Analog CRT TV Electric static noise
noise_blender	1 min. sound of a kitchen blender
noise_cable	1 min. static cable noise
noise_no-signal	1 min analog TV sound with no signal
noise_ocean	1 min. recording of ocean waves
noise_saturn	1 min recording of electromagnetic waves scaled to audible frequency recorded from planet Saturn
noise_steps-on-snow	1 min recording of someone's footsteps when walking in the snow
noise_leaves-rustling	1 min recording of wind passing thru tree leaves

**Table 5.** Analog noise

## Vocal (Acapella)

734

The Table 6 contains recorded vocal performed music sequences, classified as Vocal in the scatter plot.

735

736

Name	Description
modern	1 min sequence from "When I Was Your Man" song performed by Bruno Mars
arabic	1 min sequence from "Dari Ya Alby" song performed by Hamsa Namira arabic singer
indian	1 min sequence from "Thaniye" song performed by Vishnu Vijay indian singer
navajo	1 min sequence from "Apache Sunrise Song" performed by native Navajo singers

**Table 6.** Vocal music

## Therapy

737

The Table 7 contains music pieces proven empirical to have therapeutic effect, classified as Therapy in the scatter plot.

738

739

Name	Description
circulator	Blood Circulation enhance (337 Hz) by The Unexplainable Store
heart	1 min sequence from "Aqua body and Mind relaxation" song by Mike Vickerage
kidney	1 min sequence from "Kidney Function (625 Hz)" by The Unexplainable Store
snervos	1 min sequence from "Nervous system (764 Hz) The Source" by Deuter
well_being	1 min sequence from "Spiritual Well Being (1565 Hz)" by The Unexplainable Store

**Table 7.** Therapy music

## Classical

740

The Table 8 contains music pieces from classical music, classified as Classical in the scatter plot.

741

742

Name	Description
beethoven	1 min sequence from "Beethoven quartet no.15 in A minor"
mozart	1 min sequence from "Eine kleine Nachtmusik"
bach	1 min sequence from "Chaconne, Partita No. 2 BWV 1004"
vivaldi	1 min sequence from "Vivaldi The Four Seasons, Violin Concerto No. 3 in F Major, RV 293"

**Table 8.** Classical music

## Traditional

743

The Table 9 contains music pieces from traditional folk music, instrumental only and also voice and instrumental, classified as Traditional in the scatter plot.

744

745

Name	Description
folk-ro-voice	1 min sequence from Romanian folk song "Ciuleandra" performed by Maria Tanase
folk-ro-instr	1 min sequence from Romanian folk song 'Ca din bucium / Carpathianhorn alike tune'
folk-ch-instr	1 min sequence from Chinese folk song 'Authentic Chinese Classical Music - Ming Dynasty Court and Taoist music'
folk-ch-voice	1 min sequence from Chinese folk song 'Liu Zi Ling Song - Farewell [Traditional China]'
folk-bg-instr	1 min sequence from Bulgarian folk song '100 Kaba Gaidi'
folk-bg-voice	1 min sequence from Bulgarian folk song 'A beetle has landed on a blackthorn'
folk-celt-instr	1 min sequence from Irish folk song 'Flying Feet - LIVE Authentic Irish Music & Dance Show - 2021'
folk-celt-voice	1 min sequence from Scottish folk song 'Donald, Where's Your Trousers'

**Table 9.** Traditional music

### Religious

The Table 10 contains music pieces from important five religions we identified, classified as Religious in the scatter plot.

Name	Description
cristian	1 min sequence from the Christian song 'How Great Thou Art A Cappella'
aramaic	1 min sequence from the Judaic song 'Psalm 50 In Aramaic: Have mercy on me, O God'
islamic	1 min sequence from the Islamic song 'Beautiful Islamic Call To Prayer'
indianveda	1 min sequence from the Indian song 'Veda   Indian Mantra Vocals   Royalty Free Music'
tibetan	1 min sequence from the Tibetan song '(152) Overtone Singing & Deep Voice Chant with Tibetan Monks'

**Table 10.** Religious music

### 3.5. Running DiMuSe on the Selected Sound Sequences

DiMuSe has been implemented in Matlab to generate vectors for each sound sequence, followed by PCA reduction and visualization using Orange Data Mining. This process made it possible to map musical sequences into a three-dimensional space where clustering could emerge naturally. The goal was not only to confirm whether genres separated meaningfully, but also to test whether validated therapeutic music clustered together, indicating shared structural patterns.

Prior to vectorization and evaluation by the DiMuSe framework, all sound sequences underwent a standardized pre-processing pipeline implemented in Matlab, designed to ensure signal comparability while preserving intrinsic structural and dynamical properties. For each audio file the silent part was trimmed from the beginning and end of the sound sequences. Then they were converted to single-channel (mono) time-domain signals by averaging the stereo channels when necessary, thereby eliminating spatial effects not relevant for structural analysis. The resulting sequences were normalized to unit standard deviation and mean-centered. To reduce data dimensionality while preserving global

temporal structure, the effective sampling frequency was subsequently reduced by a factor of 100 using decimation, followed by a final centering step.

Following segmentation, the audio signals were normalized by removing the mean value and scaling by their standard deviation, thereby reducing the influence of recording gain and overall loudness. The normalized signal was subsequently transformed to its absolute value representation, emphasizing amplitude variations independently of signal polarity. To stabilize short-term fluctuations and highlight slower structural variations, a moving-average smoothing filter was applied. The resulting signal was then mean-centered again and temporally decimated, substantially reducing data dimensionality while preserving global temporal organization. A final mean-removal step was applied after decimation to eliminate residual offsets. The resulting pre-processed signal constituted a continuous, normalized nonlinear time series suitable for the computation of scalar evaluators and subsequent principal component analysis.

The functionalities were described in Table 11.

Matlab Function	Description
DimuseVectorise	Evaluates the sound sequences thru the 24 estimators, creating a 24 axis space. Has as input a folder path with subfolders containing MP3 files and creates a '.csv' file with a row of information for each MP3 file. The columns are "Name" containing the name of the MP3 file, the "Class" containing the genres extracted from the subfolder structure of the input path, and another 24 columns containing the scalar evaluators we proposed
DimusePCA	Runs the PCA. Has as input a '.csv' file structured as the one created by the DimuseVectorise function. The function runs the Principal Component Analysis on 3 axis and creates 3 new files: one with the resulting PCA values for each of the sound sequences, one with the resulting Eigenvectos and one with the Standard Deviation values
DimuseProjectToPCA	Positions sound sequences already evaluated thru DimuseVectorise in a previously defined PCA space. Has as input a '.csv' file containg data structured as a file created by DimuseVectorise, and files containg Eigenvectos and Standard deviation inforation of a previously defined PCA space.
DimuseGenerateSignal	Generates the vectors described in section 3.3.

**Table 11.** DiMuSe Functions

The first step in executing the DiMuSe method was to generate evaluation vectors for each selected sound sequence using the Matlab function *DimuseVectorise*. This function produced the file *Dimuse-MusicSequence-Mar25.csv*, containing 32 rows of data.

Each row included:

- *Name column* – The corresponding MP3 file name.
- *Class column* – The assigned genre.
- *24 scalar evaluator columns* – The statistical and fractal metrics proposed in section 3.2.

To achieve a representation similar to that used for the generated signals, a .csv file combining all melody classes was created. The *DimusePCA* function was applied to this dataset, extracting the three Principal Component (PC) scalars for each vector, with a conserved data variance of 71.52%.

The resulting file was imported into Orange Data Mining, where the k-Means unsupervised clustering method was applied, with 20 reruns and 300 maximum iterations. Similar

to the method used for the Generated Signals<sup>3</sup>, the SIL and DBI values were used, along with visual evaluation for choosing the optimal clustering  $k$  value.

The Table 12 contains combined sound sequences classes PCA plots.

**Table 12.** Base Set PCA (PC1, PC2, PC3) plot, k-Means unsupervised clustering (Orange Data Mining v3.40 Scatter Plot).

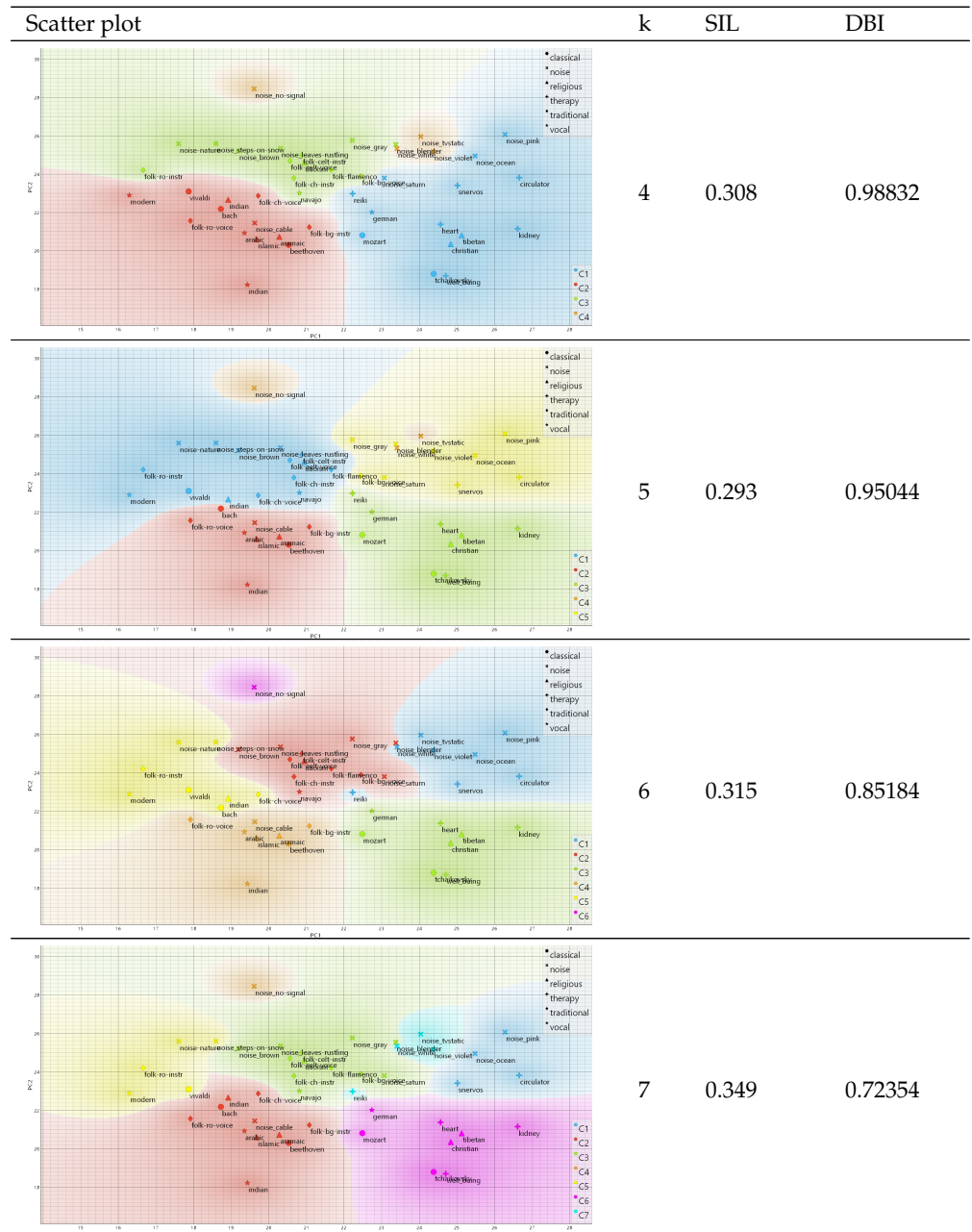
Scatter plot	k	SIL	DBI
	3	0.352	0.92931
	4	0.321	0.96864
	5	0.301	0.93239
	6	0.296	0.83049
	7	0.322	0.72269

791

792

793





After this step, the Base Set sound sequences configuration identified remained as the optimal choice, with k=4 clustering. 817

Another sound configuration was tested, by removing one sound from each of the classes, result being displayed in Table 16. The removes sound were: 818

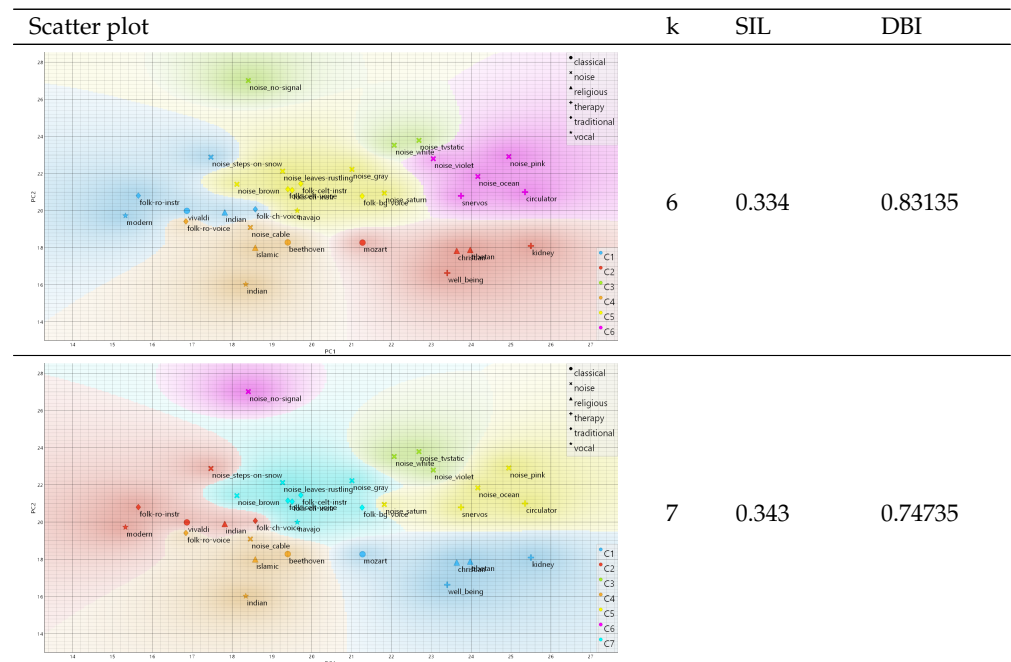
- *classical* class, bach 819
- *noise* class, noise\_blender 820
- *religious* class, aramaic 821
- *therapy* class, heart 822
- *traditional* class, folk-bg-instr 823
- *vocal* class, arabic 824

Name	Class	Description
tchaikovsky	classical	1 min sequence from 'Tchaikovsky: Symphony No. 6 Pathetique   Dresden Philharmonic & Marek Janowski'
noise-nature	noise	1 min natural noise recording of birds and bees
daoism	religious	1 min sequence from the Daoist chant 'Wuji (Daoist Chanting)'
reiki	therapy	1 min sequence from 'Reiki Music, Emotional & Physical Healing Music, Natural Energy, Stress Relief, Meditation Music'
german	vocal	1 min sequence from the German song 'Ganz Schön Feist - Enten (acapella)'
folk-flamenco	traditional	1 min sequence from the a Spanisch flamenco song performed by Alegra Sevilla band

**Table 14.** Added sound sequences for testing optimal cluster configuration

**Table 16.** PCA on Base Set where sounds we removed plot, k-Means unsupervised clustering (Orange Data Mining v3.40 Scatter Plot).

Scatter plot	k	SIL	DBI
	3	0.349	0.95540
	4	0.315	0.91648
	5	0.332	0.84043



Again, the Base Set sound sequences configuration identified remained as the optimal choice, with  $k=4$  clustering. The cluster configuration was saved under the name *Base Set 4*. 827

The next step in the DiMuSe process tried to identify an indicator that could reveal for a new sound sequence if it would be recommendable to be used for its therapeutic effect. 828

The process consisted of vectorizing the sound sequence and then projecting it on the *Base Set 4* PCA plot. This would reveal the projected sound sequences characteristics regarding similarity with the *therapeutic cluster C2*. 829  
830  
831  
832  
833

### 3.7. Correlation of DiMuSe Results Against Other Research 834

The clustering results reveal distinct groupings of sound sequences based on their intrinsic informational structures. However, to determine whether these classifications align with empirical observations of therapeutic music, it was necessary to compare our findings with previous studies. By evaluating established research on the effectiveness of specific musical pieces in stress reduction and relaxation, this paper assessed the accuracy and practical relevance of the DiMuSe method. 835  
836  
837  
838  
839  
840

At the time of writing, two relevant studies were identified in which researchers conducted clinical tests to evaluate the immediate stress-relieving effects of specific songs, while specifying the song titles in the papers. The papers concluded the research with a ranking of the songs analyzed. Using the same songs, a therapeutic effect ranking method was defined on the DiMuSe results as: 841  
842  
843  
844  
845

1. From each song a segment of 1 min. was extracted. Using the DiMuSe functions[11] the music segments were vectorized and projected onto the DiMuSe PCA plane generated from the *Base Set*. The scope of this step was to visualize how the different song segments were distributed in the DiMuSe *Base Set 4* clusters. 846  
847  
848  
849
2. For each of the *Base Set 4* resulting point cloud clusters a *centroid point* was calculated, resulting from the average of the coordinate values on each of the PCA axis, PC1, PC2 and PC3 for all the cluster points, creating a central point for each cluster: *C1*, *C2*, *C3* and *C4*. 850  
851  
852  
853
3. As the *C2 cluster* contained all of the songs considered to have therapeutic effect, the centroid point *C2\_center* was used as reference for creating the therapeutic effect 854  
855

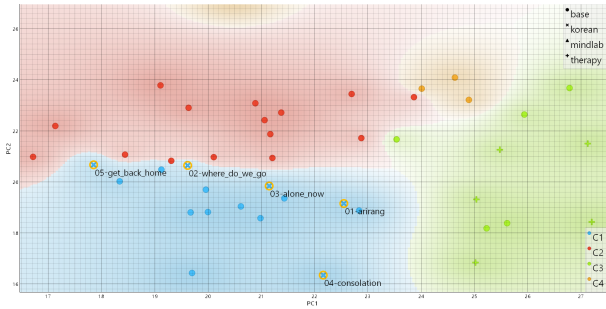
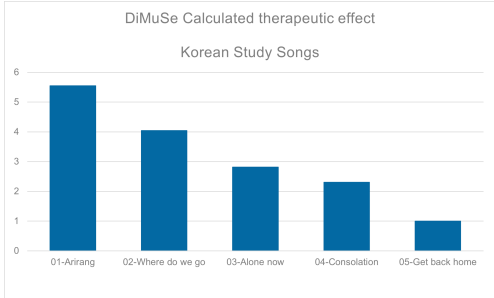
ranking. The distances between each song sequence point and the  $C2\_center$  point were calculated in the PC1, PC2, PC3 space using the formula:

$$songDistance = \sqrt{(C2\_center.PC1 - song.PC1)^2 + (C2\_center.PC2 - song.PC2)^2 + (C2\_center.PC3 - song.PC3)^2} \quad (25)$$

- A song ranking ( $songRank$ ) was calculated for each song by getting the maximum value from the list of  $songDistance$  of all the songs under evaluation, adding 1 to this value and decreasing it by the corresponding  $songDistance$ . A ranking value was obtained between the songs, where the minimum rank is 1. The resulting ranking values were sorted in an ascending order, and displayed in a graphical chart.

### 3.7.1. The Korean Population Study

In the first study[30], the researchers selected five of the most effective stress-relieving songs identified in prior research, using a trained neural network to assess their impact on Korean participants. Among the songs analyzed in the study, "Cozy Arirang" exhibited the highest therapeutic effect.

DiMuSe on Korean songs	Description
	<p>Korean songs "01-Cozy Arirang", "02-Where do we go", "03-Alone now", "04-Consolation" and "05-Get Back Home" one minute sound sequences vectors, classified as <i>Korean</i>, projected on the <i>Base Set</i> PCA space, classified as <i>base</i>, while the <i>center</i> points were calculated, but not included in the clusterization(<math>k=4</math>, <math>SLI = 0.305</math>, <math>DBI = 0.97077</math>).</p>
	<p>Resulting DiMuSe therapeutic score ranking. Song "01-Cozy Arirang" was ranked with the highest therapeutic score. For this study no other information was available about the songs ranking.</p>

**Table 17.** Applying DiMuSe method on Korean Study therapeutical songs.

Applying DiMuSe method on the songs used in the Korean paper[30] by selecting one minute sound sequences resulted that song "01-Cozy Arirang" had the highest DiMuSe calculated therapeutic score. The result matched the result from the Korean paper by identifying the same song with the highest therapeutic effect. The researchers could not match the other songs therapeutic ranking with the DiMuSe calculated ranking, as this information was not available.

### 3.7.2. The Mindlab Study

In the second study[31], a team of musicians and music therapy practitioners from the Mindlab Research Institute (named in this paper *Mindlab*) created a song named "Weightless" that was measured to have a high stress decreasing therapeutic effect. The

researchers measured the therapeutic effect thru biometric sensors and also thru users subjective ratings, by comparing the effect of listening to 17 different songs and the effect of a massage session. The *Mindlab* research conclusion was that "*Weightless*" song had the higher therapeutic effect comparing to the other songs and to the massage session.

The same 17 songs were analyzed thru this paper's proposed ranking method.

**Table 18.** Applying DiMuSe method on the Mindlab Study therapeutical songs.

DiMuSe on Mindlab songs	Description
	<p>Songs "01-Someone Like You", "02-Please Dont Go", "03-Undress Me Now", "04-Electra", "05-Strawberry Swing", "06-Teardrop", "07-Saeglopur", "08-All I Need", "09-We Can Fly", "10-Upside Down", "11-Porcelain", "12-Pure Shores", "13-Watermark", "14-Nine Million Bicycles" "15-Canzonetta Sull aria", "16-Mellomaniac", "17-Weightless" one minute sound sequences vectors, classified as <i>mindlab</i>, projected on the <i>Base Set</i> PCA space, classified as <i>base</i>, while the <i>center</i> points were calculated, but not included in the clusterization (<math>k = 4</math>, <math>SLI = 0.291</math>, <math>DBI = 1.01471</math>).</p>
	<p>Resulting DiMuSe calculated therapeutic effect ranking. Song "17-Weightless" was ranked with the highest therapeutic score, matching the <i>Mindlab</i> research. However, the other songs ranking did not match the <i>Mindlab</i> ranking order.</p>

The researchers investigated if there was correlation between the identified ranking and the *Mindlab* paper reported ranking. The Spearman's Rank Correlation  $\rho$  and Kendall's  $\tau$  values were extracted from the ranking values from Table 19. Google Sheet was used to compute the two indicators,

Considering all the 17 songs from the *Mindlab* paper, the following indicators were calculated.

1. DiMuSe ranking VS Mindlab Biometric ranking :  $\rho = 0.262$  and  $\tau = 0.205$
2. DiMuSe ranking VS Mindlab Subjective ranking:  $\rho = -0.189$  and  $\tau = 0.185$
3. Mindlab Biometric ranking VS Mindlab Subjective ranking:  $\rho = 0.617$  and  $\tau = 0.438$

The ranking indicators reported weak correlation between the DiMuSe ranking and the Mindlab reported rankings. However the correlation between the Mindlab rankings was moderate,  $\tau < 0.5$  and  $\rho \approx 0.5$ .

When comparing only the first 10 songs from the DiMuSe rating, the indicators changed:

1. DiMuSe ranking VS Mindlab Biometric ranking :  $\rho = 0.667$  and  $\tau = 0.511$
2. DiMuSe ranking VS Mindlab Subjective ranking:  $\rho = 0.536$  and  $\tau = 0.404$

Song	DiMuSe rank	Biometric rank	Subjective rank
17-Weightless	1	1	1
02-Please Dont Go	2	6	11
16-Mellomaniac	3	3	14
08-All I Need	4	11	6
05-Strawberry Swing	5	5	3
03-Undress Me Now	6	14	13
07-Saeglopur	7	18	15
12-Pure Shores	8	7	16
11-Porcelain	9	15	15
09-We Can Fly	10	10	10
04-Electra	11	2	2
06-Teardrop	12	16	12
14-Nine Million Bicycles	13	12	17
10-Upside Down	14	17	14
01-Someone Like You	15	8	4
13-Watermark	16	4	5
15-Canzonetta Sull aria	17	9	16

**Table 19.** Comparison of DiMuSe-derived rankings with biometric and subjective rankings from the Mindlab dataset

### 3. Mindlab Biometric ranking VS Mindlab Subjective ranking: $\rho = 0.537$ and $\tau = 0.404$

For the first 10 melodies ranked by DiMuSe in increasing order, the correlation between DiMuSe ranking and the *Mindlab* one was calculated as moderate, with higher correlation when comparing with the *Mindlab* Biometric score ranking.

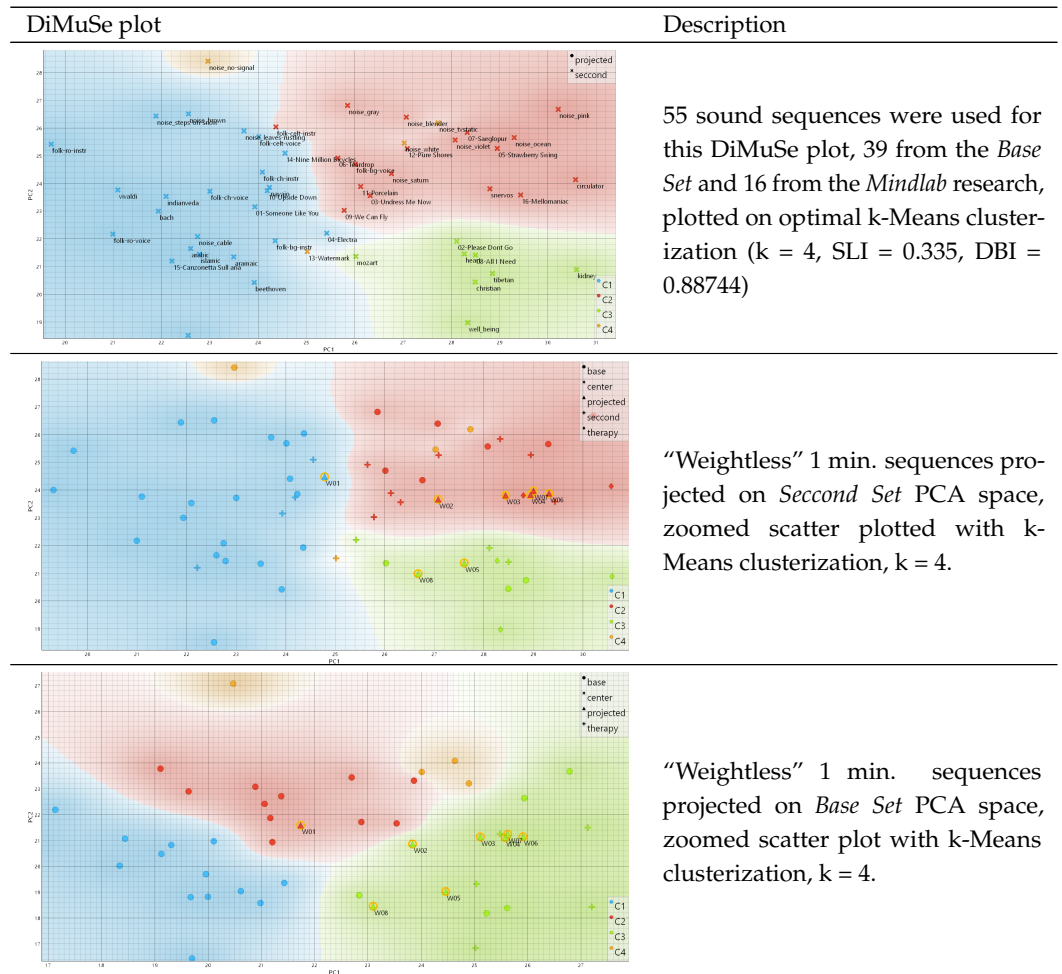
#### 3.7.3. The “Weightless” song as reference

The “Weightless” creators mentioned during an interview that by design the song was intended to have an increasing therapeutic effect during the first part of the song, and gradually decrease it towards the end. . For this paper, the researchers intended to further test the DiMuSe method to see if such a behaviour could be detected by creating the following protocol:

1. Create a new projection plan, named *Second Set* by adding to the PCA analysis all the songs used by the *Mindlab* research, except “Weightless”, meaning another 16 one minute song sequences.
2. Split the “Weightless” song in up to 1 minute sequences (named cronologically from W01 to W08), convert them in DiMuSe vectors and project those on both *Base Set* and *Second Set* PCA plans.
3. Interpret the projected plots

**Table 20.** Weightless song 1 min. sequences projected

DiMuSe plot	Description
-------------	-------------

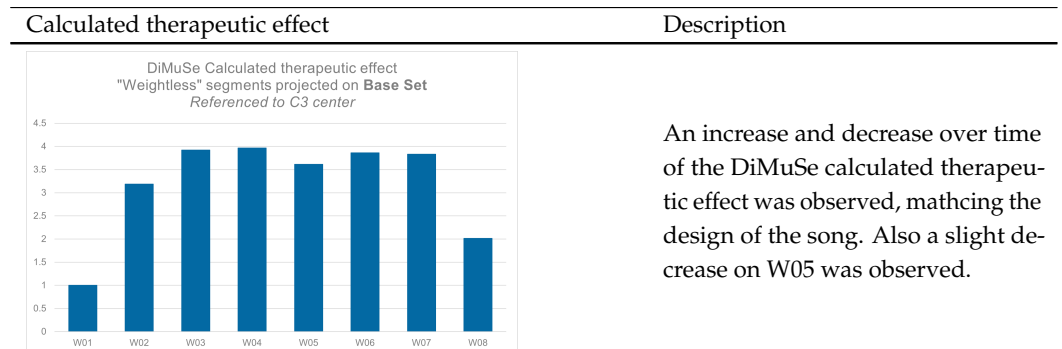


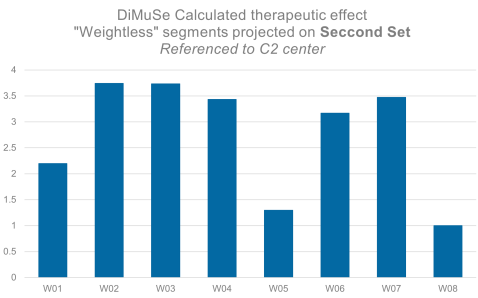
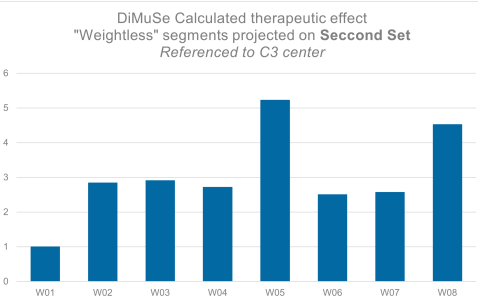
For all plots the k-Means clusterization method was used, with  $k=4$ , and the SLI and DBI values indicating stable clusters. 916  
917

Comparing the Table 20 projections on the *Base Set* and *Second Set*, it was observed that even that the clusters were modified, the projected points representing the “Weightless” sequences were grouped in clusters containing *therapy* class sound sequences, with the exception of segment W01. 918  
919  
920  
921

The DiMuSe calculated therapeutic effect was computed in Table 21, between the “Weightless” segment points and the clusters containing *therapy* class sound sequences, for *Base Set* and *Second Set*. For the *Base Set* the cluster C1 was considered. For the *Second Set*, clusters C2 and C3 contained *therapy* class sound sequences, so both were considered in the evaluation. 922  
923  
924  
925  
926

**Table 21.** DiMuSe calculated therapeutic effect for “Weightless” segments



DiMuSe plot	Description
 <p>DiMuSe Calculated therapeutic effect "Weightless" segments projected on <b>Second Set</b> Referenced to C2 center</p>	<p>A similar increase and decrease as above was observed, with a higher difference for segments W01, W05 and W08.</p>
 <p>DiMuSe Calculated therapeutic effect "Weightless" segments projected on <b>Second Set</b> Referenced to C3 center</p>	<p>A similar pattern was observed as above, with the exception that the segments W05 and W08 displayed higher values than the other segments.</p>

The results suggested that the DiMuSe calculated therapeutic effect score could capture the composers' intention to gradually build a relaxing effect when the segments were projected on the *Base Set*. This case study illustrated that DiMuSe has the potential not only to detect the song with the highest therapeutic properties in a group, but may also track intra-song variations in effectiveness.

#### 4. Conclusions and future work

The aim of this paper was to adapt a mathematical framework originally used for discriminating geophysical nonlinear signals to the analysis of music sequences, with the goal of identifying potential therapeutic effects. To this end, the study introduced the novel *DiMuSe* method, which applies 24 scalar evaluators from diverse scientific domains—including statistics, fractal geometry, nonlinear physics, and complex systems—to transform each music sequence into a multidimensional vector.

By applying Principal Component Analysis (PCA), these vectors were reduced to three principal components (PC1, PC2, and PC3) and projected into a three-dimensional space. This procedure enabled the identification of clusters of sound sequences that shared similar informational structures. Importantly, the therapeutic music sequences consistently grouped together, suggesting that DiMuSe can reveal structural signatures linked to therapeutic efficacy. The method's use of centroid points and distance-based ranking provided a systematic way to evaluate therapeutic similarity.

This study demonstrates, for the first time, that a compact set of interdisciplinary, non-music-specific non-linear scalar evaluators—drawn from statistics, fractal geometry, non-linear physics, and complex systems—can consistently separate music empirically associated with therapeutic effects from heterogeneous sound categories in an unsupervised feature space. Unlike conventional Music Information Retrieval approaches, which rely on stylistic, perceptual, or genre-dependent descriptors, the DiMuSe method identifies clustering patterns based solely on the intrinsic informational structure of sound sequences. The repeated observation that therapeutic music forms a coherent cluster across multiple configurations, while non-therapeutic genres disperse more broadly, provides empirical support for the hypothesis that therapeutic relevance correlates with persistent structural regularities rather than stylistic attributes.

An important and nontrivial outcome of the analysis is the proximity of certain natural and noise-based signals—most notably pink noise and ocean sounds—to the therapeutic music cluster. This convergence is consistent with independent literature reporting regulatory and relaxation effects of such stimuli and strengthens the interpretation that DiMuSe captures structure related to regulation and coherence, rather than culturally learned musical categories. The inclusion of a small subset of religious vocal pieces within this cluster further suggests that slow temporal organization, reduced complexity, and long-range correlations may transcend genre boundaries and function as shared informational markers.

**Limitations.** At the same time, the study’s limitations have significant implications for the interpretation of these results. Most notably, DiMuSe has not been validated in clinical or experimental therapeutic settings, and therefore no causal claims can be made regarding therapeutic efficacy. The present findings should be interpreted as identifying structural correlates of music commonly labeled as therapeutic, rather than as predictors of individual therapeutic outcomes. Furthermore, the method deliberately excludes personalization factors such as listener preference, cultural background, or personal musical history—variables known to exert strong influence on therapeutic engagement and effectiveness.

This exclusion represents both a limitation and a boundary condition of the current work. If individual preference dominates therapeutic response, it is conceivable that subjective factors could override or mask the influence of intrinsic informational structure. However, the observed structural clustering despite diverse cultural origins and genres suggests that informational structure may act as a baseline constraint within which personalization operates, rather than being entirely overwhelmed by it. In this sense, DiMuSe should be understood as identifying a structural potential for regulation or relaxation, which may be necessary but not sufficient for therapeutic impact at the individual level.

From a theoretical standpoint, the contribution of this work lies in reframing therapeutic music analysis away from outcome-driven prediction and toward structure-oriented exploration. DiMuSe provides a reproducible, interpretable framework for detecting latent informational patterns in music signals without reliance on labeled datasets or black-box models. This positions the method as a complementary tool to both clinical music therapy research and supervised machine-learning approaches, offering a means of systematic screening and comparison prior to personalized or experimental validation.

**Future Work.** Future research will focus on extending DiMuSe from a purely structural screening tool toward experimentally grounded and personalized applications, while preserving its interpretability.

First, clinical and non-clinical validation studies are required. Controlled experiments involving physiological and behavioral measurements—such as heart-rate variability, galvanic skin response, or cortisol levels—should be conducted while participants are exposed to music ranked by DiMuSe. This would allow direct testing of whether proximity to the therapeutic cluster correlates with measurable regulatory responses.

Second, the integration of biometric feedback will be pursued through two complementary pathways. In one approach, biometric signals such as EEG, ECG, or respiration patterns would be analyzed independently and used as outcome variables or labels, enabling supervised learning that maps DiMuSe-derived structural features to physiological responses. In an alternative approach, selected biometric descriptors—such as EEG spectral entropy, coherence, or long-range temporal correlations—could be incorporated directly into the PCA feature space alongside audio-derived scalars, allowing joint clustering of stimulus–response pairs. Comparing these two strategies would clarify whether biometric data are better suited as explanatory targets or as structural extensions of the signal space.

Third, personalization strategies will be explicitly addressed by modeling deviations from the DiMuSe baseline. Rather than replacing structural analysis with preference-based selection, future work will investigate how individual responses diverge from structurally similar stimuli. This may involve computing subject-specific distance metrics within the PCA space or adapting clustering thresholds based on listener sensitivity profiles, cultural background, or therapeutic context.

Finally, future iterations of the method will explore the inclusion of sound-specific evaluators—such as rhythm stability or tonal coherence—alongside the current domain-agnostic scalars, in order to assess how perceptual features interact with deeper informational structure. This hybrid approach could further clarify the relationship between signal organization, perception, and therapeutic regulation.

By combining unsupervised structural analysis with biometric validation and personalization layers, future developments of DiMuSe aim to bridge the gap between theoretical signal analysis and applied, data-driven music therapy practice.

**Author Contributions:** Emil A. Canciu (EAC)

Florin Munteanu (FM)

Valentin Muntean (VM)

Dorin-Mircea Popovici (DMP)

Author contributions EAC, FM, VM, and DMP contributed to conceptualization; EAC contributed to writing; EAC, FM contributed to methodology and resources; FM, VM and DMP contributed to reviewing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets generated and/or analysed during the current study are available in the [DiMuSe] repository, [<https://github.com/emilbrasov/dimuse>]. The file Readme.md contains the steps to run the Orange Data Mining code.

**Acknowledgments:** No funding was received to assist with the preparation of this manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism
DIMUSE	Discriminating Music Sequences Method
PCA	Principal Component Analysis
SLI	Silhouette Score
DBI	Davies–Bouldin index

## References

1. Machado Sotomayor, M.J.; Arufe-Giráldez, V.; Ruíz-Rico, G.; Navarro-Patón, R. Music Therapy and Parkinson's Disease: A Systematic Review from 2015–2020. *International Journal of Environmental Research and Public Health* **2021**, *18*, 11618. <https://www.mdpi.com/1660-4601/18/21/11618>, <https://doi.org/10.3390/ijerph182111618>.
2. Franco, L.S.; Shanahan, D.F.; Fuller, R.A. A Review of the Benefits of Nature Experiences: More Than Meets the Eye. *International Journal of Environmental Research and Public Health* **2017**, *14*, 864. <https://www.mdpi.com/1660-4601/14/8/864>, <https://doi.org/10.3390/ijerph14080864>.

3. Thompson, N.; Odell-Miller, H.; Underwood, B.R.; Wolverson, E.; Hsu, M.H. How and Why Music Therapy Reduces Distress and Improves Well-Being in Advanced Dementia Care: A Realist Review. *Nature Mental Health* **2024**, pp. 1–11. <https://www.nature.com/articles/s44220-024-00342-x>, <https://doi.org/10.1038/s44220-024-00342-x>. 1047–1050
4. Moore, K.S. A Systematic Review on the Neural Effects of Music on Emotion Regulation: Implications for Music Therapy Practice. *Journal of Music Therapy* **2013**, *50*, 198–242. <https://dx.doi.org/10.1093/jmt/50.3.198>, <https://doi.org/10.1093/jmt/50.3.198>. 1051–1053
5. Hernando Requejo, V. Epilepsy, Mozart and his sonata K.448: is the "Mozart effect" therapeutic? *Revista de Neurología* **2018**, *66*, 308. <https://www.neurologia.com/articulo/2017460>, <https://doi.org/10.33588/rn.6609.2017460>. 1054–1056
6. T. Zaatar, M.; Alhakim, K.; Enayeh, M.; Tamer, R. The Transformative Power of Music: Insights into Neuroplasticity, Health, and Disease. *Brain, Behavior, & Immunity - Health* **2024**, *35*, 100716. <https://www.sciencedirect.com/science/article/pii/S2666354623001308>, <https://doi.org/10.1016/j.bbih.2023.100716>. 1057–1060
7. Särkämö, T.; Sihvonen, A.J. Golden Oldies and Silver Brains: Deficits, Preservation, Learning, and Rehabilitation Effects of Music in Ageing-Related Neurological Disorders. *Cortex* **2018**, *109*, 104–123. <https://www.sciencedirect.com/science/article/pii/S0010945218302867>, <https://doi.org/10.1016/j.cortex.2018.08.034>. 1061–1064
8. Schaffert, N.; Janzen, T.B.; Mattes, K.; Thaut, M.H. A Review on the Relationship Between Sound and Movement in Sports and Rehabilitation. *Frontiers in Psychology* **2019**, *10*. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.00244/full>, <https://doi.org/10.3389/fpsyg.2019.00244>. 1065–1068
9. Wang, X.; Lu, T.; Bendor, D.; Bartlett, E. Neural Coding of Temporal Information in Auditory Thalamus and Cortex. *Neuroscience* **2008**, *154*, 294–303. <https://linkinghub.elsevier.com/retrieve/pii/S0306452208004764>, <https://doi.org/10.1016/j.neuroscience.2008.03.065>. 1069–1071
10. Patterson, R.D.; Uppenkamp, S.; Johnsrude, I.S.; Griffiths, T.D. The Processing of Temporal Pitch and Melody Information in Auditory Cortex. *Neuron* **2002**, *36*, 767–776. <https://linkinghub.elsevier.com/retrieve/pii/S0896627302010607>, [https://doi.org/10.1016/S0896-6273\(02\)01060-7](https://doi.org/10.1016/S0896-6273(02)01060-7). 1072–1075
11. Bonneville-Roussy, A.; Rentfrow, P.J.; Xu, M.K.; Potter, J. Music through the Ages: Trends in Musical Engagement and Preferences from Adolescence through Middle Adulthood. *Journal of Personality and Social Psychology* **2013**, *105*, 703–717. <https://doi.org/10.1037/a0033770>. 1076–1078
12. Schäfer, T.; Mehlhorn, C. Can Personality Traits Predict Musical Style Preferences? A Meta-Analysis. *Personality and Individual Differences* **2017**, *116*, 265–273. <https://www.sciencedirect.com/science/article/pii/S0191886917303215>, <https://doi.org/10.1016/j.paid.2017.04.061>. 1079–1081
13. Li, H.; Lin, X.; Wu, X. Impact of Neural Network-Quantified Musical Groove on Cyclists' Joint Coordination and Muscle Synergy: A Repeated Measures Study. *Journal of NeuroEngineering and Rehabilitation* **2025**, *22*, 233. <https://doi.org/10.1186/s12984-025-01778-7>, <https://doi.org/10.1186/s12984-025-01778-7>. 1082–1085
14. Montavon, G.; Samek, W.; Müller, K.R. Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing* **2018**, *73*, 1–15. <https://www.sciencedirect.com/science/article/pii/S1051200417302385>, <https://doi.org/10.1016/j.dsp.2017.10.011>. 1086–1088
15. Duda, A.T.; Clarke, A.R.; Barry, R.J.; De Blasio, F.M. Changes in Pink and White Noise in Global EEG Power Following Six Weeks of Daily Breath-Focused Mindfulness Meditation or Classical Music Listening. *International Journal of Psychophysiology* **2023**, *188*, 40–41. <https://www.sciencedirect.com/science/article/pii/S0167876023001824>, <https://doi.org/10.1016/j.ijpsycho.2023.05.102>. 1089–1093
16. Meijer, D.K.F.; Jerman, I.; Melkikh, A.V.; Sbitnev, V.I. Biophysics of Consciousness: A Scale-Invariant Acoustic Information Code of a Superfluid Quantum Space Guides the Mental Attribute of the Universe. In *Rhythmic Oscillations in Proteins to Human Cognition*; Bandyopadhyay, A.; Ray, K., Eds.; Springer: Singapore, 2021; pp. 213–361. [https://doi.org/10.1007/978-981-15-7253-1\\_8](https://doi.org/10.1007/978-981-15-7253-1_8), [https://doi.org/10.1007/978-981-15-7253-1\\_8](https://doi.org/10.1007/978-981-15-7253-1_8). 1094–1098
17. Farhood, M.M.; Heeger, D.J.; Marcus, G.; Hasson, U.; Lerner, Y. The Neural Processing of Hierarchical Structure in Music and Speech at Different Timescales. *Frontiers in Neuroscience* **2015**, 1100 1099–1100

9. <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2015.00157/full>, <https://doi.org/10.3389/fnins.2015.00157>. 1101
18. McDonough, J.; Herczyński, A. Fractal Patterns in Music. *Chaos, Solitons & Fractals* **2023**, *170*, 113315. <https://www.sciencedirect.com/science/article/pii/S0960077923002163>, <https://doi.org/10.1016/j.chaos.2023.113315>. 1102
19. Hespanha, J.P. *Linear Systems Theory, 2018*, 2 ed.; Princeton University Press: Princeton, NJ, 2018. 1103
20. Strogatz, S.H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2 ed.; CRC Press: Boca Raton, 2018. <https://doi.org/10.1201/9780429492563>. 1104
21. Murthy, Y.V.S.; Koolagudi, S.G. Content-Based Music Information Retrieval (CB-MIR) and Its Applications toward the Music Industry: A Review. *ACM Comput. Surv.* **2018**, *51*, 45:1–45:46. <https://doi.org/10.1145/3177849>, <https://doi.org/10.1145/3177849>. 1105
22. Ramírez, J.; Flores, M.J. Machine Learning for Music Genre: Multifaceted Review and Experimentation with Audioset. *Journal of Intelligent Information Systems* **2020**, *55*, 469–499. <http://link.springer.com/10.1007/s10844-019-00582-9>, <https://doi.org/10.1007/s10844-019-00582-9>. 1106
23. Munteanu, F.; Zugravescu, D.; Ioana, C.; Suteanu, C. Methodology for Geophysical Systems Discrimination / Evaluation. In Proceedings of the 58th EAGE Conference and Exhibition, Amsterdam, 1996; p. cp. <https://www.earthdoc.org/content/papers/10.3997/2214-4609.201408810>, <https://doi.org/10.3997/2214-4609.201408810>. 1107
24. Munteanu, F.; Ioana, C.; Suteanu, C.; Cretu, E. Smoothing Dimensions For Time Series Characterization. *Fractals* **2011**, *03*. <https://doi.org/10.1142/S0218348X95000254>. 1108
25. Takayasu, H. *Fractals in the Physical Sciences*; Manchester Univ Pr: Manchester, 1991. 1109
26. Chicco, D.; Campagner, A.; Spagnolo, A.; Ciucci, D.; Jurman, G. The Silhouette Coefficient and the Davies-Bouldin Index Are More Informative than Dunn Index, Calinski-Harabasz Index, Shannon Entropy, and Gap Statistic for Unsupervised Clustering Internal Evaluation of Two Convex Clusters. *PeerJ Computer Science* **2025**, *11*, e3309. <https://peerj.com/articles/cs-3309>, <https://doi.org/10.7717/peerj-cs.3309>. 1110
27. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53–65. <https://www.sciencedirect.com/science/article/pii/0377042787901257>, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). 1111
28. Idrus, A.; Tarihoran, N.; Supriatna, U.; Tohir, A.; Suwarni, S.; Rahim, R. Distance Analysis Measuring for Clustering Using K-Means and Davies Bouldin Index Algorithm. *TEM Journal* **2022**, pp. 1871–1876. [https://www.temjournal.com/content/114/TEMJournalNovember2022\\_1871\\_1876.html](https://www.temjournal.com/content/114/TEMJournalNovember2022_1871_1876.html), <https://doi.org/10.18421/TEM114-55>. 1112
29. Helmenstine, A. Colors of Noise - White, Pink, Brown and More. <https://sciencenotes.org/colors-of-noise-white-pink-brown-and-more/>, 2022. 1113
30. Choi, S.; Park, J.I.; Hong, C.H.; Park, S.G.; Park, S.C. Accelerated Construction of Stress Relief Music Datasets Using CNN and the Mel-scaled Spectrogram. *PLOS ONE* **2024**, *19*, e0300607. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0300607>, <https://doi.org/10.1371/journal.pone.0300607>. 1114
31. Weightless – Marconi Union Ft Lyz Cooper - The British Academy of Sound Therapy. <https://britishacademyofsoundtherapy.com/research/weightless-marconi-union-ft-lyz-cooper/>. 1115