Exploratory analysis of sound datasets in musical acoustics: the case of drumheads

Michael Starakis[1,*], Chrisoula Alexandraki[1], Rolf Bader[2], Maximos Kaliakatsos-Papakostas[1]

[1] Department of Music Technology and Acoustics, Hellenic Mediterranean University, 74133 Rethymnon, 5 Greece
[2] Institute of Systematic Musicology, University of Hamburg, 20354 Hamburg, Germany
[*] ddk216@hmu.gr (corresponding)

ABSTRACT

The increasing progress in the development of high-performance computing systems and artificial intelligence reveals new perspectives in correlating timbral texture to the structural characteristics of musical instruments. Training correlating models, e.g. deep learning neural networks, requires aggregating large volumes of data and developing tools for Exploratory Data Analysis (EDA). This article presents AudioInsight, a web application for the audiovisual exploration of sound datasets in music acoustics. This application is a byproduct of our research in drumhead acoustics, which aims at the computational estimation of tuning and damping strategies for a membrane to reproduce a desired timbre. Through the automatic rendering of histograms for parameter distributions, scatter plots for parameter correlations, and interactive cluster visualizations, AudioInsight enables the assessment of the suitability of a sound dataset for the computational task at hand. Compared to conventional data analysis tools, AudioInsight enables the identification of areas where data may be sparse, dense, or inaccurate, and it highlights complex, nonlinear relationships between physical properties and perceptual characteristics that might otherwise remain unnoticed.

ΠΕΡΙΛΗΨΗ

Η αυξανόμενη πρόοδος στην ανάπτυξη υπολογιστικών συστημάτων υψηλών επιδόσεων και στις τεχνολογίες τεχνητής νοημοσύνης αναδεικνύει νέες προοπτικές στη συσχέτιση της ηχοχρωματικής υφής με τα κατασκευαστικά χαρακτηριστικά των μουσικών οργάνων. Η εκπαίδευση συσχετιστικών μοντέλων, όπως τα νευρωνικά δίκτυα βαθιάς μάθησης, απαιτεί τη συγκέντρωση δεδομένων μεγάλου όγκου και την ανάπτυξη εργαλείων για Εξερευνητική Ανάλυση Δεδομένων (Exploratory Data Analysis - EDA). Το παρόν άρθρο παρουσιάζει το AudioInsight, μια διαδικτυακή εφαρμογή για την οπτικοακουστική εξερεύνηση συνόλων ηχητικών δεδομένων στη μουσική ακουστική. Η εφαρμογή αυτή αποτελεί παραπροϊόν της έρευνάς μας στην ακουστική των τυμπάνων, η οποία στοχεύει στην υπολογιστική εκτίμηση του τρόπου χορδίσματος και απόσβεσης μιας μεμβράνης προκειμένου αυτή να αναπαράγει ένα επιθυμητό ηχόχρωμα. Μέσω ιστογραμμάτων για κατανομές παραμέτρων, διαγραμμάτων διασποράς για συσχετίσεις παραμέτρων και διαδραστικών *απεικονίσεων συστάδων, η εφαρμογή AudioInsight επιτρέπει την εκτίμηση της καταλληλόλητας ενός συνόλου ηχητικών δεδομένων για το υπολογιστικό πρόβλημα για το οποίο προορίζεται. Συγκριτικά με τα συμβατικά εργαλεία ανάλυσης δεδομένων, το AudioInsight επιτρέπει τόσο τον εντοπισμό περιοχών στις οποίες τα δεδομένα μπορεί να είναι αραιά, πυκνά ή ανακριβή, όσο και την ανάδειξη σύνθετων, μη γραμμικών σχέσεων ανάμεσα σε φυσικές ιδιότητες και αντιληπτικά χαρακτηριστικά που σε διαφορετική περίπτωση θα παρέμεναν απαρατήρητες.*

Introduction

Advances in AI-driven techniques have improved the ability to process large datasets of acoustic signals, offering new insights into the physical and perceptual aspects of musical instruments. These techniques require large, high-quality sound datasets to estimate physical sound traits from sound and to train deep learning models for sound recognition. While large collections of audio data are currently available, finding datasets with accurate annotations that link physical properties to perceptual traits remains challenging and often requires human curation and expert knowledge. Relevant methodologies assisting this task are presented in the domain of Exploratory Data Analysis (EDA). EDA is a fundamental methodology in data science designed for summarizing, visualizing, and understanding complex datasets, by employing techniques such as data visualization, statistical summaries, and dimensionality reduction [1 - 5]. In musical acoustics, and especially in large multidimensional audio datasets, EDA tools aim to reveal underlying patterns, detect anomalies, test hypotheses, and validate assumptions through visual and statistical methods, allowing researchers to efficiently identify trends and correlations within complex data structures that might otherwise go unnoticed. This paper presents the AudioInsight application [6], which was developed to address these challenges in our recent research endeavors focusing on drumhead acoustics [7].

1. Motivation: Drumhead dataset

The development of the AudioInsight application emerged from our research on drumhead acoustics [7, 8], which aimed to computationally infer the amount and distribution of damping material to apply on the surface of a membrane in order to achieve a desired sound texture. This inverse acoustic problem presented unique challenges, particularly in handling and exploring vast amounts of data. AudioInsight was conceived as a solution to these challenges, providing a versatile visual and auditory exploration tool capable of revealing complex relationships between the physical properties and sound characteristics of drumheads. The drumhead dataset consists of approximately 11,000 synthetic sounds generated using a Finite Difference Time Domain (FDTD) algorithm [9] that models the behavior of a vibrating circular membrane. This model assumes the distribution of malleable, paste-like material on the surface of the membrane to alter its vibrational behavior and hence the sound it generates. Six paste distribution patterns were applied to vary the sounds. These patterns were inspired by commercial and custom drum dampeners, commonly used by percussionists, as shown in Fig. 1.1. By varying the area covered by paste (i.e., the mass per unit area of the paste in each pattern) and the strike position of membrane excitation, the model generated different sounds.

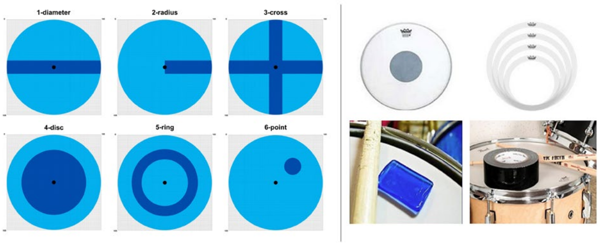

Figure 1.1 Pattern cases (left). Common dampening practices (right).

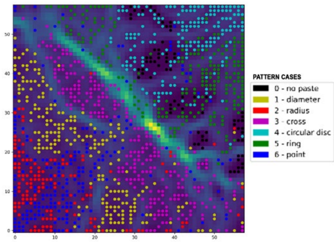

Figure 1.2 Trained SOM. Dark background colors reveal similarities, and light colors reveal strong neuron similarities.

This resulted in a large sound dataset, which was deemed appropriate for training deep neural networks to address the intended computational task. To explore the resulting dataset, for example, in terms of understanding whether similar damping schemes produce similar sounds, several clustering methods were leveraged. Initially, Self-Organizing Maps (SOMs) were used to map the distribution of the frequency ratios of sound overtones (i.e., characterizing sound texture) against the damping patterns on a 2D plane. As shown in Fig. 1.2 these representations revealed a distinguishing line between line patterns and circular patterns [8]. This fact motivated further investigation through additional clustering techniques, e.g. PCA [10], t-SNE [11, 12], and LDA [13], among others. These methods allowed visualizing high-dimensional data, namely the raw audio signal in two and three dimensions, thus permitting the identification of distinct sound clusters and providing insight into the influence of paste patterns and impact points on the resulting sound signals. An example of clustering membrane sounds is shown on Fig. 1.3. Each point represents a different sound signal and each color depicts a different damping pattern. For such a large dataset with high-dimensional parameter spaces, where each sound sample is associated with numerous parameters, conventional audio analysis methods fall short in providing a deeper understanding of the dataset, including the identification of interesting sounds, their complex relationships and data patterns that may be easily go unnoticed. To address these challenges, AudioInsight was developed as a demo application focused on drumheads. Through its interactive visualizations and clustering capabilities, it became possible to uncover complex, non-linear relationships between physical properties and perceptual features that might otherwise remain hidden. These insights not only advanced our research objectives but also demonstrated the potential of AudioInsight as a powerful tool for systematic EDA of large audio datasets in musical acoustics.

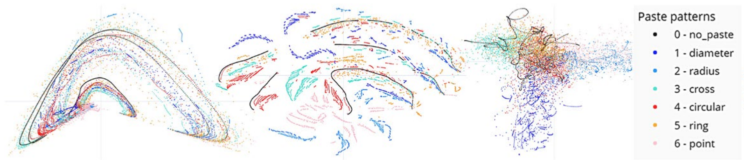

Figure 1.3 Clustering methods plots. From left to right: PCA, T-SNE, LDA.

2. The AudioInsight Application

The design of AudioInsight focuses on delivering a powerful, yet user-friendly interactive system for comprehensive data exploration and analysis, accessible to researchers, instrument manufacturers, and musicians. Its features include visual and audio analysis tools, clustering representations, audio previews, and feature graphs that help users to uncover patterns and correlations that are not obvious through regular methods, making it particularly valuable for machine learning tasks. For instrument manufacturers and musicians, AudioInsight serves as a valuable tool guiding instrument design and tuning. It enables users to explore and understand the acoustic properties of their instruments, search for similar sounds, and explore relationships as well as timbral discrepancies. The user interface of the application is organized into six pages. Except from a home page, the other pages present the dataset according to selected descriptive parameters. Users select their desired dataset file and choose specific parameters for analysis, through dropdown menus and checkboxes. For categorical parameters, which in the case of membranes are the paste patterns and the strike positions, the graphical representations can be filtered to display data distributions for specific labels. For continuous variables, like the mass of added paste and the fundamental frequency of the sounds, users can scale, zoom, and pan the graphs to focus on particular value ranges. The AudioInsight application is freely available online at: http://musicolab.hmu.gr:8050. In the following sections, we provide a brief description of the functionality of each page.

2.1 The dataset Page

The Dataset page (Fig. 2.1) features an interactive and customizable table for exploring the dataset contents, with options to filter and adjust the display of the

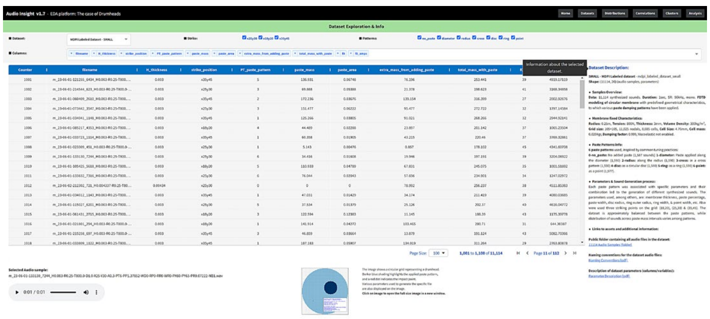

Figure 2.1 The Dataset Page

data. Rows correspond to dataset instances, i.e., sound files and columns to descriptive parameters. In this case, columns correspond to the physical parameters that describe the membrane and the damping material, such as strike position, mass of the applied paste, and more. The table supports data pagination and allows users to show or hide specific columns, as well as to sort, filter, and search for parameter values. Clicking on a row triggers audio playback of the corresponding sound file and displays its drumhead grid image, which visually represents the paste pattern applied to the drumhead. The right-side panel presents a description of the dataset and provides detailed information about the data being explored.

2.2 Distributions Page

The Distributions page (Fig. 2.2) helps users visualize and analyze the statistical distributions of various parameters within the sound dataset. This page allows users to select a parameter and display its distribution across the entire dataset or by filtering out categorical parameters, such as strike positions and pattern cases. For example, the page can display the distribution of paste mass when the distribution of paste mass when distributed in circular patterns, such as rings and discs. Hovering over the plot reveals the y-axis value and the range of x-axis values represented by each bar in the histogram. Users can customize the appearance of the plots through scaling, zooming, and panning, and they can also save the plot as an image to disk. A right-side panel provides context about the parameter being visualized, including its calculation formula and statistical metrics like minimum, maximum, median, and standard deviation. Using the Distributions page, researchers can assess whether the dataset is balanced with respect to different parameters, helping them weigh the suitability of the data for their investigation.

2.3 Correlations Page

The Correlations page (Fig. 2.3) offers tools for exploring and visualizing relationships between pairs of parameters in the sound dataset. It features an interactive scatter plot with a range of customization options. Users can select the parameters for the x and y axes and filter out specific values of categorical data. The plot updates in real time, allowing for immediate visualization of correlations. Additional features include a Kernel Density Estimation (KDE) line, histograms and colormaps based on a third parameter, which enhances the depth of the analysis.

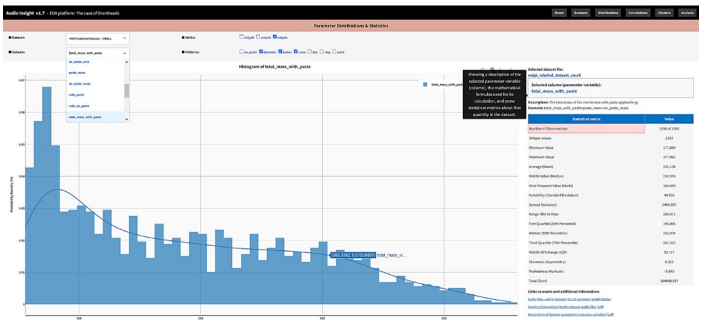

Figure 2.2 The Distributions Page

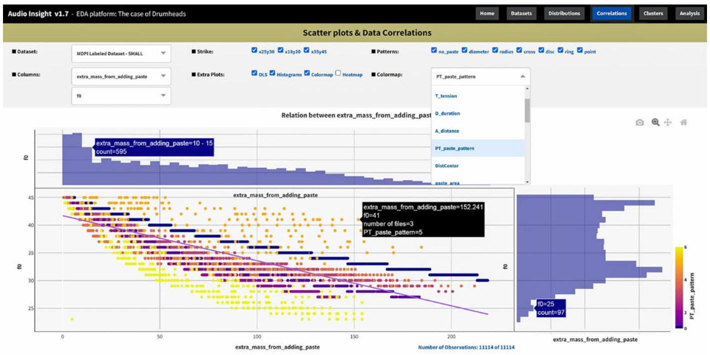

Figure 2.3 The Correlations Page

Users can also toggle a heatmap view for density visualization. For example, in Fig. 2.3, the page illustrates a correlation plot showing how the additional mass applied via paste relates to the fundamental frequency (f0) of the resulting sound. Each paste pattern case is represented by a different color on the plot. According to the diagram, various hypotheses can be formulated regarding how increasing the mass of the membrane alters its perceived pitch and how different patterns result in varying correlations between fundamental frequency (f0) and mass. The plot supports zooming and panning, and provides hover functionality that allows users to view detailed information about individual data points.

2.4 Clusters Page

The Clusters page (Fig. 2.4) provides a tool for audiovisual navigation within the dataset by using advanced clustering techniques. It features a central plot showing the data points resulting from the applied clustering method on the input data set, that in this case are the raw values of wav audio data. The clustering plot can be viewed in 2D or 3D and the currently supported clustering methods include PCA, t-SNE, and others such as PaCMAP, UMAP, and LDA. Users can specify the number of files to be previewed and filtered according to categorical data. Hovering over data points on the plot triggers real-time audio playback of the corresponding sound file, providing an auditory dimension to the visual exploration. The right-side panel displays details about the parameters of each sound file, as well as information about the utilized clustering method. The Clusters Page is designed to combine visual, auditory, and textual information, for a multisensory experience of data exploration.

2.5 Analysis Page

The Analysis page (Fig. 2.5) is designed as a comprehensive tool for real-time sound analysis and digital signal processing. It combines the visualization of data clusters with a set of spectral analysis graphs. The page is split into two sections: a clustering plot area on the left and an analysis panel on the right. The interactive clustering plot allows for audio playback when hovering over a data point, while also performing an audio analysis of the corresponding sound file. The analysis panel displays the sound waveform, its Short-Term Fourier Transform (STFT) spectrogram, and a phase-space portrait visualization. Each updates in real-time as

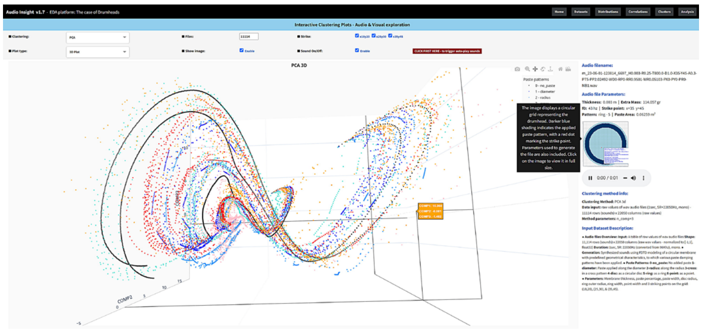

Figure 2.4 The Clusters Page

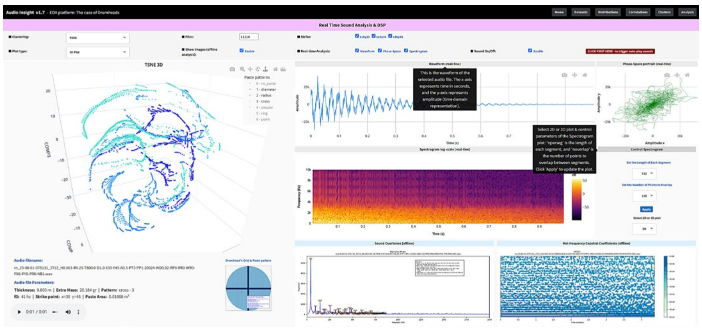

Figure 2.5 The Analysis Page

users explore different data points in the cluster area. Additionally, the analysis panel features an overtone visualization and an MFCC spectrogram. These specific visualizations are generated offline, prior to loading the dataset into the application.

3. Conclusions and future perspectives

AudioInsight presents an example of a web application aiding research in data-driven musical acoustics. Its development was motivated by our research on an inverse acoustic problem, that of computationally inferring how to damp or tune a membrane to produce a desired sound texture. Such computational problems require large amounts of data that are difficult to handle and challenging to assess in terms of accuracy, adequacy and balance for addressing the required task effectively. Through histograms for parameter distributions, scatter plots for parameter correlations and interactive clustering visualizations, the application allows gaining a deeper understanding of data by: (a) revealing areas where data may be sparse, congested, or inaccurate, and (b) uncovering complex, non-linear relationships between physical properties and perceptual features that might otherwise remain hidden within the dataset. By revealing such issues, researchers are guided to refine their datasets, whether by augmenting with more data, applying data augmentation techniques, or rejecting outliers with faulty parameter values. Through uncovering relationships, researchers can develop hypotheses that may lead to alternative data representations, such as audio features, to guide the analytical process for data modeling. AudioInsight has been designed to allow loading alternative sound datasets without requiring re-engineering of its exploratory functionalities. Current development efforts are focused on the specification of a protocol describing the expected data hierarchies and parameter mappings through JSON files. To validate the specification, we are working on integrating a sound dataset of over 200,000 cymbal sounds from well-known manufacturers. In the future, we aspire that AudioInsight can serve as a central hub for researchers, musicians, and instrument manufacturers, facilitating knowledge sharing and collaboration on perceptually informed modifications in musical instrument design.

4. References

[1] J.W. Tukey, "Exploratory Data Analysis." Addison-Wesley, Reading, MA (1977).

[2] J.T. Behrens, "Principles and Procedures of Exploratory Data Analysis," *Psychological Methods*, vol. 2, no. 2, pp. 131-160 (1997).

[3] C. Ahlberg and B. Shneiderman, "Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1994).

[4] H. Hochheiser and B. Shneiderman, "Interactive Exploration of Time Series Data," in *Proceedings of the Fourth International Conference on Information Visualization* (2003).

[5] H.L.M. Lam, "Visual Exploratory Analysis of Large Data Sets: Evaluation and Application," T, University of British Columbia (2008).

[6] M. Starakis and C. Alexandraki, "AudioInsight: Online Exploration of Large Audio Datasets for Musical Acoustics," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, Erlangen, Germany, 2024, pp. 1-10.

[7] C. Alexandraki, M. Starakis, P. Zervas, and R. Bader, "Inferring Drumhead Damping and Tuning from Sound Using Finite Difference Time Domain (FDTD) Models," *Acoustics*, vol. 5, no. 3, pp. 798-816 (2023).

[8] C. Alexandraki, M. Starakis, R. Bader, and P. Zervas, "Machine Learning of Finite-Difference Time Domain (FDTD) Physical Modelling Sound Simulations of Drumhead Paste Pattern Distributions," in *Proc. European Acoustics Association Forum Acusticum 2023*, Torino, Italy, Sept. 2023, pp. 2217-2224.

[9] R. Bader, "Finite-Difference Model of Mode Shape Changes of the Myanmar Pat Wain Drum Circle Using Tuning Paste," in *Proc. Meetings on Acoustics 2016*, AIP Publishing, New York, NY, USA, 2016, vol. 29, p. 035004.

[10] T. Jolliffe and J. Cadima, "Principal Component Analysis: A Review and Recent Developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202 (Apr. 2016).

[11] G. Hinton and S. Roweis, "Stochastic Neighbor Embedding," in *Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS'02)*, Cambridge, MA, USA, 2002, pp. 857-864. MIT Press.

[12] L. van der Maaten and G.E. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605 (2008).

[13] R.A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, pp. 179-188 (1936).