ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/353328100

A Survey of Music Visualization Techniques

Article *in* ACM Computing Surveys · July 2021

DOI: 10.1145/3461835

CITATIONS
24

READS
2,219

3 authors:

Hugo Lima
Federal University of Pará
4 PUBLICATIONS **29** CITATIONS

SEE PROFILE

Carlos Gustavo Resque dos Santos
Federal University of Pará
50 PUBLICATIONS **239** CITATIONS

SEE PROFILE

Bianchi Serique Meiguins
Federal University of Pará
122 PUBLICATIONS **502** CITATIONS

SEE PROFILE

All content following this page was uploaded by Hugo Lima on 17 April 2022.

The user has requested enhancement of the downloaded file.

A Survey of Music Visualization Techniques

HUGO B. LIMA, CARLOS G. R. DOS SANTOS, and BIANCHI S. MEIGUINS,
Federal University of Pará, Brazil

Music Information Research (MIR) comprises all the research topics involved in modeling and understanding music. Visualizations are frequently adopted to convey better understandings about music pieces, and the association of music with visual elements has been practiced historically and extensively. We investigated papers related to music visualization and organized the proposals into categories according to their most prominent aspects: their input features, the aspects visualized, the InfoVis technique(s) used, if interaction was provided, and users' evaluations. The MIR and the InfoVis community can benefit by identifying trends and possible new research directions within the music visualization topic.

CCS Concepts: • **Human-centered computing** → **Visualization techniques**; **Visualization design and evaluation methods**; *Visualization toolkits;*

Additional Key Words and Phrases: Music visualization, information visualization, music information retrieval, human-computer interaction

ACM Reference format:
Hugo B. Lima, Carlos G. R. dos Santos, and Bianchi S. Meiguins. 2021. A Survey of Music Visualization Techniques. *ACM Comput. Surv.* 54, 7, Article 143 (July 2021), 29 pages.
https://doi.org/10.1145/3461835

1 INTRODUCTION

Music Information Research (MIR) comprises "all the research topics involved in modeling and understanding music" [78]. It is a popular and active field [25], especially nowadays when many feature extraction techniques have been proposed [22, 23, 47, 69, 70]. This evokes the necessity for the development of tools and techniques to explore the data extracted. **Information Visualization (InfoVis)** has been consistently employed to augment human analytical capabilities. It enables insights and better understandings for data analysis [59], and this can be equally extended to the MIR field. **Music Visualization (MusicVis)** can be defined as "the visual representation of a musical performance on a static or dynamic canvas expressively with computer graphics" [29]. In this MusicVis survey, we analyzed 51 proposals, investigated trends, general structures, and open research problems.

Despite being an aural phenomenon, music is commonly depicted by visual means. A musical picture can convert the unidirectional sound signal into many more dimensions spatially represented [31]. Theories relating colors with music have been developed early in human history, e.g.,

Authors' address: Hugo B. Lima, Carlos G. R. dos Santos, and Bianchi S. Meiguins, Federal University of Pará, Augusto Correa St., 01, Belém, Pará, Brazil, 66075-110; emails: {hugoblim, gustavo.cbcc, bianchism}@gmail.com.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
0360-0300/2021/07-ART143 $15.00
https://doi.org/10.1145/3461835

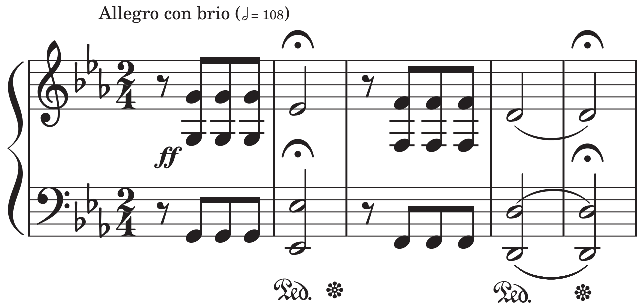

Fig. 1. Beethoven's Fifth Symphony—the five first measures for piano solo. The visual elements comprise the *Common Music Notation* (CMN) standard.

the ancient Greeks associated the seven colors of the rainbow with music notes [77]. Music Visualization encourages the listener to speculate relations, visualize similarities, and extract more information about a performance. Music pieces frequently show some coherence or similarity over their duration; visual representations make it easier to perceive certain patterns within a music piece [9]. However, according to Isaacson [31], the visualizations should be grounded on both musical knowledge and understanding of human cognition.

Audio is traditionally visually represented as a monolithic block that spans through time-domain, for example: the **Common Music Notation (CMN)**, as in the *sheet music* (Figure 1), widely used by western music. CMN reflects many centuries of accumulated user feedback and collective wisdom synthesized in a flexible and adaptive representation [31]. It comprises signs representing tonal and rhythmic parts of music and it aims to communicate the most crucial aspects to perform a piece of music. The degree of details varies, and in the end the execution of a music represent invariably an interpretation of the written piece. Aspects like articulation or *tempo*[1] changes are often derived from the performer's perspective.

A visualization usage can be summarized in terms of *why the user needs it*, *what data are shown*, and *how the idiom is designed* [59]. Considering music visualizations have to similarly take into account these terms, we sorted the papers into topics, according to their prominent aspects, i.e., on each visualization, we checked what *input* was used; what feature(s) was/were visualized; which InfoVis technique(s) was employed; what was the proposal's goals; how the users interacted; and evaluations of the visualizations proposed.

Music visualizations can be divided into two fundamental types: *augmented scores* and *performance visualization* [28]. The augmented score extends CMN's expressiveness by employing more intuitive visual representations, which can aid the process of analysis by an expert or aid a beginner in music notation. A performance visualization is primarily used in the analytical process. It comprehends the representation of music features such as volume, pitch, mood, melody, instruments, tempo, and so on. In this survey, we considered both, augmented scores and performance visualization cases.

Visualization is frequently considered the final step in a process chain, thus it requires an initial step, that is, a proper *input*. The papers collected tended to focus on Signal Processing techniques and the ***Music Instrument Digital Interface* (MIDI)**. Signal Processing relies primarily on the Fourier Transform as a way to extract features from an audio signal (e.g., **Mel-Frequency Cepstrum Coefficients (MFCC)** [91], Chroma [58]). MIDI is a symbolic representation of notes on/off and other attributes, such as: velocity (MIDI term for "note volume"), music notation, pitch,

[1] *Tempo*: Time or measure [87].

vibrato, panning, and clock signals (which set tempo) [58]. The first contribution of this survey is to investigate the inputs used to produce the visualizations.

Considering no visualization can represent all the possible interesting features, the focus of each visualization tends to be context-dependent [31]. Nonetheless, most of the features represented are related to harmony (e.g., chords visualization), timbre, instrument visualization, synchronization of representations (e.g., implementing the Dynamic Time Warping algorithm), and so on. Hence, most of the proposals are inclined to display features that are not traditionally (i.e., not represented by the CMN) or directly represented (e.g., harmony).

Between a rich and detailed sheet music and the most basic line graph there are myriad possibilities for visual representation with varying degrees of previous/hard knowledge (i.e., each person's preconceptions), learning curves, and possible insights. InfoVis offers many techniques aiming at representing data and the development of new visualizations is an active research area. Music, as temporal data, can be expressed by flow graphs, line charts, colored gradients, glyphs, avatar representations, faces, and so on [11, 44, 59, 82]. Interaction is additionally provided by many InfoVis proposals, as they can provide more possibilities in the aspect of analysis and insights. Tools frequently rely on the combination of many already established visualizations, for example, Cui et al. [11] use the combination of flow graphs and glyphs to represent trends and events in text streams.

The music visualization topic is at a calm pace when compared with the InfoVis topic alone. In addition, there is not a substantial connection between the InfoVis and the MIR community, and we argue one could benefit from the other. Therefore, this survey also intends presenting the state-of-the-art of Music Visualization. We aim to benefit the MIR community by providing the missing spots in which new researches can be made; which trends are promising and a historical evolution of how music is visually represented in the digital era.

This survey is structured as follows: Section 2 describes our methodology and scope. We describe how the papers were collected, what we considered when collecting the papers, and what is not part of our scope. Next, Section 3 presents the results of the research by dividing the papers into categories, making it easier to grasp how the proposals were organized. Section 4 provides a discussion about the papers, what remains the essential topics, trends, and voids in the literature. Finally, Section 5 enlists the contributions, understandings and possible future works.

2 SCOPE AND METHODOLOGY

In this section, we describe the scope and our research methodology. In this survey, we investigated MusicVis papers, collected from both types of music visualization: augmented scores and performance visualizations. The proposals considered are music visualizations for the visual representation of music recordings and real-time performances. We excluded papers related to the visualization of music collections (e.g., online music libraries). We additionally did not include visualization of music playlists and playing history. Therefore, this survey investigates only visualizations of a single music piece, recorded or performed in real-time.

We collected papers from four repositories: ACM digital library (24 papers), IEEE Xplore digital library (53 papers), DBLP computer science bibliography (42 papers), and EG digital library (3 papers), using the snowball sampling method with the keyword "music visualization" for query; selecting papers based on the title, keywords, and abstract. In that case, we removed duplicates and papers that were not focused on the proposal of a music visualization technique. Later, we selected more papers from the references of the papers already gathered. After filtering out all the papers that did not treat proposals of visualization, we ultimately achieved the number of 51 papers concerning proposals of techniques for music visualization, focusing on recording or live performance of a sole piece.

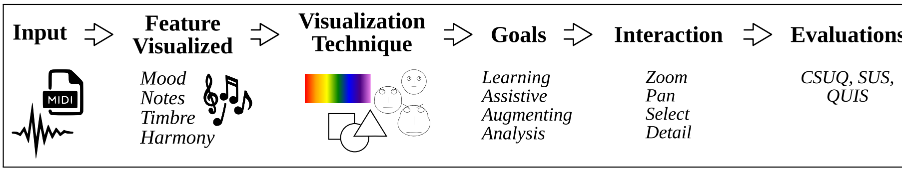

Fig. 2. Survey's pipeline.

3 A SURVEY OF MUSIC VISUALIZATION

After collecting and filtering out the papers, we analyzed how to divide the papers into classes according to their characteristics. We noticed research familiarities and used that as a starting point for the organization of this survey (our pipeline: Figure 2). We verified that, in general, the proposals had the following lineament: description of the input(s) used, description of which music property was represented, notes about the visualization technique used, the goal of the proposal, and finally, if executed or not, an evaluation with users. As follows, we adopted that structure to organize the papers collected for this survey. In the following subsection, we analyzed the most common inputs used.

3.1 What Data Input Was Used?

In this section, we analyze the papers according to their data input. The choice of a particular input affects the following possible mappings (i.e., input feature mapped to higher-level characteristics), usually the step before generating visualizations [59]. Table 1 reveals the distribution of papers according to the input used. The use of an input is unrestricted, as some proposals use both MIDI and audio recordings, e.g., Foote [18], Kim et al. [37], Lim and Raphael [42], Machida and Itoh [46], Mardirossian and Chew [51], Politis et al. [68], Snydal and Hearst [84], Valbom and Marcos [92]. Nevertheless, audio and MIDI data are the most used inputs (48 (94%) papers).

3.1.1 MIDI. MIDI is a protocol initially implemented to allow the communication between instruments from distinct brands. Withal, the public found other usages, such as: music editing on ***Digital Audio Workstations*** **(DAW)** (e.g., ProTools[2] and Cubase[3]), composition aided by computers (e.g., Sibelius[4] and MuseScore[5]) and music theory analysis. Twenty-eight proposals used MIDI as input with varied intentions. Musicologists can benefit from proposals such as Hayashi et al. [24], Malandrino et al. [49], Snydal and Hearst [84], Wattenberg [95], as they augment the traditional score, allowing a more precise analysis regarding music structure (e.g., harmony and melodic patterns). MIDI is additionally used for performance comparison (i.e., analysis of a player's performance): Foote [18], Hiraga [26], Hiraga et al. [28], Jin et al. [34], Ox [66], Smith and Williams [83]. Despite already established for a while [45], MIDI is nonetheless widely used and still maintains a growing community, despite the many extraction techniques focused on audio recordings published recently [47, 69, 70]. The growing of MIDI usage occurred due to the popularization of software for music analysis and the development of modern hardware with MIDI communication [73].

3.1.2 Audio Signal. For live and recorded audio, 20 papers extracted audio features such as: magnitude-vectors (e.g., a narrower window sampling gives better time resolution, useful for beat

[2] https://www.avid.com/pro-tools.
[3] https://www.steinberg.net/en/products/cubase/start.html.
[4] https://www.avid.com/SIBELIUS.
[5] https://musescore.org/en.

Table 1. Paper Distribution According to the Input Used for Visualization: MIDI (29), Audio (18), Others (13)

Input	#	Reference
MIDI	29	Bergstrom et al. [3], Cantareira et al. [5], Chan et al. [6], Chew and François [7], Ciuha et al. [8], Fonteles et al. [17], Foote [18], Fujishiro et al. [19], Gumulia et al. [21], Hayashi et al. [24], Hiraga [26], Hiraga and Matsuda [27], Hiraga et al. [28, 29], Jin et al. [34], Lim and Raphael [42], Lima et al. [43], Malandrino et al. [49], Mardirossian and Chew [51], McLean et al. [54], Nanayakkara et al. [61], Ox [66], Politis et al. [68], Smith and Williams [83], Snydal and Hearst [84], Tagliolato [89], Toiviainen [90], Valbom and Marcos [92], Wattenberg [95]
Audio	18	Ferguson et al. [16], Foote [18], Kamolov et al. [35], Kawahara et al. [36], Kim et al. [37], Kosugi [38], Lehtiniemi and Holm [40], Lim and Raphael [42], Machida and Itoh [46], Mardirossian and Chew [51], McLeod and Wyvill [55], Nakano et al. [60], Ohmi [64], Politis et al. [68], Sapp [75], Shi and Yang [79], Snydal and Hearst [84], Soraghan et al. [85]
Other	13	Farbood et al. [15], Kim et al. [37], Lehtiniemi and Holm [40], Machida and Itoh [46], Malandrino et al. [50], Miller et al. [56], Mitroo et al. [57], Prisco et al. [71, 72], Sapp [76], Shirzadian et al. [80], Taenzer et al. [88], Valbom and Marcos [92]

detection), fundamental frequency (f_0), MFCC, Chroma, volume, pitch, harmonics, and so on. Lim and Raphael [42], Snydal and Hearst [84] used audio to augment the score notation, Lim and Raphael [42] used a pitch estimation algorithm to locate, in the score, the notes executed in a recording. Snydal and Hearst [84] used famous recordings from musicians such as Coltrane and Miles Davis' to visualize pitch contour [67] and harmonic patterns [87].

Audio recordings were used by Soraghan et al. [85] to map low-level features (i.e., extracted using signal processing) to semantic descriptors (i.e., higher-level descriptors/adjectives) related to timbre (quality of tone or sound [87] visualization). Lima et al. [43] exploited the audio signal as input for their visualization and extracted features using the Essentia API [4]. Taenzer et al. [88] as well used audio, for instance, to visualize the distinct tracks within a song.

MFCC integrates the Mel-scale and "relates to the perceived frequency, or pitch, of a pure tone to its actual measured frequency" [91]. Foote [18], Kosugi [38], Politis et al. [68] extracted MFCC and employed it as input (e.g., Figure 3 represents audio features provided an audio stream, denoted visually as a colored waveform). *Chroma* was also considered by Politis et al. [68]. This is considered a trend, as Humphrey et al. [30] (one of the pioneers at extracting musical features using deep-learning architectures) notices that the MIR community slowly converged towards a reduced set of features, using predominantly MFCC and *Chroma*.

3.1.3 Other Inputs. Other proposals (8 papers) used different data input. Shirzadian et al. [80] used ***Galvanic Skin Response*** **(GSR)** as a way to measure audience's engagement. Kim et al. [37] used sensors integrated with *Arduino* to assist children with hearing impairment. The popularization of bio-sensing products and its adoption by the scientific community make its use potential for future works [39].

Farbood et al. [15] developed *Hyperscore* to be "a composition system for users with limited or no musical training takes freehand drawing as input. They allowed users to literally sketch their

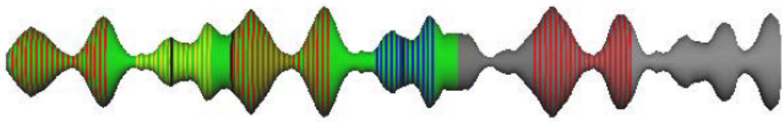

Fig. 3. Misual's visualization uses audio signal as input. As clearly shown in the illustration, the visualization resembles the classic waveform visualization [38] ©Republished with permission of ACM (Association for Computing Machinery), from "Misual: music visualization based on acoustic data," by Naoko Kosugi. 2010. Permission conveyed through Copyright Clearance Center, Inc.

pieces"; they make use of a sketchpad to input notes directly into a *gridded* layout. Lehtiniemi and Holm [40] used song *lyrics* and audio signal processing to pick a background picture that represents a song adequately.

Malandrino et al. [50], Miller et al. [56], Prisco et al. [72] used MusicXML as input for their customized applications, aiming at visualizing the music score and annotate them.

3.1.4 *Summary.*

- MIDI is still widely used as input for music visualization;
- There is a tendency in recent years to employ other types of data input (e.g., MusicXML, bio-sensors, sketchpad);
- MFCC and Chroma are present in most of the recent proposals that use audio as input;

3.2 Which Element Was Visualized?

In this section, we investigate which features were mostly selected to be visually represented. Audio features can be divided generally as low-level and high-level features, depending of the used abstraction. Low-level features are related to the simplest features extracted using signal-processing techniques, such as: frequency, MFCC, Chroma, amplitude, pitch, and so on. High-level features, however, represent a more abstract type of feature; they are: timbre, semantic descriptors, harmony, melody, and so on. Low-level features are frequently used to describe high-level features, e.g., MFCC is customarily used for timbre description; amplitude is used to measure loudness. Thus, we separate into general classes of representations. Finally, Table 2 organizes the references.

3.2.1 Pitch. "Pitch" or "tone" is defined as the position of a sound with reference to the number of vibrations that produce it; it is also defined as the relative height of a sound [87]. Tagliolato [89] represented adjacent pitch classes aiming at presenting pitch evolving through time in a music piece. Aiming at visualizing a whole piece, Hiraga et al. [28] proposed a a visualization capable of representing many features, including pitch.

3.2.2 Timbre. "Timbre" was represented by Soraghan et al. [85] using semantic descriptors. They mapped audio features to semantic descriptors, i.e., adjectives used to describe timbre, such as: "bright-dark, full-empty, dull-sharp, colorful-colorless, and compact-diffused" [94]. Smith and Williams [83] also proposed a visualization of timbre and considered it the most difficult characteristic of a tone to represent. They used MIDI input and grouped instruments with similar timbres: distinct colors differentiating each group. Interestingly, to assist children with hearing impairment, Kim et al. [37] represented timbre with movement: the dynamics of a sound coded as a plant moving its leaves. Fonteles et al. [17] also used colors to represent classes of instruments with similar timbre (e.g., violins, viola, and cellos belonging to the same timbre grouping).

Table 2. Paper Distribution According to the Element Visualized: Pitch (16), Timbre (4), Mood (4), Structure (18), Harmony (10), Melody (9)

Feature	#	Reference
Structure	18	Bergstrom et al. [3], Cantareira et al. [5], Chan et al. [6], Chew and François [7], Ferguson et al. [16], Foote [18], Hayashi et al. [24], Hiraga [26], Hiraga et al. [29], Kim et al. [37], Kosugi [38], Lima et al. [43], Malandrino et al. [49, 50], Ohmi [64], Prisco et al. [71], Sapp [75, 76]
Pitch	16	Chew and François [7], Ciuha et al. [8], Ferguson et al. [16], Fonteles et al. [17], Hiraga et al. [28], Kawahara et al. [36], Lim and Raphael [42], McLeod and Wyvill [55], Mitroo et al. [57], Nanayakkara et al. [61], Ox [66], Sapp [75], Shi and Yang [79], Smith and Williams [83], Snydal and Hearst [84], Tagliolato [89]
Harmony	10	Cantareira et al. [5], Chew and François [7], Ciuha et al. [8], Ferguson et al. [16], Malandrino et al. [49], Miller et al. [56], Prisco et al. [71, 72], Sapp [75], Snydal and Hearst [84]
Melody	9	Cantareira et al. [5], Fonteles et al. [17], Hayashi et al. [24], Kim et al. [37], Politis et al. [68], Prisco et al. [71], Sapp [76], Snydal and Hearst [84], Taenzer et al. [88]
Mood	4	Hiraga and Matsuda [27], Hiraga et al. [28], Lehtiniemi and Holm [40], Lima et al. [43]
Timbre	4	Fonteles et al. [17], Kim et al. [37], Smith and Williams [83], Soraghan et al. [85]

3.2.3 Mood. To visualize "mood," Hiraga et al. [28] classified parts of the music structure with semantic data, instead of simply showing concrete features such as pitch and volume. Hiraga and Matsuda [27] visualized mood using similar-sized squares whose textures represent the mood; they aimed at implementing this visual representation to allow a user to retrieve music based on mood figures. Lehtiniemi and Holm [40] developed a music player with background pictures based on the mood of the song's genre. More recently, Lima et al. [43] extracted mood and other semantic descriptors (they employed SVMs, assisted by the Essentia API [4]) to represent abstract features of music; abstract features are harder to represent due to their subjective nature.

3.2.4 Music Structure. Music pieces often show some coherence or similarity over their durations, and this is an example of how visual representations can be used to visualize patterns within a music piece [9]. Foote [18] (Figure 4), visualized music similarities using a matrix; similar regions are represented with vivid colors while dissimilar regions are darker. Cantareira et al. [5] developed a visualization capable of representing similar/dissimilar structures between songs.

3.2.5 Harmonic Features. Harmonic features relate to the "fitting together" effect when a proper arrangement is applied, based on a "scale," or known as a "system of tuning" [87]. Ciuha et al. [8] visualized harmonic structures by representing intervals[6] with colors and varied saturation: notes' dissonance with low saturation and notes' consonance with high saturation. ***Music Overview, Stability, and Harmony Visualization*** **(MoshViz)**, offered a detail+overview approach; they

[6] *Interval*: The distance between any two sounds [87].

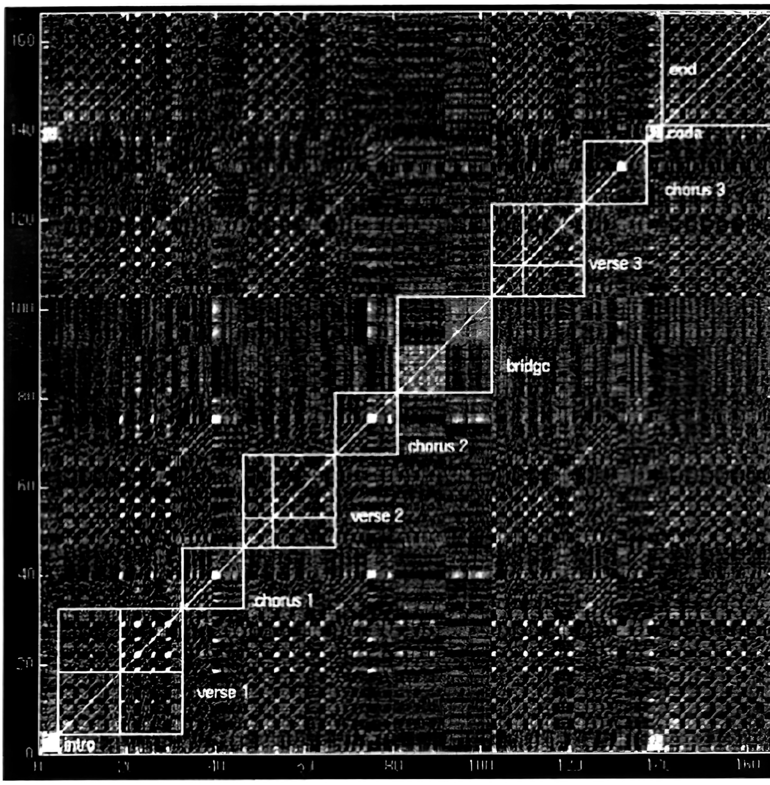

Fig. 4. *Day Tripper,* by Lennon/McCartney, represented using a matrix. The diagonal (bottom-left to up-right) shows the music structure (e.g., verse, chorus) and the brightness represents similar parts: "the brightness of a point (i, j) is proportional to the audio similarity at instants i and j" [18] ©Republished with permission of ACM (Association for Computing Machinery), from "Visualizing music and audio using self-similarity," by Jonathan Foote. 1999. Permission conveyed through Copyright Clearance Center, Inc.

produced a visual metaphor aiming at a faster interpretation (i.e., music analysis); harmonic features were represented using colors and highlights on a timeline layout. Snydal and Hearst [84] created a harmonic palette showing the notes played every measure of a 12-bar blues. To analyze the harmonic structure, Sapp [75] plotted the music's keys on a logarithm vertical scale; allowing to verify the key of more minor parts.

Malandrino et al. [50] and Prisco et al. [72] also approached the harmonic representation by augmenting the score notation with the addition of colored shapes to typify harmony in musical passages. Miller et al. [56] represented harmony by means of the circle of fifths. Their method supported distant and close reading in sheet music, exposing harmonic relationships while maintaining the CMN structure [56].

3.2.6 Melodic Features. Melodic lines were visualized by Politis et al. [68]; they conceived a visualization based on a bar divided into colored-blocks, according to their chromatic content and employed an algorithm to "extract the dominant frequencies of a melody" [68]. Fonteles et al. [17] achieved better perception of rhythms and melodies by making use of colored particles indicating pitches and rhythm patterns. Bergstrom et al. [3] used the *Tonnetz* grid representation (Figure 5) to convey information about chord quality, interval quality, and chord progression. Taenzer et al.

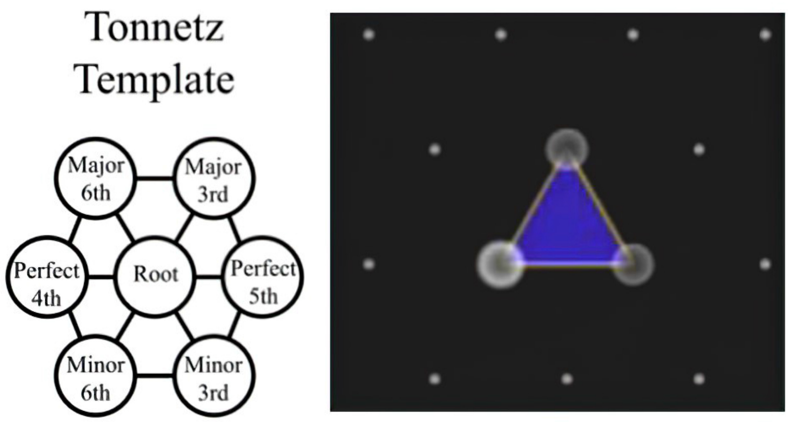

Fig. 5. The *Tonnetz* template/Isochords by Bergstrom et al. [3] presents in a gridded layout characteristics related to the interval, chord quality, and chord progression. The structure is a regular hexagon presenting each interval as a vertex, and the root note in the center. ©Republished with permission of ACM (Association for Computing Machinery), from "Isochords: visualizing structure in music," by Tony Bergstrom, Karrie Karahalios, and John C. Hart. 2007. Permission conveyed through Copyright Clearance Center, Inc.

[88] used source separation techniques to later detect the principal melodies and represented them employing spectrograms.

3.2.7 Summary.

- Note's pitches are the most visualized elements;
- Timbre was timidly approached and recently semantic descriptors were used to describe them;
- Beyond feature representation for musicologists or score readers, mood was equally visually represented by a modest number of papers;
- To assess music analysis many papers aimed to represent harmonic and melodic structures;
- Most of the visualizations represented features that are not directly expressed by the score notation.

3.3 Which Visualization Techniques Were Used?

In this section, we analyze the visual signs used to represent musical features on the papers we surveyed. Some proposals are focused on extending the score capabilities by providing features that are not clearly presented [6, 29, 42, 49, 56, 71, 76, 95], especially for non-readers of the **Common Music Notation (CMN)**. Other proposals convey diverse visual representations by using other types of layouts and visual cues [7, 15, 27, 34, 40, 43, 54, 79]. CMN is considered ideal for performing a music piece, but it obfuscates information about harmony, tempo changes, and so on, that can solely be perceived by a specialist (e.g., an experienced musician). New visualizations can provide better representations for features that are not presented in the CMN, but they can fail to deliver proper metaphors, diminishing the visualization's usefulness. Table 3 organizes the references.

3.3.1 Colors. The use of color to visualize music has been applied extensively, and nearly all the papers surveyed handled some sort of coloring, except: Bergstrom et al. [3], Ferguson et al. [16], Hiraga [26], Hiraga et al. [29], Kamolov et al. [35], Machida and Itoh [46], McLean et al. [54], McLeod and Wyvill [55]. Those exceptions relied on other visual metaphors to convey music's

Table 3. Paper Distribution According to the Visualization Technique Used: Color (34), Particles (3), Shapes (16), Line graphs (9), Score notation (9), Glyphs (4), Pictures (3)

Feature	#	Reference
Colors	34	Cantareira et al. [5], Chan et al. [6], Chew and François [7], Ciuha et al. [8], Farbood et al. [15], Fonteles et al. [17], Foote [18], Hayashi et al. [24], Hiraga and Matsuda [27], Hiraga et al. [28], Kim et al. [37], Kosugi [38], Lim and Raphael [42], Lima et al. [43], Malandrino et al. [49, 50], Mardirossian and Chew [51], Miller et al. [56], Mitroo et al. [57], Nanayakkara et al. [61], Ohmi [64], Politis et al. [68], Prisco et al. [71, 72], Sapp [75, 76], Shi and Yang [79], Shirzadian et al. [80], Smith and Williams [83], Snydal and Hearst [84], Soraghan et al. [85], Taenzer et al. [88], Tagliolato [89], Toiviainen [90]
Shapes	16	Bergstrom et al. [3], Cantareira et al. [5], Chew and François [7], Fujishiro et al. [19], Gumulia et al. [21], Jin et al. [34], Kawahara et al. [36], Lima et al. [43], Malandrino et al. [50], McLean et al. [54], Miller et al. [56], Nanayakkara et al. [61], Prisco et al. [72], Shi and Yang [79], Shirzadian et al. [80], Valbom and Marcos [92]
Line graphs	9	Gumulia et al. [21], Kawahara et al. [36], Lim and Raphael [42], McLeod and Wyvill [55], Nakano et al. [60], Prisco et al. [71], Snydal and Hearst [84], Taenzer et al. [88], Wattenberg [95]
Score notation	9	Chan et al. [6], Hiraga et al. [29], Lim and Raphael [42], Malandrino et al. [49, 50], Prisco et al. [71, 72], Sapp [76], Snydal and Hearst [84]
Glyphs	4	Chan et al. [6], Hiraga [26], Hiraga et al. [29], Machida and Itoh [46]
Particles	3	Ferguson et al. [16], Fonteles et al. [17], Kamolov et al. [35]
Pictures	3	Lehtiniemi and Holm [40], Machida and Itoh [46], Ox [66]

content. To represent notes, the number of colors varied between 7 and 12 colors, representing, respectively, diatonic and chromatic tones; other proposals relied on a more flexible approach, by utilizing a color wheel allowing the representation of combination of tones. Ohmi [64] used colors to "visualize expression, with the blue color indicating higher levels of frequency power spectrum while the yellow and red indicating lower and medium levels, respectively." Malandrino et al. [49] assigned the circle of the fifths with a discrete number of colors aiming at representing harmonic structure.

Hayashi et al. [24] used the score representation as background and highlighted with colored blocks similar structures within a music piece. Sapp [75, 76] cared about representing the music key with a pyramidal structure. Foote [18], instead, provided a colored matrix displaying similar structures with similar colors. *Misual* (Figure 3) was also developed to display on a waveform repeated parts of a song, varying the color saturation according to the note volume [38]. Ciuha et al. [8] proposed a key spanning circle of thirds assigned to a color wheel, aiming at representing every tone and also the volume of the notes through time (Figure 6).

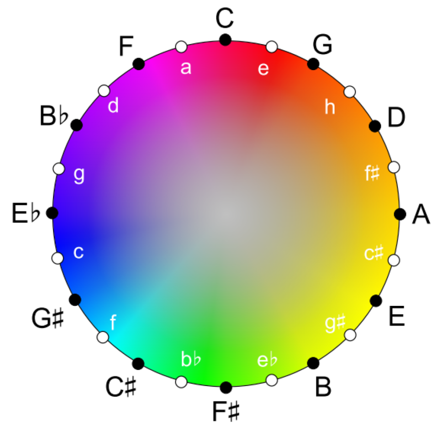

Fig. 6. Circle of thirds in a color wheel by Ciuha et al. [8]. It allows a continuous color selection according to the tone of choice. ©Republished with permission of ACM (Association for Computing Machinery), from "Visualization of concurrent tones in music with colours," by Peter Ciuha, Bojan Klemenc, and Franc Solina. 2010. Permission conveyed through Copyright Clearance Center, Inc.

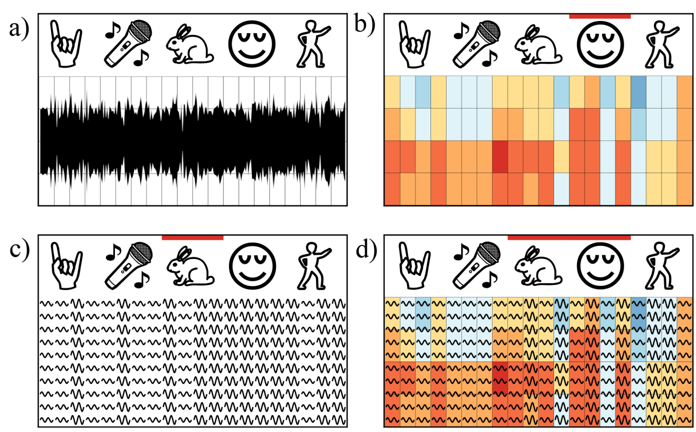

Fig. 7. *SongVis* by Lima et al. [43]. Each song is represented by a group of glyphs and a bottom section, named "wavesection," denoting more details (*details-on-demand*) whenever the user hovers/clicks the glyphs representing mood and tempo: (a) The default state presents the glyphs and the waveform; (b) By hovering the mood glyph, the state changes and the wavesection presents a heat map; (c) By hovering the *tempo* glyph, the state changes and the wavesection presents shapes denoting fast/slow passages; (d) By combining *mood+tempo,* the user has another view and he can analyze *tempo* comparing with its respective *mood* [43]. ©2019 IEEE. Reprinted, with permission, from "Visualizing the Semantics of Music," by Hugo Lima, 2019.

Lima et al. [43] used colors to denote mood (sad-to-happy scale) in their visualization: blueish tones related to sad passages and reddish tones to denote happier passages (Figure 7). Taenzer et al. [88] represented instruments melodic lines with colored traces, varying saturation according to the volume. Malandrino et al. [50] (VisualHarmony) and Prisco et al. [72], as well, employed colors to highlight harmonically similar passages aiding composition tasks: They aimed to help users find the best harmonic succession of chords.

3.3.2 Particles. On the score notation, Fonteles et al. [17] extended CMN's capability by providing a visualization of the instrument section by using particles, altering their movements and size

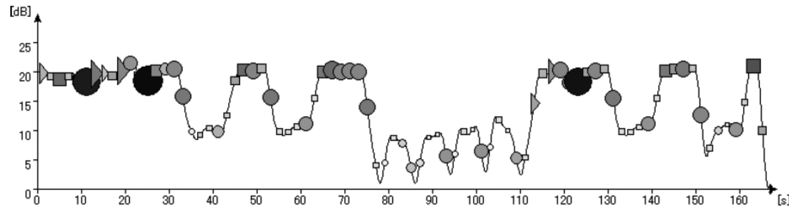

Fig. 8. Visualizationof *Hungarian Dances No. 5* by Brahms. The line graph relates time on the x-axis with loudness, in dB, in the y-axis. Shapes are used to represent the perceptual changes in music tempo/speed, e.g., *squares* represent slower passages than the *circles* [21]. ©Republished with permission of ACM (Association for Computing Machinery), from "Music visualization: predicting the perceived speed of a composition—Misual project,-" Puzon, and Naoko Kosugi. 2011. Permission conveyed through Copyright Clearance Center, Inc.

according to the dynamics of a musical passage and with the structure based on the "arrangement of a string quartet in a chamber orchestra" [17]. Ferguson et al. [16] also used particles to show, in real time, the performance of a musician. They presented, in a 3-dimensional environment, the first four harmonics, noisiness,[7] and also aiding the intonation of the musician by monitoring the sound signal.

3.3.3 Shapes. Shapes are also adopted to represent sound features. For example, Nanayakkara et al. [61] developed a music visualizer in which many aspects of sound, such as pitch, note duration, how hard a key is struck, and which instrument is playing; all features from MIDI data. *SeeGroove2* [34] used shapes to describe rhythm patterns; notes were represented as orbiting circles. Circles and colored rectangular blocks were used by Cantareira et al. [5] to typify notes and patterns within a music piece. *Celestia*, a vocal interaction music game, used geometrical shapes and a 3D environment based on the space and stars to symbolize pitch and sound volume; the movement triggered by the user's vocal notes in real-time.

Interestingly, Kawahara et al. [36] modeled the vocal tract shape to allow the visualization of "the intricate body-voice relations." Valbom and Marcos [92] modeled a 3D environment (using Virtual Reality) and represented notes and rhythms with shapes and movements. Lima et al. [43] employed sine-waved shapes to depict *tempo* variations (Figure 7). Each sine wave depicts a range of *tempo*—less than 80 BPM, between 80 and 120 BPM, and higher than 120 BPM—and the shape is placed below each respective music passage.

3.3.4 Line Graphs. To indicate fast and slow passages in music, Gumulia et al. [21] introduced a line graph with shapes outlined upon it, each shape representing whether a certain moment is perceptively fast or not (Figure 8). Pitch contour, defined as "time and pitch continuous sequences of salience peaks" [74], was approached by Kawahara et al. [36], Lim and Raphael [42], McLeod and Wyvill [55], Snydal and Hearst [84]. Taenzer et al. [88] depicted melodic lines with line graphs.

3.3.5 Score. Augmenting the score is a common approach [42, 76, 84]. Figure 9 presents the analysis of three soloing performances using the score as layout/grid by the visualization proposed by Snydal and Hearst [84]. Prisco et al. [71] used the score as base to show above it shapes and lines representing the melodic lines and music intervals. Colored rectangles were drawn above the score notation to represent similar structures with similar colors. Similar approach was used by Malandrino et al. [50] and Prisco et al. [72].

[7] *Noisiness*: "comparison between periodic and arbitrary elements within the data stream" [16].

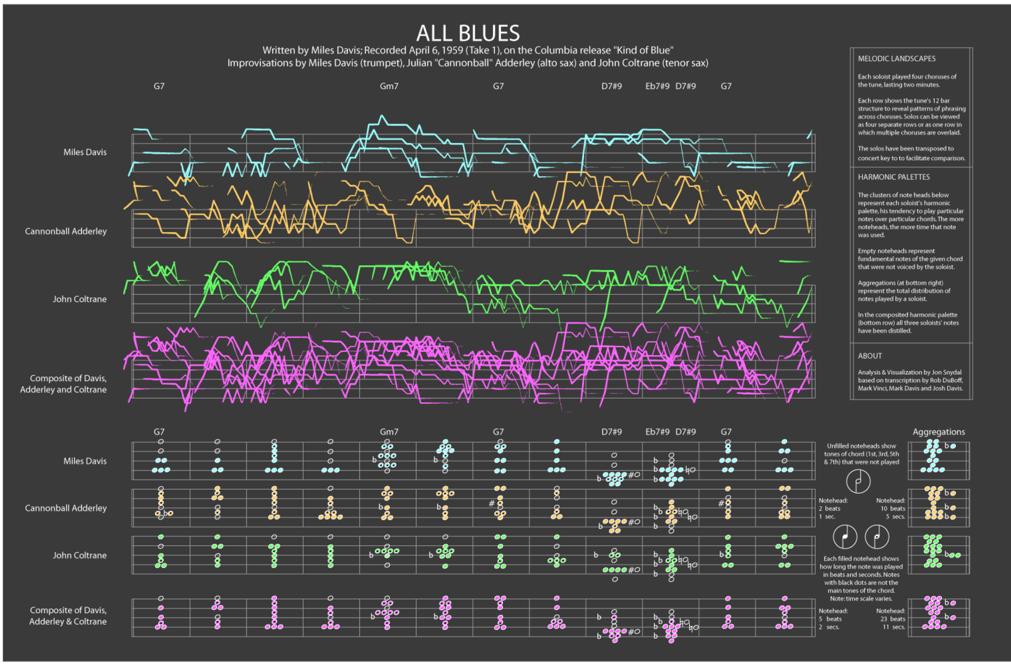

Fig. 9. *ImproViz* presents a "composite melodic landscapes (top) and shows the stylistic differences between three solos" [84]. "Harmonic palettes for the soloists appear below that illustrate their harmonic differences and similarities in each measure of the 12 bar blues" [84]. ©Permission of reuse gently granted by Jon Snydal.

3.3.6 Glyphs. To visualize classical music works, glyphs, typography, and colors were used by Chan et al. [6] to produce a visualization similar to the score notation, aiming at representing music with simpler visual metaphors, as the authors argue: reading a score is demanding, especially for a beginner. *Chernoff faces* were used by Hiraga [26], Hiraga et al. [29]; Hiraga et al. [29] represented expressiveness ("tempo, articulation, and sound level" [29]) and reasoned that: "since human beings are familiar with faces, we can easily detect changes."

Hiraga [26] employed the score layout trimmed to fit the whole music piece in the screen and used faces to show selected sections' notes, articulations (e.g., *legato*, *staccato*), and volume; they argue that with faces, more dimensions can be represented. Machida and Itoh [46] used mood icons to represent automatically the content of a music according to the semantics provided by the lyrics. Miller et al. [56] used a a non-stacked Nightingale Rose Chart to depict the circle of fifths, allowing readers to quickly harmonically analyze musical passages. Lima et al. [43] adopted faces (representing *mood*) and other glyphs, by means of *emoji* icons, to represent abstract features.

3.3.7 Pictures. Pictures were employed by Lehtiniemi and Holm [40] to represent the mood of a music according to specified configurations. Each mood picture revealed some general aspect about a music and it is another way to classify songs instead of a genre tag. Ox [66], with an artistic intention, used images of landscapes to represent the instruments that are playing, using color and saturation to convey aspects related to notes and volume (Figure 12).

3.3.8 Summary.

- Colors are still present in most of the visualizations proposals;
- The use of glyphs to represent music is still timid, despite the CMN's extensive usage of glyphs, pointing as a possible course;

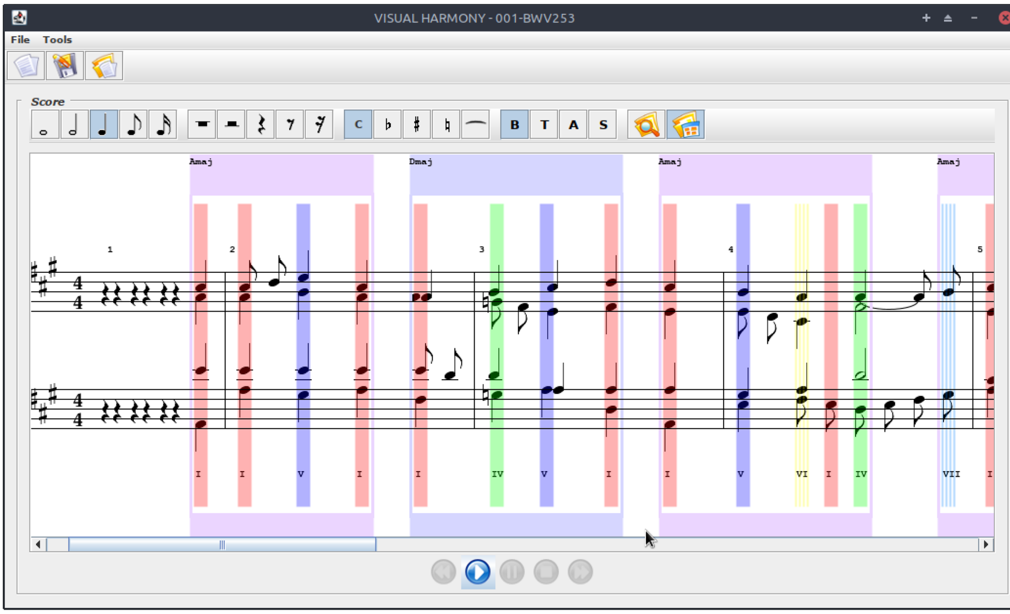

Fig. 10. *VisualHarmony* is an editor capable of assisting composers in harmonic related tasks [50]. Music passages highlighted with colors, the tonalities are based in the circle of fifths. The colors for the degrees are based in their function (primary, secondary, and tertiary)[50]. Screenshot from tool available at: http://www.di-srv.unisa.it/~delmal/research/usability/VisualHarmony/Tool/. ©2018 IEEE. Reprinted, with permission, from "Visualization and Music Harmony: Design, Implementation, and Evaluation," by Delfina Malandrino et al., 2018.

- The use of the score as layout base represent a convenient way to deliver other information while preserving the traditional score notation;
- Virtual reality proposals, despite being scarce, tended to use geometrical shapes to represent sound features;

3.4 What Was the Goal?

To investigate the various goals within the papers assessed, we created three *personas* [62] (role-based), which summarize the probable type of user a proposal is suited. A *persona* represents a description of a fictitious user; the aim is to envisage the needs of such users through investigation or experience in the field [62]. The adoption of the *personas* is simply to stereotype the visualization usage, according to the groups of users in the Music field. In no way is the intention to isolate the potential of each proposal, i.e., one work destined to a specific persona might be of interest to other(s) persona(s). Table 4 organizes the references.

3.4.1 The Music Engineer. The first persona is the Music Engineer. They may incorporate equalization, reverb, delay, and compression into the sound [86]. They work with signal processing techniques and might find convenient the visualization of each effect applied or how the techniques affect the waveform and the spectrogram. They can work with audio synthesis as well [32] and can use visualizations to aid their workflow. Visualizations can provide the visual feedback necessary for decision-making.

Ferguson et al. [16] developed a visualization to give feedback about a musician's performance, representing aspects such as intonation, loudness, harmonic content (the first four harmonics of

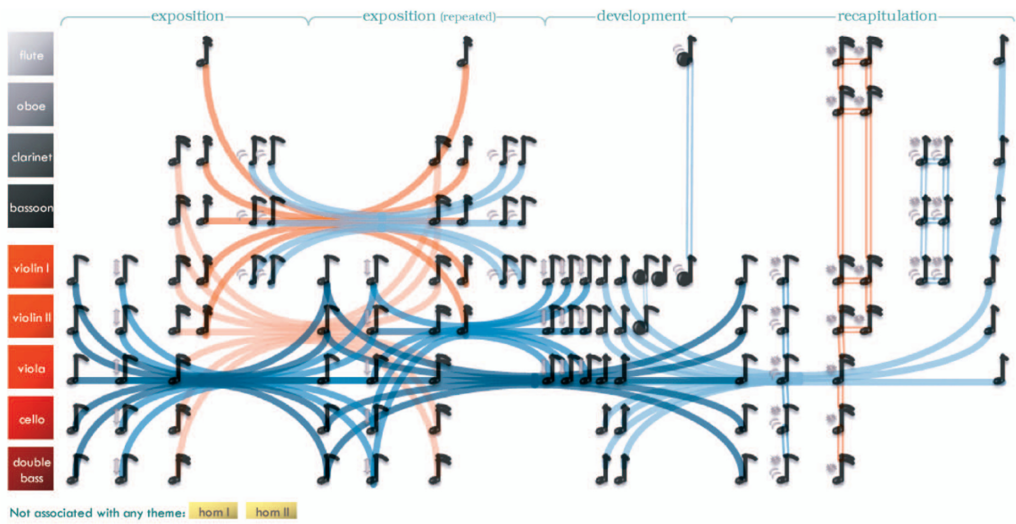

Fig. 11. "Theme fabric in bundled style for the first movement of Mozart's Symphony No. 40. Each theme occurrence is represented by a musical symbol glyph encoding its variation. Identical glyphs are connected by bundled threads" [6]. ©2010 IEEE. Reprinted, with permission, from "Visualizing the Semantic Structure in Classical Music Works," by Wing-Yi Chan, 2010.

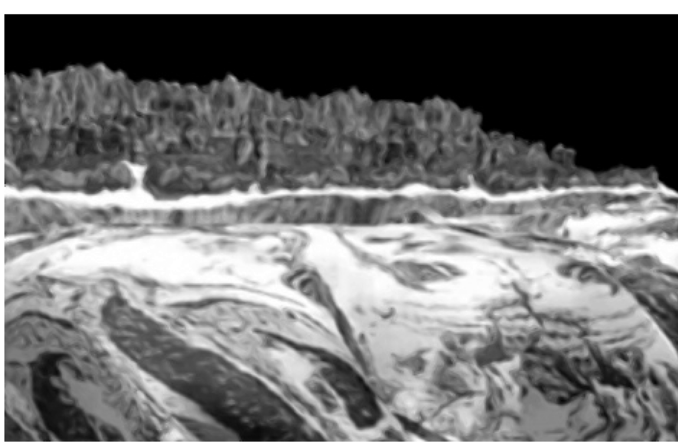

Fig. 12. Stillshot with music from the Color Organ [66]. ©Republished with permission of ACM (Association for Computing Machinery), from "2 performances in the 21st century virtual color organ," by Jack Ox. 2002. Permission conveyed through Copyright Clearance Center, Inc.

the fundamental frequency), and noisiness. Their proposal might be useful for session engineers to identify features that might need to be modified. Kawahara et al. [36] showed information about pitch and vocal tract shape in real-time for students of music therapy. Their feedback system can support the correction of posture of performers.

To evaluate intonation again, Lim and Raphael [42] used the score notation layout and allowed an instrumentalist to compare the note played in real-time with the music score. The visualization of Timbre by Soraghan et al. [85] helps users, through a feedback loop, to identify features that

Table 4. Paper Distribution According to the Goals of the Visualizations, Based on the *personas* Created: Music Engineer (12), Music Composer (10), Music Academic (46)

Goal	#	Reference
Music Academic	46	Bergstrom et al. [3], Cantareira et al. [5], Chan et al. [6], Chew and François [7], Ciuha et al. [8], Ferguson et al. [16], Fonteles et al. [17], Foote [18], Fujishiro et al. [19], Gumulia et al. [21], Hayashi et al. [24], Hiraga [26], Hiraga and Matsuda [27], Hiraga et al. [28, 29], Kamolov et al. [35], Kim et al. [37], Kosugi [38], Lehtiniemi and Holm [40], Lim and Raphael [42], Lima et al. [43], Machida and Itoh [46], Malandrino et al. [49], Mardirossian and Chew [51], McLean et al. [54], McLeod and Wyvill [55], Miller et al. [56], Mitroo et al. [57], Nakano et al. [60], Nanayakkara et al. [61], Ohmi [64], Ox [66], Politis et al. [68], Prisco et al. [72], Sapp [75, 76], Shi and Yang [79], Shirzadian et al. [80], Smith and Williams [83], Snydal and Hearst [84], Soraghan et al. [85], Taenzer et al. [88], Tagliolato [89], Valbom and Marcos [92], Wattenberg [95]
Music Engineer	12	Cantareira et al. [5], Ferguson et al. [16], Fujishiro et al. [19], Gumulia et al. [21], Jin et al. [34], Kawahara et al. [36], Kosugi [38], Lim and Raphael [42], Lima et al. [43], Nakano et al. [60], Soraghan et al. [85], Taenzer et al. [88]
Music Composer	10	Farbood et al. [15], Hayashi et al. [24], Hiraga [26], Hiraga et al. [28], Jin et al. [34], Malandrino et al. [50], Mitroo et al. [57], Prisco et al. [71], Smith and Williams [83], Valbom and Marcos [92]

need to be changed. Timbre is a high-level feature, i.e., it is defined by a bunch of different low-level features.

In SeeGroove2 [34], the authors created an interactive rhythm visualization. *Groove* is defined as "the sense of rhythmic feel or musical swing" [34], and it is a significant part of a musical performance. SeeGroove2, designed for beginner drummers, allows the user to visualize his performance and match the song in sync. In addition, it can be a valuable tool for engineers, as they can have the visual perception of what they are hearing.

3.4.2 The Music Composer. The second persona is the Music Composer, who may work directly with the **Common Music Notation (CMN)**. The Music Composer also finds more straightforward to use MIDI, or the piano-roll visualization, inputting notes with a keyboard [86]. Composers are no longer limited by the tangible aspects of composing: They can extend their capabilities by embracing innovative technologies released each year [2], inputting digitally and visually all the parts necessary.

The CMN is considered demanding, and "beginners have to spend considerable time learning the fundamental notations and technical terms of music theory before mastering a score for in-depth understanding" [6]. This way, many proposals aimed at providing a new visualization based on other visual metaphors to ease the analysis task. For live performances, the most used argument motivating the usage of visualizations was in regards to feedback [3, 6, 16, 17, 36, 42, 55, 60].

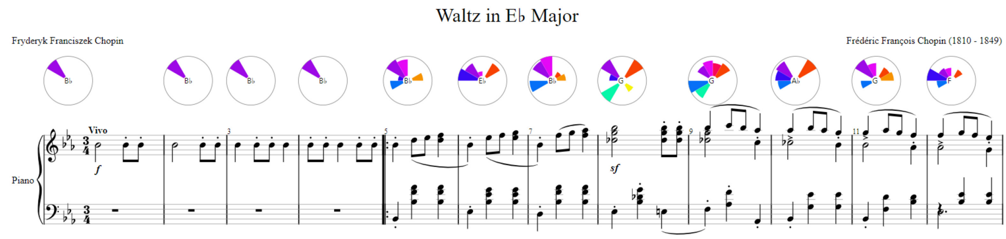

Fig. 13. Nightingale rose of the circle of fifths, by Miller et al. [56] (Screenshot). Each musical compass has a Nightingale-rose above, summarizing the notes below. Tool available at: https://musicvis.dbvis.de/app/fingerprint/.©Permission of reuse gently granted by Mathias Miller.

McLeod and Wyvill [55], for instance, provide the visualization of a real-time performance implementing a directer representation than that of the CMN.

With Hyperscore [15], composers can produce and explore compositions in an interactive way. Hyperscore is designed for users with limited or no musical training and lets them draw in freehand to input notes. According to Hayashi et al. [24], the most common way to learn music is through the study of music scores, although this practice is demanding and the learning curve is considered slow. To alleviate the struggle of novice composers, Hayashi et al. [24] designed *VisualHarmony*. Their tool focused in the harmonic analysis problem: "given a musical composition, the objective is to find the best harmonic successions of chords" [24].

VisualMelody was developed by Prisco et al. [71] as a way to allow new musicians to compose music in chorale style (four-voice music), which follows specific rules. Their proposal is aimed to "people lacking a strong knowledge in music theory" [71]. Shapes and colors were used to represent melodic lines and intervals as a way to visually separate bass lines, tenor, alto, and soprano voices. Valbom and Marcos [92] developed an "immersive musical instrument, aiming at expanding the concepts of traditional musical elements, and allows the integration of a spatial dimension using 3D music and sound objects" [92]. Their experiments showed users can integrate auditory and visual responses to enable the localization in the virtual environment.

To assist composers, Malandrino et al. [50] developed *VisualHarmony* (Figure 10); they focused on the harmonic analysis problem: "given a composition, find the best harmonic successions of chords" [50].

3.4.3 The Music Academic. The Music Academic Researcher can be consisted of a mixture of the previous two personas, however, his intentions are more academically oriented. The works are focused on the analysis of the music structure, chord progressions, as well as groove, using MIDI or not [86]. They can also work with more subjective aspects, such as: music charts, demographics, record categorization and stylistic preferences [2]. The researcher can, as well, work with Machine Learning techniques and prepare visualizations to convey abstract features of music.

Miller et al. [56] supported music analysis by adding the circle of fifths above each bar of the score notation (Figure 13). Prisco et al. [72] prepared a visualization for harmonic analysis of musical compositions, by which they allowed users to visually annotate the compositions with colored rectangles, correlated with the tonality degree (i.e., tonic, dominant, subdominant). To assist children with hearing impairment, Kim et al. [37] created an unusual visualization of melodies and rhythm, by capturing notes of a violin's fingerboard. The data are displayed using a plant; its petals wave, signaling a response to the music being played.

Hiraga et al. [28] proposed a performance visualization that aimed to visually aid users to "understand, analyze and compare performances and their musical structures" [28]. Politis et al. [68]

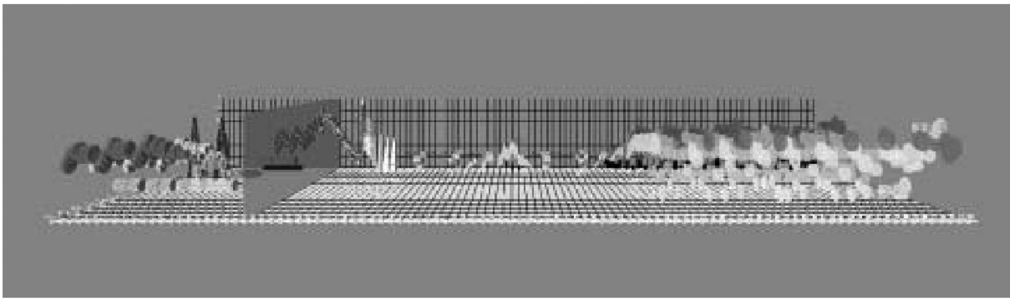

Fig. 14. Hiraga et al. [28] displayed the whole song in the screen and allowed the user to select notes represented by geometrical shapes. ©Republished with permission of ACM (Association for Computing Machinery), from "Performance Visualization – A New Challenge to Music Through Visualization," by I. Fujishiro; R. Miyazaki; R. Hiraga. 2002. Permission conveyed through Copyright Clearance Center, Inc.

conceived a visualization to aid the Chromatic analysis a song. For that, they extracted the dominant frequencies and associated them with colors. To represent a whole song, Lima et al. [43] extracted low-level and high-level features, allowing users to quickly gaze and understand which aspects are the most prominent in each song.

Some proposals even tried to convey an artistic visual, e.g., artistic installation in Mitroo et al. [57] and the entertaining, immersive visualization provided by Ox [66] aimed at being executed in a ***Audio-Visual Experience Automatic Virtual Environment*** **(CAVE)** [10].

3.4.4 Summary.

- Most of the papers (88%) aimed at assessing the music analysis task (Music Academic persona);
- Despite its usefulness regarding assisting learning, few papers cared at improving learning;
- To aid a music notation beginner, most of the visualizations used the excuse that the CMN is "demanding";
- Most of the bibliography surveyed focused to benefit the academic domain.

3.5 How the Users Interacted?

Interaction is the communication between user and system [12], i.e., how an individual performs a task aided by a computer. In this section, we enlist designs that, in some way, allowed the interaction on the visualizations surveyed. Only 20 proposals (39%) implemented some kind of interaction in the visualizations, and here they are divided according to the interaction taxonomies defined by Yi et al. [97]. Table 5 is a compilation of this section, relating paper proposals with the taxonomy complied.

3.5.1 Select. Nakano et al. [60] allows the selection of the fundamental frequency (f_0) to adequate intonation Hiraga et al. [28] allowed selection to track notes through time (Figure 14). Also selecting notes, but for a following analysis, was the proposal by Lim and Raphael [42]. In a virtual environment, Valbom and Marcos [92] allowed selection task of notes. Miller et al. [56] allowed the user to hover over notes to isolate them in the score (Figure 13). Also, Lima et al. [43] let users hover and click on the glyphs (*emojis*) to reveal more details about mood and tempo.

3.5.2 Explore. Panning was the most employed technique to allow the exploration of the visualizations. Prisco et al. [71] employed panning to explore the visualization above the score notation layout. With an artistic aim, Ox [66] allowed the exploration of music using a CAVE environment

Table 5. Paper Distribution According to Interaction Taxonomies by Reference [97]: Select (8), Explore (9), Reconfigure (4), Encode (6), Abstract/Elaborate (11), Filter (8), Connect (4)

Taxonomy	#	Reference
Abstract/ Elaborate	11	Cantareira et al. [5], Chan et al. [6], Chew and François [7], Hiraga et al. [28, 29], Lim and Raphael [42], Lima et al. [43], Machida and Itoh [46], Malandrino et al. [50], Mardirossian and Chew [51], Valbom and Marcos [92]
Explore	9	Chew and François [7], Farbood et al. [15], Hiraga et al. [28, 29], Machida and Itoh [46], Malandrino et al. [50], Miller et al. [56], Ox [66], Prisco et al. [71]
Filter	8	Cantareira et al. [5], Chan et al. [6], Hiraga et al. [28], Lehtiniemi and Holm [40], Machida and Itoh [46], Mardirossian and Chew [51], Miller et al. [56], Politis et al. [68]
Select	8	Cantareira et al. [5], Hiraga et al. [28], Lim and Raphael [42], Lima et al. [43], Malandrino et al. [50], Miller et al. [56], Nakano et al. [60], Valbom and Marcos [92]
Encode	6	Chan et al. [6], Farbood et al. [15], Malandrino et al. [50], Nakano et al. [60], Prisco et al. [71], Shi and Yang [79]
Connect	4	Cantareira et al. [5], Chan et al. [6], Hiraga et al. [29], Lim and Raphael [42]
Reconfigure	4	Hiraga et al. [28], Lima et al. [43], McLean et al. [54], Politis et al. [68]

[10]. On the virtual space created by Hiraga et al. [28]; exploration is allowed by panning through a 3D environment. Moving the view to a different angle employed by Chew and François [7], Hiraga et al. [29] produced a visualization to overview a music sheet and allowed the exploration by panning the visible area. *Lyricon* [46] also used pan to explore the icons created by analyzing music lyrics. Malandrino et al. [50] and their editor allowed to add and remove notes, as well as panning.

3.5.3 Reconfigure. The reconfigure task was used by the Apollonius diagrams [54] to alter the notes' properties (volume, pitch, stereo panning, and modulation), directly changing the visualization according to the new parameters. The user is allowed to adjust the segmentation factor (i.e., user defines the beginning and extremity of each segment according to his will/perception) on the visualization of chromatic descriptions provided by Politis et al. [68].

3.5.4 Encode. Chan et al. [6] allowed the customization of the colors used by the threads of instruments on their visualization. Changing the visualization by inputting notes was done by Prisco et al. [71]. *MiruSinger* [60], intended to offer real-time feedback for singers and allowed the correction of the fundamental frequency (f_0). *Celestia* [79] is a game that permits various encodings based on the user's voice. To aid the composition task by novices, *Hyperscore* [15] allowed different color encoding based on user's taste. Malandrino et al. [50] let color-blind users reconfigure the color schemes that better satisfied them.

3.5.5 Abstract/Elaborate. The combo *focus+context* was the most used in the visualizations that incorporated tasks of Abstraction/Elaboration (i.e., "the ability to adjust the level of abstraction of a data representation" [97]). *Zoom* was used by Chan et al. [6] to gather more information regarding instrument icons; also, the score was showed on demand. By hovering the mouse above the grid of keys, it is possible to distinguish the names of the keys visually represented on Mardirossian and Chew [51]. Cantareira et al. [5] allowed zooming into attractive parts of a song for following analysis. The geometrical shapes used by Valbom and Marcos [92] to represent notes make possible, by clicking, to listen to the item selected. As already mentioned, Lima et al. [43] allowed users to hover over the icons to get more detailed information.

3.5.6 Filter. Filters were used by Chan et al. [6] to reduce the number of instruments threads (Figure 11). Fine-tuning the segmentation size was a filter task used by Mardirossian and Chew [51]. The music suggestion varied according to the mood picture chosen on the visualization created by Lehtiniemi and Holm [40].

3.5.7 Connect. Linking and brushing was used by Chan et al. [6] to allow users to establish connections, for possible future insights, on the visualization offered. The multiple views designed by Cantareira et al. [5] allowed the visualization of various details regarding a selected part. Lim and Raphael [42] also permitted connection of insights by employing multiple views: score view, pitch-trace and a spectrogram view. Moreover, all the proposals that aimed at providing the identification of similar structures within a music piece can also be referenced in this subsection for the "connect" taxonomy, with credits to: Foote [18], Malandrino et al. [49], Ohmi [64].

3.5.8 Summary.

- Only 39% (20 papers) of the surveyed papers implemented some sort of interaction;
- Most of the exploration tasks rely on panning the visualization;
- The most used interaction was abstract/elaborate (11/20 papers);
- Interactions were in majority used to assess an analysis task.

3.6 How the Users Reacted?

In this section, we describe how the authors evaluated their proposals according to the user's perspectives. Formal studies have been conducted exclusively by: Jin et al. [34], Malandrino et al. [49, 50], Miller et al. [56], Prisco et al. [71, 72], Shirzadian et al. [80]; the others relied on informal interviews with users or did not executed any sort of evaluation. Shirzadian et al. [80] concluded that "there is a tendency for the participants to feel more immersed when they appreciate the music which the experiment is employing" [80], and this can be utilized as a reference for future evaluations regarding engagement.[8]

Malandrino et al. [49] evaluated a color-based visualization to understand harmonic structures of musical compositions. They conducted a study assessing the effectiveness of their proposal. They interviewed participants, most of whom had a music knowledge for analysis (75%) and "67% spent more than 2 hours per day playing music" [49]. The results reported that most of the users found the colored-based visualization useful (4.4 score in a 5-point Likert-scale) and aesthetically pleasing (4.2 score in a 5-point Likert-scale). The conclusion was that "augmenting musical scores with graphical elements, individuals improved their performance in all the analyzed tasks" [49].

The Prisco et al. [71] proposal supported people lacking a keen knowledge of music theory. They exploited visual features to draw attention on composition errors. To evaluate their visualization, they employed a modified version of **Computer System Usability Questionnaire (CSUQ)** [41] and "verified system usability and user satisfaction when using *VisualMelody*." Those who tried the visualization were capable to "compose in less time when compared with those that did not examined the tool with the visualization" [71].

[8] *Engagement*: a quality of the user experience that emphasizes the appealing aspects of interaction—in particular, the fact of being captivated by a resource [1].

Users evaluated Valbom and Marcos [92] visualization by answering a survey based on the **unified theory of acceptance and use of technology (UTAUT)** [93] and also usability tests incorporating *ISO 9241-10* principles [20]. They reported that all the users interviewed answered positively in regards to the visualization proposed, but no formal evaluations were made or more details provided by the authors referring to their proposal.

Jin et al. [34] developed *SeeGroove2* to assist musicians at learning groove. Users were examined and questioned if they identified any difference while using SeeGroove2 in the aspect of perception between a track with groove and another without. They additionally noted "interactivity is vital to effective music visualization" and this requires a reasonable temporal resolution, avoiding delays.

More recently, Miller et al. [56] developed a visualization capable of representing harmonic content. They evaluated their proposal by conducting a study, which analyzed the performance of experts and non-experts, using their visualization. They concluded that all non-experts could identify patterns of the themes without needing to verify their choices using the CMN score. The fingerprints developed by Miller et al. [56] facilitate the access to sheet music, reducing cognitive labor. Malandrino et al. [50] and Prisco et al. [72] employed CSUQ and QUIS forms to evaluate their proposals. Both got a favorable feedback from users about the effectiveness and pleasantness of their visualization. Users are willing to use in the future as a way to reduce the time bestowed while analyzing harmonic content of music. Malandrino et al. [50], Prisco et al. [71, 72] were the only proposals to use the CSUQ.

3.6.1 Summary.

- Most of the proposals do not evaluate their visualizations;
- Most of the evaluations are informal, lacking methods and metrics adopted by other InfoVis branches (e.g., Likert-scale, CSUQ, Cronchabch's alpha);
- Usability is not considered a crucial aspect;
- Users are mostly interviewed by answering simple and informal questionnaires.

4 DISCUSSION

In this section, we discuss about the general overview of the 51 Music Visualization (MusicVis) papers surveyed. We enlist trends and missing spots that can be used as motivation for future works. Visualization is broadly used to convey better insights regarding music features, and it was noticed that the elaboration of new music visualizations remains a recurrent theme; as new extraction algorithms and advancements in the field of Signal Processing are produced, new possibilities for visualization will arise. The topic of music visualization is at a relative steady pace when compared with other topics from the **Information Visualization (InfoVis)** field. Therefore, many open research problems are still around.

It is noticeable that MIDI is still the most used input for music visualizations, and we hypothesize this is explained by the easiness provided by the MIDI protocol (i.e., structured and tabulated data). Second, the many already established techniques from the Signal Processing field, active since the early days of computing [65], are equally relevant contenders. Other proposals [50, 56, 72] used MusicXML as input. Others used unusual types of inputs, such as the **Galvanic Skin Response (GSR)** sensors to measure user's engagement [80], and Kim et al. [37] inserted sensors on a violin fingerboard to input notes. We predict this is going to be adopted universally in the future, as the cost of sensors is lowering and their usage can bring new potential researches [39].

In regards of the features visualized, the papers tended to represent characteristics that are not clearly present, especially for a novice in **Common Music Notation (CMN)**, thus the probable trend is that this continues to happen, but including different aspects that were not treated extensively, such as: semantic descriptors, mood, rhythmic features (e.g., tempo changes), and so on.

Many works treated extraction techniques focused on timbre extraction [13, 33, 48, 52, 53, 70, 85] and this can be perceived as a trend, as timbre represents an aspect that incorporates low-level sound features and more abstract ones (e.g., semantic descriptors [43]: bright-dark).

A common excuse for the use of visualizations is to provide useful visuals for music analysis. New visualizations, specific to a context, can provide direct representations for structures that are not clearly present in the score notation. The analysis of harmony, melodic structures, and patterns within a music piece can be aided by employing visualizations, as observed throughout this work. One way to highlight patterns and structures is by implementing an association between music and colors—which was used extensively—presenting itself as a continuous tendency due to its natural relation. Current proposals are extending this relation by providing more colors, using the combination of tones in a color wheel. The InfoVis field is withal a likely contributor, as many InfoVis techniques could be integrated to represent music.

Size and shape of items were repeatedly used to represent loudness and features, respectively. The association of loudness with size/height can equally be considered a natural mapping, thus its usage will continue to be seen. Few works covered glyphs, despite its usage being essential in the score notation.[9] Therefore, a void could be filled by proposals studying the representation of sound features using glyphs, as the score notation does, but with an alternative approach (i.e., the use of glyphs to represent more abstract features). Evans [14] listed "basic considerations in producing an effective score for listeners," and we believe these reflections can nonetheless be implemented to provide augmented scores visualizations:

- "Sonic events should be simple, but visually identifiable";
- "Temporal logic should match the spacial logic";
- "The full score should be visible at a glance";
- "Score reading is not the most critical thing";
- "Scores are for listening (in service to the ear)."

Interaction was poorly considered and certainly deserves more attention, since much more insights could be acquired by combining interactive tasks. *Pan, Zoom+Detail* were the most utilized interactions, as they can ease the analysis task, in which a user pans the whole music piece in the screen and *zooms in* to obtain more details (i.e., details on demand). Taxonomies provided by Yi et al. [97] and Shneiderman [81] serve as guides to incorporate distinctive features in the visualizations, leaving behind static pictures. Future works might implement brushing as interaction or the use of multiple visualizations; both were poorly addressed in our literature investigation.

Considering that most of the proposals surveyed aimed at useful visualizations, lengthy investigations are necessary to evaluate how appropriate these visualizations are. The majority did not apply any sort of usability or appropriateness evaluation. Only three proposals used the **Computer System Usability Questionnaire (CSUQ)** [50, 71, 72] and a more formal research. Jin et al. [34], Malandrino et al. [49, 50], Prisco et al. [72], Shirzadian et al. [80] used the Likert-scale to measure user's evaluations involving satisfaction, easiness, and so on. Therefore, we argue for more studies on how music visualizations embrace fundamental design principles: "Discoverability, Feedback, Conceptual model, Affordances, Signifiers, Mappings, Constraints" [63]. Interaction is, in addition, another open research problem; most of the visualizations allowed basic zoom+detail (e.g., video game controllers in Chew and François [7]), but most omitted if there was any kind of interaction.

[9]Out of the music visualization field, but still relevant to be noted, is the proposal by Xie et al. [96], which takes inspiration in the music notation (i.e., the note symbol) to improve their, once cluttered, designed tool for analysis of E-transactions.

Table 6. Considerations for Data Input, Feature Visualized, Visualization Technique, Goals, Interaction, and Evaluation

Data Input	• New inputs can be promising: the use of sensors and other controls can provide interesting data to be treated and visualized. • The MIR community can use the many extraction algorithms (e.g., extraction using *Convolutional Neural Networks*) to provide more data to be visually represented. • Few works combined MIDI and audio signals, and this can also be exploited to overcome the limitations of each input denoted.
Feature Visualized	• Timbre, despite being extensively studied in the MIR community, still lacks proposals for visualization. • Not many visualizations represented Mood, and this can be pointed as interesting for future works, since many systems nowadays deal with custom user profiles (i.e., the music visualization can be adapted to represent users' mood).
Visualization Technique	• MusicVis should adopt InfoVis techniques for representing streaming data and adapt it to include already known music visualization metaphors (e.g., stacked area graphs plus notes symbols from the CMN). • The combination of InfoVis techniques was rarely seen, and might be promising, e.g., stream graphs plus glyphs. • Non-temporal oriented visual representations could benefit other types of structural analysis for music, e.g., Treemap visualization can represent key predominance.
Goals	• The proposals are mostly academically oriented, and future works could benefit other groups of users. • Only one proposal focused on assisting children with hearing impairment, and more proposals could be developed with the social inclusion intent. • Few proposals aimed at assisting learning. Visualizations can be beneficial, as they can clearly visually express the essential ideas about a concept/topic facilitating the learning process.
Interaction	• Interaction remains an important concept, and should be adopted by more proposals aiming at allowing even further possibilities for insights. • Research of taxonomies of interaction specific for Music Visualizations. • Back to basic: adopt the *mantra* by Shneiderman [81]: "Overview first, zoom and filter, then details-on-demand."

(Continued)

Table 6. Continued

Evaluation	• Evaluate the segments generated by visualizations in future investigations. • More formal evaluations are necessary to attest the benefits of the visualizations produced. • The use of HCI metrics should be adopted as a way to permit that new visualizations can be produced based on successful old proposals. • More extensive interviews (to identify the benefits on long term usage) with users are necessary. • Verify the right audience to be interviewed, i.e., for whom is this visualization aimed? Musicians relate differently with music compared with a non-musician.

This survey did not intend to approach collections of music, and it can be cited as a potential future work. Plus, we highlight the potential of proposals that incorporate the so called *feedback loop*, in which a player can instantly observe his performance and is able to adapt according to the visual cues provided. Also, we did not survey commercial software, due to many restrictions, such as: copyright limitations, paid softwares, and lack of documentation. However, for future works, a comparison could me made between the academic production and the commercial ventures of many well-known market contenders: iZotope,[10] Native Instruments,[11] Avid,[12] Steinberg,[13] and so on.

Frequently, the aim is to output meaningful representations; the contrary, however, can as well be experimented, especially in future works: generate music by means of visual representations. Valbom and Marcos [92], in their proposal, concluded that the concept of representing sounds with 3D shapes "enables the user to expand musical thoughts and sound combinations in new ways of composition and performance." However, the generated pieces open up a discussion on model interpretability. Visual cues can be new, presenting the user a comprehensive range of possibilities for interpretation for what kind of sound he might expect to be generated.

Future works could evaluate music generated by visualizations. Few proposals attained that goal in our investigation, as it was not the scope of our survey. Such process highly depends on the adequate integration between visual signs and music features. The objective is customarily to *shallow* the learning curve; however, we should be cautious on the quest for the "right" representation, which must, naturally, be derived from the context (problem).

More investigations are necessary to evaluate interaction in music visualizations, and perhaps explain some of the following questions: what interactions from the music field can be borrowed by the visualization? What interactions can be created or borrowed from other fields? Are the interactions effective? Does previous knowledge in music help to undertake a task on music visualization? Table 6 summarizes our findings.

5 CONCLUSION

In this work, we surveyed 51 papers related to Music Visualization. They were analyzed and divided into classes of the most prominent topics. We verified that many proposals still rely on color,

[10] https://www.izotope.com/.
[11] https://www.native-instruments.com/en/.
[12] https://www.avid.com/.
[13] https://www.steinberg.net/en/home.html.

shape, and the score notation layout to represent music features. Most of the visualizations proposed aimed at representing music features that are not clearly depicted in the **Common Music Notation (CMN)**. Also, not many works cared about formally investigating if the visualization proposed was enthusiastically accepted by the users employing scientific research methods, such as those already extensively used in the **Human-Computer Interaction (HCI)** field. Some papers presented as novelty the use of music visualizations combined with **Information Visualization (InfoVis)** techniques, such as: line graphs, glyphs, and colors. The representation of music features by visual means comprises a typical combination, and with the advancement of extraction techniques, we expect that more interesting visualizations will be produced in the future.

REFERENCES

[1] Simon Attfield, Gabriella Kazai, and Mounia Lalmas. 2011. Towards a science of user engagement (position paper). In *Proceedings of the WSDM Workshop on User Modelling for Web Applications.* DOI: https://doi.org/978-1-4503-0493-1/11/02

[2] David Baskerville. 2017. *Music Business Handbook and Career Guide* (11th ed.). SAGE Publications, Inc.

[3] Tony Bergstrom, Karrie Karahalios, and John C. Hart. 2007. Isochords: Visualizing structure in music. In *Proceedings of the Graphics Interface Conference*, Christopher G. Healey and Edward Lank (Eds.), Vol. 234. ACM Press, 297–304. DOI: https://doi.org/10.1145/1268517.1268565

[4] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José R. Zapata, and Xavier Serra. 2013. ESSENTIA: An open-source library for sound and music analysis. In *Proceedings of the ACM Multimedia Conference*, Alejandro Jaimes, Nicu Sebe, Nozha Boujemaa, Daniel Gatica-Perez, David A. Shamma, Marcel Worring, and Roger Zimmermann (Eds.). ACM, 855–858. DOI: https://doi.org/10.1145/2502081.2502229

[5] Gabriel Dias Cantareira, Luis Gustavo Nonato, and Fernando Vieira Paulovich. 2016. MoshViz: A detail+overview approach to visualize music elements. *IEEE Trans. Multim.* 18, 11 (2016), 2238–2246. DOI: https://doi.org/10.1109/TMM.2016.2614226

[6] Wing-Yi Chan, Huamin Qu, and Wai-Ho Mak. 2010. Visualizing the semantic structure in classical music works. *IEEE Trans. Vis. Comput. Graph.* 16, 1 (2010), 161–173. DOI: https://doi.org/10.1109/TVCG.2009.63

[7] Elaine Chew and Alexandre R. J. François. 2005. Interactive multi-scale visualizations of tonal evolution in MuSA.RT Opus 2. *Comput. Entertain.* 3, 4 (2005), 1–16. DOI: https://doi.org/10.1145/1095534.1095545

[8] Peter Ciuha, Bojan Klemenc, and Franc Solina. 2010. Visualization of concurrent tones in music with colours. In *Proceedings of the 18th International Conference on Multimedia*, Alberto Del Bimbo, Shih-Fu Chang, and Arnold W. M. Smeulders (Eds.). ACM, 1677–1680. DOI: https://doi.org/10.1145/1873951.1874320

[9] Matthew Cooper, Jonathan Foote, Elias Pampalk, and George Tzanetakis. 2006. Visualization in audio-based music information retrieval. *Comput. Music. J.* 30, 2 (2006), 42–62. DOI: https://doi.org/10.1162/comj.2006.30.2.42

[10] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. 1992. The cave—Audio visual experience automatic virtual environment. *Commun. ACM* 35, 6 (1992), 64–72. DOI: https://doi.org/10.1145/129888.129892

[11] Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. 2011. TextFlow: Towards better understanding of evolving topics in text. *IEEE Trans. Vis. Comput. Graph.* 17, 12 (2011), 2412–2421. DOI: https://doi.org/10.1109/TVCG.2011.239

[12] Alan Dix, Janet Finlay, Gregory D. Abowd, and Russel Beale. 2004. *Human–Computer Interaction* (3rd ed.). Pearson Education Limited.

[13] Patrick J. Donnelly and John W. Sheppard. 2013. Classification of musical timbre using Bayesian networks. *Comput. Music. J.* 37, 4 (2013), 70–86. DOI: https://doi.org/10.1162/COMJ_a_00210

[14] Brian Evans. 1969. The graphic design of musical structure: Scores for listeners : Incantation and mortuos plango, vivos voco first 8 partials from bell analysis key centers for 8 sections. 1–6. Electroacoustic Music Studies Network. http://vassar-sound-design.pbworks.com/f/The+Graphic+Design+of+Musical+Structure.pdf.

[15] Morwaread M. Farbood, Egon C. Pasztor, and Kevin Jennings. 2004. Hyperscore: A graphical sketchpad for novice composers. *IEEE Comput. Graph. Applic.* 24, 1 (2004), 50–54. DOI: https://doi.org/10.1109/MCG.2004.1255809

[16] Sam Ferguson, Andrew Vande Moere, and Densil Cabrera. 2005. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In *Proceedings of the Ninth International Conference on Information Visualisation (IV'05).* 97–102. DOI: https://doi.org/10.1109/IV.2005.114

[17] Joyce Horn Fonteles, Maria Andréia Formico Rodrigues, and Victor Emanuel Dias Basso. 2013. Creating and evaluating a particle system for music visualization. *J. Vis. Lang. Comput.* 24, 6 (2013), 472–482. DOI: https://doi.org/10.1016/j.jvlc.2013.10.002

[18] Jonathan Foote. 1999. Visualizing music and audio using self-similarity. In *Proceedings of the 7th ACM International Conference on Multimedia (Part 1) (MULTIMEDIA'99)*. Association for Computing Machinery, New York, NY, 77–80. DOI: https://doi.org/10.1145/319463.319472
[19] Issei Fujishiro, Naoki Haga, and Masanori Nakayama. 2015. SeeGroove: Supporting groove learning through visualization. In *Proceedings of the International Conference on Cyberworlds*. IEEE Computer Society, 189–192. DOI: https://doi.org/10.1109/CW.2015.65
[20] Günther Gediga and Kai-Christoph Hamborg. 1997. Heuristische evaluation und isometrics: Ein vergleich. In *Software-Ergonomie'97: Usability Engineering - Integration von Mensch-Computer-Interaktion und Software-Entwicklung. Gemeinsame Fachtagung des German Chapter of the ACM, der Gesellschaft für Informatik (Gl) und der Technischen Universität Dresden vom 3. bis 6. März 1997 in Dresden*, Rüdiger Liskowsky, Boris M. Velichkovsky, and Wolfgang Wünschmann (Eds.). Berichte des German Chapter of the ACM, Vol. 49. Teubner, 145–155. Retrieved from http://dl.mensch-und-computer.de/handle/123456789/1084.
[21] Anastasia Gumulia, Bartlomiej Puzon, and Naoko Kosugi. 2011. Music visualization: Predicting the perceived speed of a composition—Misual project. In *Proceedings of the 19th International Conference on Multimedia*, K. Selçuk Candan, Sethuraman Panchanathan, Balakrishnan Prabhakaran, Hari Sundaram, Wu-chi Feng, and Nicu Sebe (Eds.). ACM, 949–952. DOI: https://doi.org/10.1145/2072298.2071910
[22] Yoonchang Han, Jae-Hun Kim, and Kyogu Lee. 2017. Deep convolutional neural networks for predominant instrument recognition in polyphonic music. *IEEE ACM Trans. Audio Speech Lang. Process.* 25, 1 (2017), 208–221. DOI: https://doi.org/10.1109/TASLP.2016.2632307
[23] Andrea Hanke. 2017. Tools for feature extraction: Exploring essentia. Topics in Computer Music, RWTH Aachen, 1–9. https://hpac.cs.umu.se/teaching/sem-mus-17/Reports/Hanke.pdf.
[24] Aki Hayashi, Takayuki Itoh, and Masaki Matsubara. 2011. Colorscore—Visualization and condensation of structure of classical music. In *Proceedings of the 15th International Conference on Information Visualisation*. 420–425. DOI: https://doi.org/10.1109/IV.2011.19
[25] Dorien Herremans and Ching-Hua Chuan. 2017. A multi-modal platform for semantic music analysis: Visualizing audio-and score-based tension. In *Proceedings of the IEEE 11th International Conference on Semantic Computing (ICSC)*. 419–426. DOI: https://doi.org/10.1109/ICSC.2017.49
[26] Rumi Hiraga. 2002. Case study: A look of performance expression. In *IEEE Visualization (VIS'02)*. 501–504. DOI: https://doi.org/10.1109/VISUAL.2002.1183815
[27] Rumi Hiraga and Noriyuki Matsuda. 2004. Graphical expression of the mood of music. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE Computer Society, 2035–2038.
[28] Rumi Hiraga, Reiko Mizaki, and Issei Fujishiro. 2002. Performance visualization: A new challenge to music through visualization. In *Proceedings of the 10th ACM International Conference on Multimedia*, Lawrence A. Rowe, Bernard Mérialdo, Max Mühlhäuser, Keith W. Ross, and Nevenka Dimitrova (Eds.). ACM, 239–242. DOI: https://doi.org/10.1145/641007.641054
[29] Rumi Hiraga, Fumiko Watanabe, and Issei Fujishiro. 2002. Music learning through visualization. In *Proceedings of the 2nd International Conference on WEB Delivering of Music*, Christoph Busch, Michael Arnold, Paolo Nesi, and Martin Schmucker (Eds.). IEEE Computer Society, 101–108. DOI: https://doi.org/10.1109/WDM.2002.1176199
[30] Eric J. Humphrey, Juan Pablo Bello, and Yann LeCun. 2012. Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR'12)*. 403–408. Retrieved from http://ismir2012.ismir.net/event/papers/403-ismir-2012.pdf.
[31] Eric J. Isaacson. 2005. What you see is what you get: On visualizing music. In *Proceedings of International Symposium in Music Information Retrieval*. 389–395. Retrieved from http://ismir2005.ismir.net/proceedings/1129.pdf.
[32] Brian M. Jackson. 2018. *The Music Producer's Survival Guide: Chaos, Creativity, and Career in Independent and Electronic Music* (2nd ed.). Routledge. https://www.routledge.com/The-Music-Producers-Survival-Guide-Chaos-Creativity-and-Career-in/Jackson/p/book/9781138697850.
[33] Kunal Jathal. 2017. Real-time timbre classification for tabletop hand drumming. *Comput. Music. J.* 41, 2 (2017), 38–51. DOI: https://doi.org/10.1162/COMJ_a_00419
[34] Nobuhiko Jin, Naoki Haga, and Issei Fujishiro. 2016. SeeGroove2: An orbit metaphor for interactive groove visualization. In *Proceedings of the International Conference on Cyberworlds*, Alexei Sourin (Ed.). IEEE Computer Society, 131–134. DOI: https://doi.org/10.1109/CW.2016.26
[35] Ruslan Kamolov, Penousal Machado, and Pedro Cruz. 2013. Musical flocks. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference*. ACM, 93. DOI: https://doi.org/10.1145/2503385.2503487
[36] Hideki Kawahara, Eri Haneishi, and Kaori Hagiwara. 2017. Realtime feedback of singing voice information for assisting students learning music therapy. In *Proceedings of the International Conference on Orange Technologies (ICOT)*. IEEE, 99–102. DOI: https://doi.org/10.1109/ICOT.2017.8336098

[37] Jeeeun Kim, Swamy Ananthanarayan, and Tom Yeh. 2015. Seen music: Ambient music data visualization for children with hearing impairments. In *Proceedings of the 14th International Conference on Interaction Design and Children (IDC'15).* 426–429. DOI: https://doi.org/10.1145/2771839.2771870

[38] Naoko Kosugi. 2010. Misual: Music visualization based on acoustic data. In *Proceedings of the 12th International Conference on Information Integration and Web-based Applications and Services*, Gabriele Kotsis, David Taniar, Eric Pardede, Imad Saleh, and Ismail Khalil (Eds.). ACM, 609–616. DOI: https://doi.org/10.1145/1967486.1967581

[39] Jonathan Lazar, Jinjuan Feng, and Harry Hochheiser. 2017. *Research Methods in Human-computer Interaction, 2nd Edition.* Morgan Kaufmann. Retrieved from http://www.sciencedirect.com/science/book/9780128053904.

[40] Arto Lehtiniemi and Jukka Holm. 2012. Using animated mood pictures in music recommendation. In *Proceedings of the 16th International Conference on Information Visualisation*, Ebad Banissi, Stefan Bertschi, Camilla Forsell, Jimmy Johansson, Sarah Kenderdine, Francis T. Marchese, Muhammad Sarfraz, Liz J. Stuart, Anna Ursyn, Theodor G. Wyeld, Hanane Azzag, Mustapha Lebbah, and Gilles Venturini (Eds.). IEEE Computer Society, 143–150. DOI: https://doi.org/10.1109/IV.2012.34

[41] James R. Lewis. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Int. J. Hum.-comput. Interact.* 7, 1 (1995), 57–78. DOI: https://doi.org/10.1080/10447319509526110

[42] Kyung Ae Lim and Christopher Raphael. 2010. InTune: A system to support an instrumentalist's visualization of intonation. *Comput. Music. J.* 34, 3 (2010), 45–55. DOI: https://doi.org/10.1162/COMJ_a_00005

[43] Hugo Lima, Carlos Santos, and Bianchi Serique Meiguins. 2019. Visualizing the semantics of music. In *Proceedings of the 23rd International Conference on Information Visualisation*, Ebad Banissi, Anna Ursyn, Mark W. McK. Bannatyne, Nuno Datia, Rita Francese, Muhammad Sarfraz, Theodor G. Wyeld, Fatma Bouali, Gilles Venturini, Hanane Azzag, Mustapha Lebbah, Marjan Trutschl, Urska Cvek, Heimo Müller, Minoru Nakayama, Sebastian Kernbach, Loredana Caruccio, Michele Risi, Ugo Erra, Autilia Vitiello, and Veronica Rossano (Eds.). IEEE, 352–357. DOI: https://doi.org/10.1109/IV.2019.00066

[44] Shixia Liu, Weiwei Cui, Yingcai Wu, and Mengchen Liu. 2014. A survey on information visualization: Recent advances and challenges. *Vis. Comput.* 30, 12 (2014), 1373–1393. DOI: https://doi.org/10.1007/s00371-013-0892-3

[45] Gareth Loy. 1985. Musicians make a standard: The MIDI phenomenon. *Comput. Mus. J.* 9, 4 (1985), 8–26. Retrieved from http://www.jstor.org/stable/3679619.

[46] Wakako Machida and Takayuki Itoh. 2011. Lyricon: A Visual music selection interface featuring multiple icons. In *Proceedings of the 15th International Conference on Information Visualisation*, Ebad Banissi, Stefan Bertschi, Remo Aslak Burkhard, Urska Cvek, Martin J. Eppler, Camilla Forsell, Georges G. Grinstein, Jimmy Johansson, Sarah Kenderdine, Francis T. Marchese, Carsten Maple, Marjan Trutschl, Muhammad Sarfraz, Liz J. Stuart, Anna Ursyn, and Theodor G. Wyeld (Eds.). IEEE Computer Society, 145–150. DOI: https://doi.org/10.1109/IV.2011.62

[47] Esteban Maestre, Panagiotis Papiotis, Marco Marchini, Quim Llimona, Oscar Mayor, Alfonso Pérez, and Marcelo M. Wanderley. 2017. Enriched multimodal representations of music performances: Online access and visualization. *IEEE Multim.* 24, 1 (2017), 24–34. DOI: https://doi.org/10.1109/MMUL.2017.3

[48] Akira Maezawa and Hiroshi G. Okuno. 2015. Bayesian audio-to-score alignment based on joint inference of timbre, volume, tempo, and note onset timings. *Comput. Music. J.* 39, 1 (2015), 74–87. DOI: https://doi.org/10.1162/COMJ_a_00286

[49] Delfina Malandrino, Donato Pirozzi, Gianluca Zaccagnino, and Rocco Zaccagnino. 2015. A color-based visualization approach to understand harmonic structures of musical compositions. In *Proceedings of the 19th International Conference on Information Visualisation*, Ebad Banissi, Mark W. McK. Bannatyne, Fatma Bouali, Remo Burkhard, John Counsell, Urska Cvek, Martin J. Eppler, Georges G. Grinstein, Weidong Huang, Sebastian Kernbach, Chun-Cheng Lin, Feng Lin, Francis T. Marchese, Chi Man Pun, Muhammad Sarfraz, Marjan Trutschl, Anna Ursyn, Gilles Venturini, Theodor G. Wyeld, and Jian J. Zhang (Eds.). IEEE Computer Society, 56–61. DOI: https://doi.org/10.1109/iV.2015.21

[50] Delfina Malandrino, Donato Pirozzi, and Rocco Zaccagnino. 2018. Visualization and music harmony: Design, implementation, and evaluation. In *Proceedings of the 22nd International Conference Information Visualisation*, Ebad Banissi, Rita Francese, Mark W. McK. Bannatyne, Theodor G. Wyeld, Muhammad Sarfraz, João Moura Pires, Anna Ursyn, Fatma Bouali, Nuno Datia, Gilles Venturini, Giuseppe Polese, Vincenzo Deufemia, Tania Di Mascio, Marco Temperini, Filippo Sciarrone, Delfina Malandrino, Rocco Zaccagnino, Paloma Díaz, Fragkiskos Papadopoulo, Antonio Fernández Anta, Alfredo Cuzzocrea, Michele Risi, Ugo Erra, and Veronica Rossano (Eds.). IEEE Computer Society, 498–503. DOI: https://doi.org/10.1109/iV.2018.00092

[51] Arpi Mardirossian and Elaine Chew. 2007. Visualizing music: Tonal progressions and distributions. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07).* 189–194. Retrieved from http://ismir2007.ismir.net/proceedings/ISMIR2007_p189_mardirossian.pdf.

[52] Georgios Marentakis and Kristoffer Jensen. 2014. The timbre engine - progress report. January 2001 (2014).

[53] Stephen McAdams. 2013. *Musical Timbre Perception* (3rd ed.). Elsevier Inc., 35–67. DOI: https://doi.org/10.1016/B978-0-12-381460-9.00002-X

[54] Alex McLean, Frederic F. Leymarie, and Geraint A. Wiggins. 2007. Apollonius diagrams and the representation of sounds and music. In *Proceedings of the 4th International Symposium on Voronoi Diagrams in Science and Engineering*. IEEE Computer Society, 276–281. DOI: https://doi.org/10.1109/ISVD.2007.7

[55] Philip McLeod and Geoff Wyvill. 2003. Visualization of musical pitch. In *Proceedings of the Computer Graphics International Conference*. IEEE Computer Society, 300–305. DOI: https://doi.org/10.1109/CGI.2003.1214486

[56] Matthias Miller, Alexandra Bonnici, and Mennatallah El-Assady. 2019. Augmenting music sheets with harmonic fingerprints. In *Proceedings of the ACM Symposium on Document Engineering (DocEng'19)*. 17:1–17:10. DOI: https://doi.org/10.1145/3342558.3345395

[57] J. B. Mitroo, Nancy Herman, and Norman I. Badler. 1979. Movies from music: Visualizing musical compositions. In *Proceedings of the 6th Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'79)*. Association for Computing Machinery, New York, NY, 218–225. DOI: https://doi.org/10.1145/800249.807447

[58] Meinard Müller. 2015. *Fundamentals of Music Processing - Audio, Analysis, Algorithms, Applications*. Springer. DOI: https://doi.org/10.1007/978-3-319-21945-5

[59] Tamara Munzner. 2014. *Visualization Analysis and Design*. A. K. Peters. Retrieved from http://www.cs.ubc.ca/%7Etmm/vadbook/.

[60] Tomoyasu Nakano, Masataka Goto, and Yuzuru Hiraga. 2007. MiruSinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data. In *Proceedings of the 9th IEEE International Symposium on Multimedia - Workshops*. DOI: https://doi.org/10.1109/ISMW.2007.4475948

[61] Suranga Chandima Nanayakkara, Elizabeth Taylor, Lonce Wyse, and S. H. Ong. 2007. Towards building an experiential music visualizer. In *Proceedings of the 6th International Conference on Information, Communications Signal Processing*. 1–5. DOI: https://doi.org/10.1109/ICICS.2007.4449609

[62] Lene Nielsen. 2019. *Personas—User Focused Design*. Springer London. DOI: https://doi.org/10.1007/978-1-4471-7427-1

[63] Donald A. Norman. 2013. *The Design of Everyday Things*.

[64] Kazuo Ohmi. 2007. Music visualization in style and structure. *J. Vis.* 10, 3 (2007), 257–258. DOI: https://doi.org/10.1007/BF03181691

[65] Allan V. Oppenheim and Ronald W. Schafer. 1998. *Discrete-time Signal Processing* (2nd ed.). Vol. 1. Prentice-Hall, Inc.

[66] Jack Ox. 2001. 2 performances in the 21st century virtual color organ: GridJam and im januar am Nil. In *Proceedings of the 7th International Conference on Virtual Systems and Multimedia*. DOI: https://doi.org/10.1109/VSMM.2001.969716

[67] Larry Polansky and Richard Bassein. 1992. Possible and Impossible Melody: Some Formal Aspects of Contour. *Journal of Music Theory* 36, 2 (1992), 259–284. DOI: https://doi.org/10.2307/843933

[68] Dionysios Politis, Dimitrios Margounakis, and Konstantinos Mokos. 2004. Visualizing the chromatic index of music. In *Proceedings of the 4th International Conference on WEB Delivering of Music*. IEEE Computer Society, 102–109. DOI: https://doi.org/10.1109/WDM.2004.1358106

[69] Jordi Pons and Xavier Serra. 2017. Designing efficient architectures for modeling temporal features with convolutional neural networks. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17)*. 2472–2476. DOI: https://doi.org/10.1109/ICASSP.2017.7952601

[70] Jordi Pons, Olga Slizovskaia, Rong Gong, Emilia Gómez, and Xavier Serra. 2017. Timbre analysis of music audio signals with convolutional neural networks. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO'17)*. 2744–2748. DOI: https://doi.org/10.23919/EUSIPCO.2017.8081710

[71] Roberto De Prisco, Delfina Malandrino, Donato Pirozzi, Gianluca Zaccagnino, and Rocco Zaccagnino. 2017. Understanding the structure of musical compositions: Is visualization an effective approach? *Inf. Vis.* 16, 2 (2017), 139–152. DOI: https://doi.org/10.1177/1473871616655468

[72] Roberto De Prisco, Delfina Malandrino, Donato Pirozzi, Gianluca Zaccagnino, and Rocco Zaccagnino. 2018. Evaluation study of visualisations for harmonic analysis of 4-part music. In *Proceedings of the 22nd International Conference Information Visualisation*, Ebad Banissi, Rita Francese, Mark W. McK. Bannatyne, Theodor G. Wyeld, Muhammad Sarfraz, João Moura Pires, Anna Ursyn, Fatma Bouali, Nuno Datia, Gilles Venturini, Giuseppe Polese, Vincenzo Deufemia, Tania Di Mascio, Marco Temperini, Filippo Sciarrone, Delfina Malandrino, Rocco Zaccagnino, Paloma Díaz, Fragkiskos Papadopoulo, Antonio Fernández Anta, Alfredo Cuzzocrea, Michele Risi, Ugo Erra, and Veronica Rossano (Eds.). IEEE Computer Society, 484–489. DOI: https://doi.org/10.1109/iV.2018.00090

[73] Markus Rovito. 2016. The MIDI Association Launches at NAMM 2016. Retrieved from https://www.emusician.com/gear/the-midi-association-launches-at-namm-2016.

[74] Justin Salamon and Emilia Gómez. 2012. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Trans. Speech Audio Process.* 20, 6 (2012), 1759–1770. DOI: https://doi.org/10.1109/TASL.2012.2188515

[75] Craig Stuart Sapp. 2001. Harmonic visualizations of tonal music. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.2001.029.

[76] Craig Stuart Sapp. 2005. Visual hierarchical key analysis. *Comput. Entertain.* 3, 4 (2005), 1–19. DOI: https://doi.org/10.1145/1095534.1095544

[77] Susanne Scheel. 2006. Music visualization—The interplay of color and sound. Ars Electronica 2006. Hatje Cantz Verlag, Ostfildern-Ruit, pp 273–289.
[78] Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, S. Dixon, Arthur Flexer, Emilia Gómez, F. Gouyon, Perfecto Herrera, Sergi Jordà, Oscar Paytuvi, G. Peeters, Jan Schlüter, H. Vinet, and G. Widmer. 2013. *Roadmap for Music Information ReSearch* Geoffroy Peeters (Ed.). Creative Commons BY-NC-ND 3.0 license.
[79] Yang Shi and Cheng Yang. 2013. Celestia: A vocal interaction music game. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Wendy E. Mackay, Stephen A. Brewster, and Susanne Bødker (Eds.). ACM, 2647–2650. DOI:https://doi.org/10.1145/2468356.2479485
[80] Najereh Shirzadian, Judith A. Redi, Thomas Röggla, Alice Panza, Frank Nack, and Pablo César. 2017. Immersion and togetherness: How live visualization of audience engagement can enhance music events. In *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology*, Adrian David Cheok, Masahiko Inami, and Teresa Romão (Eds.). Lecture Notes in Computer Science, Vol. 10714. Springer, 488–507. DOI:https://doi.org/10.1007/978-3-319-76270-8_34
[81] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. IEEE Computer Society, 336–343. DOI:https://doi.org/10.1109/VL.1996.545307
[82] Artur Silic and Bojana Dalbelo Basic. 2010. Visualization of text streams: A survey. Lecture Notes in Computer Science, vol. 6277. Springer, Berlin, Heidelberg. DOI:https://doi.org/10.1007/978-3-642-15390-7_4
[83] Sean M. Smith and Glen N. Williams. 1997. A visualization of music. In *Proceedings. Visualization'97 (Cat. No. 97CB36155)*. IEEE, 499–503. DOI:https://doi.org/10.1109/VISUAL.1997.663931
[84] Jon Snydal and Marti A. Hearst. 2005. ImproViz: Visual explorations of jazz improvisations. In *Extended Abstracts on Human Factors in Computing System (CHI'05)*. 1805–1808. DOI:https://doi.org/10.1145/1056808.1057027
[85] Sean Soraghan, Felix Faire, Alain Renaud, and Ben Supper. 2018. A new timbre visualization technique based on semantic descriptors. *Comput. Music. J.* 42, 1 (2018). DOI:https://doi.org/10.1162/comj_a_00449
[86] Beyond Sound. 2014. Beyond sound: The college and career guide in music technology. *Choice Rev. Online* 51, 05 (Jan. 2014). DOI:https://doi.org/10.5860/CHOICE.51-2405
[87] John Stainer and William Barret. 2009. *A Dictionary of Musical Terms*. Cambridge University Press.
[88] Michael Taenzer, Burkhard C. Wünsche, and Stefan Müller. 2019. Analysis and visualisation of music. In *International Conference on Electronics, Information, and Communication (ICEIC'19)*. 1–6. DOI:https://doi.org/10.23919/ELINFOCOM.2019.8706365
[89] Paolo Tagliolato. 2009. Synaesthetic analysis, exploring music structures by multimedia representation. prometheusz12, a novel 3D visualization tool. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.2009.022.
[90] Petri Toiviainen. 2005. Visualization of tonal content with self-organizing maps and self-similarity matrices. *Comput. Entertain.* 3, 4 (2005), 1–10. DOI:https://doi.org/10.1145/1095534.1095543
[91] Srinivasan Umesh, Leon Cohen, and Douglas J. Nelson. 1999. Fitting the Mel scale. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. 217–220. DOI:https://doi.org/10.1109/ICASSP.1999.758101
[92] Leonel Valbom and Adérito Marcos. 2007. An immersive musical instrument prototype. *IEEE Comput. Graph. Applic.* 27, 4 (2007), 14–19. DOI:https://doi.org/10.1109/MCG.2007.76
[93] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quart.* 27, 3 (2003), 425–478. Retrieved from http://misq.org/user-acceptance-of-information-technology-toward-a-unified-view.html.
[94] Gottfried von Bismarck. 1974. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica United with Acustica* 30, 3 (1974), 146–159.
[95] Martin Wattenberg. 2002. Arc diagrams: Visualizing structure in strings. In *IEEE Symposium on Information Visualization (INFOVIS'02)*. 110–116. DOI:https://doi.org/10.1109/INFVIS.2002.1173155
[96] Cong Xie, Wei Chen, Xinxin Huang, Yueqi Hu, Scott Barlowe, and Jing Yang. 2014. VAET: A visual analytics approach for e-transactions time-series. *IEEE Trans. Vis. Comput. Graph.* 20, 12 (2014), 1743–1752. DOI:https://doi.org/10.1109/TVCG.2014.2346913
[97] Ji Soo Yi, Youn ah Kang, John T. Stasko, and Julie A. Jacko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graph.* 13, 6 (2007), 1224–1231. DOI:https://doi.org/10.1109/TVCG.2007.70515

Received November 2019; revised February 2021; accepted April 2021

View publication stats