2019 23rd International Conference Information Visualisation (IV)

Visualizing the Semantics of Music

Hugo Lima
Federal University of Pará (UFPA)
Belém – Pará – Brazil
Email: hugo.lima@icen.ufpa.br

Carlos Santos
Federal University of Pará (UFPA)
Belém - Pará - Brazil
Email: carlosresque@ufpa.br

Bianchi Meiguins
Federal University of Pará (UFPA)
Belém - Pará - Brazil
Email: bianchi@ufpa.br

***Abstract*—We introduce SongVis, a visualization that represents music's semantic descriptors. SongVis uses emojis, colors, lines and shapes to embody the semantic content of a song. It aims to aid users on tasks related to exploration/browsing of music libraries and at queries for similar tracks based on visual characteristics. We first collected the descriptors after surveying papers on the topic of "music visualization", and used a questionnaire to rank the terms by consulting the public. Then, the features: mood, danceability, tempo, music genre and instrument, were extracted using state-of-the-art music information retrieval algorithms and we considered their visuals. Then, we discuss potential improvements to be made. With SongVis we perform a step forward towards visually representing semantic descriptors and expect that the music information research, the information visualization fields and the public can benefit from this work.**

***Index Terms*—music visualization, information visualization, semantic descriptors.**

I. Introduction

Music is a ubiquitous social phenomenon present arguably in all human societies. Despite being an aural phenomenon, many attempts aimed to represent music visually. Some intended to convey its content, others did merely as an artistic installation. Although music visualization is not a new topic (e.g., ancient Greeks associated music tones with colors), few tried to represent the subjective content of it.

Music needs to be listened to be comprehended, but a good visual representation can quickly deliver information of what's in a song, and, perhaps, introduce characteristics that were not noticed by listening. The problem lays in how to visually present, meaningful and natural, analogies regarding the semantics of a phenomenon that is strictly aural and also perception-dependent in its nature.

The most common visual representation is the *waveform*. It presents two dimensions, one being *time* (present in almost every music visualization conceived), the other is *amplitude* (associated with "loudness"), and it's used on popular websites such as FreeSound [1] and Soundcloud [2]. However, the waveform representation does not give details about the subjective content. Another common representation is the common music notation, used widely by western-music. It presents a variety of notes and information concerning how to play those notes, leaving to the reader/musician the task of interpreting/executing the piece.

No visualization can capture and represent all data content, therefore specialized visualizations depending on each niche or necessity are conceived. On tasks related to the exploration, for instance browsing music libraries, some descriptors are more desired to be visualized. We base our proposal on the premise that when users are looking for music, they are interested on certain features, such as: mood, danceability, genre, how fast a music is, etc. We consider the previous features after consulting the most represented characteristics in papers related to the subject of music visualization.

This proposal offers users a chance to gather better understandings when staring quickly to a visual representation of a song, leaving behind less expressive visual representations. Music catalogs can benefit, as music is represented more interestingly, based primarily on its abstract content –beyond the basic 'time x amplitude' used by the classic waveform representation. We hope our visualization proposal can aid users when doing tasks like: browsing music libraries searching for new songs, group similar songs based on its visual representation (similar songs will have similar visuals), recognize patterns in a song (e.g., checking for repetitive structures), filtering databases according to the desired visual features, and many others tasks related to the exploration of music collections and music analysis.

Our contributions are the following:

- We investigate which features the general public consider the most important when they wish to describe a song;
- A visualization prototype is built comprised by abstract features;
- Information Visualization techniques, such as: glyphs, patterns, colors and shapes, are broadly used to represent the features selected;
- Emojis are used to represent the majority of the features as a way to exploit the "knowledge in the world" in a novel way;
- Songs can be compared by their visuals, which represent directly an abstract feature;
- Interaction is provided, and highlights, in general, features that can be important at browsing and analysis tasks;
- A step forward is given aiming at solving the existing semantic gap between features and visual representations [3].

II. Related Work

The volume of papers published in the "visualization of music's semantic descriptors" topic is small, when compared with the amount of research done to represent quantitative

2375-0138/19/$31.00 ©2019 IEEE
DOI 10.1109/IV.2019.00066

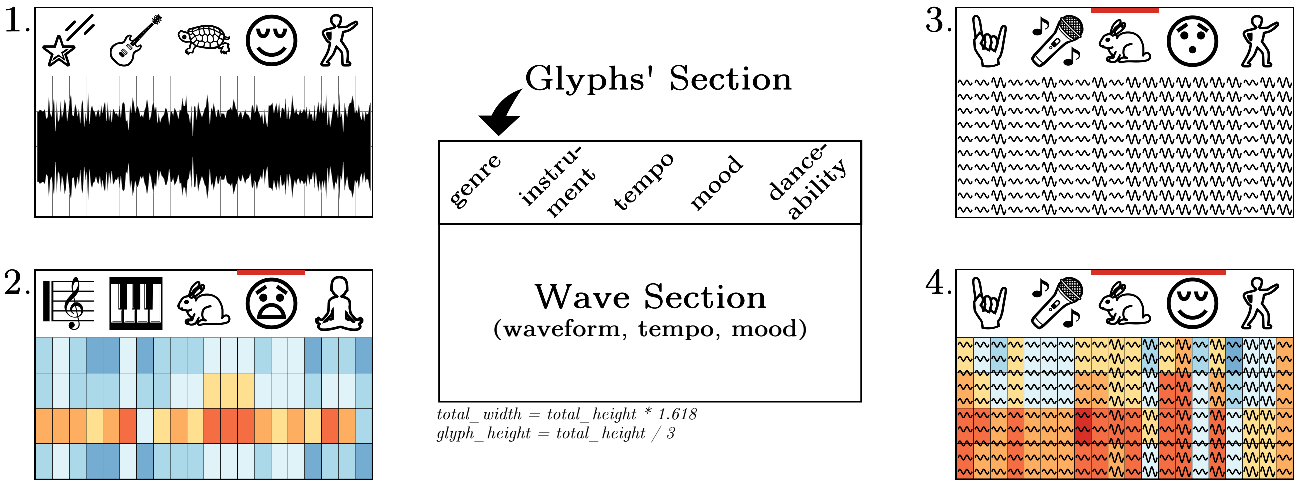

Fig. 1. Compilation of music visualizations using SongVis. In the center, the layout of SongVis, displaying the names of each section and at the bottom the calculation of the layout's dimensions. **1.** the first example depicts the song "I Want You" by Bob Dylan, primarily showing the waveform of the music file and *emojis* at the top representing 5 characteristics; **2.** representation of the Prelude "Op. 28 - 12. In G Sharp Minor" by Frederic Chopin, with the "mood" highlighted in the tiled map at the bottom of the picture: the large presence of *blue hues* indicates that the song is rather sad; **3.** representation of the song "Wonderwall" by Oasis with the bottom section (named "wave section") presenting changes in BPM; **4.** the song "The Prettiest Star" by David Bowie, with the "mood" selector and the "BPM/Tempo" selectors combined;

features (i.e., somewhat stable features, extracted using signal-processing algorithms). From the few proposals that aimed to visually represent the semantics of music, we can reference the work by Ciuha et al. [4], in which they assign colors to music intervals and their harmonies, saturation is used to denote loudness.

Wing-YiChan et al. [5] developed a modified version of the Common Music Notation (CMN); and mapped color, shape and texture to high-level features like timbre, melodic contour and harmonic tension; this way, users can notice patterns in the music structure more easily. Despite their initiative to represent semantic relations, they represented only a small variety of high-level features. Also, they primarily aimed to make the sheet music notation easier, so they were motivated by a music analysis task for musicians with knowledge of music notation. Our proposal is different because it represents many more high-level features and it is focused on search/browsing and simple analysis tasks for the general public.

Other proposals to be noted are: Gumulia et al. [6], which represent as circles and squares the perceived speed of a composition along the time domain; Wattenberg et al. [7] suggested the use of arcs to point out repetitive parts within a music; Malandrino et al. [8] published "VisualHarmony", a tool that aids users on analysis tasks, mainly tasks regarding the study of harmony in tonal music, highlighting passages and notes/chords considered more important. Grill et al. [9] represented perceptual qualities of textural sounds, which are stationary sounds, i.e., do not evolve overtime. They were assembled in a tiled map, forming a visual representation of sound diversity. The authors mapped high and low volume to variations on brightness and color hue; tonal and noisy auditory constructs were mapped to changes in saturation; and other features, aiming to represent both the structure of an entire collection and the properties of its individual elements.

Commercial platforms, like Apple Music [10], Spotify [11] and Google Play Music [12], use the cover of the album or a photo of the artist for illustration purpose only. We argue that a visual representation would help users to find similar tracks or new tracks based on their visuals.

For music libraries in the literature, however, [13] represented each song using a colored rectangular bar highlighting the most prominent features; they designed a feature extractor that captures the most relevant chord sequences to be represented. [14] used a connected corpus graph to map similarities between songs, based on their melodic structures; they aimed at allowing visual exploration of melodic relationships within traditional tune collections; the tunes were represented by means of vertices connected. [15] also used graphs and each song was represented by a node, placed in the screen by its mel-frequency cepstral coefficients (MFCCs); similar tunes, in that regard, were placed closed to each other. As seen, most of the proposals only provide basic representations which are not concerned whether they are useful to their targeted users, specially at tasks related to comparing songs within a music collection.

All the references denoted in this section, despite their limitations, are reference to SongVis, as many of their goals are similar to ours, i.e.: they highlighted patterns, represented the perceptiveness of features, visually represented subjective characteristics, represented the content using visuals and employed a wide range of Information Visualization (InfoVis) techniques, simple to complex ones, to meet their ends. The majority of the music visualizations consulted aimed to represent some few features, most lack formal studies and others lack practicality. We work on all those issues by implementing a useful music visualization; practical enough to be widely adopted by online music libraries, incorporating a wide range of semantic descriptors.

III. Visualizing the Semantics of Music

In this section we introduce SongVis, a music visualization that represents the semantic content of music. A group of semantic descriptors was selected, coming from a wide range of terms used to describe music, after a brief investigation with the general public. The features were extracted using state-of-the-art techniques and visual metaphors associated to them, such as: glyphs, colors, shapes, texture and other InfoVis' techniques. We begin this section by selecting what the general public consider the most important semantic descriptors. Latter their visual representation is discussed, i.e., we evaluate information visualization techniques that are useful at representing semantic descriptors.

A. Selecting the Semantic Descriptors

To investigate which features are the most important to represent the content of music, we surveyed 50 papers related to "semantic representation" and "feature visualization" of music. The words used to describe music were collected and an *online questionnaire* was prepared, letting users vote the most important features. This way, instead of arbitrarily defining which features we desire to represent, we asked users' opinions, and we assume that *the most voted features are the most interesting features to be represented.*

1) Methodology: From the papers consulted in the literature we found 48 terms used to describe the content of music. Then, we grouped the terms into similar categories, e.g., the words "sad" and "happy" were grouped into the same group "mood"; "chords", "notes" and "keys", for instance, can also be part of the same category, although each term describes something different; other categories needed exemplification, such as "instrument", allowing the respondent to clearly understand that the questionnaire is about *music* description; that was also the case for "genre", that needed exemplification to not be confused with other definitions of genre; "danceability" also needed some clarification in our view, we considered the tuple "danceable/not-danceable" much easier to be grasped.

The final result is a selection of *16 categories*, as listed at the "Features" column in Table I. Despite the assumption that the previous categorization was fair, we consider that the final result (i.e., the questionnaire) is much easier to be assimilated after our modifications. The next paragraphs describe how the questionnaire was structured and how the interviews were conducted.

Questionnaire. An online questionnaire was written to investigate which are the most important features used to describe music. The online address of the questionnaire was spread around social media and some interviews were made physically. 72 people responded the questionnaire and they came from unknown backgrounds, i.e., we did not focused on a specific group, such as musicians or non-musicians, but let everybody participate as a way to encompass the most general responses. This way a more general visualization can

TABLE I
Table for features *x* occurrences. The columns "Plays an Instrument" and "Does not play an instrument" are the average of the responses, the column "Ocurrences" is the sum of the averages.

Occurrences		**Features**	**Plays an Instrument**	**Does not play an instrument**
72	100.00%		39	33
49.50	68.75%	genre: pop, rock, funk, jazz, classic	26.50	23.00
40.50	56.25%	instrument: guitar, piano, bass, drums	25.50	15.00
32.75	45.49%	tempo: slow, fast	17.50	15.25
29.25	40.63%	mood: angry, cheerful, sad, happy, relaxed, tired	15.75	13.50
26.50	36.81%	danceable/not-danceable	9.25	17.25
20.50	28.47%	timbre: bright/warm/dull	13.25	7.25
20.00	27.78%	chords, notes, keys, pitch	15.00	5.00
16.50	22.92%	acoustic/electric	9.50	7.00
15.75	21.88%	harmonious/non-harmonious	7.75	8.00
15.00	20.83%	structures: verse, chorus, bridge	9.00	6.00
11.25	15.63%	rhythm: 3/4, 4/4	9.00	2.25
10.50	14.58%	smooth/coarse	4.50	6.00
9.25	12.85%	volume: high/low	2.25	7.00
6.75	9.38%	ordered/chaotic	4.25	2.50
3.50	4.86%	tonal/atonal/noisy	2.50	1.00
2.50	3.47%	homogeneous/heterogeneous	1.50	1.00

be made, reducing the necessity of a background knowledge to understand the features visually represented.

The first section of the questionnaire was comprised by information regarding the individual's age, sex and if he/she plays an instrument (39 played an instrument and 33 did not). The second section had three questions, they were in summary: a general question about the most important features according to his/her opinion, and other two questions involving music description. The respondents were allowed to mark down how many options they wanted, but a recommendation for selecting "up to five categories" was always given. The features, on each question, were ordered randomly whenever the respondent loaded the page, forcing them to read all the 16 categories listed, searching for the desired option to be selected. The songs were also randomly loaded from a range of 9 different genres, 2 sample songs for each genre.

2) Results: The responses collected previously resulted in a "occurrence *X* feature" table (Table I). We considered for visualization only features that were marked down by at least $1/3$ of the users consulted, this way, we reduce the amount of features to represent and focus on the study for the adequate visual representation.

B. Extracting the Semantic Descriptors

We used *Essentia* [16] to extract the features previously selected. Essentia is an Open-source library for audio analysis, it has an extensive collection of algorithms for music information retrieval (MIR), capable of extracting low-level and high-level features. Low-level features are mostly acoustic descriptors, which rely on signal-processing techniques, and their values are stable. On the other hand, High-level descriptors are extracted using pre-trained data (Support-Vector Machines (SVM) in the case of Essentia), and the accuracy of the classifier changes according to the technique and the dataset used.

A Python script organizes the many calls to the Essentia API, and outputs a JSON file containing rough information for three sections of SongVis: "waveform", "wavesection" and "glyphsection". Audio samples are extracted (10 samples per second) and later used to plot a waveform representation in SongVis. For the wavesection, the total samples are divided in 20 blocks, each block containing the values for the average bpm (avg_bpm), average mood (avg_mood), strong peak (strong_peak) and the average energy for 4 frequency bands: *low* [20,150] Hz, *middle_low* [150,800] Hz, *middle_high* [800,4000] Hz and *high* [4000,20000] Hz. Finally, the glyphsection contains probability values for *genre*, *predominant instrument*, *bpm*, *danceability*, *mood*. More about how each feature was extracted in the following paragraphs.

1) Genre: The dataset *genre_rosamerica* [17] was used for genre, with an accuracy of 87% and the following classes considered: classical, dance, hip-hop, jazz, pop, rhythm and blues, rock.

2) Instrument: We trained a custom SVM model for detection of the "predominant instrument" in music files using the IRMAS dataset [18]. We achieved a general accuracy of 74% for the following classes: cello, clarinet, flute, acoustic guitar, electric guitar, organ, piano, saxophone, trumpet, violin and human singing voice.

3) Tempo: Essentia has a BPM detector on its *rhythm extractor*. According to the classical subdivision of music tempo, in a scale from *Larghissimo* (less than 24 BPM) to *Prestissimo* (up to 200 BPM) [19], we divided between two glyphs, one for "slow" tempo (less than 108 BPM) and another for "fast" tempo (above 108 BPM).

4) Mood: To detect "mood", we used the the *mood_happy* default model provided by Essentia. The probability, from 0 to 1, is used in SongVis to pick the corresponding icon. The result is a probability considering "how happy" is a music passage. Thus, we used the probability to correlate directly with the sentiment scale for emojis by Novak et al. [20]. Normalized according to the following equation:

$$ScoreNormalized_i = \frac{|Score_{min}| + Score_i}{|Score_{min}| + Score_{max}} \quad (1)$$

5) Danceability: Also in the Essentia rhythm extractor, we extracted the *danceability* of a song. The value is in the range [0...3] and the *global mean* for danceability, as reported by [21], is 0.863, thus it is used as point to decide if a song is "danceable" or not.

C. Selecting the Visual Metaphors

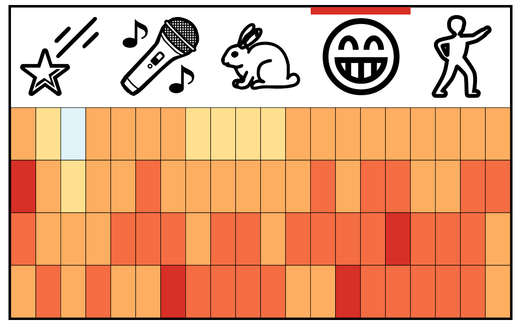

Fig. 2. The song "Mamma Mia" by ABBA represented in SongVis with the "mood" mapping selected.

In this section we study how to visually represent the semantic descriptors selected in the previous section. We considered many types of representations, such as: glyphs, shapes, hue, saturation, waveform, etc. Each representation is carefully selected and organized in the layout to avoid misrepresentations.

1) Glyphs: The features selected formerly are independent, i.e., they represent an autonomous set of attributes related to music description. A unique graphical representation for each feature would suffice the former characteristic, and data glyphs conform nicely with such idea.

Glyphs, as defined by Ware [22], are "graphical objects designed to represent some entity and convey one or numerical attributes of that entity". They are often used to reduce the visual trade-off existent between the number of data dimensions and the easiness to find items in potentially crowded visualizations [23]. Although the mapping *many-to-one* ("many features to one 'icon'") is the most used approach, *unique* representations (i.e., *one-to-one*) are also common in the literature, specially for "faces", according to Fuchs et al. [24]. For SongVis, we used glyphs at a one-to-one mapping to take advantage of the preattentive processing provided by the saccadic movements [22], the intention is to make the user notice the ideogram after quickly gazing through the visualization.

We reason that the glyphs' graphical design should be widely known (i.e., adequate for the general public, perhaps taking advantage of the common knowledge), expressive (i.e., they should convey an abstract idea with ease) and should be easily recognizable (i.e., compatible with the *preattentive processing*).

With the previous restrictions, Emojis were considered adequate candidates for SongVis. *Emojis* are graphic symbols and ideograms that represent objects, facial expressions, concepts and ideas [20], [25]. They are used globally, millions of times a day, in social networks and other communication tools. They are expressive, known and recognized by billions of users,

TABLE III
SELECTED ICONS FOR GENRE, INSTRUMENT, TEMPO, DANCEABILITY AND MOOD.

Genre				Instrument			
Feature	Description	Unicode	Icon	Feature	Description	Unicode	Icon
classical	musical score	U+1F3BC		voice	microphone	U+1F3A4	
dance	robot	U+1F916		guitar	guitar	U+1F3B8	
hip-hop	cd	U+1F4BF		piano	musical keyboard	U+1F3B9	
jazz	trumpet	U+1F3BA		violin	violin	U+1F3BB	
pop	shooting star	U+1F320		drums	drum with drumsticks	U+1F941	
r&b	sunglasses	U+1F576		saxophone	saxophone	U+1F3B7	
rock	sign of the horns	U+1F918					

Tempo				Danceability			
Feature	Description	Unicode	Icon	Feature	Description	Unicode	Icon
slow	turtle	U+1F422		danceable	man dancing	U+1F57A	
fast	rabbit	U+1F407		not-danceable	person in lotus position	U+1F9D8	

Mood							
Scale value [-1...1]	Scale value normalized [0...1]	Unicode	Icon	Scale value [-1...1]	Scale value normalized [0...1]	Unicode	Icon
0.568	1.000	U+1f600		-0.063	0.346	U+1f627	
0.557	0.989	U+1f603		-0.08	0.328	U+1f62a	
0.482	0.911	U+1f60c		-0.118	0.289	U+1f61e	
0.449	0.877	U+1f601		-0.146	0.260	U+1f614	
0.421	0.848	U+1f604		-0.212	0.192	U+1f623	
0.409	0.835	U+1f606		-0.368	0.030	U+1f629	
0.194	0.612	U+1f62c		-0.388	0.009	U+1f610	
0.123	0.539	U+1f62f		-0.397	0.000	U+1f615	

and performed an important role in our work, as the selected features (Section III-A) were represented as emojis.

Genre. We considered glyphs for the following music genres: classical, dance, hip-hop, jazz, pop, rhythm'n'blues, rock [17], [26]. According to Holm et al. [27], [28], "icons" are appropriate candidates to represent music genre. From their investigation we take the guidelines to design ours, i.e., we redrew the icons they used with emojis. Table III display the final result for the icons used to represent music genre. One exception was

Instrument. Icons for instruments were straightforward as the most popular had an emoji. For now we integrate icons for vocal, guitar, piano, violin, drums and saxophone (Table III).

Tempo. "The Hare and The Tortoise" fable by Aesop is used as metaphor for "fast" and "slow" tempo respectively (Table III).

Danceability. A "person dancing" and a "person in lotus position", motionless, were the emojis selected to represent if a music is danceable or not (Table III).

Mood. To represent "mood" with emojis, we used the sentiment ranking for Emojis by Novak et al [20]. In their work, the sentiments of the emojis were evaluated according to the sentiment of the message in which they were part, i.e., they collected *tweets* and annotated categorically as a *negative, neutral, positive* message. In the end, each emoji had a sentiment value in a scale [-1...1].

Donato et al. [29] and Capallo et al. [25] highlighted the predominant *non-redundancy* between *tweets x emojis*, i.e., the content of the tweet is, most of the time, *different* with the semantic value of the emoji it has. However, we only adopted the sentiment scale for a positive or negative sentiment, leaving apart more complex semantical interpretations. In any case, Donato et al. themselves admitted the broad range of readings an emoji can encompass [29].

"Faces" were selected to represent mood and they were filtered to remove complex emotions, e.g, a "smiling face" (😃) was allowed, but a "face screaming in fear" (😱) was not allowed. The goal is to leave only emojis that display "emotion" in the scope sadness-happiness. Table III lays out the final icons chosen to depict "mood".

2) Hue and Saturation: In the wavesection, when the "mood" option is selected, the tiled map is colored according to the *strong peak*, *energy* of the frequency band and the general *mood* value. The strongest the peak, the energy and the general mood, *reddier* is the color using the following color scale provided by [30]:

Fig. 3. Color scale to represent mood: "non-happy" to "happy".

3) Patterns: Patterns represent how fast a music passage is and the appropriate sinewave pattern is selected for the whole column. We picked three possible representations for three considered intervals of BPMs:

TABLE IV
BPM CATEGORIES CONSIDERED.

$BPM < 80$ (*andante*)	$80 <= BPM < 120$ (*moderato*)	$BPM >= 120$ (*allegro*)

4) Waveform: The classical waveform is again used to depict sound, giving users the whole picture and a *sense of duration*.

D. General Principles

The visualization is divided in two parts: One represents 5 semantic descriptors using *glyphs* (glyphs's section) and is positioned above the visualization in five squares. Below the glyph's section we represent the wave section, representing three types of visuals and one combination. The first representation is of the music's waveform, the second is the BPM representation using sinewave patterns, the third is the Mood mapping using a fixed number of colored blocks (4x20: 4 energy bands and 20 equally-sized slices of the music samples). Users can also combine the BPM selection with the

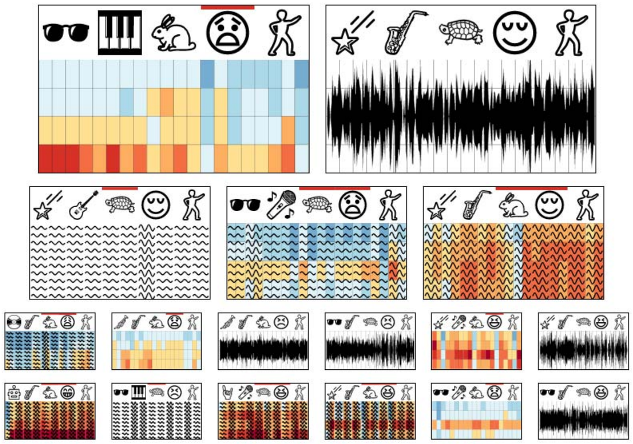

Fig. 4. A compilation of a bunch of songs represented using SongVis.

Mood selection to obtain a more customized visual to aid their tasks.

IV. Conclusion and Future Works

In this paper we proposed a new visualization to represent music's content, which can potentially aid users at exploration tasks by using pre-attentive processing. However we only designed the general guidelines in which a more profound research can be addressed. Thus the most obvious following step is to implement the visualization and evaluate with users, musicians and non-musicians the efficacy and appropriateness of the choices here made.

Although the visual representation of music is not a new topic, not many rigorous works have been made and, although we employed formal questionnaires evaluating the initial selection of the features, we know that much work need to be done regarding the final steps of the design of this visualization. The challenge is to create a visualization natural enough so users with no previous contact can assert some correct interpretations about the music content represented.

Therefore, the contributions here made were in the aspect of acknowledging the lack of music's visualizations that considered the general public preference for which feature to represent, and we solved that by conducting a formal study by means of a questionnaire. Also, potential widely recognizable visual metaphors for music were selected (the case of emojis). Finally the picture is assembled and some interaction (details on demand and filtering) is proposed.

References

[1] "Freesound.org," www.freesound.org, accessed: 2019-02-02.

[2] "Soundcloud," www.soundcloud.com, accessed: 2019-02-02.

[3] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jorda, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer, *Roadmap for Music Information ReSearch*, 2013.

[4] P. Ciuha, B. Klemenc, and F. Solina, "Visualization of concurrent tones in music with colours," *Proceedings of the international conference on Multimedia - MM '10*, p. 1677, 2010.

[5] Wing-Yi Chan, H. Qu, W.-H. Mak, Huamin Qu, and Wai-Ho Mak, "Visualizing the Semantic Structure in Classical Music Works," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 1, pp. 161–173, jan 2010.

[6] A. Gumulia, B. Puzon, and N. Kosugi, "Music visualization," in *Proceedings of the 19th ACM international conference on Multimedia - MM '11*. New York, New York, USA: ACM Press, 2011, p. 949.

[7] M. Wattenberg, "Arc diagrams: Visualizing structure in strings," *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, vol. 2002-Janua, pp. 110–116, 2002.

[8] D. Malandrino, D. Pirozzi, and R. Zaccagnino, "Visualization and Music Harmony : Design , Implementation , and Evaluation," in *Information Visualization*, 2018.

[9] T. Grill and A. Flexer, "Visualization of Perceptual Qualities in Textural Sounds," *ICMC Proceedings*, 2012.

[10] "Apple," www.apple.com/apple-music/, accessed: 2019-05-07.

[11] "Spotify," www.spotify.com, accessed: 2019-05-07.

[12] "Google play music," www.freesound.org, accessed: 2019-05-07.

[13] A. Soriano, F. Paulovich, L. G. Nonato, and M. C. F. Oliveira, "Visualization of Music Collections Based on Structural Content Similarity," in *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*. IEEE, aug 2014, pp. 25–32.

[14] C. Walshaw, "A Visual Exploration of Melodic Relationships within Traditional Music Collections," in *Information Visualization*, 2018, pp. 478–483.

[15] C. Muelder, T. Provan, and K.-L. Ma, "Content Based Graph Visualization of Audio Data for Music Library Navigation," in *2010 IEEE International Symposium on Multimedia*. IEEE, dec 2010, pp. 129–136.

[16] D. Bogdanov, X. Serra, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, and J. Zapata, "Essentia," *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pp. 855–858, 2013.

[17] E. Guaus, "Audio content processing for automatic music genre classification: descriptors, databases, and classifiers," Ph.D. dissertation, Universitat Pompeu Fabra, 2009.

[18] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, "A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals," in *ISMIR*, 2012.

[19] J. Stainer and W. Barret, *A Dictionary of Musical Terms*. Cambridge University Press, 2009.

[20] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PLoS ONE*, vol. 10, no. 12, p. e0144296, 2015.

[21] S. Streich and P. Herrera, "Detrended Fluctuation Analysis of Music Signals Danceability Estimation and further Semantic Characterization," *118th Audio Engineering Society Convention – AES 118*, no. January 2012, 2005.

[22] C. Ware, *Information Visualization: Perception for Design: Third Edition*. Elsevier, 2012.

[23] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proceedings 1996 IEEE Symposium on Visual Languages*. IEEE Comput. Soc. Press, 1996, pp. 336–343.

[24] J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim, "A systematic review of experimental studies on data glyphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 7, pp. 1863–1879, 2017.

[25] S. Cappallo, S. Svetlichnaya, P. Garrigues, T. Mensink, and C. G. Snoek, "New Modality: Emoji Challenges in Prediction, Anticipation, and Retrieval," *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 402–415, 2019.

[26] "Acousticbrainz," https://acousticbrainz.org/datasets/accuracy.

[27] J. Holm and H. Holm, "Associating icons with musical genres," in *Proceedings of the 12th IASTED International Conference*, vol. 624, no. 005, 2008, p. 81.

[28] J. Holm and H. Siirtola, "A Comparison of Methods for Visualizing Musical Genres," in *2012 16th International Conference on Information Visualisation*. IEEE, jul 2012, pp. 636–645.

[29] G. Donato and P. Paggio, "Investigating redundancy in emoji use: Study on a twitter based corpus," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, 2017, pp. 118–126.

[30] "Color schemes," https://observablehq.com/@d3/color-schemes, accessed: 2019-03-20.