THUMBNAILDJ: VISUAL THUMBNAILS OF MUSIC CONTENT

Ya-Xi Chen

Media Informatics, University of Munich
Amalienstr. 17, 80333 Munich, Germany
`yaxi.chen@ifi.lmu.de`

René Klüber

Media Informatics, University of Munich
Amalienstr. 17, 80333 Munich, Germany
`klueber@cip.ifi.lmu.de`

ABSTRACT

Musical perception is non-visual and people cannot describe what a song sounds like without listening to it. To facilitate music browsing and searching, we explore the automatic generation of visual thumbnails for music. Targeting an expert user groups, DJs, we developed a concept named ThumbnailDJ: Based on a metaphor of music notation, a visual thumbnail can be automatically generated for an audio file, including information of tempo, volume, genre, aggressiveness and bass. We discussed ThumbnailDJ and other 3 selected concepts with DJs, and our concept was preferred most. Based on the results of this interview, we refined ThumbnailDJ and conducted an evaluation with DJs. The results confirmed that ThumbnailDJ can facilitate expert users browsing and searching within their music collection.

1. INTRODUCTION

People can easily gain an overview of a photo by glimpsing at its thumbnail. With assistance of multiple thumbnails, people can browse many photos in parallel and locate the desired ones quickly. On the contrary, music carries no visual information and people cannot describe what a song sounds like without listening to it. Cover art is commonly used as a visual assistance for music. However, it only visually encodes the relevant artist and album, and has no reflection on the intrinsic music content. Will a visualization of musical content help, and who will benefit from it? To answer these questions, we explore the automatic generation of visual thumbnails for music content and develop a concept named ThumbnailDJ. We conducted several rounds of survey and interview, and the results confirmed that our concept can help expert users browsing and searching within their music collections.

2. RELATED WORK

A map-based representation is widely used to display music collections. Islands of Music [6] and MUSICtable [9] cluster songs based on their acoustic similarity. In Artist Map [13] the user can chose any two of the four criteria of mood, genre, year and tempo to display music on a map. In iCandy [3], songs are displayed in a grid layout and their order is determined by the selected criteria such as genre, most played artist or album. Besides map-based representation, there are other visualizations of music collections. Torrens et al. [10] presented three visualization concepts: a disc, a rectangle and a tree. In MusicRainbow [7] artists are displayed in rings of a rainbow.

The aforementioned visualizations focus on the representation of an entire music collection, in which single songs are either displayed with the cover art or as the name of artist, album or song, and none of them reflect the intrinsic music content. Some work addressed this issue by producing visualization for the file content. Semanticons [8] produce semantically meaningful icons for different file types. Music Icon [5] is a similar concept using a blossom metaphor: Each music file is represented as a blossom icon with two rings of petals (see Figure 1d for a simplified version). The music feature is reflected by the color, shape and number of petals.

Besides these metaphoric visualizations, some researchers focus on the sequential representation. Beat Histogram [11] is a temporal representation of beat strength in audio signals (see Figure 1a). It helps to gain an overall impression of how beat strength changes over time. TimbreGrams [12] represents an audio file as sequential stripes. Bright colors correspond to speech or singing, and purple and blue ones associate to classical music (see Figure 1c). In Arc Diagrams [14], instances of the identical notes are connected by arcs, thus depicting the repetitive structure in a music file (see Figure 1b).

3. ONLINE SURVEY

In order to assess the understandability and suitability of the concept of visualizing music content, we conducted an online survey with four selected concepts (see Figure

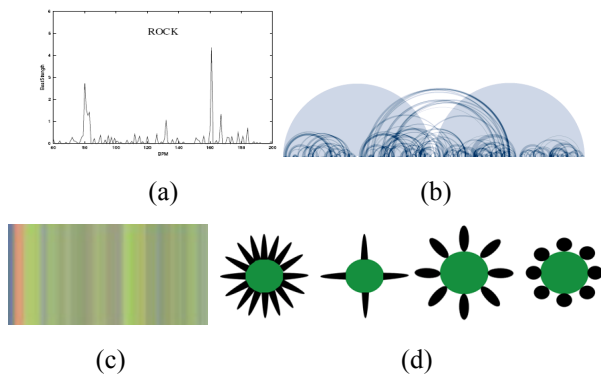1): Beat Histogram, Arc Diagram, TimbreGrams and simplified Music Icon.

(a)   (b)

(c)   (d)

Figure 1. Four selected concepts. a) Beat Histogram [11]. b) Arc Diagram [14]. c) TimbreGrams [12]. d) A simplified Music Icon has one ring of petals. The number and shape of petals represent tempo and aggressiveness respectively.

3.1 Survey Design

The participants first filled out a questionnaire about their personal information and general experience with music. Then the four visualization concepts were briefly introduced. The participants answered 15 questions about these concepts: We chose 8 popular songs and asked the participants about their familiarity with these songs. They could follow the corresponding links to listen to these songs online. For each concept, they were required to map one visualization to a correct song out of 4 candidates, then map one song to a correct visualization out of 4 candidates. After this, they were asked about information that can be derived from each concept. Then we asked them about their preference between and comments about these concepts. All scores were rated on a 5-point Linkert-scale where 5 represented the highest score.

3.2 Participants

In total we received 38 complete questionnaires, 9 female and 29 male. Their age ranged from 18 to 55 with an average age of 26.6 years. 31 out of 38 participants were students and employees from Europe.

3.3 Results

Regarding the reflection on music content, the participants thought attributes such as melody, mood, rhythm, instrument and genre were more important than the general information of lyrics, length and release year. Concerning the usefulness of different information in helping gaining an overall impression of a song, the 30-second preview clip received the highest score (M=3.84, SD=0.17). Similar artists/tracks (M=3.16, SD=0.19) and top tags (M=2.78, SD=0.19) were scored lower. Unfortunately, our concept of visual thumbnails of music content was rated lowest (M=2.22, SD=0.16).

Although the participants were generally familiar with the tested songs (M=3.50, SD=0.74), the correctness of their answers was quite low: To map a visualization to a correct song out of 4 candidates, 14 participants (36.8%) chose the correct song for Music Icon, 12 (31.6%) for TimbreGrams, 9 (23.7%) for BeatHistogram and 8 (21.1%) for Arc Diagrams. To map a song to a correct visualization out of 4 candidates, the performance was slightly better but still low: 17 (44.7%) for Music Icon, 16 (42.1%) for BeatHistogram, 10 (26.3%) for Arc Diagrams and 8 (21.1%) for TimbreGrams.

Concerning the information that can be derived from each concept, the results illustrated that BeatHistogram helped to learn tempo and volume, and Music Icon tempo and volume. TimbreGrams and Arc Diagrams facilitated gaining song structure and information of harmony. In general, the scores for easiness of deriving each information were rather low (all below 3). The usefulness of each concept was also rated quite low: BeatHistogram (M=2.17, SD=1.40), Arc Diagrams (M=2.09, SD=1.17), Music Icon (M=1.97, SD=1.32) and TimbreGrams (M=1.40, SD=0.55).

3.4 Discussion

All the participants commented that these visualizations were overall too complicated for them. Although they agreed that attributes such as melody and rhythm are important features of a song, they did not think such information can help them to gain the overall impression of a song. Instead, they would prefer direct and non-trivial assistance, such as 30-second preview clips.

4. CONCEPT DEVELOPMENT

The results of the survey revealed an overall low appreciation for the concepts of visualizing musical content. The normal music listeners seem not be unsuitable as the consumers of those thumbnails, as they have neither requirements nor efficient knowledge in understanding technical details of a song. Therefore, we shift our focus to more professional users, Disk Jockeys (DJs). We develop a concept named ThumbnailDJ. We discussed ThumbnailDJ and 3 other concepts with DJs. ThumbnailDJ was preferred most, and we derived implications for the refinement of this concept.

4.1 Four Tested Concepts

We first conducted a preliminary discussion with 7 DJs with the four concepts presented in Figure 1. They were overall appreciated the idea of visualizing musical content. Arc Diagram and TimbreGrams were commented as helpful to gain an overall impression of a song, but lacking of precise values, such as those shown in Beat Histogram. Beyond the single criterion displayed in Beat Histogram, more attributes were requested, such as aggres-

siveness and volume. Music Icon was generally preferred most among these four tested concepts.

The results of the preliminary test illustrated that expert users require multiple attributes with precise values, and simplicity is vital for mental perception [1]. Therefore, we selected four simple and compact visualizations and tested them with DJs. Havre et al. [4] introduced a river metaphor to represent topical changes within a document collection. We employ this concept to describe temporal changes of attributes (see Figure 2a). Border Community[1] offers hand-drawn graphs to show the composition of songs. The amplitude depicts intensity and gray color represents bass strength. We name this visualization TensionDiagram and map its background color to genre (see Figure 2b). The third concept is Music Icon, which has already been tested in the previous online survey (see Figure 1d).
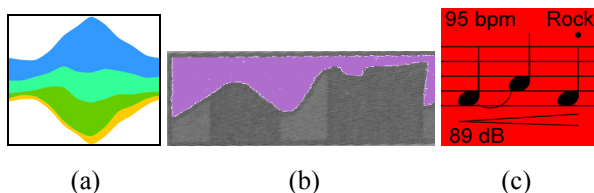

(a)   (b)   (c)

Figure 2. Tested concepts with DJs. a) ThemeRiver [4]. b) TensionDiagram[1]. c) Initial concept of ThumbnailDJ.

Besides these three selected visualizations, we develop our own concept ThumbnailDJ (see Figure 2c). Our design is built on the metaphor of music notation, the most common symbolic representation of music, from which abundant information can be read out: Pitch is shown as the vertical position of notes on a five-line staff. Duration is illustrated as the note value and additional symbols such as dot and tie. A dot extends the value of a note and a tie connects two notes with same pitch. Tempo and dynamics are shown above or below the staff. Tempo is normally represented as Beats per Minute (BPM), and dynamics as the overall volume of the whole piece. By reading the staff from left to right, the overall temporal impression can be gained.

In our initial concept, we made some modification of the original metaphor. The lengthy notation contradicts to the fundamental characteristic of compactness, and thus we decided to employ only three notes, each representing 1/3 of the song. The vertical location of a note in the staff depicts the aggressiveness. Two notes are connected by a tie if they share similar aggressiveness. On the contrary, a dot on the top of a note stands for change in aggressiveness. Tempo and volume are shown in the top and bottom left corners respectively. Crescendo (<) and decrescendo (>) describe the increase/decrease of volume. The genre is shown in the top right corner and associated with the background color of the entire graphics. Figure 2c shows

a medium-tempo calm rock song which gets louder over time. Aggressiveness keeps constant in the first two parts and changes in the third part.

4.2 Discussion with DJs

In order to test the suitability of the four selected concepts, we conducted a second round of discussion with the same 7 DJs. The open questions covered mainly their routine tasks and general impression of the tested concepts, which were introduced in a blind fashion. The participants have DJ experience about 10 years in average. They are all male, and their age range from 24 to 37 with an average age of 28.4 years. Two of them play mainly Hip-Hop music, three play Electronic, and the other two play diverse genres. Two DJs play analogue music and the others play digital. All DJs can read music notation.

Analogue DJs organize their collections on shelves, either sorting them by alphabet or genre, or no ordering at all. Digital DJs store their collections on hard disk and sort them by folders and ID3tags. They rely heavily on the search functionality in DJ software to look for music.

Concerning the desired music attributes, DJs playing multiple genres required a general impression of a song. DJs playing only one or few genres requested detailed temporal information. The generally important attributes were genre, tempo, aggressiveness and volume.

The idea of visualizing music content was overall appreciated. TensionDiagram was well accepted, as it is similar to the signal histogram shown in most of the DJ software. Consistent with the preliminary test, Music Icon and ThemeRiver were commented as lacking of precise values. ThumbnailDJ was preferred most, as it uses the general metaphor of music notation and reflects both overview and precise information for the most desired attributes of genre, tempo, aggressiveness and volume. The participants' comments indicated further improvement of ThumbnailDJ. Representing the entire piece of song as three separated parts was commented as simple and easy to learn. But a continuous flow description was desired, which helps to gain temporal changes and facilitate observation of the representative parts, such as peaks and gaps. Bass value was desired, as it helps to achieve smooth transition between songs. The symbols of crescendo, decrescendo, dot and tie were commented as less important and poor readable in a size-limited thumbnail.

5. CONCEPT REFINEMENT

Based on the implications derived in the discussion with DJs, we refined ThumbnailDJ (see Figure 3a): We excluded the symbols of crescendo, decrescendo, dot and tie. We associated pitch (namely the vertical location of a note in the staff) with bass value. For example, a note on the bottom line implies a higher bass value. The bass value is also represented by the position of note head: A note with head on the bottom represents higher bass

value, and head on the top lower bass value. Aggressiveness is depicted as note value, for example, half note for calm and sixteenth for aggressive (see Figure 3b). A light gray flow is drawn under the staff, indicating changes of aggressiveness. Genre is displayed in the top right corner and also represented as the background color of the entire graphics. Tempo is represented as BPM in the top left corner, and volume as decibel (dB) in the bottom left corner. Currently we define 6 main categories of genres (see Figure 3c), and more genres can be easily included. Figure 3a shows an example thumbnail for a rock song, with average tempo of 108 BPM and volume of 94 dB. It is quite aggressive and very heavy on bass in both beginning and end parts. The background flow illustrates that aggressiveness descends constantly in the middle part, rises again and becomes quite fluctuate in the last part.
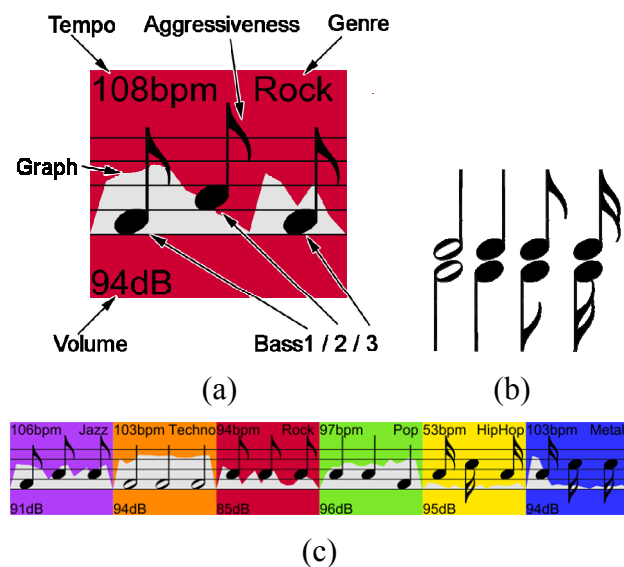

(a)   (b)

(c)

Figure 3. The refined ThumbnailDJ. a): Rock song " Biffy Clyro-As dust dances". b) 8 kinds of notes. The values of aggressiveness and bass are both mapped to a discrete value between 1 and 4. The note with head on the bottom represents bass value of 3 or 4, and on the top 1 or 2. Half to sixteenth notes indicate aggressive values from 1 to 4 respectively. c) Some example thumbnails for different genres.

5.1 Implementation

ThumbnailDJ is implemented in Java. All songs and their relevant data are saved in a SQLite database. We use Tritonus[1] to read in the ID3tags of a song. JLayer[2] is used to decode the audio file. After comparing the performance of different audio features, we decided to use zero crossings and Fast Fourier Transformation (FFT) to calculate values of aggressiveness and bass. We first use mp3splt[3] to cut the audio file into 15 snippets with equal length.

We then apply jAudio[4] to extract the corresponding low-level features from each snippet. We compute the value of zero crossings over each snippet. The average value of each successive 5 snippets represents aggressiveness for each 1/3 part of the song. Bass is determined by the FFT frequencies. For each snippet, the frequencies are sorted from low to high. The bass value is represented as the sum of lower 1/4 frequencies divided by the sum of all frequencies. The average value of each successive 5 snippets represents bass value for each 1/3 part of the song. Both aggressiveness and bass are normalized to a discrete value between 1 and 4, in order to map them to one of the 8 note shapes (see Figure 3b). Volume is determined by the average value of Root Mean Square (RMS) and tempo by the average value of beats in the beat histogram.

6. EVALUATION

We conducted a user study with DJs to evaluate the performance of ThumbnailDJ. We were specifically interested in how it helps gaining an overall impression of a song, and facilitating browsing and searching in a music collection.

6.1 Settings and Procedure

As DJ equipments were required, the evaluation was conducted in the work places of the participants. Each participant was asked to offer a collection of 100 songs. For each collection, the thumbnails were generated before hand and shown in Windows Explorer in another laptop. On average the user study lasted about 90 minutes per participant. It was recorded on video using the Think-Aloud protocol was applied. All scores were rated on a 5-point Linkert-scale where 5 represented the highest score.

Since the participants already joined the former interview and they preferred ThumbnailDJ most, in this evaluation we focused on ThumbnailDJ and did not compare it with other concepts. After a brief introduction of refined ThumbnailDJ, the participants were asked to describe their impression of two unfamiliar songs by viewing the corresponding thumbnails. Then they were shown thumbnails of two familiar songs and asked to rate how well these thumbnails describe these songs. After that, they executed a routine task with their own methods (two participants with analogue music and the other three with digital) and through browsing the corresponding thumbnails respectively: Finding some appropriate songs as intro, bridge and outro respectively for an X (the genre the participant often plays) party. The order of their own methods and ThumbnailDJ was counterbalanced between the participants to minimize learning effects. After com-

---

[1] http://www.tritonus.org/

[2] http://www.javazoom.net/javalayer/javalayer.html

[3] http://mp3splt.sourceforge.net/mp3splt_page

[4] http://jmir.sourceforge.net/jAudio.html

pleting all tasks, they filled out a questionnaire concerning their overall impression of ThumbnailDJ.

6.2 Participants

We recruited 5 DJs, who took part in the earlier discussions. Their age ranged from 24 to 31, with an average age of 27.4 years. They are all experience DJs with average experience about 10 years. Two participants play mainly Electronic music, two play Hip-Hop, and the other one plays multiple genres. Two participants play analogue music and the other three digital. All participants can read musical notation.

6.3 Results

The meaning of attributes shown in the thumbnails was clear to all participants. They claimed that these thumbnails reflected efficiently the features of the tested songs. However, the combination of multiple attributes was still insufficient to help gaining the sense of an unfamiliar song. The participants were reluctant to "guess" the feeling of an unfamiliar song without listening to it. ThumbnailDJ was thus considered rather useful as a quick visual reminder of a familiar song. Some additional attributes were desired, such as vocal and instrument.

Song selection was overall very subjective, which was also influenced by the context, such as the audience feedback, the performance duration, the order of DJs in the same show and music played by the previous DJ. Songs played as intro were characterized as moody and relaxing, and thus slow and calm songs were selected. Outro songs were similar to the intros, and those with similar tempo were composed in a block. Bridge song should fit the tempo and bass intensity of the connecting songs.

Using their own methods and ThumbnailDJ respectively, there were on average 5 songs selected in each session. Among these selections each DJ picked 2 to 3 same songs and other songs were quite similar. When asked about their selection criteria, the participants pointed out the decisive factors such as song attributes and other contextual considerations. However, they could not formulate formally why they chose a specific song, as "music is kind of sense that can not be precisely described" (DJ 6). Therefore, with the general open tasks of song selection, we could not collect details about how ThumbnailDJ assisted search and browsing in certain aspects.

Considering the completion time, digital DJs were overall faster. Their tag-based search was comparable with browsing thumbnails. Analogue DJs were slightly slower by flipping through the analogue collections. They claimed the time difference would become more noticeable with a larger collection.

The concept of visualizing music content was generally appealing (M=4.20, SD=0.84). The impression of ThumbnailDJ was quite positive in the aspects of ease of use (M=4.0, SD=1.40), learnability (M=4.80, SD=1.84) and understandablity (M=4.20, SD=1.27). Enjoyment was rated lower (M=3.40, SD=0.63). Most participants believed that the performance of ThumbnailDJ was promising, but needed more graphical and acoustical appealing effects. All participants expressed high willingness to have ThumbnailDJ as plug-in in their DJ software, (M=4.80, SD=0.45).

6.4 Discussion

The final feedback was quite encouraging, and we received valuable implications for further improvement. Besides the included attributes of tempo, volume, genre, aggressiveness and bass, some additional information is desired, such as vocal and instrument, which are important for a smooth transition between songs. Currently, attributes are directly extracted from low-level features, and more elaborate mapping algorithms should be integrated to achieve a better association between low-level features and high-level perception. Music taste is subjective and different users may have different requirements. Besides, it is not practical to display too much information in a compact thumbnail. Therefore, we suggest employing a personalization mechanism, and thus the user can produce personalized thumbnails, for example, by defined the desired attributes shown in the thumbnail.

Concerning the evaluation method, the collected data was mainly qualitative. Detailed information about the performance of ThumbnailDJ could not be collected with the open tasks. To derive deeper understanding of how such visualization facilitates browsing and decision making, more controlled tasks should be considered. With a possible integration in existing DJ software, a field study in a real DJ working environment will help to gain more insights on the practical usage of such a tool.

7. CONCLUSION AND FUTURE WORK

We explore how to facilitate browsing and searching within music collections with the assistance of visual thumbnails of music content. Based on the metaphor of music notation, we developed ThumbnailDJ to generate thumbnails for music content, which include information of tempo, volume, genre, aggressiveness and bass. We discussed ThumbnailDJ and other 3 selected concepts with DJs, and our concept was preferred most. We then refined ThumbnailDJ and conducted an evaluation with DJs. Our concept received overall positive feedback, especially towards its helpfulness for quickly recalling of a familiar song. Moreover, DJs showed high willingness to have such a plug-in in their DJ software.

We discussed with a senior developer of Traktor[1] the potential integration of ThumbnailDJ into their DJ software. The feedback was quite encouraging. The different

---

[1] http://www.native-instruments.com/en/products/dj/traktor-pro

representations of the same information, such as genre, bass and aggressiveness, were especially appreciated. Considering a possible integration into a commercial product, some modification of ThumbnailDJ is necessary, for example, displaying more meaningful information for tempo and volume beyond the currently discrete values.

Besides DJs, online audio searchers might be another potential user group. For example, in iStockPhoto[1], to determine which audio file to buy, the user has to listen to many retrieval results, which is quite time-consuming. We expect that browsing and decision-making can be facilitated by assisting the audio files with visual thumbnails of their content. Then the user can exclude unmatching files quickly, or scan all thumbnails in parallel while looking for candidates satisfying certain graphical features. We discussed ThumbnailDJ with 10 online audio searchers. They were offered a set of 20 thumbnails, covering diverse genre, tempo, volume and aggressiveness. They first categorized these thumbnails and sorted them in each category. They then selected an audio file for a coffee advertisement. All participants could manage these tasks and commented the assistance of ThumbnailDJ as helpful.

We agree that "Music is more of an art than science" [2]. We also believe that visualizations can help people gaining more musical insights. We hope our exploration can shed some light on facilitating browsing and searching within music collections by bridging the perceptions of vision and acoustics.

8. ACKNOWLEDGEMENT

9. REFERENCES

[1] M. D. Byrne: "Using Icons to Find Documents: Simplicity is Critical," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 446–453, 1993.

[2] S. J. Cunningham, D. Bainbridge and A. Falconer, " 'More of an art than a science': Supporting the Creation of Playlists and Mixes, " *Proceedings of the International Symposium on Music Information Retrieval*, 2006.

[3] J. Graham and J. J. Hull: "iCandy: A Tangible User Interface for iTunes," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ext. Abstracts,* pp. 2343-2348, 2008.

[4] S. Havre, B. Hetzler and L. Nowell: "ThemeRiver: Visualizing Theme Changes over Time," Proceedings *of the IEEE symposium on Information Visualization*, pp. 115-123, 2000.

[5] P. Kolhoff, J. Preuß and J. Loviscach: "Music Icons: Procedural Glyphs for Audio Files," *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing*, pp. 289-296, 2006.

[6] E. Pampalk: "Islands of Music: Analysis, Organization, and Visualization of Music Archives," *Journal of the Austrian Society for Artificial Intelligence*, 2003.

[7] E. Pampalk and M. Goto: "MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling," *Proceedings of the International Symposium on Music Information Retrieval*, 2006.

[8] V. Setlur, C. Albrecht-Buehler, A. A. Gooch, S. Rossoff and B. Gooch: "Semanticons: Visual Metaphors as File Icons," *Computer Graphics Forum*, Vol. 24, pp. 647-656, 2005.

[9] I. Stavness, J. Gluck, L. Vilhan and S. Fels: "The MUSICtable: A Map-based Ubiquitous System for Social Interaction with a Digital Music," *Proceedings of the International Conference on Entertainment Computing*, pp. 291-302, 2005.

[10] M. Torrens, P. Hertzog and J. L. Arcos: "Visualizing and Exploring Personal Music Libraries," *Proceedings of the International Symposium on Music Information Retrieval*, 2004.

[11] G. Tzanetakis, G. Essl and P. Cook: "Human Perception and Computer Extraction of Musical Beat Strength," *Proceedings of the International Conference on Digital Audio Effects*, pp. 257-261, 2002.

[12] G. Tzanetakis and P. Cook: "3D Graphics Tools for Sound Collections," *Proceedings of the International Conference on Digital Audio Effects*, 2000.

[13] R. van Gulik and F. Vignoli: "Visual Playlist Generation on the Artist Map," *Proceedings of the International Symposium on Music Information Retrieval*, 2005.

[14] M. Wattenberg: "Arc Diagrams: Visualizing structure in strings," *Proceedings of the IEEE symposium on Information Visualization*, pp. 110-116, 2002.

---

[1] http://www.istockphoto.com