Exploring the Benefits of 2D Visualizations for Drum Samples Retrieval

Chloé Turquois chloeturquois@live.fr
Martin Hermant martin.hermant@upf.edu
Daniel Gómez daniel.gomez@upf.edu
Sergi Jordà sergi.jorda@upf.edu
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138
08018 Barcelona, Spain

ABSTRACT
This paper explores the potential benefits of using similarity-based two-dimensional classifications and their corresponding GUIs, for drum samples retrieval in a creativity-oriented context. Preliminary user studies with professional electronic music producers point up the frustration and laboriousness of finding suitable drum samples in the increasingly large libraries of sounds available, and suggest the need for alternative interfaces and approaches. To address this issue, two novel spatial visualizations (respectively organized by name and by timbre-similarity) are designed as potential alternatives to the traditional 1D list-based browsers. These visualizations are implemented and compared in a music creation task, in terms of both the exploration experience and the resulting production quality, within a system for drum kit configuration. Our study shows that spatial visualizations do improve the overall exploration experience, and reveals the potential of similarity-based arrangements for the support of creative processes.

Keywords
Music interaction, creativity, spatial visualization, user studies.

1. INTRODUCTION
The creation of music is an essential creative human activity and it can be seen as the most paradigmatic of the creative arts. In contemporary music production, machines constitute an integral part of this creativity. However, most current tools for digital music production still lack of support for creative exploration and inspiration. Music production professionals would like to work faster; creative musicians and producers envision free-flowing and inspiring tools that could boost their creativity by minimizing technical difficulties and personal limitations. Nowadays, one of the most well identified bottlenecks to creativity in digital music production is the difficulty of finding suitable sound samples. Samples are short sound files such as a drum hit, often arranged rhythmically and put together to form a track. Producers dispose of increasingly large libraries of samples (often several thousands) from numerous databases. The current interfaces, displaying the samples in long scroll lists of sometimes poorly indicative file names, seem unadapted for browsing such large sound libraries. Interviews performed with 16 expert music producers [1] confirmed the tediousness of this type of interfaces and the need for more intuitive and inspiring modes of exploration. On the other hand, the current state of Music Information Retrieval (MIR) provides methods for automatically classifying sounds according to criteria, such as their pitch or timbre, which are potentially more meaningful than their file name.

We are interested in the potential of two-dimensional (2D) visualizations, and especially similarity-based 2D visualizations (using timbre classification tools), in replacing the traditional 1D list for the exploration of samples libraries. Therefore we examine the suitability of: a) using a 2D visualization for browsing drum samples, and b) using perceptual similarity algorithms for organizing this 2D visualization, so that two samples that sound similar are placed close to each other. In this paper, we present a comparative study using different visualizations of a drum samples library. The visualizations are used by participants in a creative task, where they have to choose a set of drum samples to compose a rhythmic track. We explore how these interfaces affect the overall exploration, user satisfaction and outcome quality. The rest of this document is structured as follows: Section 2 reviews some related work in the psychology and sound computing research communities, Section 3 presents the alternative spatial visualization we designed, which is compared within a study described in Section 4 and further discussed in Section 5, before concluding in Section 6.

2. BACKGROUND

2.1. Creativity and Flow
There is currently a growing interest in HCI research for designing tools that promote creativity and not only productivity [12]. Creativity, according to its more agreed definition, is the generation of something that is both original and valuable [13]. In psychology, the most recognized theories depict creativity as a combination of divergent, free-flowing processes, where the subject generates many solutions in a rather unconstrained manner, and of convergent, analytical processes, where she interprets, evaluates and refines these solutions[4][7]. To promote creativity, a system should therefore support both types of creative processes: unrestricted, divergent processes on the one hand, and convergent, critical processes on the other hand. On another hand, flow, an “almost autotelic, effortless, yet highly focused state of consciousness” [3], has been positively correlated with higher performance in creative activities [10]. It is thus important, when supporting creativity, to try and preserve the flow that the user might be experiencing. Indeed, flow is a very fragile state that is easily disrupted by tedious or unrelated tasks.

CHIIR '16, March 13 - 17, 2016, Carrboro, NC, USA

ACM 978-1-4503-3751-9/16/03...$15.00
DOI: http://dx.doi.org/10.1145/2854946.2854996

2.2. Timbre, MIR and Timbre Spaces

Timbre is usually described as the "color" of a sound, or all the qualities of a sound that are not its pitch, level, duration and location [8]. It is a multidimensional property that is related to the changes of amplitude and phase of the spectral components of the sound. Music Information Retrieval allows to extract many attributes from a sound to describe its timbre. These descriptors can be used to classify sounds by similarity. Although we cannot yet guarantee which dimensions are the most relevant for perceptual similarity in a given group of sounds, mapping perceptual similarity to spatial proximity, so that timbres that are similar are placed close to each other, is called a timbre-space [6].

To help handling the increasingly large amounts of multimedia content, researchers have developed several tools for browsing audio materials by similarity, using spatial representation and content-based information retrieval. The Snare-Drum Navigator [5] is a 2D timbre-space interface of snare drum sounds represented as colored rectangles. The efficiency of this interface was evaluated in a study where users had to find a given sample as fast as possible. They showed that users' performance improved faster when the samples were sorted by similarity than randomly. However, until now, user studies on such tools have focused mostly on efficiency, only involving non-creative tasks.

3. DRUMSPACE

We designed Drumspace as an alternative to the aforementioned scroll lists for browsing drum samples. Using 2D representation and Music Information Retrieval, Drumspace aims to improve the exploration and retrieval of samples, and in particular to better support the creative workflow in these activities.

3.1. Objectives

To support the creative workflow in the search of sounds, the interface should support the two "directions" of creative processes (convergent and divergent). Considering the divergent processes, the interface should allow for an unrestricted exploration and facilitate direct access to a diversity of materials. On the other hand, and with consideration to the convergent/critical side, we need to facilitate the evaluation of samples in context (fitness with the other drum sounds and the overall composition) and the reduction of the solution space. More generally, despite not being a musical instrument in itself, the system should be suited for a creative utilization. That is, the musicians should be able to use it for coming up with novel and relevant sounds or ideas. To support the state of flow, the system should require as little cognitive effort as possible, be easy to learn and to use. The number of features and commands should be limited to what is strictly necessary in order not to disrupt the flow or obstruct the user from her goal and her ongoing creation.

3.2. 2D representation of samples

The first advantage of a 2D representation over the list is to offer an instant overview of all the available samples. Unlike the linear exploration afforded by 1D lists, the exploration can be guided by other various strategies. Finally, scattering the samples on 2D takes advantage of the human ability to deal with spatial locations and relationships [11].

Current list browsers often display sample names ordered alphabetically. The correspondence between name and sound is mostly arbitrary and the large number of samples does surely not allow memorizing all these correspondences. Nor does it allow inferring where a particular sample could be located. Timbre, on the other side, is probably the most relevant quality for describing a drum sound. If sound samples are arranged by timbre on a 2D plane, their proximity to each other will suggest their similarity, while their position on the plane will directly inform on some perceptual timbral qualities. This can guide the exploration such that if a user listens to a sample that happens to be very different from what she needs, she should go and explore another zone, and inversely, refining her choice progressively towards the sounds she likes. With time, she can memorize the overall spatial arrangement and develop expectations as to where she is most likely to find a certain type of sound.

4. COMPARATIVE STUDY

Can spatial visualization improve the exploration of sound libraries and support better creativity in music production? The goal of this comparative study is to explore the suitability and the convenience of a) using 2D interfaces for exploring sounds in a database, and more particularly of b) using timbre similarity algorithms for automatically clustering these sounds in 2D. The hypothesis is that these solutions will ease the process of finding suitable sounds, thus improving the user's experience and creative outcome.

4.1. Experiment prototype

To test this hypothesis, we integrated Drumspace into a browser application for the creation of drum kits. Users can browse a library of 1226 drum sounds and select 4 of them to be used together on a given rhythmic track. As shown in Figure 1, the interface is split into 4 panels, one for each sample family that we have selected as voices for the rhythmic tracks, namely: Kick, Snare, Open Hi-Hat and Closed Hi-Hat. Each panel contains all the samples of a family, which can be displayed in three different views, as later explained in *Visualizations*. The central red button allows playing and stopping the drum track, which consist of a predefined rhythm to which the selected samples are applied in real-time.

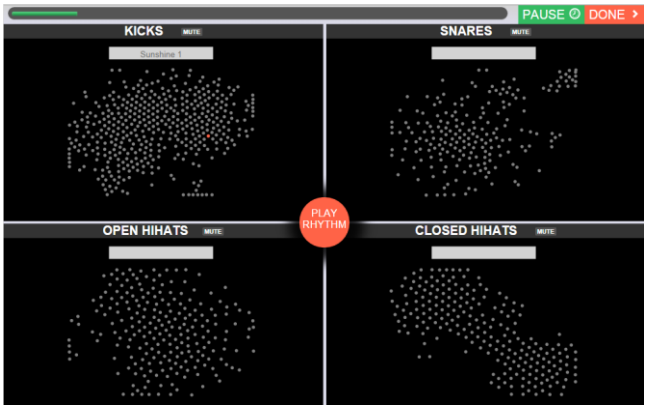


Figure 1. Interface for drum kit creation created for this study.

4.1.1. Interaction

Users interact with the application using a normal desktop mouse and a keyboard. Each sample is represented by a circle. The user can listen to any sample by just pointing the mouse over it. The user selects a sample (e.g. a *kick*) by clicking on its circle; the circle is highlighted in red and the sample becomes the current *kick* voice within the rhythm.

4.1.2. Visualizations

For the purpose of our comparative study we implemented three visualizations of the library, which differ in dimensionality and ordering: 1D list; alphabetical 2D; and similarity based 2D. As shown in Fig. 2a, the list view resembles as closely as possible the library browsers used in most DAWS: 1 dimension, alphabetical

order, with the file names visible. The second view (Fig. 2b) differs from the list by its dimensionality (2 dimensions), but maintains the alphabetical order. Samples are arranged on lines based on the first letter of their names. Finally, the third view (Fig. 2c) differs from the list in both its dimensionality and in the spatial arrangements of samples: samples are scattered on the 2D plan according to their timbre similarity. This timbre similarity space was built using the open-source C++ library Essentia for audio analysis and audio-based MIR [2], based on the list of audio descriptors used in Timbre Toolbox [9], and a Student-t Stochastic Neighbor Embedding (tSNE) 2D dimension reduction [14].

4.2. Methods

The experiment was composed of two parts: Drum kit creation, in which participants used alternatively the three different views of the samples to compose drum sets for different given rhythmic patterns, and a Drum kit evaluation, in which listeners were asked to evaluate the quality of the loops created in the first phase.

4.2.1. Part I: Drum kit creation

20 participants took part in this first phase (18 musicians and 2 nonmusicians). The experiment was composed of 9 consecutive trials with a maximum duration of 3 minutes each, with each of the 3 views being used on 3 respective trials. In each trial, participants were given a rhythmic pattern with 4 very basic "default" sounds (kick, snare, open and closed hi-hat), playing on loop. Participants were asked to browse the drum samples and choosing one sample for each family (1 kick, 1 snare, 1 open hi-hat and 1 closed hi-hat) in order to replace the initial default sounds within the given rhythm. At the end of each trial, participants were asked to rate how satisfied they were with 1) the given rhythmic pattern, and 2) the drumkit they had created (i.e. the sounds they chose). The rating was made on a discrete 3-point Likert scale from 0 ("not at all") to 2 ("very satisfied"). Upon completing the 9 different patterns of the experiment, participants were also invited to comment briefly on the three different views in an open-interview format. They commented freely on their overall experience, and were finally asked which view they preferred to use and why.

4.2.2. Part II: Drum kit evaluation

In this test, the drum loops created by the participants in the first part were listened and evaluated by 24 listeners, both musicians (10 of them experienced in digital music production) and non-musicians. The evaluation was done remotely using an online application. Each evaluator was assigned a set of 21 rhythmic loops to evaluate (out of the 180 -- 9 patterns x 20 participants -- created in the first part). Each set comprised 7 different rhythmic patterns, and 3 different drumkit sound combinations per pattern, created by different creators using different views. The sets were presented in a randomized order, and evaluators were asked to base their rating on any criteria that they found relevant, such as overall quality of the loop, fitness of the chosen samples to the rhythm, or originality.

4.3. Results

4.3.1. Part I

First, the number of listened samples per trial was significantly lower with the list view (M=46.9) than with either of the two 2D views (M=93.2 with the alphabetical 2D view, and M=101.2 with the similarity-based 2D view), $\chi2(2) = 30.9$, $p<0.001$ at the $p<0.017$ significance level.

For studying the alphabetic repartition of samples chosen, a score was assigned to each selected sample according to the alphabetic index of its name's first letter (e.g. "Beastie" gets 2 points, "Zyklotron" gets 26 points). There was a significant effect of the view on the alphabetic position of the chosen samples, $F(2,38)=14.6$, $p<0.001$. Scheffe post hoc analyses indicated that the List view score (M=29.5, SD=22.4) was significantly lower than the Alphabetical 2D view (M=43.7, SD=12.6) or the 2D-timbre view (M=47.4, SD=13.9), which did not differ from each other. Regarding *Result satisfaction*, a Friedman test of differences on repeated measures showed a close to significant effect, with participants being more satisfied with their creation when using the Alphabetical 2D view: $\chi2(2)=5.9$, $p=0.052$.

During the post-hoc interviews, 8 participants out of 20 declared preferring the 2D-timbre view, 8 preferred the 2D-alphabetical view, 3 preferred the list view, and 1 expressed no preference. Considering the dimensionality exclusively, chi-square goodness-of-fit showed a significant preference for 2D views over 1D List, $\chi2 (2, N=20) = 19.9$, $p < 0.005$.

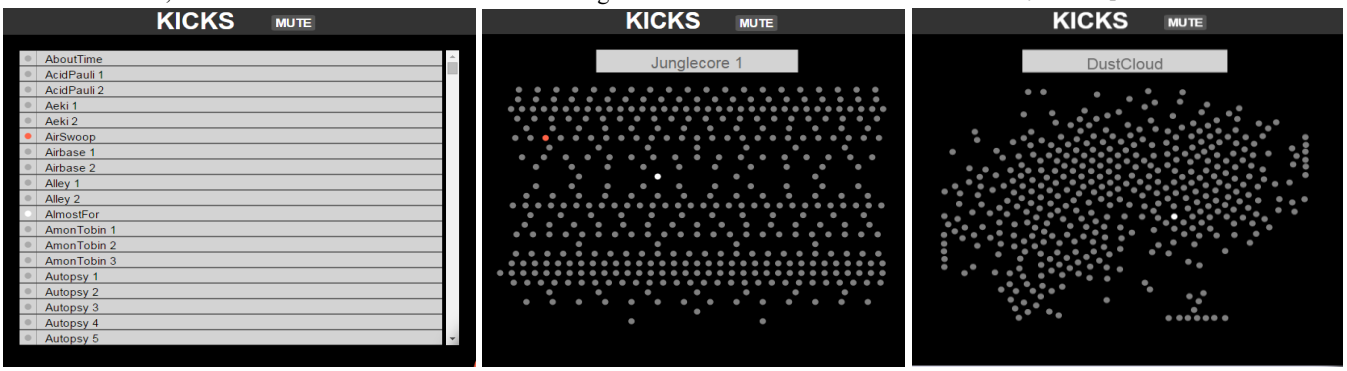
4.3.2. Part II

We did not measure any significant difference between the scores of the drum kits created by the three views. To isolate the appreciation of the chosen sounds from that of the rhythm itself, we subtracted to each score the score given by the same evaluator to the same rhythm rendered with default sounds. Again the view used for creating the drum kits did not have any significant effect on their assessed quality, $X(2)=2.155$, $p=0.34$. Although we randomized the repartition and order of presentation of the kits, a much larger number of evaluators would probably be required to counterbalance the subjectivity of the appreciation of drum loops.

5. DISCUSSION

5.1. Improving the exploration with 2D

The results from the first part of our study support our first hypothesis that a 2D visualization provides a better experience than a traditional list interface and seems to improve the exploration. 2D visualizations increase the diversity of samples chosen. This can be



(a) 1D scroll-list visualization **(b) Alphabetical 2D visualization** **(c) 2D visualization based on timbre similarity**

Figure 2. Three visualizations of the Kick library

considered a direct benefit of the global "bird view" on the whole library: the user gets instant access to all the available samples. The follow-up interviews revealed that this encouraged a non-linear exploration rather than following the vertical distribution of the list. We also found that the 2D views let users listen to more samples over the same time span. The short distance between the samples facilitates their playback, providing the direct feedback necessary in the support of flow states, and seemingly improving the user experience (participants said it was "easier", "faster", "more pleasant", and "more fun"). This is further attested by the large preference reported by users for the 2D views over the list, and by the satisfaction ratings showing that participants tended to be more satisfied with their creation when using the 2D alphabetical view. Participants' feedback also highlighted the "experimental" aspect of the exploration on 2D interfaces. Many of them reported having discovered unexpected and inspiring sounds thanks to the non-linear and rapid browsing. For these reasons, we suggest that 2D visualizations, as they facilitate and broaden the exploration, provide an improvement in the support of flow states and, by favoring inspiration, of creativity. This however is a purely theoretical interpretation and calls to future experimentations that focus on these phenomena.

5.2. Issues with the Timbre Space

Our results do not show any significant improvement brought by the arrangement by perceptual similarity. In the light of participants' feedback, it appears that the similarity-based arrangement was unexpectedly difficult to understand and to use efficiently. This visualization represents all samples identically and displays no axis or any indicator of the timbre similarity. It lacks transparency and does not allow the user to understand the organization of the sample space. Experimenting with more explicit representations would be necessary in order to design a more intuitive timbre space interface. For instance, the use of colors to indicate some timbral features might improve the understanding of the similarity clustering.

5.3. Supporting Flow and Creativity

Participants who preferred the 2D-timbre view appreciated the freedom and "implicit" guidance provided by the similarity-based organization. They described their navigation as progressing from random exploration to "fine-tuning". This closely resembles the processes described in the theories on creativity (the divergent, unrestricted navigation, followed by examining and fine-tuning processes), suggesting that the 2D-timbre view is a good candidate for supporting the musicians' creative workflow.

Our results suggest that a 2D visualization of samples facilitates the exploration and improves the user experience. This suggests that spatial visualizations are beneficial for supporting the states of flow, where fluency, direct feedback and enjoyment are important factors. Therefore it would be highly relevant to design a more formal evaluation of how much this type of visualization sustains flow.

6. CONCLUSION

As an alternative to traditional list browsers we proposed a 2-dimensional visualization, where algorithms for audio-based music information retrieval allow displaying the samples in a timbre space according to their perceptual properties. Although some studies have already examined the potential of such visualizations for audio materials retrieval, they have rather been focused on efficiency and accuracy. Instead we considered the importance of the qualitative aspects in music creation: quality of the experience and quality of the outcomes. Our comparative study examined the qualitative benefits of using perceptual similarity and 2D visualizations for the exploration of samples libraries in a creative task. As predicted, users reported a strong preference for the 2D visualizations. Although we could not measure any effect on the quality of the drum loops created, 2D visualizations by facilitating the navigation and offering a broader overview, seem to improve the exploitation of the available sounds, and to promote discovery and inspiration. Contrary to our hypothesis, the arrangement by perceptual similarity did not appear more advantageous than an arbitrary arrangement by sample name. However, users' feedback on the experience suggests that perceptual similarity arrangement promotes a different, more musical exploration. It opens many perspectives for the design and the investigation of novel spatial visualizations and their potential for supporting flow and creative processes in music creation.

7. ACKNOWLEDGEMENTS

This research has been partially supported by the EU funded GiantSteps project (FP7-ICT-2013-10 Grant agreement 610591).

8. REFERENCES

[1] Andersen, K. et al. 2015. "Update on Qualitative and Quantitative Methodological Frameworks". Deliverable 2.3, GiantSteps FP7-610591. Retrieved from: http://assets.contentful.com/xrzr1u3na612/5lpsdiQKuQoOKEWwgmiQSK/04b084584af567aa04b696ee00365db7/GiantSteps_D2.3_WP2_2015_30June_STEIM_v1.0.pdf.

[2] Bogdanov, D. et al. 2013. Essentia: an open-source library for sound and music analysis. *Proceedings of the 21st ACM international conference on Multimedia* (2013, October), 855-858.

[3] Csikszentmihalyi, M. 1997. Flow and the Psychology of Discovery and Invention. *HarperPerennial,New York*, *39*.

[4] Finke, R. et al. 1992. *Creative cognition*. MIT Press.

[5] Fried, O. et al. 2014. AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting.

[6] Grey, J. 1977. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*. 61, 5 (1977), 1270.

[7] Guilford, J. 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3-14.

[8] Krumhansl, Carol L. 1989. Why is musical timbre so hard to understand. *Structure and perception of electroacoustic sound and music.* 9: 43-53.

[9] Peeters, G. et al. 2011. The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902-2916.

[10] Perry, S. 1999. *Writing in flow: Keys to enhanced creativity*. Writers Digest Books.

[11] Robertson, G. et al. 1998. Data Mountain: using spatial memory for document management. *Proceedings of the 11th annual ACM symposium on User interface software and technology* (1998), 153-162.

[12] Shneiderman, B. et al. 2006. Creativity support tools: Report from a US National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction*, *20*(2), 61-77.

[13] Stein, M. 1953. Creativity and culture. *The journal of psychology*, *36*(2), 311-322.

[14] Van der Maaten, L., & Hinton, G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research 9,* (2008), 2579-2605.