giant_steps

GiantSteps

Seven League Boots for Music Creation and Performance

FP7-610591

D2.5

Third Annual Report on User Involvement

Dissemination level:	Public
Contractual date of delivery:	M36, 31-10-2016
Actual date of delivery:	29-11-2016 (with update on 15-03-2017)
Workpackage:	WP2 Iterative user-centred Design & Evaluation
Task:	T2.2 Generating and testing requirements and prototypes with users T2.3 Planning and conducting workshops T2.4 Analysing and evaluating user data
Type:	Report
Approval Status:	Released
Version:	2.0
Number of pages:	54
Filename:	GiantSteps_D2.5_WP2_2017_15March_NI_v2.0.docx

Abstract

This document reports on the user sessions during the third year of the GiantSteps project. It is the continuation of D2.2 “First Annual Report of User Involvement” and D2.4 “Second Annual Report of User Involvement”, and follows the principles and standards outlined in D2.1 and D2.3 (the methodological framework for the user involvement in the GiantSteps project).

co-funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement n° 610591

History

Version	Date	Reason	Revised by
0.1	10.11.2016	Initial draft	Kristina Andersen
0.2	22.11.2016	Second draft	Kristina Andersen
0.3	28.11.2016	Review document	Kristina Andersen
0.4	29.11.2016	Final document ready for submission	Kristina Andersen
0.5	29.11.2016	Revision and formatting	Roman Kaurson
1.0	29.11.2016	Released version	Roman Kaurson
1.1	24.02.2017	Confidential information removal	Steinunn Arnardottir
1.2	14.03.2017	Confidential information removal	Kristina Andersen
1.3	15.03.2017	Revision and formatting	Roman Kaurson
2.0	15.03.2017	Public version release	Roman Kaurson

Author list

Organization	Name	Contact Information
STEIM	Kristina Andersen	kristina@steim.nl
NI	Steinunn Arnardottir	steinunn.arnardottir@native-instruments.de
NI	Mickael Le Goff	mickael.legoff@native-instruments.de
JPC	Roman Kaurson	roman.kaurson@jcp-connect.com
JKU	Peter Knees	peter.knees@jku.at
JKU/DMU	Tom Collins	tomthecollins@gmail.com
JKU	Richard Vogl	richard.vogl@jku.at
RS	Andrés Bucci	andres@reactable.com
MTG	Ángel Faraldo	angel.faraldo@upf.edu
MTG	Cárthach Ó Nuanáin	carthach.onuanain@upf.edu
MTG	Daniel Gómez	daniel.gomez@upf.edu

Executive Summary

This document gives an overview of the user sessions during the third and final year of the GiantSteps project. It is the continuation of D2.2 “First Annual Report of User Involvement” and D2.4 “Second Annual Report of User Involvement”.

The goal of these user sessions is to maintain a strong and consistent link between the project’s technical and academic research process on the one hand, and music practitioners on the other hand. This is achieved by exposing the scientific and technological developments in real-life scenarios. The user involvement described in this document follows the principles and standards outlined in D2.1 and D2.3 (the methodological framework for the user involvement in the GiantSteps project).

In the first two years the user involvement in the project is focused on notions of qualitative user testing and participatory design. In the third year, as the project draws to a close, the user involvement has increasingly focussed on local testing of prototypes and products along with reflection in the form of academic publications and talks. However as certain prototypes become products (like ROTOR) the user-facing events tend to cross over from feature-testing to gauging satisfaction and interest, and the difference between user-involvement and dissemination events blur. As a result, many events carry elements of both and are mentioned both in this document and D7.7.

Having identified the in-depth user interviews as a valuable resource in itself and the underlying methodology as a possible exploitation area (identified at the Exploitation Booster session in Berlin), we have continued these interviews into the third year.

Table of Contents

1 Introduction

This document provides an overview of the user sessions conducted in Work Package 2 (WP2) during the third year of the GiantSteps project. As with the previous two, it is meant to be read as a report on the various ways in which the consortium engaged with users during the final year of the project.

In WP2, the stated goal is to maintain strong and consistent links between the project's technical and academic research process on the one hand, and music practitioners on the other hand. This is achieved by exposing the scientific and technological developments into real-life scenarios. As the project has progressed the form of this work has shifted:

During the first year of the project, the user sessions provided a baseline understanding of existing work practices, current interface frustrations, desired developments, and future scenarios of use. We began our practise of exposing the project's developments to expert users at the annual Red Bull Music Academy, which marks the end of a year-long cycle and a milestone of the project.

During the second year of the project, we developed a set of user personas, a system for managing prototypes, generated keyword requirements and methods for exploring ideas in functional and non-functional prototypes. We shifted towards executing smaller, focused user studies locally to facilitate quicker feedback on the functionality of prototypes and ideas.

In the third and final year we have continued the local testing, alongside more complex events such as hack-days and festivals, where user exposure can be seen as both dissemination and a kind of meta-feedback. Throughout the three years we have returned to the RBMA for more in-depth engagement with users.

In effect, this is a work package that spans all partners and work packages and as a result the work conducted and outcomes gathered must serve many different workflows and individual partner needs. At the end of the three years, some prototypes have shipped to the market in products, while others remain matters of ongoing investigation and development. This is reflected in the user involvement reported here, some tests are focussing on evaluating specific features, while others are concerned with entire products (such as ROTOR).

In a somewhat similar manner, the meta concerns of our users have shifted too, where the initial interviews in Tokyo revealed much data and search concerns, these have abated somewhat in favour of a focus on collaboration and communication inside the creative process. This is driven in part by the subtle shift towards physical music machines, we described in [Grote et al, 2015] and a shifting cultural environment, where increasingly music is created in collaborations, and these may occur online as tentatively described in [Grote, 2014].

This report will detail the work conducted in WP2 this year and outline some of the outcomes and results.

2 Overview of User Involvement in Year Three

In year three the user involvement centered on three types of sessions:

1. Smaller local tests of prototypes, conducted by local partners and aimed towards testing functionality and efficiency of ideas and execution. This last category also includes online tests, managed and executed by local partners.
2. Dissemination and participatory design sessions on location at various events.
3. In-depth interviews and user feedback from RBMA 2016 in Montreal.

These events are related to the following milestones:

- MS6 (M28): Expert user workshop. Event in Berlin with RBMA and NI alumni.
- MS7 (M32): Expert user workshop evaluating prototypes and interfaces. Local and web-based testing.
- MS8 (M36): Red Bull Music Academy workshop. Conducted in Montreal.

In the following sections, we give an overview over the various user sessions throughout year three of the GiantSteps project (number of users involved given in parentheses):

Berlin RBMA event:

- Feb. 2016. Berlin: Beat Straightening/Beat Regularization Prototype (11)
- Feb. 2016. Berlin: RhythmCAT (16)
- Feb. 2016. Berlin: House-Harmonic-Filler and Dr. Drums (13)
- Feb. 2016. Berlin: EDM Note Suggestion Web Prototype (12)
- Feb. 2016. Berlin: In-depth Interviews (8)

Local and Web-based Testing:

- Feb-May. 2016. NI: Confidential Research (75)
- Jun.-Oct. 2016. JKU: Rhythm Pattern Variation App (10)
- Feb. 2016. MTG: RhythmCAT (5)
- Feb. 2016. MTG: House-Harmonic-Filler and Dr. Drums (7)
- Mar. 2016. Online: RhythmCAT Listening Survey (21)
- Apr. 2016. Online: EDM Note Suggestion Output Evaluation (4)
- Mar. 2016. Online: Mad competition for discriminating Rhythm Spaces (20)
- Ongoing. 2016. Online: ROTOR application beta testing (34)

Demos, hackathons and workshops:

- May. 2016. Festival: Music Tech Fest (~12)
- June. 2016. Festival: Sonar (>100)
- Oct. 2016. Conference: Workshop at ADE 2016 (8)
- Oct. 2016. Festival: Waves Festival Vienna (4)

RBMA Montreal:

- Oct. 2016. RBMA: In-depth Interviews (31)
- Oct. 2016. RBMA: Prototypes (20)

Eight different personas were created in year 2 and presented in D2.4, Second Annual Report on User Involvement. They have been used to identify the target groups for each test session and are listed in a table at beginning of each test section. The personas within the GiantSteps project all come from our central user group: the expert users and were created with close collaboration between STEIM, JKU and NI, which itself uses a persona-guided design process. The following personas were identified:

- Headliner DJ
- Apprentice Producer, Semi-Pro DJ
- Laptop/Home Studio DJ
- Tablet/Multi-device Producer
- Professional Producer
- Serious Modular Controller Hobbyist
- Music Technology Hacker
- Ground Breaker

The sheets with further description for each persona can be found in D2.4, section 3.

2.1 RBMA alumni event in Berlin

To make up for the cancelled RBMA event in Paris we collaborated with Yadastar/RBMA to conduct in-depth user testing on location at NI in Berlin. The invited participants were all professional music producers and/or DJs recruited from the pool of former RBMA participants and NI in-house musicians and producers.

A large room was setup into five separate “stations”, each home to a prototype or activity, which allowed us to test in parallel during the day. In all five, prototypes were tested and a set of in-depth interviews were conducted.

This event provided a large amount of data and feedback to the participating researchers. The individual setups and outcomes are outlined in the following.

Fig. 1: Testing "station".

2.1.1 Beat Straightening/Beat Regularization Prototype

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Headliner DJ (offline/prep) Apprentice Producer Semi-Pro DJ (offline/prep) Laptop/Home Studio DJ Professional Producer Music Technology Hacker Ground Breaker	Berlin	11	D3.8

The Beat Straightening/Beat Regularization Prototype is a prototype for automatic beat straightening developed for WP3. The aim of the prototype is to provide means to automatically detect the beats in a loop or song and align the beats to a fixed and regular tempo grid using pitch invariant time stretching. DJs and musicians working with samples require such a tool when working with samples or songs that are recorded without using a click track or metronome which should be remixed with other material. Ableton offers rudimentary automation of this task, but works unreliable and so this task is, in most cases, very work intensive.

To test this, early prototype users at the test session were introduced to the tool and asked to use it on tracks for which they have needed such a tool in the past and tracks that come to mind for which such a tool would be useable. Results, suggestions, and further use-cases were discussed after that.

In total 11 users participated in the feedback session for this prototype.

Outcomes

The prototype was received positively by the majority of the participants. Besides some feedback regarding the user interface (which was very rudimentary at that point) suggestions for further applications of the technology were provided.

"I can imagine that being a real game changer for a lot of people."
- JKU-NI-06

"You know what would be great? Export this tempo changes automation, so you could import it back [into the DAW as tempo track]."
- JKU-NI-19

More details about the prototype and the result discussion can be found in deliverable D3.8.

2.1.2 RhythmCAT

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Headliner DJ Apprentice Producer Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Serious Modular Controller Hobbyist Music Technology Hacker Ground Breaker	Berlin	16	Ó Nuanáin et al. 2016 Ó Nuanáin et al. 2016a Ó Nuanáin et al. 2016b

Formal user evaluation for *RhythmCAT* comprised a quantitative evaluation of the listener's perception of the similarity metrics and concatenative algorithm, in addition to a qualitative evaluation that sought to gather rich thematic analysis of the user's experience in using the software. Our findings were substantial enough to warrant discussion in two separate publications, at ISMIR 2016 [Ó Nuanáin et al., 2016a] and MUME 2016 [Ó Nuanáin et al., 2016b]. Here we summarise the user evaluation and interviews that took place at RBMA at Native Instruments HQ in Berlin.

With each participant we explained briefly the instrument and guided them through the process of generating sounds with the instrument. Mostly, the participants were eager to start playing with the instrument as soon as possible, which we permitted. As all other test stations at the venue, the RhythmCAT station was set up with a laptop, monitor and headphones.

While the interviews themselves were kept informal, we at least tried to steer the individual sessions with some common questions or themes in order to elicit conversation. These included questions such as:

- Did the overall concept make sense to you?
- Was the interface intuitive? What elements were confusing?
- Would you use this system to make music?
- Would you use this system in production scenarios, live performance or both?
- What did you like, what didn't you like?
- What improvements would you make?
- What features would you like to see?

Fig.2: Word cloud.

Positive Reactions

Figure 2 above shows a wordcloud depicting the most frequent positive descriptions participants attached to the tool during the course of the interviews.

Some specific remarks that people had in general:

"It's an excellent tool for making small changes in real time. The interface for me is excellent. This two dimensional arrangement of the different sounds and its situation by familiarity, it's also really good for making these changes."

"I'm really interested in more visual, more graphic interface. Also the fact that you can come up with new patterns just by the push of a button is always great."

"It's inspiring because this mix makes something interesting still, but also I have the feeling I can steal it."

"The unbelievable thing is that it can create something which is so accurate. I wouldn't believe that it's capable of doing such a thing."

Usage Scenarios

The participants proposed some interesting scenarios where they could see themselves using the tool creatively. In particular there was a marked interest in live recording and analysis for working with instruments and beatboxing:

"This is great! Ah, but wait. Does it mean I could like beat box really badly some idea that I have... and then bring my samples, my favourite kits and then it will just work? "

Other users were not really interested in the targeted capabilities of the tool and preferred the exploratory and experimental activity of creating new sounds from scratch without any guide.

"I've got this fully on wet straight away, which tells you the direction I'd be going with it."

" ...you just want to drag in 100 different songs and you just want to explore without having this connection to the original group. Just want to explore and create sound with it."

Shaping the Sounds

Other than generating the sequences and rearranging individual units in the sequence, the synthesiser offers no additional ways to modify the output sound (discounting the ability to mix between the target sequence and the generated sequence). Many users agreed it would be useful to be able to manipulate these individual sounds sonically somehow. Most crucially they desired the option to be able to control the envelopes of the individual units via drawable attack and decay parameters, which is currently being implemented.

" ... an attack and decay just to sort of tighten it up a little bit. Get rid of some of the rough edges of the onsets and offsets "

"Yeah, the thing is if you listen to it now, there's kind of a rhythm going, but it would be great if you could increase the decay of the snare for example. Which if it's a prototype, you can't expect to have all those functions there immediately, but in an end product, I think it would be a necessity."

Parameterisation and Visualisation

One of the recurring difficulties that faced participants was our presentation of the parameters. As we explained briefly in the implementation section, the user is able to control the influence of the four features in concatenation algorithm and the PCA visualisation. We relabelled the features from their objective names to what we considered a subjective equivalent that a lay user may understand. MFCCs are labelled as "Timbre", spectral centroid as "Brightness", spectral flatness as "harmonicity" with loudness unchanged.

Unfortunately, users expressed confusion at their purpose and were unable to interpret their effect on neither the arrangement of the units of sound in space or their resulting effect on the patterns generated. For instance:

"The problem is I'm a little bit lost already."

"you have four parameters, and you don't know which thing is this what"

"I would prefer not to have too much controls."

Presenting this additional complexity was a naive inclusion on our part. Clearly the typical user is content with the overall output from the system and would rather not delve into these specifics. However, at least in terms of the visualisation there is a ”sweet spot” for feature weightings in the arrangement of the units of sound in the timbre space and this is why the controls were made available. The weightings can vary greatly depending on the corpus, though in our experience MFCCs alone often provide the best separation and clustering.

The challenge will be to find the best approach to arranging the units in space with the best separation and shielding these parameters from the user. Potentially, a way forward could be to remove the sliders and replace them with a number of options including an “advanced” mode with the ability to select specific parameters for the axes like CatArT in addition to “automatic” arrangement presets made possible using dimensionality reduction techniques. An area for study would be to gather many different sound sets and try various combinations of feature selections and weightings to find the best visual separation. At present PCA is used for dimension reduction but there are other algorithms that can be integrated [Frisson 2015]. We are in the process of integrating the t-SNE algorithm which has exhibited good performance in many musical applications [Turquois et al. 2016, Flexer 2015, Frisson et al. 2014].

2.1.3 House-Harmonic-Filler and Dr. Drums

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Apprentice Producer Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Music Technology Hacker Ground Breaker	Berlin	13	D5.3 and D5.5

For Berlin, we prepared a joint demo/interview of the House Harmonic Filler and Dr. Drums, two prototypes for harmonic exploration of house music, drum patterns of various styles of Electronic Dance Music, respectively (see deliverables 5.3 and 5.5 for a description of these two prototypes). The setup consisted in a session in Ableton Live, to where the two prototypes sent midi data in real time. This way, we wanted to emphasise the symbolic domain of our prototypes, and the users were free to select among their preferred instrumental sounds and drumkits in the Ableton session.

For ease of use, an iPad interfacing with two prototypes was offered, but this did not prove very helpful, as every participant preferred to interact directly with the Graphical User Interface on the computer screen.

The experiments proceeded in the following pace: first, the participant would receive a 5 minutes introduction about the general concept of the prototypes and the basic functioning of the system, after which they were left to experiment with the instrument themselves for

about 10 minutes. Another 15 to 20 minutes were devoted to an interview about the prototypes, their features and visual layout.

As these prototypes had less refined user interfaces, within the 13 total interviews, we can clearly observe different tendencies between users being concerned and critical with issues related to the interface design and architecture while the others are more concerned with aspects of artistic authorship.

Regardless of this difference, users were generally satisfied with the GUIs shown. Three interviewees showed some concern about the degree of detailed control one could have with the parameters on the interface, but agreed that in general, an interface with endless options is poorly operational. One solution proposed by several users was to hide some of the finer controls and only let them appear in a sort of *advanced* mode.

An interesting observation is that all interviewees seemed exclusively biased toward one or another musical aspect, represented by the two prototypes (rhythm vs. pitch/harmony): users interested in rhythm were notably less interested in harmonic variations and vice versa.

Regarding *Dr. Drums* in particular, most users found the *commonness* and *density* dials interesting as a way to distort a given musical pattern. On the contrary, some of them were dissatisfied with the apparent randomness of the generative algorithm and one user considered the degree of randomness completely unacceptable, for this would impede him to use such a system in a performance or even to compose music. Regarding the generative algorithms used, two modes were hiddenly presented to the users. *Mode A* was a 2nd order markov chain, whereas *Mode B* was a 3rd order one. Users could not show a preference for one or the other in the time given for experimentation.

With the House Harmonic Filler, users were generally satisfied, showing interest regarding the two-dimensional map of chord sequences. Some users coincided in seeing the prototypes as a sketching tool, allowing composers to sketch out simple harmonic ideas before entering the studio and make everything on their own.

2.1.4 EDM Note Suggestion Web Prototype

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Apprentice Producer Laptop/Home Studio DJ Professional Producer Music Technology Hacker Ground Breaker	Berlin	12	D4.4, [Collins et al. 2016]

The EDM note suggestion API described in D4.4 was tested in a Web-based prototype during the Berlin RBMA alumni session.

In total, 12 expert users used the interface shown in Figure 3 to compose two four-measure EDM-style loops accompanying a given drum track. One loop was composed in the presence of a “suggest” button and the other loop was not. The aim of the study was to investigate composers' opinions and edit behaviour under suggestion-enabled versus suggestion-free (or baseline) conditions, and to shed light on a particular instance of using MIR to shape musical creativity.

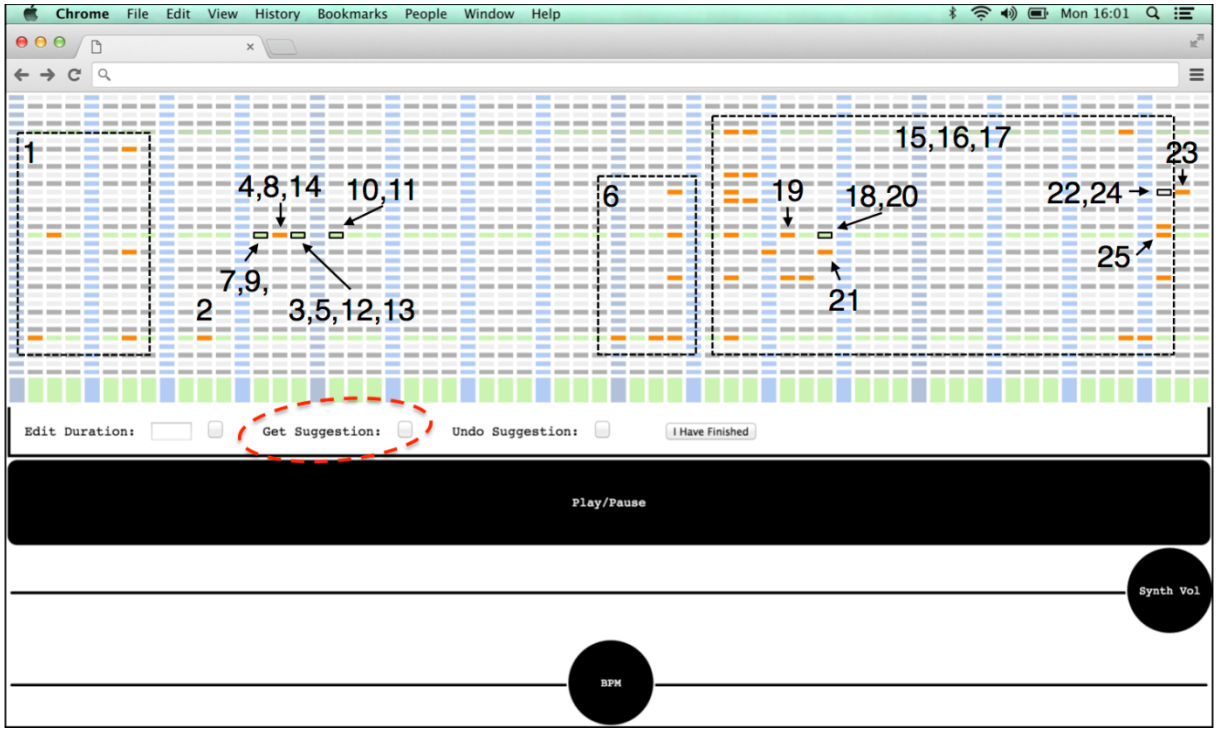


Fig. 3: Browser-based, piano-roll interface, with orange oblongs of certain x- and y- values corresponding to notes with certain start times and pitches. Numbers 1-25 indicate the time-lapse data for this composition session by participant 2, with dashed black lines bounding suggestions and black-edged oblongs indicating added-then-removed notes.

Participants

We spoke to 12 international, professional music producers. They reported mean 12.1 years of experience with music production software (sd = 5.8 years) and mode “Daily” use of such applications.

Stimuli

The composition of one loop was accompanied by a drum track excerpt from “The age of love” by Scooter, and the other by a drum track excerpt from “We fly...tonight” by De/Vision. The tracks were both in common time and well matched in terms of rhythmic characteristics.

Apparatus

A JS package called NexusUI was used to build the clickable piano-roll interface depicted in Fig. 3. Pitch was depicted on the y-axis using variegated rows (light gray for white notes on the piano keyboard and darker gray for black notes, green for pitch-class C and darker green for "middle C"). Measures and main beats were depicted on the x-axis using variegated columns (dark blue for beat one of a measure and lighter blue for subsequent beats). The drum track was heard but not seen in the interface, and progress through the loop was indicated by orange coloring of the taller row at the bottom of the piano roll. A second JS package called Tone.js was used to link the clicking of piano-roll cells to synth pad sounds of appropriate ontime, pitch, and duration, allowing real-time interactive editing of the loop and accurate synchronisation with the drum track. Clicking an empty cell turned it orange (adding a note), clicking it a second time turned it black for duration editing, and clicking it a third time turned it back to empty (removing a note). When a cell was black, the duration of the corresponding note appeared in a text box next to an "Edit Duration" label, as depicted in Fig. 3 in the panel immediately below the piano roll (duration 0.25 quarter-note beats by default). The button to the right of this text box could be used to update note duration edits. The "Get Suggestion" button is circled by a red dashed line in Fig. 3. Next to this is an "Undo Suggestion" button and an "I Am Finished" button for submitting a completed loop. Each time a user added, removed, or edited the duration of a note, or requested/undid a suggestion, a combination of HTML, JS, and PHP resulted in this edit being time-stamped and stored for analysis. Toward the bottom of the interface are a play/pause button and sliders for controlling volume of the synth pad, drums (not shown in Fig. 3), and tempo of the loop (default 120 BPM).

In-depth Interaction Example

The black numbers 1–25 in Fig. 3 indicate note-by-note edits of participant 12 in the "suggest" condition, and this among other sessions can be seen/heard in the online material, substantiating the claims made in the introduction about ease of collection and analysis of time-lapse composition data. The user begins by requesting a suggestion (dashed black bounding box labelled 1), then makes some additions/removals following the suggested events (2–5), requests another suggestion (6), followed by further edits of intervening material (7–14), etc. The way in which these suggestion requests are integrated among ordinary composition edits (adding/removing notes) is typical of other users too, and is encouraging, because it suggests that users were able to easily assimilate the suggestion functionality into their creative processes.

Experiment Procedure

Participants read through instructions inviting them to compose EDM loops and explaining the piano-roll interface. They proceeded to the piano-roll interface and were assigned a condition (baseline or suggestion-enabled) and drum track ("The age of love" or "We fly...tonight") at random. This counterbalancing minimized the effects of learning and differing drum tracks on the results. After five minutes of composing, a message appeared to prompt users to make any final adjustments and submit their loop. They moved on to compose a second loop, and then to complete a short questionnaire that asked about number of years experience with music software, current regularity with which they used

such applications, and enjoyment of the suggestion button. Ten participants also agreed to be recorded in verbal post-study interviews.

Results

Questionnaire responses indicated that participants enjoyed the suggestion function (mean rating 5.167 on a scale 1–7, sd = 1.528). The suggestion function enjoyment ratings are plotted in Fig. 3 as purple bars, linearly scaled to 0–100%. The total edits performed by each user in the suggestion-enabled condition are plotted as blue bars, ranging from 10 for participant 11 to 82 for participant 3. The percentage of these edits that were suggestion requests is plotted in turquoise. The mean suggestion requests as a percentage of total edits is 15.9%. The percentage of suggestion requests that were undone is plotted in green. The mean undos as a percentage of total requests is 19.25%. In other words, a suggestion might be considered successful just over 80% of the time. The remaining bars for each participant relate to mean creativity ratings observed in the study described in the next section.

On the qualitative side, we get further encouraging results. When asked if they would consider incorporating such suggestion functionality into their compositional practice, all but one interviewee responded positively through statements such as “Something I would use, definitely” (participant 8) or “It’s definitely got a place. . .yeah for sure” (participant 12).

In terms of evaluating the quality of the resulting EDM hook compositions, we performed an additional, independent study using an online questionnaire, described in Section 2.3.2**.**

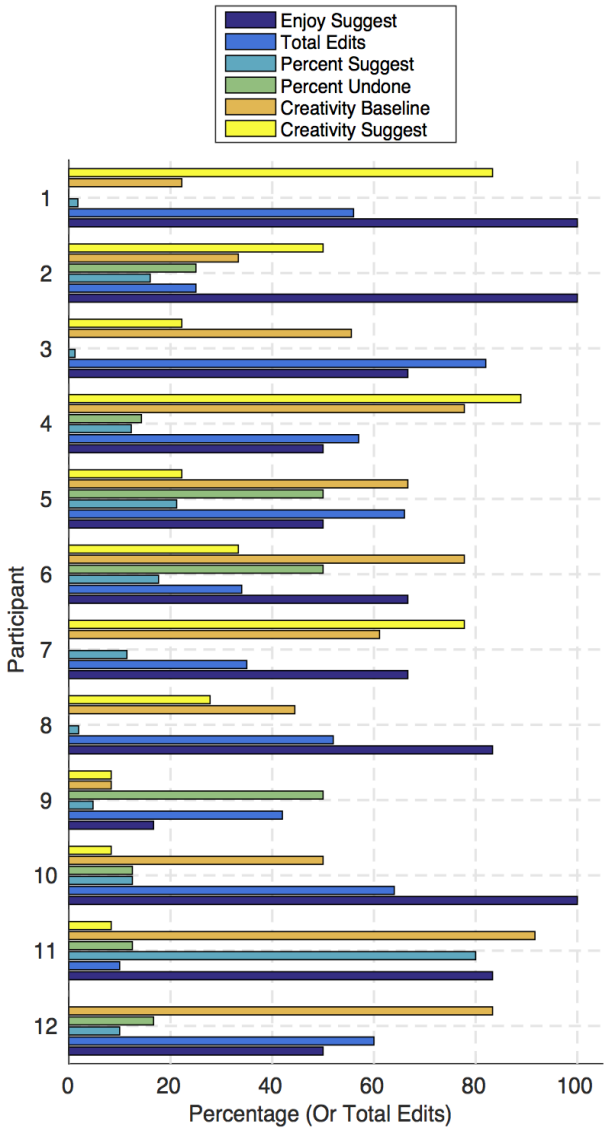


Fig. 4: User ratings of suggest functionality (purple bars), quantifications of edit behavior (blue, turquoise, and green bars), and mean listener ratings of creativity (orange and yellow bars). Likert scales are mapped linearly to percentages for ease of comparison.

2.1.5 In-depth interviews with test session participants

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Headliner DJ Semi-Pro DJ Tablet/Multi-device Producer Professional Producer Ground Breaker	Berlin	8	Appendix

During the Berlin user testing event, 8 in-depth interviews were conducted. The outcomes of these interviews were analysed alongside the datasets from other RBMA.

2.2 Local testing

Following the Berlin user testing event the researchers from each partner organisation continued to test prototypes locally. These tests are reported in the following section.

2.2.1 Rhythm Pattern Variation App

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Headliner DJ Apprentice Producer Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer	JKU Linz	10	D5.5

The Rhythm Pattern Variation iOS app developed in WP5 (see Deliverable D5.5) was tested in user test sessions locally at JKU. This was done in order to evaluate the interaction with the prototype as well as the quality of the generated patterns compared to the old windows based prototype.

To this end, ten experts (musicians and producers which are familiar with using DAWs in the context of electronic music production or performance) were interviewed using a questionnaire as guideline.

In such a session, participants were introduced to the two prototypes (windows-based and iOS/touch-based) and the aim and functionality was explained. They were asked to input rhythm patterns they usually work with and let the prototypes generate a list of variations for these patterns. After browsing through the patterns and exploring the features of the systems, users were interviewed about their experience with the prototypes and their preferences.

Specifically, they were asked to rate the following properties on five point Likert scales:

- The usability of the prototypes,
- The application of such a tool in a live performance or
- in a studio production environment.
- The preferred input method (MIDI controller and mouse, touch interface, or touch interface combined with physical knobs), and
- usefulness of the additional features in the touch-interface-based prototype.

Additionally to the UI evaluation, the differences between the pattern variation algorithms were also evaluated. To this end, both algorithms were available in the touch-interface-based prototype. Participants were asked to browse through variation lists generated by

both algorithms for seed patterns of their choice. After that they were asked to rate both algorithms in the following categories on five point Likert scales:

- Consistency of the variations with the seed pattern,
- musicality and meaningfulness of created patterns,
- difference of created patterns to the seed pattern,
- creativity and interestingness of created patterns,
- suitability of created patterns for continuous beat, and
- suitability of patterns for fills or breaks.

These categories correspond roughly to the ones used in the web survey in our previous evaluation which compared different variation algorithms (see deliverable D2.4) The order in which the algorithms were tested was randomized to avoid experimenter bias. The Likert scale for the *difference* rating ranged from ``too similar'' to ``too different'', therefore the optimal answer ``just right'' was placed in the middle. This is also reflected in the evaluation: For the *difference* ratings additionally root mean square errors (RMSE) towards the optimal value (3) are provided.

A more general discussion including topics like the prototype's concept and applicability, positive and negative experiences during the experiment, UI details, missing features, and the participant's usual workflow and preferred tools concluded the sessions.

Results

The interviews were conducted during the period between June and October of 2016. Participants (9 male, 1 self-identified as “other”; mean age 31.1) were selected according to the following criteria: They were required to have experience in *i.* using DAWs or similar music production software, *ii.* producing or performing electronic music live, and *iii.* using drum step sequencers and/or drum roll editors.

	Consistency	Musicality	Difference		Interestingness	Substitute	Fill
			mean	RMSE			
Algo. A	3,7	4,2	3,2	0,6	3,8	3,8	4,0
Algo. B	3,9	4,4	2,9	0,3	4,0	4,4	3,6

Table 2.2.2: Mean values of participant rating for the two algorithms. For *difference* additionally the RMSE to the neutral value (3) is provided.

Table 2.2.2 shows the mean values for the participants' ratings of the comparison between the old variation algorithm (top) and the updated one (bottom). Since the number of participants of ten is too low for statistical significance analysis the numbers merely indicate tendencies. Nevertheless, the ratings in combination with in-depth discussions with the participants show a clear preference towards the updated variation generation algorithm. The only exception being the suitability of the generated patterns for fills. This can be explained by taking into account that outlier patterns are discarded by the variation algorithm. For this reason the algorithm produces patterns more similar to the seed pattern, which are less suitable for fills or breaks.

The tendencies regarding the ratings for the UI are similarly consistent. Ratings for usability are higher for the touch interface (mean: 4.7/4.3). The difference is even higher for the suitability in live scenarios (mean: 4.0/3.5). While 50% of the participants uttered concerns about the practicality of using a mouse to enter rhythm patterns on stage, only two participants were concerned that the touch device is not suitable for a live performance. One participant's reservations regarding the touch interface did not concern the way of interaction but rather if the hardware (iPad) would survive the harsh conditions on stage and on the road (heat, mechanical strain, spilled drinks, etc.). Another participant raised concerns regarding the touch interface's precision and reliability in a live environment. Regarding the applicability of the prototypes in a studio or production setting, the difference was smaller, but still in favor of the touch based prototype (mean: 4.7/4.6). The comment of one participant summarizes the tenor of the users:

"Using the touch interface is definitely faster and easier [...] compared to entering patterns with a mouse." - Participant03

Regarding the preferences of the input method, a clear tendency towards the touch interface was observable: Six participants preferred the touch interface, three were undecided, and only one voted in favor of the physical controller and mouse system. Regarding the touch interface with attached physical knobs, seven participants preferred the touch only approach, one was undecided, and two preferred using the physical knobs.

"Regarding the physical knobs I prefer just the touch interface. If there are small knobs I would just loose them." - Participant03

The additional features were generally received very positively. Only two participants were unsure if the feature to start new patterns only with a new bar was useful. All other participants were in favor for all three additional features.

In the discussions with the participants several key-takeaways were identified. Three participants considered the arrangement of the patterns in the one-dimensional list of the variation wheel as unclear or ambiguous:

"It seems a bit random to me. I can browse through the list [...] but I cannot look for something specific." - Participant04

While the idea of a simple one-dimensional variation dial was well suited to conduct experiments regarding the quality of variation algorithms, it might be an over-simplification for a real tool. After all, two different properties (sparseness and similarity) are projected into one dimension. Participants suggested to solve this by adding the option to change the sorting of the list or by using an x/y variation pad similar to the one used for the Drummer plugin of the Logic Pro DAW. One participant requested the option to switch between the two algorithms on the interface, to be able to generate fills and rhythms.

While the visual preview was a well received feature, two participants missed an acoustic preview or "pre-listen" mode. Finding suitable audio material for remixing by listening to it on separate headphones is a common technique used by DJs in live situations.

"I would like [to have] an option to listen to a preview on headphones." - Participant01

In this context two participants also requested the possibility of creating and working with patterns while another drum pattern is looped in the background:

"An interesting feature would be to have two patterns you can switch. One is running, and the other one can be modified." - Participant03

Almost all participants (8/10) mentioned that they use a piano or drum roll editor within their DAW to produce drum rhythm patterns. One of them explicitly stated that he tries to avoid it:

"I use the piano roll editor in Cubase if I have to, but it is a real pain." - Participant04

2.2.2 RhythmCat

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Headliner DJ Apprentice Producer Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Serious Modular Controller Hobbyist Music Technology Hacker Ground Breaker	Barcelona	16	Ó Nuanáin et al. 2016 Ó Nuanáin et al. 2016a Ó Nuanáin et al. 2016b

We carried out a similar testing phase to that described previously 2.1.2 locally in Barcelona. Participants comprised 5 music researchers and students from Universitat Pompeu Fabra and the Escola Superior de la Música. In fact when we studied the interview transcripts and performed thematic analysis we considered the Barcelona participants and Berlin participants together. We have include some observations specific to Barcelona here however for reference.

One common thread we noticed with these users was their concern that, while the musical output was interesting, they did not feel they could predict what was going to happen, and would have preferred more granular control over parameters in contrast to the RBMA users, who were quite happy as long as it was generating interesting sounds.

"it's not predictable enough. You don't know what's going to happen."

"The results are interesting but the way to get the results, it's unpredictable."

"I would say now it's more useful for production probably, because for performance, it's not predictable enough."

Our impression was that most users would want to use the tool to explore their own sounds and samples, but we did encounter one participant who expressed no interest in this; rather

they expected some sort of preset or stock samples to be available and to be able to work from there.

"Experimenting with this for me is really pleasant and easy but having to generate corpuses of recordings really is not for me. "

Participants also wondered whether it was available for mobile devices or at least standalone (not dependent on VST hosting within DAW applications). Indeed this is something we are currently investigating, it is a challenge to have the code flexible enough to write once and distribute for desktop, VST and mobile devices but certainly not impossible.

"It's nice to have both worlds. Integrate it with a track but also standalone. The standalone makes sense, why not?"

"I want a standalone so you can use it on a tablet for example."

"For the interface, it's really good for iPad."

2.2.3 House Harmonic Filler + Dr. Drums

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Apprentice Producer Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Music Technology Hacker Ground Breaker	Barcelona	7	Jordà et al. 2016

Tests in Barcelona were conducted in preparation to RBMA user tests in Berlin. We conducted 7 in-depth interviews, of about 1 hour each, with MTG-associated experts in the domains of musical performance and/or human-computer interaction.

The procedure of interviews was identical to the one in Berlin: The two prototypes were tested together. After a brief introduction to the system the interviewees were left to experiment with the system for about 10 minutes before a longer interview of 30-45 minutes.

Opinions were generally positive with interesting feedback regarding the user interfaces and in the demoing process. A few bugs were detected, allowing us to fix some functionality before the tests in Berlin. Some interesting insights follow in the next paragraph.

One of the interviewees, an engineer working in music technology, was in favour of reducing the control parameters of the prototypes. *"I want to be surprised by the system,"* he said, instead of having to set various parameters individually.

A participant suggested to substitute musical notions in the instruments (key, chord, octave, legato, etc.) with more general notions, suitable for a novice, guiding the semantic notions

with changes in brightness and/or color on the interface. However this opinion was contrasted by another interviewee, a formally trained classical musician, who was suggesting to incorporate more musical terminology, like the names of the chords and modes.

Another user commented on the possibility of having a *Suggestion Mode*, graying out or highlighting different continuation possibilities according to the users actions.

2.3 Web- based testing

A number of local test were conducted online. The following section is an overview of these outlining setup and outcomes.

2.3.1 RhythmCAT Listening Survey

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Headliner DJ Apprentice Producer Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Serious Modular Controller Hobbyist Music Technology Hacker Ground Breaker	Online	21	Ó Nuanáin et al. 2016 Ó Nuanáin et al. 2016a Ó Nuanáin et al. 2016b

This quantitative evaluation investigated the algorithm's efficacy from a retrieval point of, i.e. does the system retrieve correctly labelled items in a predictable fashion based on its similarity/distance metric. Subsequently we reconciled this objective evaluation with the subjective impressions of the listeners by performing an online listener survey. To formulate the experiment we gathered a small dataset of 10 breakbeats - short drum solo samples taken from funk/soul recordings during the 1970s - that ranged between 75 BPM to 142BPM and truncated each to a single bar in length.

In turn, each of these 10 breakbeats was chosen as the seed loop with the remaining 9 loops used to derive the corpus (not unlike holdout validation in machine learning research). Four newly concatenated sequences were derived for each of the 10 targets using 4 different distance ranges, to generate 40 variations in total (10 targets * 4 distances). The normalised distances then chosen were at 0.0 (the closest to the target), 1.0 (the furthest from the target) and two random distances in ranges 0.0 - 0.5 and 0.5 - 1.0.

As mentioned previously, the objective evaluation considered the algorithm's ability to retrieve labelled items predictably. We labelled each onset that was segmented by the software in terms of their constituent drum sounds (kick, snare, hi-hat, cymbals and X for anything else). Labels from generated sequences were matched against corresponding labels

in the target and accuracy ratings were given by (A = correct_labels / total_labels). This produced a moderate negative correlation between accuracy and distance (distance being inversely proportional to similarity).

To compare this against what the human perceives, we used the same dataset and created a web survey (Figure 6) to gather listener ratings for three facets. On a Likert scale of 1-5 users rated each generated pattern in terms of its timbral similarity (i.e. do the drum sounds sound similar) to the target, its pattern similarity (are the temporal placement and organisation of the sounds similar) to the target and lastly, how much they liked it aesthetically.

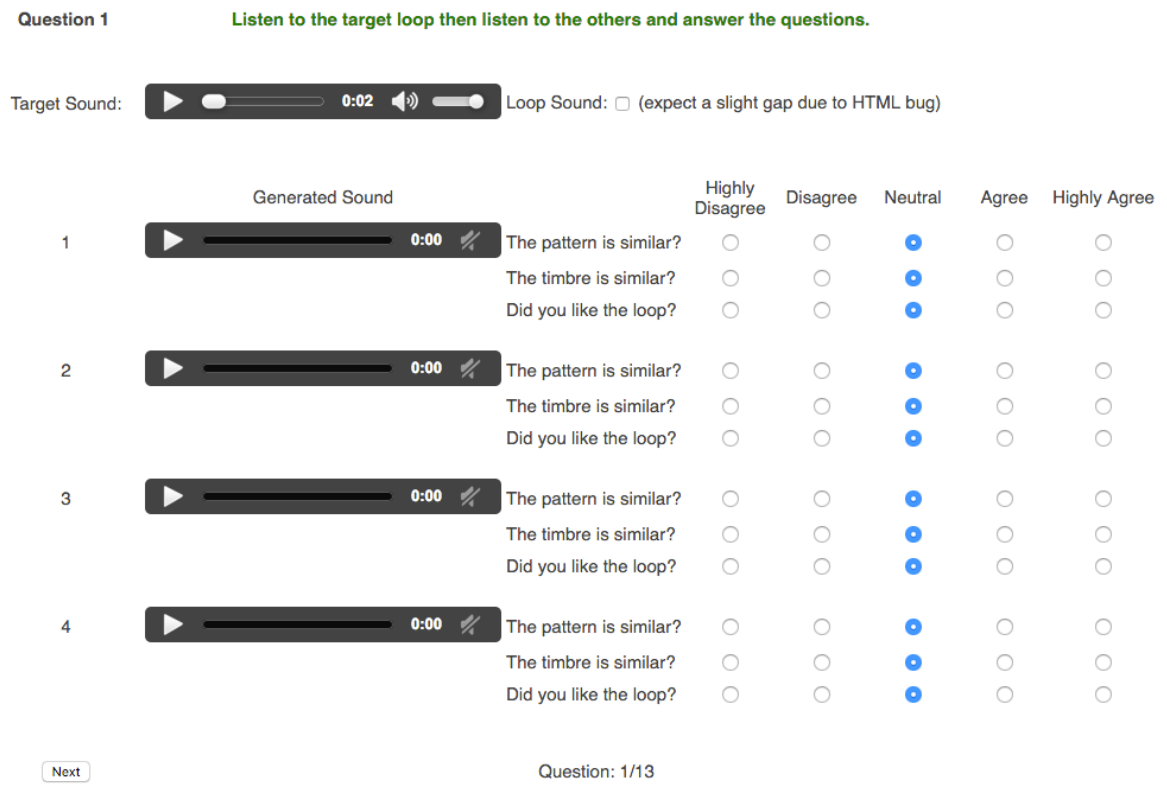


Fig. 6: Web Survey Interface

We gathered 21 participants to take part in the survey, most of whom were drawn from the institutes of Universitat Pompeu Fabra (UPF) and Escola Superior de la Música Catalunya (ESMUC) in Barcelona. Twenty of the participants played an instrument, 10 of whom specified a percussion type instrument. Nine out of the 21 participants indicated they were able to read notation.

Results were collated and analysed using Spearman's Rank Correlation in the R Statistical Environment. For each of the variables under consideration by the listener's we discovered a moderate to strong negative correlation ($p < 0.01$ in all cases) with the associated distance values of the stimuli (Figure 7). Listeners considered generated patterns with smaller

distances as more perceptually similar to the target, as well as preferring those more similar also.

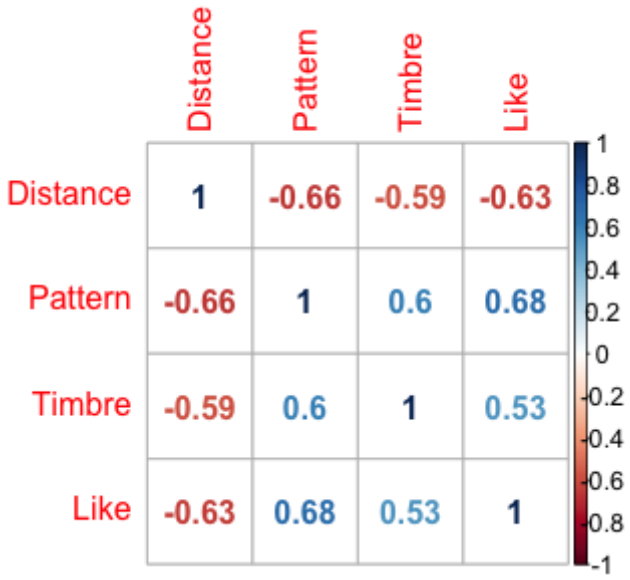


Fig. 7: Correlation Matrix of User Responses and Distances

2.3.2 EDM Note Suggestion Output Evaluation

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Apprentice Producer Laptop/Home Studio DJ Professional Producer Music Technology Hacker Ground Breaker	Online	4	[Colllins et al, 2016]

To evaluate the quality of the compositions resulting from the Berlin test session of the EDM note suggestion API and Web interface (see 2.1.4), independently of the Berlin testing session, we conducted a listening study in which participants familiar with EDM (and not present in the Berlin session) heard the completed loops created by the users of the Berlin session "blind" (without knowing whether a given stimulus was composed in baseline or suggestion-enabled conditions) and rated the creativity according to Amabile's Consensual Assessment Technique (CAT) [Amabile, 1996]. In the CAT framework, judges rate the creativity of an artistic product on a scale 1–7. The aim of the study was to investigate the perceived creativity of loops completed under suggestion-enabled versus suggestion-free conditions, and to shed light on whether suggestion functionality—and, by extension, creative MIR—acts as a negative, neutral, or positive influence on the creativity of EDM compositions.

Participants

Four graduate students were recruited from Johannes Kepler University and DeMontfort University (mean age 34.6 years, sd=2.2 years) to act as judges. They reported mode “Daily” regularity of listening to EDM. Although the origin of stimuli was not disclosed to judges, they may have been aware of the topic of creative MIR, due to familiarity with our research.

Stimuli

Stimuli consisted of complete loops from the user study described in Sec. 2.1.4, looped twice and then faded out on the third play.

Apparatus

Participants completed the listening study individually online via sound files, rating sliders, and free-text response boxes embedded in an HTML form. We did not control for volume or method (headphones/speakers) of delivery, allowing listeners to set these according to preference, since such differences were not expected to influence the results.

Procedure

Judges read through instructions introducing two examples of 90's EDM, describing musical creativity in a manner consistent with the CAT, and inviting them to listen to and judge the creativity of EDM loops on a scale 1–7. They proceeded to the ratings page, which contained an embedded sound file, 1–7 slider, and free-text response box for each of 24 excerpts (one baseline and one suggestion-enabled excerpt for each of twelve users from the Berlin session. Excerpt order was randomized to minimize ordering effects.

Results

One judge used only a limited range of the creativity scale, so his data were removed from the following analyses. For the remaining three judges, there was significant agreement between creativity ratings according to Kendall's coefficient of concordance ($W=.589$, $\chi^2(15)=26.5$, $p<.05$). This is in keeping with findings that the use of appropriate observers/judges tends to result in significant levels of inter-judge agreement [Amabile, 1996; Hickey, 2001]. We calculated the average creativity rating for each finished loop, giving twelve values for the suggestion-free loops (plotted in orange in Fig. 4) and twelve analogous values for the suggestion-enabled loops (plotted in yellow in Fig. 4), and used a paired t-test to interrogate the null hypothesis of same underlying distribution. The null hypothesis was not rejected ($t(11)=1.612$, $p=.135$). A well known issue with frequentist hypothesis testing is that non-rejection of the null hypothesis is not equivalent logically to finding evidence in favor of the null [Rouder et al, 2009]. Therefore, we conducted a Bayes factor paired t-test, which is capable of favoring the null or the alternative, with scale parameter $r=1$ chosen prior to analysis. The JZS Bayes factor for the reported t value and sample size was $B_{01}=1.525$, which favors the null slightly (odds of 3:2). In other words, we have found evidence that our judges considered the creativity of the suggestion-enabled loops to be on par with that of their suggestion-free counterparts. Despite having found this statistical evidence, we will conduct further investigations with larger sample sizes in order to support this finding.

Letting y represent user enjoyment ratings of the suggestion functionality (purple bars in Fig. 4, we explored whether some of the subsequent variables plotted in this figure could predict these ratings. For example, letting x_1 represent the percentage of times a user requests suggestions (turquoise bars), does x_1 predict y their enjoyment of the functionality? Letting x_2 represent perceived creativity of loops in the suggestion-enabled condition (orange bars), and x_3 perceived creativity of loops in the suggestion-free condition (yellow bars), do either of these in isolation or combination predict y enjoyment of the suggestion functionality? We found that the only (borderline) significant predictor of suggestion functionality enjoyment was the percentage of times a user undid suggestions ($F_{1,\ 10}$= 4.050, p=.072, r^2=.288, s=17.85), with smaller percentages of undone suggestions corresponding to greater enjoyment of the functionality and vice versa.

2.3.3 Mad competition for discriminating Rhythm Spaces

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Apprentice Producer Laptop/Home Studio DJ Professional Producer Music Technology Hacker	Online	20	Gómez Marín et al. 2016

During the development of the rhythm space infrastructure one online experiment was carried out regarding rhythm similarity on drums. At this point, different models for building rhythm spaces were available and we needed information to discriminate between them. We decided to select the space that was more aligned with similarity values reported by subjects. The goal was to select the space which presented the highest alignment with human similarity judgements for drum patterns.

Method

The methodology we used was based on a fixed set of drum patterns containing MIDI loops in four different styles (Hip-Hop, Soul, Garage House and House), which were compared pairwise with three different metrics (PAD, SAD and Euclidean) and dimensionally reduced using three different methods (MDS, PCA and t-SNE). In the end we had nine different versions of a rhythm space containing the same rhythms, and each locating the rhythms in a different position. Thus each space offers a different version on which patterns are more similar/closer and different/distant. To select the best rhythm space model we used a Maximum Differentiation competition which allows to compare two spaces by using maximally distorted pairs, that is, two elements which in one space are very close together and in the other are distant. By asking subjects if the elements produce a similarity or different sensation one of the spaces scores a point. If different maximally distorted pairs are used, some that are close in space A and distant in B and vice versa, both similarity and difference are tested, thus the space that holds the highest alignment with subject ratings has scored the maximum agreement with human ratings. A survey was then prepared for

subjects to listen online to the rendered patterns and rate them on a Likert scale from one to seven. One signifying that the patterns are very different and 7 that the patterns are exactly the same.

Results

Twenty voluntary subjects, 4 female and 16 male in an age range 24 to 54, participated in the experiment. Fourteen of the subjects reported frequent listening to EDM, 16 were instrument players and 16 had formal music training. Subjects were not paid for taking the experiment. The scores for each pair's ranking presents a chi square value of 152.5 and a p-value 2.2e-16 based on a Friedman rank Sum test. For each compared pair the mode of the results was selected as the relevant similarity value to be used. The 7 value scale by which similarity was graded between the pairs was grouped in three categories: rankings below 4 (difference between pairs), rankings equal to four (neutral) and rankings higher than 4 (similarity between pairs). The comparison between two spaces is based on two different groups of 4 rhythm pairs (similar in A / distant in B and distant in A / similar in B). The results for each group of 4 pairs are summed and the winner space is the one that has more favorable scores: if most pairs were regarded as different, the space in which the pairs were different wins and vice versa.

The Euclidean t-SNE space wins over the PAD MDS in 7 out of 8 comparisons, the other one is tied at neutral value. SAD PCA win 5 out of 8 pairs with 3 losses over Euclidean t-SNE. SAD PCA wins 7 out of 8 pairs over PAD MDS with 1 loss. Only in one cases the mode of the results was central value. The winner of this competition is thus the SAD-PCA space which has better scores than PAD-MDS and Euclidean-t-SNE when compared with them.

2.3.4 ROTOR application beta testing

Prototype relates to Personas	**Test Location**	**No. of Users**	**Prototype documented in**
Tablet/Multi-device Producer Professional Producer Serious Modular Controller Hobbyist Music Technology Hacker	Online	34	D6.2

During the final stages of development of the ROTOR application, a group of 34 users consisting of 8 advanced reactable mobile users, 23 musicians who normally include iPads in their music production and 2 and professional application reviewers were invited to take part in an early access to the new application. This served two purposes: the first and more targeted was to find out which parts of the application were not functioning as expected, and the second and more interesting would be to get feedback on the features developed during the project.

When the application was released, an anonymous survey was sent to all the participants to measure not only the general satisfaction with the app, but also how the new features developed within GiantSteps impacted their general workflow.

Some of the feedback from the users directly contributed to specific changes in the functionality and user interface of the application, such as:

- Removal of all the global objects in favor of a more streamlined workspace.
- Addition of a "Transport" bar that enabled the app to play/pause
- Addition of an Interactive help for all items inside the application.

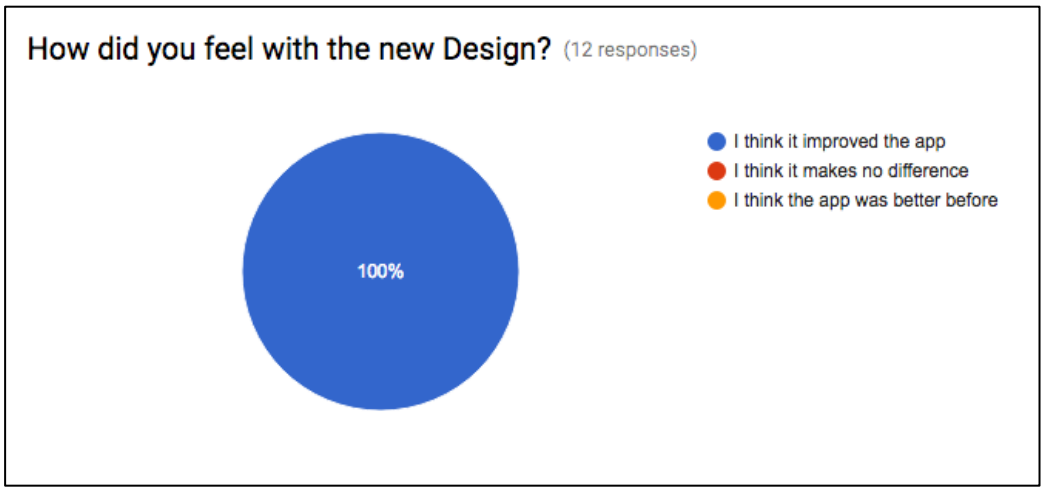
Some comments from the users about the new functionality, taken from the beta testers

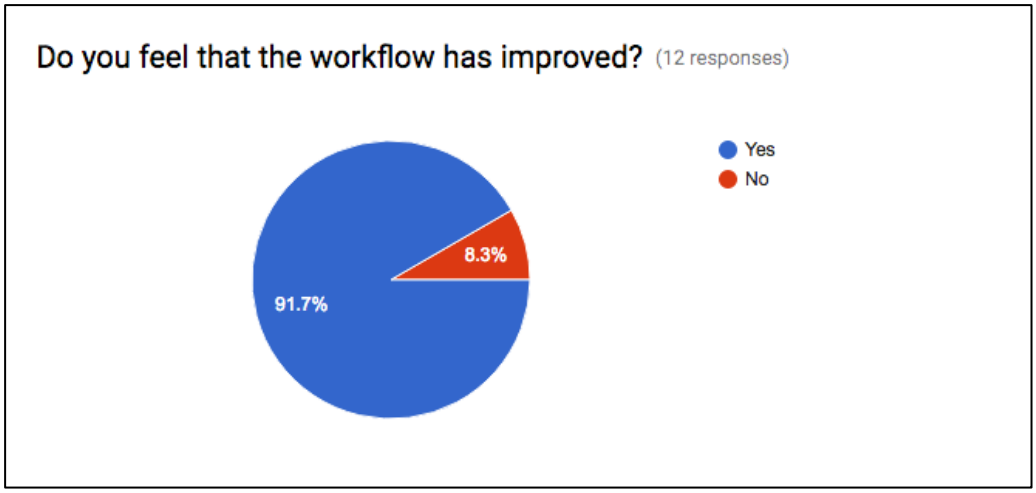
> *"Love the new Automatic tonalizer. It prevents me from having to change the scale on the oscillator objects, and I just noticed it can do it once per beat. The results are strange, but I like it" - Beta tester No.25*

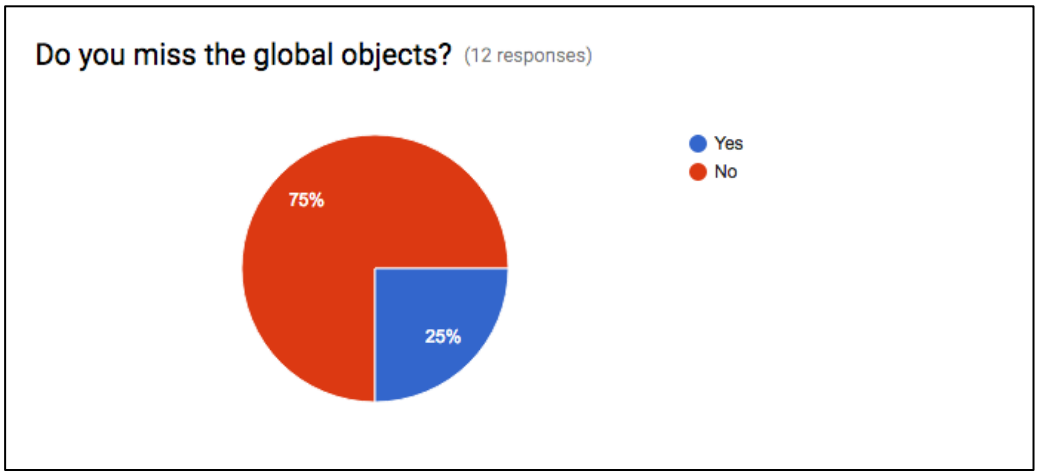
> *"I think although the application is not very suitable for a production environment [...] it does open a set of very stimulating experiences to jam or perform live, and that makes it something very unique and compact that allows you to understand the internal dynamics of a live electronic performance in a clear and concise way." - Beta tester No. 03*

From the outcomes of the online poll, only 19 users answered the final questionnaire, however we found that they were the most engaged users within the beta testing period. From this data we can extract relevant conclusions, on both their experience during this period and on the overall satisfaction with the app.

a) On the improvement from the previous apps by reactable

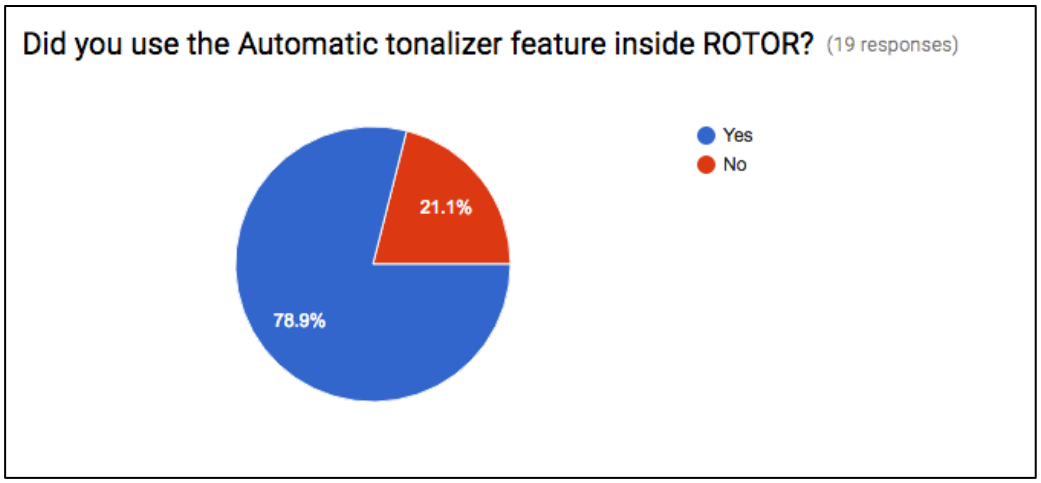


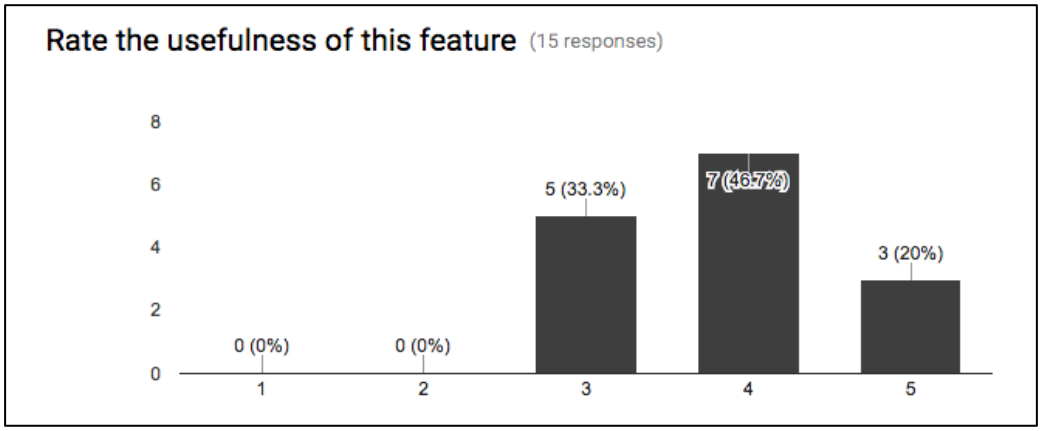


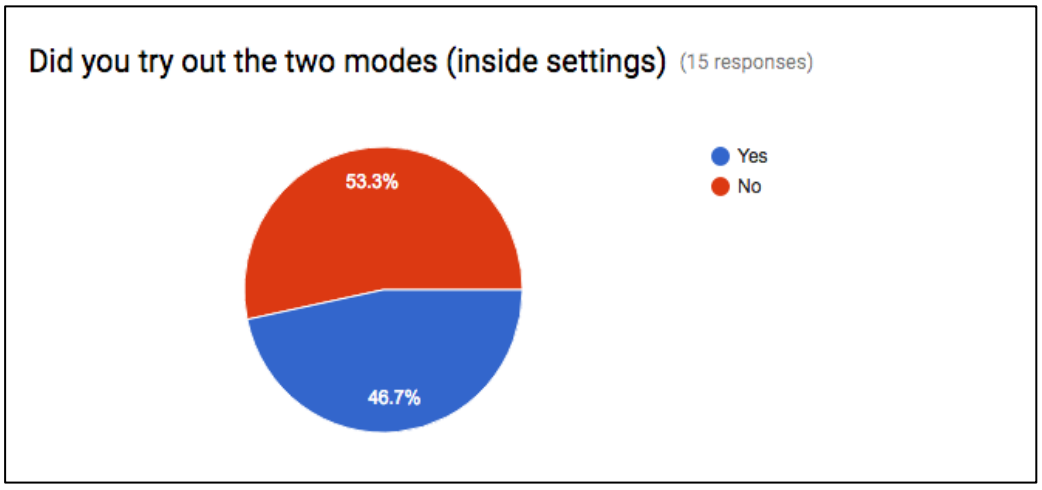


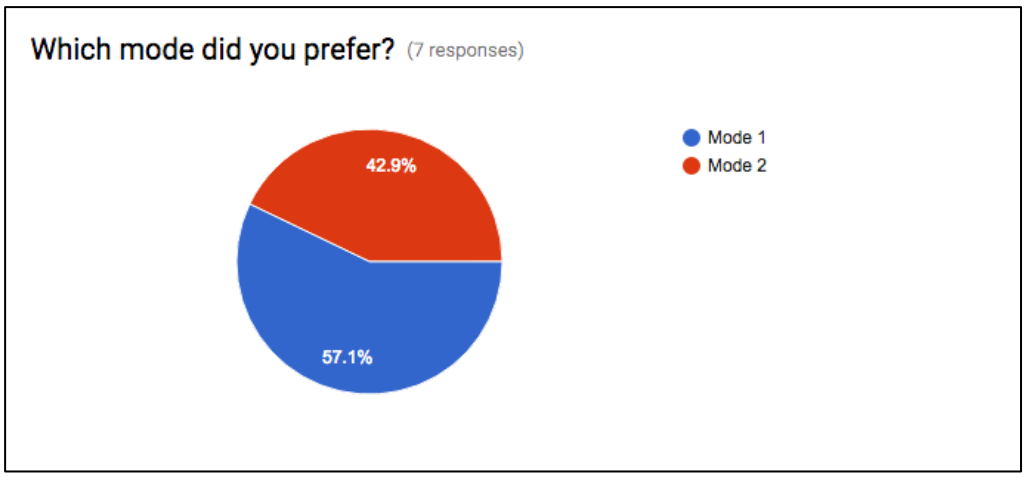
The overall workflow of the application was perceived as an improvement for the majority of the participants that had used reactable mobile before; the removal of the global objects and the perception of the new design were two metrics that were tracked because they offered the biggest change and it was important that the users felt good from a change in an application they had already learned to use.

b) On the automatic tonalizer feature



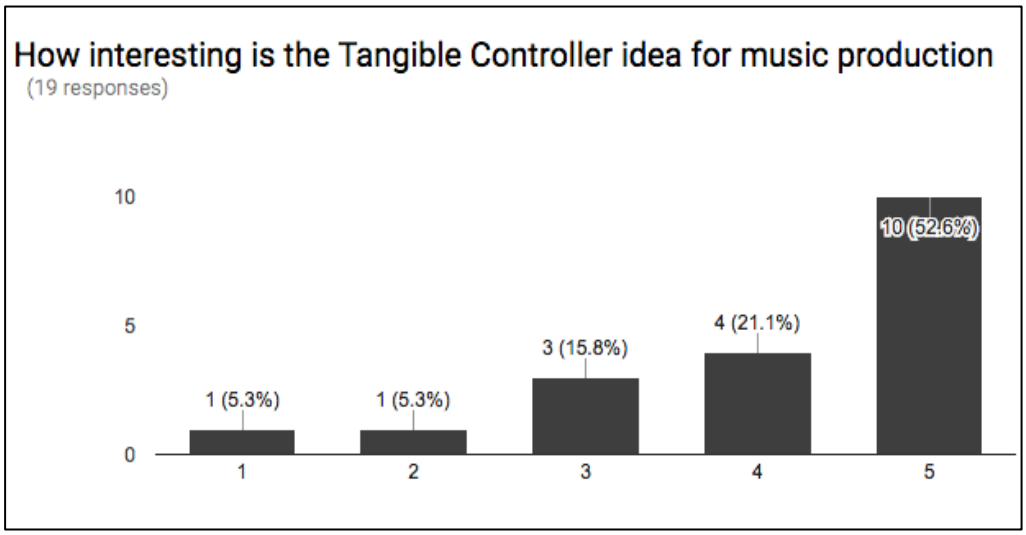






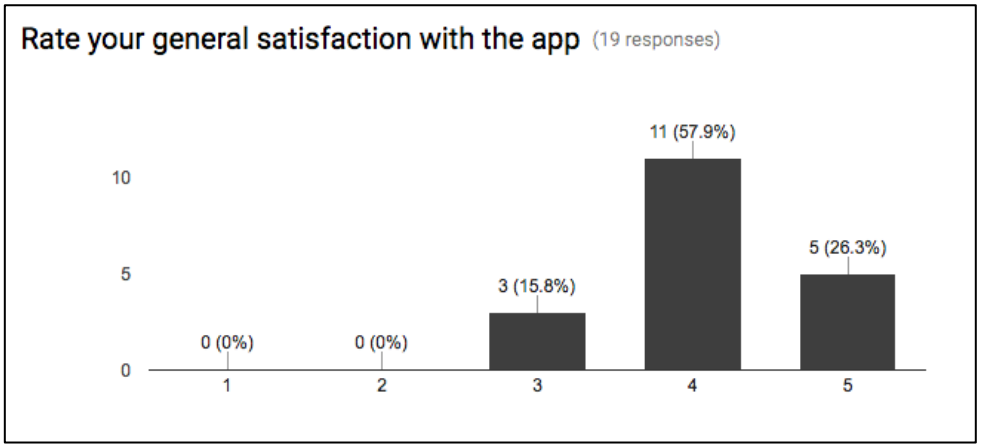
On the automatic tonalizer feature the most important feedback was that it is perceived as a useful feature (66.7% of the users gave a score of 4 or 5 in the scale) but the different modes were not discovered by most users; the ones that did try switching between them slightly preferred the mode with stylistic information (Mode 2 in the chart) over the mere probabilistic model (Mode 1 in the chart).

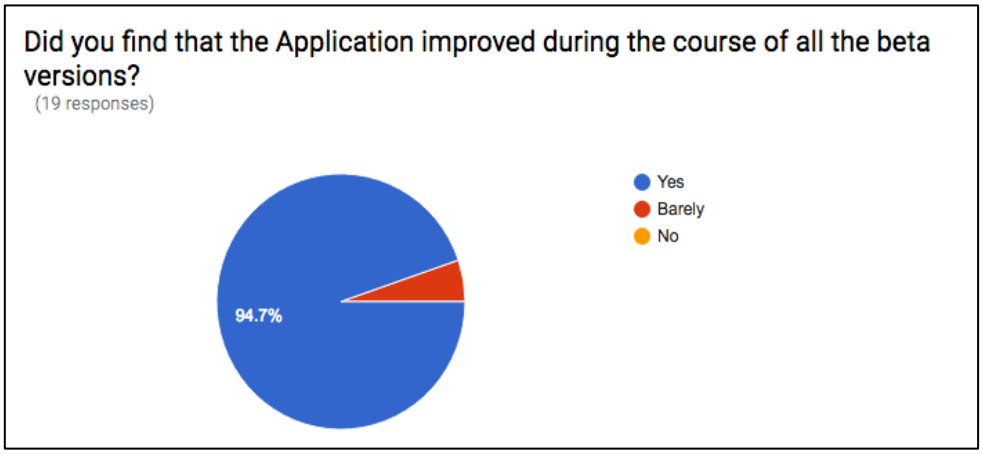
c) On the physical controllers



On the idea of using tangible controllers when performing with touchscreen devices the majority of users were interested. None of the participants used the ROTOR controllers designed to work with the application, so this is more of a general concept; one of the drawbacks of performing with these devices over dedicated music hardware is the lack of feedback 2D screens provide. Tangible objects have the potential of providing this missing experience.

d) On the overall app and beta testing experience





On the beta testing experience we did perceive positive feedback when working in a tight loop with future users of the app. All the participants perceived that the application evolved during subsequent releases and the satisfaction levels after the beta testing phase are coherent with the results obtained from the general public after the release was made public.

The information gathered during this process was a great influence on design and user experience, as well as on implementing features such as an interactive help and manual that were very well received by users as reported in D6.5 and D6.6.

2.4 Informal demos, hackathons and workshops

During the final year of the GiantSteps project we participated in a number of public events with the dual aim of exposing ideas to users and gathering more informal feedback from these interactions. The following section describes the GiantSteps participation at Music Tech Fest 2016, Sonar 2016, Waves Festival Vienna 2016 and ADE 2016.

Some of these events are also mentioned in D7.7 and we will only outline them here.

2.4.1 Music Tech Fest 2016

Prototype relates to Personas	Test Location	No. of Users	Documented in
Laptop/Home Studio DJ Serious Modular Controller Hobbyist Music Technology Hacker Ground Breaker	MTF Berlin	~ 12	D7.7

GiantSteps acted as one of the sponsors of the Music Tech Fest **#MTFHacks**. In this event some of the libraries and tools designed for the development community, such as the madmom library, Dr.Drums and House Harmonic Filler as well as some of the products and prototypes for music creation, such as a beta version of ROTOR, the Rhythm Pattern Variation app and RhythmCat were shown and offered to participants to use for their hacks.

Fig. 7: GiantSteps team setting up at Music Tech Fest.

The hack that won the GiantSteps prize was using the House Harmonic filler to generate automatic chord progressions and bass lines. The hack was an elaborate interactive system that sensed light changes via an array of sensors installed on the branches of a tree which were converted to musical sounds. The bass sounds were played by a set of piano chords that were installed on the tree trunk each having a piano hammer that was activated via an Arduino board by the information from the sensors. The chords were reproduced from a loudspeaker.

This event is described in more detail in D7.7.

2.4.2 Sonar 2016

Prototype relates to Personas	Test Location	No. of Users	Documented in
Apprentice Producer, Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Music Technology Hacker Ground Breaker	Sonar Barcelona	>100	D7.7

GiantSteps was present with a booth at Sónar festival to demonstrate and showcase the products and prototypes developed during the last year of the project. It was an opportunity to not only for dissemination but to engage with the public and get feedback from potential users. Since the musical abilities of the attendants was very broad, and a constant flow of people visited the booth for a very short period of time, it was a good event to observe trends and to confirm that some of the prototypes and products showcased during the festival have a target audience or are seen as potential future products that add value to the music-making community.

Fig. 8: Impressions from Sonar.

Certain trends that are in line with GiantSteps goals are interconnectivity and multiple-devices performances. The setup where all the applications were playing together and in sync was seen as a great feature for most of the participants. Other features that were perceived as added value was the ability to manipulate high-level controls to generate coherent musical content (such as in Dr.Drums, the Rhythm Pattern Variation App, the House-Harmonic filler), particularly in the amateur musicians.

2.4.3 Waves Festival Vienna 2016

Prototype relates to Personas	Test Location	No. of Users	Documented in
Apprentice Producer, Semi-Pro DJ Laptop/Home Studio DJ Serious Modular Controller Hobbyist Music Technology Hacker	Waves Festival Vienna	4	D7.7

GiantSteps acted as an organizer and sponsor of the Music Hackday at the Vienna Waves Festival 2016 through project partners JKU and Yadastar. Apart from using this as a dissemination opportunity of the project's results (see D7.7), the tools developed within GiantSteps were presented and offered to the hackers for their projects.

Fig. 9: Hacking at the Waves Festival Hackday 2016 in Vienna

While the large majority of the presented hacks carried out by over 100 participants was dealing with gestures and hardware, two groups built their unfinished projects around GiantSteps tools, namely the madmom library, the The Eear prototype, and Essentia. Ad-hoc interviews with questionnaires revealed the reasons why the intended hacks (an automatic video clip generator and an automatic accompaniment system) were infeasible within the short hacking period of eight hours. These were problems with platform compatibility, more precisely, a lack of support for Windows platforms (madmom), and over-optimistic expectations of functionality and integration for non-programmers, i.e., an out-of-the-box solution for real-time downbeat tracking and/or chord recognition (madmom and Essentia/The Eear, resp.).

2.4.4 ADE 2016

Prototype relates to Personas	Test Location	No. of Users	Documented in
Apprentice Producer, Semi-Pro DJ, Laptop/Home Studio DJ, Serious Modular Controller Hobbyist	ADE Amsterdam	8	D7.7

In October 2016 we were invited to conduct another version of the "Hypothetical Instruments" workshop at ADE (Amsterdam Dance Event) in Amsterdam.

We worked with 8 participants and asked them to "draw a sound, then build the machine that makes that sound" (see full initial workshop description in D2.1). Equipped with paper plates, cups, cutlery and some helper tools, the participants built their hypothetical instruments. The following are images and quotes from the conversations that were informed by the non-functional instruments:

Fig. 10: Hypothetical Instruments workshop at ADE2016

Several pieces were dealing with the live modification of time and timing:

"It is a piece of sound that can be played back in different ways. When you change the structure it sounds different, as soon as you start to manipulate the shape the whole piece changes… I was thinking of how I could just hold it in my hand and play it. The unmodified plate is the composition and when I touch it I can manipulate it and play it… I am folding the

timeline itself… By expanding the surface I am having access to different timezones on it." Participant 1 (top right in fig 11)

"my idea was to make a system that triggers sounds in different timbres…Every time you press a key, there is a delay and then the sounds come, something happens… so we are setting up a series of things in the future, ready to happen… it's a matter of milliseconds" Participant 3 (bottom left in fig 11)

Others were principally concerned with creating systems for live performance:

"It stretches out and then the sounds changes as the different objects are placed on the line. you can choose your own way how to use it: It is a performance for two or stereo for one. And you can only hear the music if you play. " Participant 2 (bottom right in fig 11)

And in one case the model itself crossed over from conceptual prototype into an instrument:

"What I am trying to do is to keep like some kind of motor that works with it, to keep this kind of constant flow of noisy FX sounds and high spectral sound… So I am gonna make more cups like this, more cups with different kinds of material, so you get this kind of peak in the sounds and that keeps it interesting. So you can modulate it and I am gonna connect it with some filters and some delays and let see what comes out of it. I will just modify the shit out of it." Participant 4 (top left in fig 11)

Fig. 11: Images of participants with workshop outcomes.

During the ADE event we also conducted a number of user evaluations with prototypes, since we had them the exact same setup as RBMA we have decided to proceed this data set alongside the RBMA data.

2.5 RBMA in Montreal

GiantSteps was also present at this year's RBMA during both sessions. With the idea to make the maximum amount of prototypes available to the participants, several iPads with all the applications installed on them were displayed in a public and available space in order to engage in informal conversations and introduce participants to the tools. This was intended

to be the first contact to then engage into more in-depth interviews about not only the products and prototypes that were present, but on the general concepts that the prototypes were trying to showcase.

Fig. 12: GiantSteps team talking with participants at RBMA

The intention of bringing the mobile devices was to encourage users to take them into the studios and use them in their musical productions during the event and to test the interaction between multiple devices since all the applications could be synchronized using Ableton Link.

2.5.1 Software and Hardware Prototype Interviews

Prototype relates to Personas	Test Location	No. of Users	Prototype documented in
Headliner DJ Semi-Pro DJ Professional Producer Serious Modular Controller Hobbyist Ground Breaker	RBMA Montreal	20	Forthcoming

During both RBMA sessions we did testing on different level prototypes and newly released products and features. In total we collected feedback from around 20 RBMA participants and visitors. Informal testing of various prototypes was conducted (RhythmCat with pitch variations, House Harmonic Filler, Dr. Drums), as well as of Reactable's ROTOR and iMASCHINE 2. It was apparent from the first interviews that different prototypes were covering a range of approaches, suiting needs of different types of practitioners.

In general lines, we observed a tendency to consider iPad musical apps as exploratory tools or amateur instruments rather than composing or performance tools proper: most users seemed happy to use iPad apps as sketching tools (while traveling, for example), but less interested in using them live, for example. One of the participants confessed that

“touchscreen apps are far from perfect for control interaction.” However, half of the interviewees recognised that they use mobile devices at some stage of their production, and all of them accepted that they will eventually do at some point (in answer to the question “do you think you will use mobile devices in your music production in the future”). Having said this, the response to the interface designs was overall “very interesting” and “highly intuitive.”

MTG prototypes (RhythmCat, House Harmonic Filler, Dr. Drums) were distributed as plugins on USB sticks and tested on a laptop. In general, participants agreed at seeing these programs as sources for inspiration, and as useful tools in “avoiding the gap of not knowing.” As observed in previous testing sessions, some participants were more akin to use harmony plugins (one of the participants: “The accessibility to interactive harmony tools is a game changer”) and others to use rhythmic generators. The RBMA proved to be a more experimental-oriented event, in which some of the participants were expecting more randomness on the output (“I would like that machines take more decisions”), a result that contrasts with opinions in other testing sessions. Perhaps the main critique to the three MTG demos was their bias to specific musical styles, what could be limiting their use in more personalised aspirations.

ROTOR was particularly attractive for participants with experience on modular synthesisers and more generally with sound synthesis techniques. This was more or less the same reason for which other type of users (singer-songwriters, instrumentalists) found the range of possibilities of ROTOR slightly overwhelming and unattractive. Regarding the new tangible objects, the feedback was mixed and many users felt the objects were not responsive enough. On a more conceptual level, one of the users wondered what the point of using objects in touch screens was, since they are clearly separate realms of operation.

For iMASCHINE 2 we collected feedback on helper tools such as the intelligent chord, scale and arpeggiator play assistant modules developed during the GiantSteps project. People showed interest in the variety of different scales available, and many found this to bring a new dimension into their music making: *“I can very easily make music on the train in a scale I didn’t even know existed”* was a quote of one participant trying out iMASCHINE 2. Participants with no music theory background found them very helpful in order to sound musical, while people with more established base in music theory said it would allow for more time in creative flow. What participants also claimed helpful with these features and made them very desired on mobile devices is that one can play complex chords with one finger only, which is very convenient when screen space is a limited resource.

Fig. 13: Feature demos and prototype testing at RBMA Montreal

Outcomes specific for ROTOR

The two main features that were evaluated as the final outcome of the project inside ROTOR because they proved to be the main differentiating elements from other existing musical applications in the market. Other features such as the normalization of the perceived loudness between different sound files using the replay gain functionality and other essentia features such as onset detection and bpm estimation were not tested against users in the current version of ROTOR.

ROTOR tangible controllers

Reactable Systems wanted to obtain quantitative data that reflected the satisfaction of the user experience using the tangible controllers developed to be used within the ROTOR application. 14 participants ranked how well the tangible controllers worked in a scale of 1 to 5 where 5 was the maximum score and 1 meant that they did not work at all. The results can be seen in table 2.5.1 below.

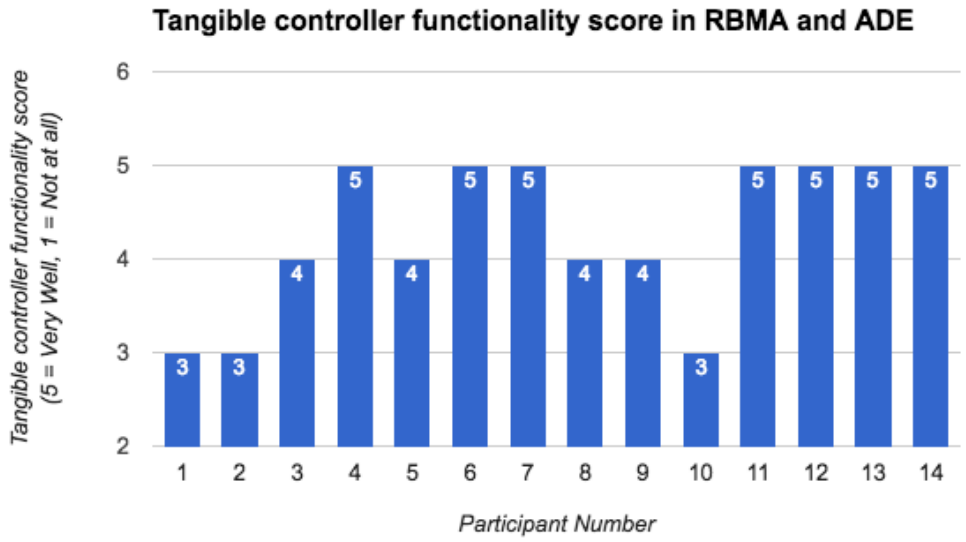


Table 2.5.1 – Tangibe controller fucntionality score in RBMA and ADE

One of the main sources of complaint was the way the controllers performed less accurately when rotating the virtual objects than when moving them, causing undesired effects when playing live and thus perceived as non fully reliable for a live situation.

With this information we proceeded to fine-tune the object detection library and decided to postpone the launch of the physical controllers to the general public, even though it is still one of the most requested features made by the existing users.

Since a considerable amount of noise in the detected touch coordinates from the iPad screen when points were in movement was observed. This noise, probably caused by the discrete nature of the projected capacitive sensing grid, caused excessive fluctuations in the computed ROTOR orientation. This was the primary culprit of the noticeable lack of performance of the controllers since the rotation angle is used for controlling the main parameters of the musical modules.

For this reason the new version of the object detection library now performs low-pass filtering of the 2D touchscreen coordinates before computing the rotation angles. The cut-off frequency of the filter has been chosen to offer an acceptable trade-off between responsivity and noisiness. Filtering is not applied for computing the object's centre since it would introduce excessive latency when moving the object, exacerbating the already existent touch-to-image positional lag of the iOS platform.

The way the ROTOR controllers work inside the application also changed. Using the user feedback from Sónar and the Beta testing sessions two new features were added to the application. Since when manipulating the object on top of the screen a certain degree of occlusion occurs we decided to "expand" the view of the user interface so that the user has a better view of the parameters he can control with both the hand using the tangible controllers and the free hand. When this new feature was tested in the first beta versions, it was seen that this feature could be even further customized depending on the hand dominance preference of the user. This led to a specific setting that allows the user further customization and personalization of the app.

Fig. 14: Rotor as presented at sonar

Fig. 15: Expanded widgets preventing occlusion when performing live with hand dominance setting: a) both sides b) right-handed c) left-handed users

We also added a momentary control feature where users can decide to modify a parameter in two ways: one as a permanent change and one as a momentary or transitional effect. Momentary changes occur only when the performer moves the controller away from the virtual element.

Since the final version of the ROTOR controllers has not been made available for the general public, a second round of user testing has not been scheduled at this point in time, however most of the complaints from interviewed users have been addressed and we feel we have improved the tracking library enough to issue a commercial release for the first days of December 2016.

ROTOR automatic Tonalizer

This feature was tested by 11 participants and ranked in a scale of 1 to 5, where 1 was the maximum score in terms of usefulness, obtaining a mean score of 2.09. It was generally perceived as a useful feature for beginners and people without musical knowledge.

Use of mobile devices in music production

During the RBMA event in Montreal one trend was observed in all the interviewed participants, regarding the use of mobile devices in music production.

Two specific questions were asked to the participants during the short interviews.

1. Do you use mobile devices in your music productions? **a) Yes b) No c) Sometimes**
2. Do you think you will in the future? **a) Yes b) No c) Maybe**

The results from these questions can be seen in the table below:

Question 1	Yes	No	Sometimes
RBMA	0	9	2
ADE	3	4	5
Total	3	13	7

Question 2	Yes	No	Maybe
RBMA	3	0	8
ADE	4	0	8
Total	7	0	16

So even if the two events hosted different kinds of users it shows that currently the majority of users do not use mobile devices for music production, but they do envision that they will use them in the near future. This appears to be logical since mobile devices are becoming increasingly more powerful in terms of processing power, they come with an increasing number of sensors that can be used in musical applications, and are very suitable for multi-user performances.

2.5.2 In-depth Interviews

Prototype relates to Personas	Test Location	No. of Users	Documented in
Headliner DJ Apprentice Producer, Semi-Pro DJ Laptop/Home Studio DJ Tablet/Multi-device Producer Professional Producer Serious Modular Controller Hobbyist Music Technology Hacker Ground Breaker	RBMA Montreal	31	Appendix

During the two Montreal events, 31 in-depth interviews were conducted. The outcomes of these interviews were analysed alongside the datasets from other RBMA.

3 User Quotes from GiantSteps

Throughout the project we have conducted in-depth interviews with the expert users, we met through the RBMA events. A large task within WP2 has been the conducting of these interviews, and the even larger task to transcribe, identify quotes and then tag and sort these into the concerns, complaints, insights and ideas.

This process of analysis and sensemaking has informed our work throughout the project, but also fed into the process of writing and publishing outcomes in academic contexts, which has so far resulting in the following publications accepted in the third year:

- The Dial: Exploring Computational Strangeness. K. Andersen, P. Knees. ACM CHI 2016 (Poster)
- Attending to Objects as Outcomes of Design Research. T. Jenkins, K. Andersen, W. Gaver, W. Odom, J. Pierce, A. Vallgårda. ACM CHI 2016 (Workshop and position paper)
- Anti-Solutionist Strategies: Seriously Silly Design Fiction. M. Blythe, K. Andersen, R. Clarke, P. Wright. ACM CHI 2016 (Paper)
- Drumming with style: From user needs to a working prototype. S. Jordà, D. Gómez-Marín, Á. Faraldo, P. Herrera. NIME 2016 (Paper)
- Conversations with Expert Users in Music Retrieval and Research Challenges for Creative MIR. K. Andersen, P. Knees. ISMIR 2016 (Paper)
- The GiantSteps Project: A Second-Year Intermediate Report. P. Knees, K. Andersen, S. Jordà, M. Hlatky, A. Bucci, W. Gaebele, and R. Kaurson. ICMC 2016 (Paper)
- A Prototype for Exploration of Computational Strangeness in the Context of Rhythm Variation. P. Knees, K. Andersen. UMAP Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems 2016 (Workshop and position paper)
- Searching for Audio by Sketching Mental Images of Sound – A Brave New Idea for Audio Retrieval in Creative Music Production. P. Knees, K. Andersen. ACM ICMR 2016 (Paper)

As we have attended events and presented the work we have often been asked to publish the full interviews. However, this has raised a sticky point: Due to the strict ethics setup of the project and the terms under which our users have agreed to engage with the project, we are unable to publish the interviews. We can however publish statistics on the quotes from each of the four main interview events: RBMA Tokyo, RBMA Paris, testing event in Berlin and RBMA Montreal.

Over the three years we have seen the concerns and interests of these expert users shift slightly: Concerns about workflow remain high, while search and sorting of samples is still a major concern, it is mentioned less than it was in 2014. At the same time we see increasingly attention paid to collaboration, often over distance and specifically focussed on how to manage workflows in collaborative situations.

The collections of quotes we have gathered are of course subjective and even more importantly products of the time-frame and technological landscape that these specific users are conducting their work in, influenced by shifting musical styles, popular technological

solutions and the cultural context. But as a whole they illustrate and give flavour to the academic and commercial concerns in a project such as this.

4 Conclusions and Reflections

In the third year we have continued to involve our core users in the work process of the project. As the GiantSteps project draws to a close, the users have provided us with three kinds of input:

- Feedback on close to market or shipped products.
- Continued test results for prototypes and work-in-progress.
- Visionary ideas documented in papers and forming the basis for post-GiantSteps work.

As a result the work reported here is multifaceted and a combination of individual tests of components, exposure of outcomes and finally as always, conversations with users.

We started the year with the Berlin RBMA event, which was an opportunity to catch up with our expert users through interviews and test experimental prototypes.

During the year, we continued the prototype evaluation and development with local and web-based testing, alongside more dissemination-like events such as demos, hackathons and workshops. The chaotic and hard to document feedback gathered at such events has been very valuable for the project, users got to try out finished product and components, and we were able to see how they use them and learn from them.

We ended the year with our last RBMA, this time in Montreal, where we continued the process of exposing users to the outcomes of the project and interview them in-depth.

Overall the work has confirmed to us the importance of continuing in our efforts to engage with users. The inherent difficulty of gathering useful feedback and insights from expert users (who have much more important things to do) has been the main underlying focus of the work in this work package throughout the three years, and along the way we have developed some ways to do this. As the project ends, we hope that alongside the other outcomes, we may be able to refine and develop the methods for user engagement that have emerged during our work here.

5 References

[Amabile, 1996] T. M. Amabile, *Creativity in Context*. Westview Press, Boulder, CO, USA, 1996.

[Collins et al., 2016] T. Collins, P. Knees, and C. Coulon. MIR Web Interface for Shaping Musical Creativity, in *17th International Society for Music Information Retrieval Conference Late-breaking/Demo Session (ISMIR-LBD 2016)*, New York, NY, USA, July 7-11, 2016.

[Flexer 2015] A. Flexer. Improving visualization of high-dimensional music similarity spaces. *16th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 547–553, 2015.

[Frisson et al, 2014] C. Frisson, W. Yvart, N. Riche, X. Siebert, & T. Dutoit. A Proximity Grid Optimization Method To Improve Audio Search for Sound Design. *15th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 349–354, 2014.

[Frisson 2015] C. Frisson. *Designing interaction for browsing media collections (by similarity)*. Université de Mons, 2015.

[Grote 2014] F. Grote. Jamming with Machines: Social Technologies in Musical Creativity. In *Proc 8th midterm Conference of the European Research Network Sociology of the Arts*, 2014.

[Grote et al, 2015] F. Grote, K. Andersen, and P. Knees. Collaborating with intelligent machines: Interfaces for creative sound. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2015.

[Gómez-Marín et al, 2016] D. Gómez-Marín, S. Jordà, & P. Herrera. Rhythm Spaces. In *Proceedings of the 4th International Workshop on Musical Metacreation*.

[Hickey, 2001] M. Hickey. An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions in *Journal of Research in Music Education*, vol 49., no. 3, pp. 234-244, 2001.

[Jordà et al, 2016] S. Jordà, D. Gómez-Marín, A. Faraldo, P. Herrera. Drumming with Style: From user needs to a working prototype. In *Proceedings of the 16th International Conference on New Interfaces for Musical Expression.*

[Ó Nuanáin et al, 2016] C. Ó Nuanáin, S. Jordà and P. Herrera. An Interactive Software Instrument for Real-time Rhythmic Concatenative Synthesis. In *New Interfaces for Musical Expression*. Brisbane, Australia, 2016.

[Ó Nuanáin et al, 2016a] C. Ó Nuanáin, P. Herrera, and S. Jordà. An Evaluation Framework and Case Study for Rhythmic Concatenative Synthesis in *Proc. 17th International Society for Music Information Retrieval Conference*, 2016.

[Ó Nuanáin et al, 2016b] C. Ó Nuanáin, S. Jordà and P. Herrera. Towards User-Tailored Creative Applications of Concatenative Synthesis in Electronic Dance Music. *MUME 2016 - The Fourth International Workshop on Musical Metacreation*, 2016.

[Rouder et al, 2009] J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey and G. Iverson. *"Bayesian t tests for accepting and rejecting the null hypothesis,"* Psychonomic Bulletin & Review, vol. 16, no. 2, pp. 225-237, 2009.

[Turquois 2016] C. Turquois, M. Hermant, D. Gómez, and S. Jordà. Exploring the Benefits of 2D Visualizations for Drum Samples Retrieval in *ACM SIGIR Conference on Human Information Interaction and Retrieval*, pp. 329–332, 2016.

APPENDIX 1

GiantSteps Quotes Collection 2014 to 2016

GiantSteps Keyword index. Interviews 2014 to 2016					
Category	Keyword	Location			
		MTRL RBMA 2016	Berlin NI 2016	Paris RBMA 2015	Tokyo RBMA 2014
Number of interviews		31	39	20	16
Number of interviewees		31	28	20	16
SEARCH					
	Finding	0	1	0	8
	Looking for	0	0	2	4
	Searching	0	0	1	2
CATEGORIES					
	Sound Characteristics	1	3	0	0
VISUALISATION					
	Visualising Colours	0	1	4	6
	Waveforms	1	1	1	4
COLOUR					
	Images	0	0	1	0
	Personalise	0	0	1	0
ORGANISATION					
	Tagging	2	0	2	3
	Filtering	0	0	0	1
	Organising	0	3	2	5
	Frequency	0	0	0	2
ASSISTANCE					
	Help	0	0	0	5
	Learning	1	1	0	4
	Teaching	1	0	0	2

WORKFLOW					
	Workflow	22	17	5	5
	Annoying Tasks	0	0	5	4
CONNECTIONS					
	Rewiring	0	0	1	1
	Connecting	2	1	0	1
CORRECTION					
	Rhythms	0	0	4	4
	Quantize	1	0	2	2
	Correcting	0	1	0	0
SUGGESTIONS					
	Creativity	1	14	5	5
	Inspirations	3	4	1	2
OBSTRUCTIONS					
	Random	0	1	0	6
	Accidents	0	0	1	1
	Unexpected Strangeness	1	0	2	3
TWEAKS					
	Tweaks	0	0	0	3
	Change	0	0	0	1
INTERFACES					
	Intuitive Interfaces	9	6	3	4
LIVE					
	Performance	13	10	7	7