Towards User-Tailored Creative Applications of Concatenative Synthesis in Electronic Dance Music

Cárthach Ó Nuanáin, Sergi Jordà, Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
Barcelona
carthach.onuanain@upf.edu

Abstract

The ecosystem of concatenative synthesis systems is large, diverse and oftentimes esoteric. In this article, we examine many of its key works, attempting to summarise, compare and categorise their characteristic traits. We observe that many existing systems have not taken into account any user other than the designer of the system. Subsequently we position our own implementation in this ecosystem as geared specifically with electronic dance music producers in mind. Through intensive discussion with producers and commercial music software practitioners we summarise their reactions, responses and impressions to the usability and musicality of this approach to music creation. We report on active and future work based on the outcomes of these discussions.

Introduction

Sampling has long been a cornerstone of electronic music production, from the early days of *musique concrète* to hip-hop through to modern Electronic Dance Music (EDM) styles that make liberal use of sample packs and loops. Paradigms and practices for exploiting samples in music are thankfully becoming less labour-intensive and humdrum. The composer no longer needs to toil over expensive and time consuming manipulation of tape machines, hardware samplers (though there will always be continued usage for aesthetic purposes). Software sampling and editing is fast and well integrated in the workflows of every computer musician.

We assert that for sampling to advance further it needs to become more intelligent. It needs to know about what it is sampling and what it is being sampled for. Music Information Retrieval (MIR) is the branch of information retrieval that strives to extract meaningful information from digital music representation forms. Combining MIR advances with the musical intentions of sampling is at the heart of concatenative synthesis. Concatenative synthesis has its origins in speech synthesis (Hunt and Black 1996), where the descriptions need to be extremely accurate in producing effective reconstructions of natural speech. Music tends to be a bit more forgiving. Indeed, depending on the goals of the composer, accuracy is not always the intended thought in mind. Informally, we begin with some music, we deconstruct it, we describe it symbolically using some numerical descriptors and then put it back together in some new form. It is related to granular synthesis (Roads 2004), but operates on longer, more descriptive orders of scale.

In the next section we present a summary of many of the concatenative synthesisers presented in the literature. Some are scarcely described, some are fully-formed commercial systems, but most are confined to academic and experimental domains and at the mercy of the designer. This article argues for concatenative synthesisers that take into account the creative aspirations and needs of other users, composers and musicians and not just the inventor. A synthesiser is described that presents one realisation of this for what we feel are the desires of the modern EDM producer. We present this prototype to the user and through interactive interviews we gather qualitative appraisal and feedback of the system and its relevance to the user's own needs as an artist and producer.

The Landscape of Conctatenative Synthesis

Probably, CataRT by Diemo Schwarz (Schwarz et al. 2006) is the most widely cited example of musical concatenative synthesis, largely due to its distinct 2D timbre space visualising the range of explorable sounds. This 2D paradigm is echoed in other systems, notably in EarGram (Bernardes, Guedes, and Pennycook 2013) and AudioGarden (Frisson et al. 2014). Figure 1 shows examples of this timbre space concept from these systems. In CataRT (top), sound units are projected into space according to selected descriptors on the X/Y axes, with a third parameter being mapped to colour. EarGram (bottom left) positions sounds in space according to clustering strategies, including K-Means, QT-Clustering and DBSCAN. AudioGarden (bottom right), in comparison, offers two unique alternative approaches to mapping in 2D. The first of which, "Disc" mode, places units by assigning the length of the audio file to the radius of the unit from the centre, with the angle of rotation corresponding to a principal component of timbre (MFCCs). In the other "flower"

The Fourth International Workshop on Musical Metacreation, MUME 2016.
www.musicalmetacreation.org.

AUTHOR (YEAR)	Name	User Evaluation	Interaction & Visualisation	Rhythmic/Tempo
Schwarz (2000)	Caterpillar	No	2D explorer	No
Zils & Pachet (2001)	Musaicing	No	?	Yes
Hazel (2001)	Soundmosaic	No	Command Line	No
Hoskinson & Pai (2001)	Soundscapes	No	Waveform/Menus	No
Xiang (2002)	Granuloop	No	2D Controller	Yes
Kobayashi (2003)	Sound Clustering Synthesis	No	?	No
Cardle et al. (2003)	Directed Soundtrack Synthesis	Videos of use cases	Waveform/Menus	No
Lazier & Cook (2003)	MoSievius	No	?	Looping
Sturm (2004)	MATConcat	No	Waveform/Menus	No
Lindemann (2004)	Synful (Commercial)	Not available	Knobs/Sliders	No
Casey (2005)	Soundspotter	Retrieval Accuracy	Native Pure Data	No
Aucouturier & Pachet, (2005)	Ringomatic	User Experiment	Lists/Menus	Drumming Tool
Simon et al. (2005)	Audio Analogies	Algorithm Performance	None	No
Jehan (2005)	Skeleton	Algorithmic evaluation	Waveform/Menus	
Schwarz (2005)	CataRT	No	2D Timbre Space	Looping option
Weiss et al. (2009)	None	No	Waveform Segments	No
Frisson et al. (2010)	AudioGarden	No	2D Timbre Space / Waveform	No
Hackbarth (2010)	AudioGuide	No	2D CataRT based	No
O'Connell (2011)	None	Yes	Native Pure Data	Looping demo
Bernardes (2014)	EarGram	Author's impressions	2D Timbre Space and Matrices	Looping

Table 1: Summary of Concatenative Synthesis Systems

mode, a point of the sound is positioned in the space according to the average timbre of the entire sound file. Segments of the particular sound are arranged in chronological fashion around this centre point.

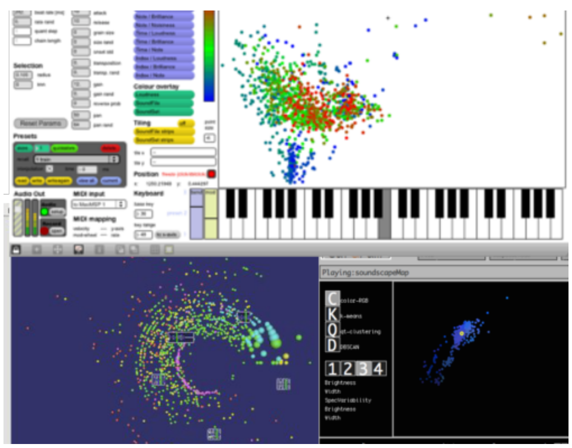

Figure 1: CataRT, AudioGarden and EarGram

In our view, these systems, with their characteristic 2D timbre spaces, represent the most recognisable encapsulations of visually-oriented concatenative synthesis. But digging deeper into the literature reveals a bewildering number of custom systems that address the broad aesthetic of concatenative systems in one way or another. (Schwarz 2005), (Sturm 2006) and later (Bernardes 2014), have gathered comprehensive descriptions of many of these systems. In Table 1 we have gathered together some of the most of the relevant ones together with some remarks and descriptions of each of the systems. These remarks reflect our impressions of how each system delivers in terms of its user-centricity and facility in creating elements of electronic dance music styles.

Zils and Pachet have carried out early work concatenative synthesis, and coined its synonymous portmanteau of music and mosaicing known as *musaicing* (in turn complementing image mosaicing) (Zils and Pachet 2001). Recognising that modern dance music styles such as techno can often entail heavily sampling, they also indentify that the manual selection and arrangment of samples is "difficult" and cumbersome. In essence, they are concerned with digitising the earlier efforts made known by composer John Oswald. Oswald's seminal paper "Plunderphonics, or Audio Piracy as a Compositional Prerogative" (Oswald 1985) equally set out his artistic intentions while also provocatively raising the natural legality issues attached. In any case, Zils and Pachet approach the problem much like Schwarz's earlier experimental contribution (Schwarz 2000) by treating this process of arranged sampling as a constraint satisfaction problem. Concatenated sequences are defined firstly by local *segment* constraints that affect qualities like mean pitch or loudness. These segment constraints are coalesced into global *sequence* constraints that influence higher level desires of the final sequence. Based on a computed global cost, this controls overarching qualities such as cardinality, continuity, and perhaps most importantly, tempo. Sturm's MATConcat (Sturm 2004) is also a frequently cited early implementation of the technique. It comprises an application written for the Matlab environment and operates by selecting feature thresholds for matching corpus sounds to the target sounds. For example, the user requests the system to find all frames of audio that lie within +/- 10% tolerance of the target for the spectral flux. The interface is quite involved, with user control over minute features such as window types and hop sizes. The website includes many audio examples and two full-length compositions.

There are a number of works that specifically address the area of rhythm. Xiang proposes Granuloop (Xiang 2002) for automatically rearranging segments of 4 different drum loops into 32 step sequence. Segmentation is done manually, without the aid of an onset detector, using Recycle [1]. Segmented sounds are compared using the inner product of the normalised frequency spectrum, supplemented with the weighted energy. These values become weights for a Markov-style probability transition matrix. Implemented in Pd, the user interacts by moving a joystick in a 2D space that affects the overall probability weightings determining which loop segments are chosen from. The system presents an interesting approach but is let down by its lack of online analysis. The FFT computation and comparison appears to be the bottleneck, and could be rectified by computing higher level features that return fewer data points. Ringomatic (Aucouturier and Pachet 2005) is a real-time agent specifically tailored for combining drum tracks, expanding on many of the constraints based ideas described by Pachet (Zils and Pachet 2001) previously. They applied the system to real-time performance following symbolic feature data extracted from a human MIDI keyboard player. A predominance of lower register notes in the performance, for example, applies a constraint that high-frequency heavy cymbal sounds should be concatenated. The drum track matches the overall energy levels of the human performer and a continuity constraint ensures subsequent sound units also follow a smooth loudness trajectory.

Concatenative synthesis has been considered useful in sound design tasks. Cardle et al. (Robinson, Brooks, and Cardle 2003) report on Directed Sound Synthesis as a means of providing sound designers and multimedia producers a method of automatically reusing and synthesising sound scenes in video. Users select one or more regions of an existing audio track, and can draw probability curves on the timeline to influence resynthesis of these regions (one curve per region) elsewhere. Soundscapes (Hoskinson and Pai 2001) in a nod to granular synthesis, refers to their segments as "natural grains" and seek to synthesise endless streams of soundscapes. The selection scheme by which segments are chosen is based on a representation of each segment as a transition state in a Markov chain. Its interface features knobs and sliders for controlling gain and parameters of multiple samples interactively. To evaluate the platform they conducted an additional study in 2007 (Hoskinson and Pai 2007) to reveal whether listening subjects found the concatenated sequences convincing compared to genuinely recorded soundscapes.

More specific and applied use cases include Hackbarth (Hackbarth 2011) who works intimately with the possibilities of concatenative synthesis in large scale music composition. He has worked with Schwarz to provide an alternative interface for exploring variations based on a force-directed graph. O'Connell describes a graphical system for Pd that demonstrate the use of higher level perceptual concepts like mood (happy vs sad) for informing selection in audio mosaics (O'Connell 2011).

[1] https://www.propellerheads.se/recycle

Commercial implementations also exist for concatenative synthesis. Of particular note is Loopmash [2], a software plugin and mobile application for automatically creating mashups from existing looped content. The interface consists of a number of tracks in a timeline arrangement. One track is set as a master, and slices in the master are replaced with matching slices from the other slave tracks. Users interact by manipulating "similarity gain" sliders that control the influence of each track in the slice selection algorithm. Other applications exist more as MIDI sampler systems attempting to model the performance qualities of natural sources such as orchestral ensembles [3] or the human voice [4].

Exploratory Rhythmic Concatenative Synthesis

Clearly, from our whirlwind tour of the many systems available and the summary in Table 1, there exist many approaches to concatenative sound synthesis, some solely theoretical, others extremely prototypical, withheld or unmaintained. Furthermore, most of those systems that provide some kind of interface to potential users don't give details on user evaluation. It is encouraging to see commercial implementations of concatenative synthesis in the wild, but the instances we reported on do not incorporate (in our view) one of the most exciting and promising aspects of the form: 2D timbre visualisation and exploration.

Our desire from the outset is to extend the experimental, exploratory possibilities of sampling-based concatenative synthesis to the needs of very specific creative users, namely electronic dance music (EDM) producers. To do this we needed to distil the important technical elements to something that is easy to use and easy to integrate, while remaining compelling visually and legitimate as a tool for music creation. We want it to be fun, but not a toy.

RhythmCAT (Ó Nuanáin, Jordà, and Herrera 2016) is a concatenative synthesis VST plugin specifically designed with the intention of assisting EDM producers to build rhythm patterns. Rhythm patterns are generated from units in a user provided corpus (referred to as a *sound palette*) of sound. Each sound intended for the sound palette is segmented using onset detection into smaller units. Features are then extracted from these units and their vectors stored in a data structure. The features we use are loudness, spectral centroid, spectral flatness and MFCCs. These features were chosen based on consultation of previous literature dealing with percussion sounds and identification tasks (Herrera, Dehamel, and Gouyon 2003), (Roy, Pachet, and Krakowski 2007) and (Tindale et al. 2004). The units of sound are arranged in such a way that they reflect the pattern and timbre characteristics of a target sound which we refer to as the *seed*. The seed sound is created by recording a four bar loop from the track on which it resides in the VST host. It is segmented using the host's tempo information and a specified

[2] www.steinberg.com/en/products/mobile_apps/loopmash.html

[3] http://www.synful.com

[4] http://www.vocaloid.com/en/

beat subdivision level (e.g. 1 bar to 1/64th notes). These target units are processed and analysed in a similar manner to the corpus units in the sound palette.

For each unit i in the segmented target sequence (e.g. 16-step) and each corpus unit j (typically many more), the concatenation *unit cost* $C_{i,j}$ is calculated by the weighted Euclidean distance of each feature k as given by Equation 1. a_k is a feature value from the target unit vector, b_k is the corresponding feature value from the corpus unit vctor with w_k providing a weighting for that feature value.

$$C_{i,j} = \sqrt{\sum_{k=1}^{n} w_k (a_k - b_k)^2} \tag{1}$$

These unit costs are stored in similarity matrix A of dimensions $m * n$, where m is the number of units in the corpus and n is the number of units in the target sequence. Next we create a matrix B of the indices of the ascendingly sorted elements of A. Finally a concatenated sequence can be generated by returning a vector of indices from this sorted matrix and playing back the associated sound file. To retrieve the closest sequence V_0 one would only need to return the indices of the first row (Equation 2).

$$V_0 = (b_{0,0}, b_{0,1} ..., b_{0,n}) \tag{2}$$

By selecting a random index from each column in the matrix, we can generate n^m possible sequences. By replacing m with an index threshold between 0 and m, the user can restrict the number of possibilities to ensure more similar sounding sequences.

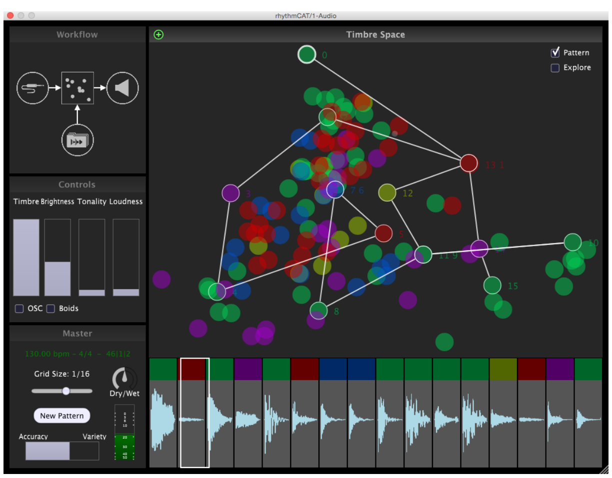

Figure 2: RhythmCAT User Interface

Visually (Figure 2) it is clear the system uses the familiar 2D timbre space concept we saw in previous state-of-the-art applications. Based on the feature vectors, each unit of sound is positioned accordingly using Principal Component Analysis. Where it distinguishes itself is in the noticeable grid layout connecting a subset of the units in space, implemented algorithmically as a linked list data structure. This graph represents a generated sequence of concatenated sound units modelled on the target sound. These units of sound can be manipulated and edited by the user selecting the graph edges and attaching them to other onsets as desired. The bottom portion of the interface represents the linear waveform representation of the current sequence. The weighting of the features can be adjusted by the sliders to the left of the interface. These weightings not only affect the influence of the features in Eq. 1 but also the influence over the PCA projection.

Expert User Reports

Testing Setup

In February 2016 we conducted several days of intensive interactive user discussions with a prototype of our system. The interviews took place at Universitat Pompeu Fabra in Barcelona and at Native Instruments Headquarters in Berlin. These interviews complement a more stringent evaluation of the system in terms of perceptual output and retrieval performance reported in (Ó Nuanáin, Herrera, and Jordà 2016). This evaluation found that the retrieval ability to of the system returned greater instances of correctly labelled polyphonic drum sounds with sequences that were more similar to the target using the metric and algorithm previously described. An accompanying listener evaluation revealed that participants' ratings correlated with our metrics in terms of similarity of pattern and timbre to the target. Additionally, the listener ratings found a correlation between closer patterns and subjective preference.

The profiles of the users could be divided into roughly four categories. In Barcelona, they were mostly researchers on one side, and students with a background in Sound and Music Computing on the other side. Nearly all of these participants worked with digital music in some form or other. In Berlin the users were drawn from Red Bull Music Academy (RBMA) associated artists based in the city as well as employees of Native Instruments. The RBMA was initiated in 1998 with the aim of gathering young music producers and DJs at worldwide events for the purposes of lectures, workshops and performances [5]. RBMA collaborate in the GiantSteps project (Knees et al. 2015) to provide access to these upcoming artists for research interaction and evaluation. These Berlin participants included producers, promoters, DJs, musicologists, product designers, engineers and graphic designers. United by the influence of the city, nearly everyone identified themselves as producers of techno and/or house, whether it be full-time or as a hobby.

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 were more than happy to oblige. Test stations were set up throughout the venue with a laptop, monitor and headphones as shown in Figure 3.

[5] http://www.redbullmusicacademy.com/

Figure 3: Testing Station Setup

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 statements 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?

Positive Reactions and Outcomes

Before delving into the specifics, we will first highlight the overall extremely positive feedback received from the participants. The word cloud in Figure 4 shows a culmination of some of the frequent positive descriptions participants attached to the system during the course of the tests.

Figure 4: Most Frequent Positive Descriptions

Some of the more detailed positive remarks give further insights into exactly why the system appealed to them:

> " 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. "

Many of the producers we spoke to reflected a particular trend in EDM at the moment for working with hardware and modular systems. This is often borne out of a desire to break out of the typical computer or digital audio workstation workflow and seek another path for inspiration.

> " I just jam for quite a while and then try to build something that I like and then bring it to computer and then add stuff from computer. You have to jam out really. The biggest issues come with recording. "

The most encouraging outcome from our studies conducted with these users was that the "interesting" and "different" design of our system offers these discouraged users a way "back in" for composing with the computer once again. This was suggested by comments such as:

> " I think something I've been looking to do in terms of experimentation and generating ideas melodically, is looking to go a bit more modular, use some modular stuff. To me, this is a digital form of it. "

> " Yeah. I use a lot of hardware, but if I'd use ... a few disco breaks or something or funk breaks that would be kind of nice, totally. "

Recurring Themes

We will now touch upon some of the common recurring themes that arose during the course of the interviews, and describe our own interpretations and plans to address them in future.

Usage Scenarios With respect to specific use cases, users provided some interesting scenarios where they could see the tool being used in their own interest. Numerous users were curious as to the ability to record and analyse live input such as instrumental performance or beatboxing for example.

> " 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? "

Live performance input was not something we had previously considered but is theoretically possible since the host would handle the capture of input audio. It would however require continuous re-analysis and computation of the similarity matrices which could be computationally costly. Still, other users have also expressed a desire for the possibility to continuously analyse incoming audio so it will be investigated.

Quite a number of users weren't interested in using the targeting capability of the synthesis engine whatsoever, and wondered if it was possible to start building patterns from scratch by selecting the onsets manually one by one. For example, referring to the fact that the dry signal is the original and the wet signal is the concatenated output:

> " 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. "

This is entirely possible, in fact, we had another "exploration" mode previously that gave the option to scan and audition the timbre space with a circular mouse radar that triggered the enclosed sounds, *CataRT* style. The motivation for this was to allow users to explore and audition the timbre space freely to identify regions of interest before proceeding to build their patterns. Merging this auditioning ability to create a sequence of patterns from zero would also stem the frustration of many of the users who wanted to create sounds straight away without capturing target input.

Traditional Navigation A very early outcome of the user testing was the realisation that although users were more than open to this new way of dealing with their sound, they still wanted a link to the familiar - the 2D waveform/timeline paradigm they are so used to dealing with in existing tools such as DAWs.

- "It's a bit hard to figure out which sixteenth you are looking for, because you are so used to seeing this as a step grid."
- "It's kind of good to see a different interface and not always follow the same timeline. .... But it could just be mirrored in a timeline"
- "You have a waveform or something... Then I know, okay, this is the position that I'm at."
- "Is there also a waveform place to put the visualisation? People are so used to having that kind of thing."

Initially this timeline was not something we had intended on offering; after all isn't it these paradigms we're seeking to break away from? After hearing these comments we realised that it was a useful option for the user and was one of the first items we implemented subsequently, as can be seen from Figure 2.

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 prototype, you can't expect to have all those functions there immediately, but in an end product, I think it would be a necessity. "

State and Feedback One of the consistent items that concern users of generative systems is the notion of *state*. On a superficial level state can refer to an effective preset management system that stores their efforts, for as one participant notes "you're always afraid of losing something.". Users are terrified of losing their progress once they've entered a pleasing "state", although this is a much bigger concern in probabilistic systems that produce - but then may never reproduce - "happy accidents".

At present our system has no state, and this was something frequently remarked upon and something we are actively considering. Users expressed a desire for complex state operations. Comparing two generated sequences visually and sonically for example, and being able to mix or find an interpolation between the two of them somehow. One artist wondered whether would it be possible to extend the graph visualisation technique to a space of "patterns". In this manner a series of stored patterns would be plotted in 2D space one after each other, and could be explored and sequenced in a more high-level arrangement or score-type interaction with the instrument. As he explained:

> " Even with this if it wasn't actually blending or interpolating between points, but just so you could save the state in a composition screen of dots and you could just jump between. "

This was actually inspired by the artist's own experiences with another experimental but commercial tool for music production: Audiomulch [6]. It includes a unique feature known as the *MetaSurface*, which allows navigation and interpolation of multiple parameter states by manipulating a colourful visual cluster space.

Also related to issue of state was the ability to initiate *feedback*, i.e. continuously assigning the concatenated output to the input, which once again can be done manually by recording to the host and re-recording in. This could be facilitated in the tool itself, but would require having the option of removing the matching sounds from the corpus itself to encourage diversity of sound.

[6] http://www.audiomulch.com/

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 sound with the best separation and shielding these parameters from the user. Potentially, a way forward could be remove these 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).

Conclusions

This paper argued for approaching concatenative synthesis from the musical user point of view, rather than the theoretical or singular composer standpoint. We discussed many existing examples of concatenative synthesis in the literature and some known commercial incarnations. Our system is offered as a user-centred implementation that uses some novel interface and interactive paradigms. In-depth user interviews provided much needed qualitative feedback on the state of the instrument, stimulating many ideas for improvements which we are actively pursuing.

Acknowledgments

We like to extend our thanks and appreciation to those participants who took part in the discussions, for their valuable time and expertise.

This research has been partially supported by the EU-funded GiantSteps project (FP7-ICT-2013-10 Grant agreement nr 610591).[7]

References

Aucouturier, J.-J., and Pachet, F. 2005. Ringomatic: A Real-Time Interactive Drummer Using Constraint-Satisfaction and Drum Sound Descriptors. *Proceedings of the International Conference on Music Information Retrieval* 412–419.

Bernardes, G.; Guedes, C.; and Pennycook, B. 2013. EarGram : An Application for Interactive Exploration of Concatenative Sound Synthesis in Pure Data. *From Sounds to Music and Emotions* 110–129.

Bernardes, G. 2014. *Composing Music by Selection: Content-based Algorithmic-Assisted Audio Composition*. Ph.D. Dissertation, University of Porto.

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

Frisson, C.; Yvart, W.; Riche, N.; Siebert, X.; and Dutoit, T. 2014. a Proximity Grid Optimization Method To Improve Audio Search for Sound Design. *15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (Ismir):349–354.

Frisson, C. 2015. *Designing interaction for browsing media collections (by similarity)*. Ph.D. Dissertation, Université de Mons.

Hackbarth, B. 2011. Audioguide : A Framework for Creative Exploration of Concatenative Sound Synthesis. *IRCAM Research Report*.

Herrera, P.; Dehamel, A.; and Gouyon, F. 2003. Automatic labeling of unpitched percussion sounds. In *Audio Engineering Society 114th Convention*.

Hoskinson, R., and Pai, D. 2001. Manipulation and Resynthesis with Natural Grains. *Proceedings of the International Computer Music Conference (ICMC)* 338–341.

Hoskinson, R., and Pai, D. K. 2007. Synthetic Soundscapes with Natural Grains. *Presence: Teleoperators & Virtual Environments* 16(1):84–99.

Hunt, A., and Black, A. 1996. Unit selection in a concatenative speech synthesis system using a large speech database. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings* 1:373–376.

Knees, P.; Andersen, K.; Jordà, S.; Hlatky, M.; Geiger, G.; Gaebele, W.; and Kaurson, R. 2015. Giantsteps - progress towards developing intelligent and collaborative interfaces for music production and performance. In *Multimedia Expo Workshops (ICMEW), 2015 IEEE International Conference on*, 1–4.

[7] http://www.giantsteps-project.eu

Ó Nuanáin, C.; Herrera, P.; and Jordà, S. 2016. A User-Centred Evaluation Framework for Rhythmic Concatenative Synthesis. In *International Symposium on Music Information Retrieval*.

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

O'Connell, J. 2011. Musical Mosaicing with High Level Descriptors. Master's thesis, Universitat Pompeu Fabra.

Oswald, J. 1985. Plunderphonics, or Audio Piracy as a Compositional Prerogative. In *Wired Society Electro-Acoustic Conference*.

Roads, C. 2004. *Microsound*. The MIT Press.

Robinson, P.; Brooks, S.; and Cardle, M. 2003. Audio and User Directed Sound Synthesis. *Proceedings of the International Computer Music Conference (ICMC)*.

Roy, P.; Pachet, F.; and Krakowski, S. 2007. Analytical Features for the classification of Percussive Sounds: the case of the pandeiro. In *10th Int. Conference on Digital Audio Effects (DAFx-07)*, 1–8.

Schwarz, D.; Beller, G.; Verbrugghe, B.; and Britton, S. 2006. Real-Time Corpus-Based Concatenative Synthesis with CataRT. *Proceedings of the 9th International Conference on Digital Audio Effects* 18–21.

Schwarz, D. 2000. A system for data-driven concatenative sound synthesis. *Digital Audio Effects (DAFx)* 97–102.

Schwarz, D. 2005. Current Research In Concatenative Sound Synthesis. In *Proceedings of the International Computer Music Conference*, 9–12.

Sturm, B. L. 2004. Matconcat: An Application for Exploring Concatenative Sound Synthesis Using Matlab. In *7th International Conference On Digital Audio Effects (DAFx)*, 323–326.

Sturm, B. L. 2006. Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Composition. *Computer Music Journal* 30(4):46–66.

Tindale, A.; Kapur, A.; Tzanetakis, G.; and Fujinaga, I. 2004. Retrieval of percussion gestures using timbre classification techniques. *Proceedings of the International Conference on Music Information Retrieval* 541–545.

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

Xiang, P. 2002. A New Scheme for Real-Time Loop Music Production Based on Granular Similarity and Probability Control. *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)* 89–92.

Zils, A., and Pachet, F. 2001. Musical mosaicing. *Digital Audio Effects (DAFx)* 1–6.