DOI: http://dx.doi.org/10.14236/ewic/EVA2017.34

Sonic Xplorer: A Machine Learning Approach for Parametric Exploration of Sound

Augoustinos Tsiros
Centre for Interaction Design
Edinburgh Napier University
10 Colinton Road, EH10 5DT
Scotland
a.tsiros2@napier.ac.uk

This paper presents Sonic Xplorer an interfaces that uses timbre adjectives for multiparametric control sound synthesis. The interface utilises an artificial neural network to create a personalised interface. Users can manipulate a large number of sound synthesis parameters without the need to learn or use the synthesiser's complex interface by utilising programed sounds by expert users. Sonic Xplorer learns a correlation based on users' ratings between timbre adjectives and the acoustic descriptors. Timbre adjectives are then used to describe the acoustic qualities of the desired sound. This paper discusses in detail the approach that has been followed to develop the system and the mapping and strategies users employed when using the interface in order to discover new sounds.

Sound synthesis. Multiparametric control. Artificial neural networks. Sound timbre. Semantic descriptors.

1. INTRODUCTION

The design of novel sounds using sound synthesis is a laborious process that requires a good understanding of (i) how the synthesis methods work, (ii) the special capabilities offered by the actual instrument being used, and (iii) how to configure the parameters of user interface in order to achieve the desired sonic output. A wide range of software tools for sound synthesis are available. These software tools have very elaborate capabilities for sound production and processing but they are also complex. The user interfaces of different sound synthesis tools vary significantly and they lack clear affordances (Couturier 2006, Hunt & Kirk 2000). The learning curve associated with each tool may be discouraging for users.

Furthermore, most commonly interaction with these tools takes place in the digital domain through Graphical User Interface (GUI). One issue with GUI controls is that they are usually implemented around a linear based model of interaction and the relationship between interface control and synthesis/processing parameter are direct one to one (Hunt & Kirk 2000, Hunt et al. 2003). This means that the user needs to manipulate the various interface controls provided by the application in a serial fashion in order to perform a set of actions. Although this model of interaction is highly suitable for some types of activities such as audio editing and general postproduction, there is evidence suggesting that serial interaction has a negative impact on creativity (Hunt et al. 2000; Hunt et al., 2003). Musicians, sound designers and users without technical expertise in a synthesis method can spend a considerable amount of time learning the different tools. Having to configure large numbers of parameters in a serial fashion before getting any desirable results hamper creativity, increases users' cognitive effort and potentially disrupts the creative process. In this paper we present *Sonic Xplorer*, a user interface that enables multiparametric control of sound synthesis parameters through the manipulation of a limited set of perceptually relevant sound parameters. The system utilises an artificial neural network. Through training, users can build a model by teaching the system a correlation between six adjectives and four different audio features. The model is employed to allow users to explore sound synthesis parameter space enabling high-level control of a large numbers of underlying sound synthesis parameters.

2. PREVIOUS WORK

Machine learning refers to a corpus of statistical methods that provide computers with the ability to

BCS Learning and Development Ltd.
Proceedings of Proceedings of EVA London 2017, UK

learn through the provision of training examples. Machine learning algorithms have been widely explored in the development of interfaces for musical interaction, to achieve tasks such as cross-modal analysis and multiparametric control of sound synthesis; for a review see (Caramiaux & Tanaka 2013, Fiebrink & Caramiaux 2016). The aim of the present investigation is to explore ways in which supervised machine learning algorithms could be utilised to reduce the large number of synthesis parameters used to program a sound to a limited set of perceptually relevant descriptors. The most closely related work to the one presented in this paper is Pardo et al. (2012), which uses machine learning to enable users to create personalised mappings, by teaching the system their subjective correlation between acoustic concepts (e.g., bright, warm, tinny) and audio equalisation gain curves. In contrast, the present system deals with sound synthesis that has a larger set of parameters in comparison with equalisation.

Previous works that used supervised machine learning techniques for the generation of mappings for continuous control of sound synthesis parameters, required that either the interface developer or the user provide training examples. These examples are used to form a model that establishes how input and output parameters values are associated. Most sound synthesis tools come with a large number of ‘presets’ installed, i.e., stored settings for creating various sounds. Presets are commonly designed by individuals that have high level of technical expertise and understand very well the sound synthesis method and given software. However, users that wish to use the synthesis tool do not necessarily have the same level of expertise, but still want to create interesting and novel sounds. The present approach aims to exploits the sounds programed by expert users in order to create new sounds.

3. SONIC XPLORER

3.1 Definition

This first implementation of *Sonic Xplorer* uses a neural network for exploration of the synthesis parameter space. The interface allows the user to discover new sounds in a controlled manner without the need for technical expertise. Sounds can be created by simply controlling six on-screen sliders. Each slider corresponds to one acoustic adjective, i.e., *Warm, Bright, Stable, Thick, Noisy, and Evolving* (see Figure 1). After providing the system with a few training examples, the only input required by the user is to control the six values that describe the qualities of the sound they want to synthesise. This minimal set of data allows sound practitioners to control many parameters (i.e., 237) but the process does not require the user to know how to control the underlying parameters.

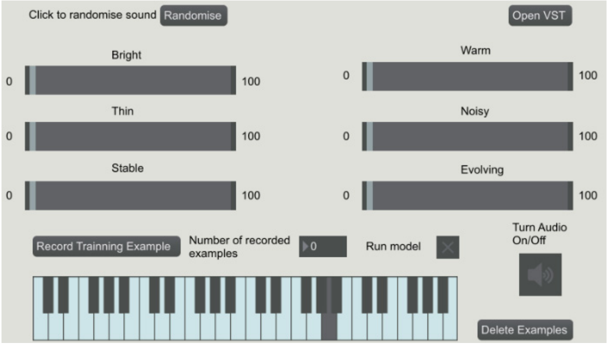

Figure 1: Sonic Xplorer user interface.

Our paradigm thus offer several distinct advantages to sound practitioners:

(i) Sound designers and musicians focus more on perceptually relevant sound attributes and less on low level attributes of the given synthesis methods.
(ii) Benefits from the custom build sounds (i.e., pre-programed sound settings) developed by expert users and enable users to obtain more sounds that lie between those presets.
(iii) This approach does not remove control of the low-level parameters, as the underlying synthesis parameters can still be accessed by the user.
(iv) Reduces the time and effort required to create new sounds.
(v) The user can learn how ‘discovered’ sounds are created by looking at the settings of the underlying parameters.

3.2 Implementation

The system has been developed in Max/MSP, utilising the Open Sound Control protocol to send data to Wekinator which is used for building the artificial neural network model. The Sylenth synthesiser was used in this first implementation of Sonic Xplorer. First, 40 pad sounds were recorded from the presets banks of the synthesiser. Each audio recordings had a duration 10 seconds. This duration was considered appropriate for pad type sounds, since this typology of sounds can have long attach and release times. Audio analysis was performed on each sound using the MIR toolbox (Lartillot et al. 2007) to extract the time-series of four descriptors, i.e., spectral brightness, spread, roughness and entropy using frame decomposition, with a frame length of 50ms and half overlapping.

Next the time-series data of the 40 sounds were standardised by using the minimum and maximum of all data points for each of the descriptors in order to adjust the values across the descriptors and fit them in a common scale ranging from 0 to 1. Each sounds' mean value was calculated for each descriptor (e.g., mean spectral brightness, entropy, roughness and spread for each sound). Then principal component analysis was performed, using the sounds' mean descriptor values as variables. The first and second component was used establish the location of each sound in a 2D spaces. The mean standard deviation between the descriptor dimensions of each sound was used to determine a radius that defines the area the sound occupies in the two dimensional feature spaces. Figure 2 shows an example of nine sounds represented by numbers (i.e., 1 to 9) plotted in the two-dimensional feature space based on the first and second components data points.

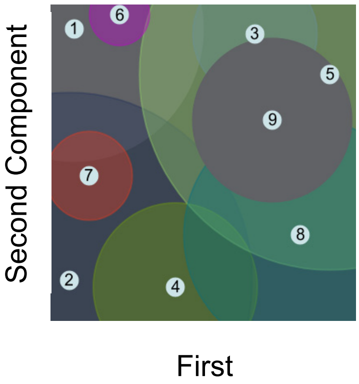

Figure 2: *shows an example of nine sounds plotted in a two-dimensional space based on the first and second components and occupying an area determined by their mean standard deviation across all four descriptors.*

Next the values for all underlying 237 synthesis parameters for each of the 40 preset sounds were extracted and stored using the *patterstorage* object in *Max/MSP*. Then the *recallmulti* function of the *patterstorage* object which permits weighted interpolation was used to enable morphing between the sound synthesis parameter configurations of the 40 sound presets. This led to the following type of mapping:

- If the user requests a sound that matches exactly the component values of a sound for instance sound 2 in figure 2, then the output of the weighted interpolation will produce the exact configuration of the synthesis parameters that create sound 2.
- If the user's input vector falls in an area of the feature space between two or more sounds, then the output synthesis parameters will be equal to the weighted sum of all the values of each preset.
- In the case where the users input vector falls between multiple presets, the distance between the actual mean of each preset across the four feature dimensions determines the weight of a particular preset, e.g., preset-1: 56%; preset-2: 18%; preset-3: 26%.

3.3 Training

The last step of the implementation consisted of employing an artificial neural network in order to correlate a set of high-level timbre adjectives to the feature space of the acoustic descriptors. A large body of research shows that musicians intuitively use adjectives such as bright or warm to describe musical timbre, (Disley et al. 2006, Fritz et al. 2012, Rioux & Västfjäll 2001, Stepánek 2006, Zacharakis et al. 2011). The terms that were selected (e.g., *Warm, Bright, Stable, Thick, Noisy, and Evolving*) derive from a previous research that examined the use of adjectives used to describe acoustic qualities of musical sounds. The process used to achieve the mapping between the adjectives and the acoustic descriptors extracted from the audio files is described below:

1. The system generates a series of random coordinates that correspond to a location in the two-dimensional feature space, the creation of which is described in section 2.2.
2. After listening to the sound, the user controlling six on screen sliders rates how well the example sound exemplifies the six verbal descriptors. This is the used as a training example.
3. Step 1 and 2 are repeated and on every repetition the model; the model gets better at understanding the correlation between the adjectives and the audio descriptors given that there is a consistency in the user responses.
4. After the user has provided several examples, the model is deployed by the user to control the sound synthesis parameters, based on the learned model of correlation between the user's timbre ratings and the underlying audio feature space.

For an overview of the system, see Figure 3.

4. EVALUATION

In order to evaluate the performance of the system, a participatory evaluation was conducted. The evaluation aimed to test:

- The effectiveness of the mapping.
- The appreciation of the user's ability to conceptualise their sonic ideas using the interface.

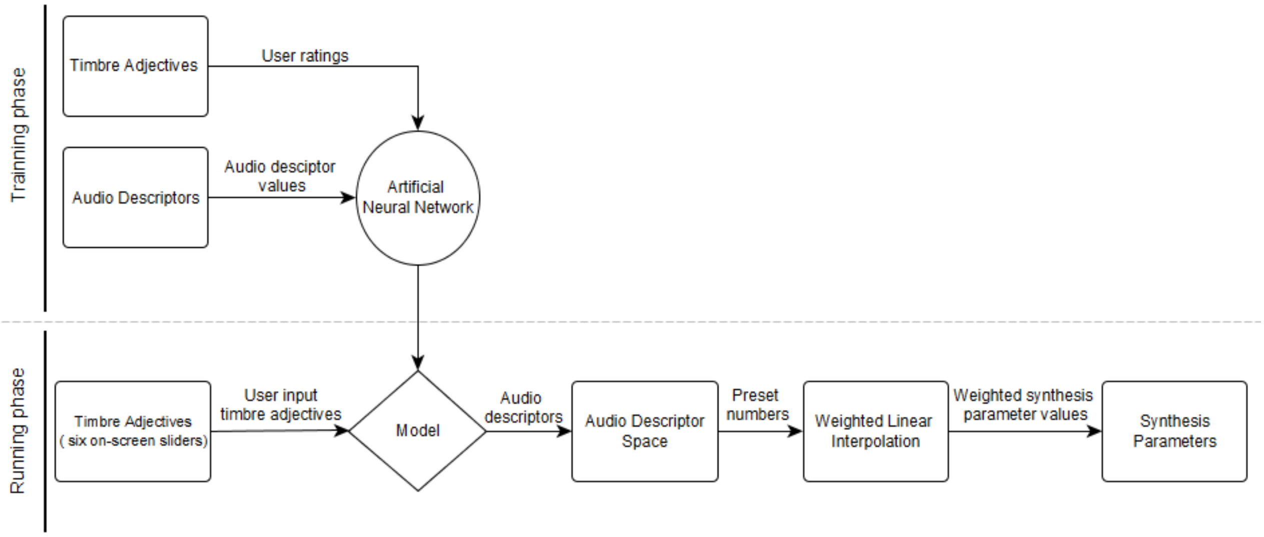

Figure 3: *System overview during training and running phases.*

- The participants' appreciation of this interaction paradigm.
- Finally, identify usability issues and gather suggestions for further development of the system.

4.1 Subjects

Four volunteer sound practitioners took part in the evaluation. All of the participants reported using analogue and digital equipment for sound synthesis, signal processing and audio sequencing.

4.2 Task

In this study, participants were asked to design two pad sounds using the sonic *Xplorer* interface for two different musical compositions of electronic music. Participants were told that in addition to using the Xplorer interface they could also open the synthesisers user interface to further refine their sounds by changing the settings of the synth, but only after they have discovered a sound they liked. After a short training session where a short demonstration of how to use the graphical user interface of the system was provided, participants freely used the graphical user interface in order to become familiar with the set-up. When participants felt confident using the interface, they were instructed to listen and rate 20 sounds across the six acoustic descriptors in order to build the model. After, users were asked to run the model and listen to the first musical and proceed with the tasks. After creating two sounds, they were asked to complete a questionnaire. The questionnaire consisted of seven Likert type (i.e., 1= strongly disagree, 5 = strongly agree) open-ended questions. The questions were aimed at assessing how users felt about the interface.

4.3 Results

Table 1 shows the results from the questionnaire that was administered after participants had used the system to accomplish the task. The first question aimed at assessing participant' satisfaction of the sound created using Sonic Xplorer. The participants' median response shows that overall they were satisfied with the sound they designed. The next three questions aimed to assess the perceived level of control over the sound parameters using the mapping and participants' understanding of how their input was affecting the output sound. Participants' responses suggest that there was a correlation between their inputs (values of timbre adjectives) and the output sound, but this correlation was not very strong. The responses also indicate that more precise control of the audio parameters would be desired.

The following three questions (i.e., 5–7) aimed at assessing the level of satisfaction and the creative potential of the system. Participants' responses indicate that Sonic Xplorer helped them discover new sound ideas that they might have not have thought of designing. Finally, participants agreed that Sonic Xplorer offers an interesting model for interaction with sound synthesis parameters and that it would be a useful addition to the sound synthesis tools they already use.

4.4 Discussion

Based on the results presented above, it can be concluded that overall Sonic Xplorer achieves a satisfactory level of performance. The level of control over the sound synthesis parameters through the manipulation of the six sliders

***Table 1:** Statistics of the Likert type questions that were answered by participants (1= strongly disagree, 5 = strongly agree).*

	Questions	Median
1	I am satisfied with the sound I designed using this mapping.	4
2	I felt there was a strong correlation between the values set in the control sliders and the sound that was synthesised by the system.	4
3	I felt I understood how the values I was setting were associated to the sound synthesis parameters.	3
4	I felt I could articulate my creative intentions using this mapping.	4
5	I felt that using the interface allowed me to discover new sounds that I might have not have come across without the aid from the system.	5
6	I believe that Sonic Xplorer offers an interesting approach to interacting with sound synthesis.	5
7	I believe that this interface would be a useful addition to other audio production tools I currently use.	5

corresponding to each adjective was very good, (see question 2 and 3). The responses above are interesting, as they suggest that even with 20 examples the mapping between the adjectives used for sound exploration were related to the audio features that were used for developing the mapping and enabled participants to describe the qualities of the sounds they wanted to synthesise.

Overall the performance of the mapping was satisfactory with regards to the ability of the participants to articulate their ideas using the interface and discover new sounds. However, most participants said that they would still want to be able to access the low-level parameters of the synthesiser. A brief interview with the participants after the study suggests that the users do not see this type of interface as removing control over the sound, instead they viewed this interface as an extension of the current interface of the synthesiser, for example:

> I do not think that this tool and the standard software interface are mutually exclusive, the opposite, the two can work well together. For instance, you want to create a sound of a certain type you configure the settings accordingly in the Xplorer once you find something you like you bring up the synth and you tweak the parameters further until it meets the requirements of your project.

> I really enjoyed exploring the sounds using the system. I think it can be used as a learning tool as well, where you discover a sound you like and then you open the synth to see what the settings for constructing that sound are.

According to the results, it seems that all of the participants thought that the interface is useful and that they consider that similar tools could be a good addition to the tools they currently use for sound synthesis (see questions 6 and 7). This is also supported by participants' comments, for example:

> Certainly, a novel idea for the control of sound synthesis.

> I can see how this approach to the control of parameters could be employed in many different types of interfaces not only sound synthesis, for example sound compression or equalisation.If I could easily do something like the sonic Xplorer does for every software synth I own, it would certainly speed up my workflow and perhaps there would be little need for purchasing presets.

Finally, all participants were asked to describe what strategy they followed to accomplish the experimental task, i.e., how they approached using the interface to design the pad sounds. Looking at the participants responses, we can distinguish between two strategies; the first is quite deterministic, while the second is more exploratory. For instance, participants reported that they first listened to the music track provided in order to identify a set of requirements for the acoustic features of the sound which would fit the music. Then they attempted to map those requirements to the high-level controls provided by the Sonic Xplorer interface, see comments below:

> I listened to the music tracks that you provided. The first recording had a lot of low end and there was a lot of space in top end of the frequency spectrum, hence, I was looking for a bright sound with more energy in the higher end of the spectrum, hence I moved the sliders to bright and low spread to find a sound that would fit. I moved the other sliders about to see how that affects the sound. Once, I found a sound the form of which I liked I opened the synth and tweaked it even further.

> After listening to the recordings you provided, I thought what type of pad would fit and had a look at the six sliders, tried to match my idea to the four parameters. Then I had a listen at the sound, and started to move the sliders around playing different keys on the keyboard until I found a sound I thought would fit with the music track.

Other participants used the interface without any preconceptions of what they need, purely as a sound explorer, see comments below:

> I looped the audio and I moved the sliders around hitting different chords to listen to the sounds the system would generate.
>
> I simply moved the sliders and listening to the results until I found sounds I liked.

5. CONCLUSIONS AND FUTURE WORK

We have presented a first implementation of the Sonic Xplorer interface and an approach to construct mapping between a large set of sound synthesis parameter values to a smaller set of perceptually relevant timbre related acoustic descriptors. The technique utilised can be employed in a wide range of sound synthesis and audio processing interface to create mappings that do not require users to necessarily understand the technical aspect of the given audio process, but instead let the users express their conception in descriptive terms.

We reported on the evaluation of the interface with sound practitioners. Their responses suggest that they view the interface as a useful addition to the tools they currently use. Furthermore, the evaluation indicates that the interface supports well the discovery of new sounds, and could be beneficial in accelerating the workflow of sound practitioners. It also highlighted the importance of balancing between the ability for sound exploration/ discovery and the level of control which the system supports. Finally, participants' responses suggest a similar interaction paradigm to the one employed in the present system could be applied to a wide array of creative signal processing applications.

Future work includes comparing different machine learning techniques for continuous control of parameters, for instance artificial neural networks and polynomial regression. Applying similar techniques to different types of creative applications for signal processing will be explored. Another direction for future research involves asking users to choose descriptive terms that best exemplify a particular typology of sounds, such as bass, leads, pads, and use these for the user interface. Finally, in order to create a more robust model that will be more representative of a range of users, multiple participants will be asked to provide training examples to enable a correlation between descriptive terms, acoustic descriptors and synthesis parameters.

6. REFERENCES

Caramiaux, B. and Tanaka, A. (2013) Machine Learning of Musical Gestures. Proceedings of the International Conference on New Interfaces for Musical Expression 2013 (NIME 2013), pp. 513–518.

Couturier, J. (2006) A model for graphical interaction applied to gestural control of sound. Sound and Music Computing Conference.

Disley, A. C., Howard, D. M., and Hunt, A. D. (2006) Timbral description of musical instruments. In International Conference on Music Perception and Cognition.

Fiebrink, R. and Caramiaux, B. (2016) The Machine Learning Algorithm as Creative Musical Tool. Handbook of Algorithmic Music.

Fritz, C., Cross, I., Woodhouse, J., and Woodhouse, J. (2012) Exploring violin sound quality : Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. *The Journal of the Acoustical Society of America*, 131, pp. 783–794.

Hunt, A., and Kirk, R. (2000) Mapping strategies for musical performance. In Trends in Gestural Control of Music (pp. 231–258).

Hunt, A., Wanderley, M., and Kirk, R. (2000) Towards a Model for Instrumental Mapping in Expert Musical Interaction. In International Computer Music Conference.

Hunt, A., Wanderley, M., and Paradis, M. (2003) The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, 32(4), pp. 429–440. DOI: 10.1076/jnmr.32.4.429.18853

Lartillot, O., Lartillot, O., Toiviainen, P., and Toiviainen, P. (2007) A matlab toolbox for musical feature extraction from audio. International Conference on Digital Audio FX, (Ii), pp. 1–8.

Pardo, B., Little, D., and Gergle, D. (2012) Towards Speeding Audio EQ Interface Building with Transfer Learning. In NIME 2012 Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 145–148.

Rioux, V. and Västfjäll, D. (2001) Analyses of verbal descriptions of the sound quality of a flue organ pipe. *Musicae Scientiae*, 6(1), pp. 55–82.

Stepánek, J. (2006) Musical Sound Timbre : Verbal Description and Dimensions. In 9th Int. Conference on Digital Audio Effects, pp. 121–126.

Zacharakis, A., Pastiadis, K., Papedelis, G., and Reiss, J. D. (2011) An Investigation of Musical Timbre : Uncovering Salient Semantic Descriptors and Perceptual Dimensions. In 12th International Society for Music Information Retrieval Conference, pp. 807–812.