EVALUATION OF A SKETCHING INTERFACE TO CONTROL A CONCATENATIVE SYNTHESISER

Augoustinos Tsiros, Grégory Leplâtre
Centre for Interaction Design
Edinburgh Napier University
10 Colinton Road, EH10 5DT
a.tsiros@napier.ac.uk,g.leplatre@napier.ac.uk

ABSTRACT

This paper presents the evaluation of Morpheme a sketching interface for the control of sound synthesis. We explain the task that was designed in order to assess the effectiveness of the interface, detect usability issues and gather participants' responses regarding cognitive, experiential and expressive aspects of the interaction. The evaluation comprises a design task, where participants were asked to design two soundscapes using the Morpheme interface for two video footages. Responses were gathered using a series of Likert type and open-ended questions. The analysis of the data gathered revealed a number of usability issues, however the performance of Morpheme was satisfactory and participants recognised the creative potential of the interface and the synthesis methods for sound design applications.

1. INTRODUCTION

Morpheme[1] is a sketching interface for visual control of concatenative sound synthesis (see [1]) for creative applications. In recent years a number of user interfaces have been developed for interaction with concatenative synthesis [2]–[5]. Furthermore, although sketching has been widely explored as a medium with interaction with sound synthesis and musical composition (see [6]–[10]) there have been very few attempts to evaluate the usability of such interfaces. Additionally, Morpheme is in our knowledge the first attempt ever made to use sketching as a model of interaction for concatenative synthesis.

The way concatenative synthesis works is different to that of most conventional sound synthesis methods. Unlike other synthesis methods were the sound is represented by low-level signal processing parameters which can be controlled in a continuous manner, in concatenative synthesis, sounds are represented using sound descriptors related to perceptual/ musical parameters, and sounds are synthesised by retrieving and combining audio segmented from a database. Although this is a very interesting way of synthesising audio, it can lead to unexpected results, particularly for users that are not familiar with this type of sound synthesis. For example, while in other synthesis methods, increasing the amplitude parameter results in changes only to the parameter that was controlled, in the context of concatenative synthesis requesting a sound of greater or smaller amplitude may result in selecting different audio units that have very different timbre characteristics. These sudden/discreet changes could potentially confuse practitioners that are not familiar with this synthesis method.

The aims of the study presented in this paper are the following:

1) Evaluate Morpheme's graphical user interface: detect usability issues and identify desired functional requirements.
2) Evaluate the mapping between the visual features of the sketches and the control parameters of the concatenative synthesiser
3) Assess whether the audio used in the corpus affects the perceived level of control of the interface, the appreciation of the system and the mapping.

2. MORPHEME

2.1 Graphical User Interface

Figure 1. shows a screenshot of *Morpheme's* main graphical user interface. We could distinguish between four main interface components in the second version of *Morpheme's* interface, the canvas, the timeline, the playback controls, the brush controls and the video display.

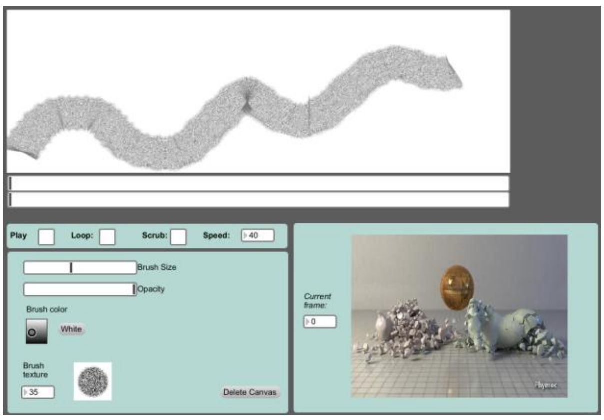


Figure 1. Morphemes' main graphical user interface.

[1] Download Morpheme: https://inplayground.wordpress.com/software/

The playback controls provide a number of function (see **Figure 2**) including:

- **Play**: starts the analysis of the sketch which results in the data used to query the database and drive the sound synthesis engine.
- **Loop:** repeats the entire length of timeline when the cursor reaches at the end of the timeline.
- **Scrub:** functions freezes the cursor in a given location of the timeline. Dragging the cursor of the timeline can move the analysis window through the sketch to a desired position.
- **Speed:** allows the user to determine the speed (in milliseconds). The speed controls the rate at which the analysis window moves from left to right though the timeline.

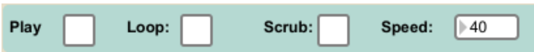
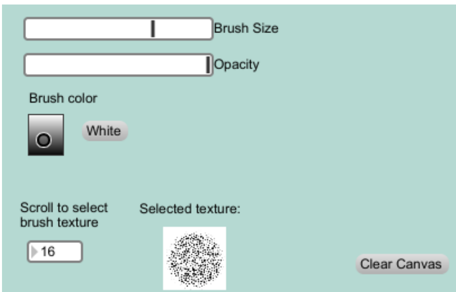


Figure 2. Screenshot of the user interface playback controls.

Brush Controls provide a number of function (see **Figure 3**) including:

- **Brush size:** size of the brush
- **Opacity:** opacity of the textured brush.
- **Brush color:** color of the textured brush.
- **White:** control can be used as an eraser.
- **Brush selection:** by clicking and scrolling on the number box users can select from 41 different textured brushes.
- **Clear Canvas:** erases the sketch from the canvas.
- 

- **Figure 3**. Screenshot of graphical interface for the control of the brush parameters.

2.2 System Architecture

Figure 4 illustrates the architecture of Morpheme. During playback windowed analysis is performed on the greyscale version of the sketch. A window scans the sketch from left to right one pixel at every clock cycle, the rate of which is determined by the user. Only the areas of the canvas that are within the boundaries of the window area are subjected to statistical analysis. The window dimensions are determined by Window width by window height. The window width can be determined by the user, however the default size of the analysis window is 9 pixel wide by 240 pixel height. The analysis of the canvas' data matrix results in a four dimensional feature vectors that describes the visual attributes of the sketch and which is used as the target for querying audio-units from the CataRT's database.

2.1.1 Mapping Visual to Audio Features for Selection and Processing of Audio Units

In the current implementation of *Morpheme,* we can distinguish between two mapping layers. The first layer consists of a mapping between visual and auditory descriptors for the selection of audio units, see **Table 1**. The second layer consists of a mapping that associates the distances between audio and visual descriptors to the synthesis parameters, see **Table 2**.

Visual Features	Audio Features
Texture compactness	Spectral flatness
Vertical position	Pitch
Texture entropy	Periodicity
Size	Loudness
Horizontal length	Duration

Table 1. Associations between audio and visual descriptors.

Audio features	Synthesis parameters
Spectral flatness	Transposition randomness
Periodicity	Grain size and amplitude randomness
Pitch	Transposition
Loudness	Amplitude

Table 2. Mapping the distances between audio and visual feature vectors to synthesis parameters.

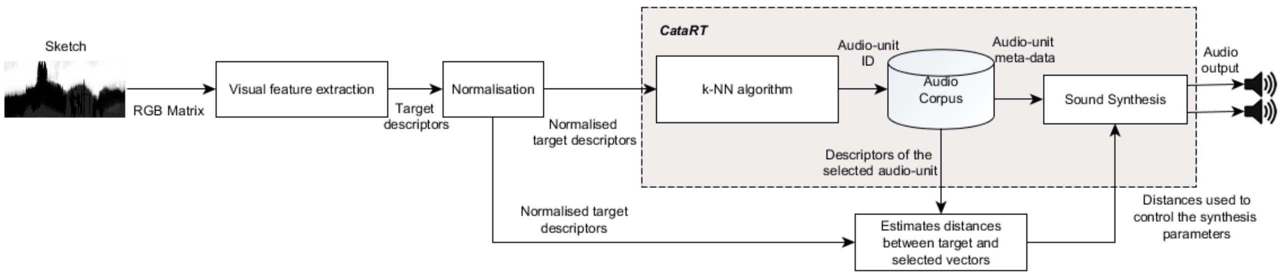


Figure 4. An overview of the architecture of Morpheme

3. MORPHEME EVALUATION

3.1 Participants

One group was recruited that consisted of eleven musician/sound practitioner volunteers. All of the participants played a musical instrument and the self-reported level of expertise was five intermediate and six advanced. Seven ofthe participants had received formal music theory training at least for six months. All of the participants reported using analogue and digital equipment for sound synthesis, signal processing and sequencing. Four participants self-reported a level of expertise regarding the use of digital and analogue equipment as intermediate and seven reported advanced skills. None of the participants in this study reported having hearing or visual impairments. All participants had first participated in the experiments described in the chapters five and six prior to taking part in the present one. All participants were male while the age group ranged from 18 to 64.

3.2 Apparatus

The experiments took place in the Auralization room at the Merchiston Campus of Edinburgh Napier University. Participants used Beyer Dynamics DT 770 Pro monitoring headphones with 20db noise attenuation to listen to the audio stimuli. An HP ENVY dv7 laptop with 17.3 inch screen was used. For sketching on Morpheme's digital canvas a bamboo tablet was used. However participants were allowed to use a computer mouse if they preferred. SurveyGismo was used to record the participants' responses after the sound design task was completed.

3.3 Procedures

In this study participants were asked to design two soundscapes using the *Morpheme* interface for two video footages. Subject responses were collected independently. In each session a single participant completed the following tasks. Participants were given a brief description of the task followed by a short demonstration of Morpheme's graphical user. After a short training session were participants were shown how to use the graphical user interface of Morpheme in order to synthesize sounds, participants were instructed to proceed with the tasks. There were two eight minutes sessions (one for each video footage) during which participants were free to produce a soundscape that best suited the video using Morpheme. At the end of the sessions, participants were asked to complete a questionnaire. The questionnaire consisted of 15 Likert type (i.e. 1= strongly disagree, 5 = strongly agree) an open-ended questions. The questions aimed at assessing experiential, cognitive and expressive aspects of the interaction as well as to detect usability issues and gather ideas regarding usability improvements of the interface.

3.4 Materials

3.4.1 Video footages

Two videos have been selected for this task. The first video footage has been captured in Bermuda during the recent hurricane Igor, see **Figure 5** top row. The duration of the hurricane video is one minute. The camera shots included in the video have been captured from several locations during the hurricane. The second footage is a 3d animated scene that last for 4 seconds which represents a simulation of two porcelain objects been shattered on a tilled floor, see **Figure 5** bottom row. Both video footage require a relatively high precision in the way the sound is synced to the video sequence. However the second video sequence is slightly more challenging in this respect in comparison to the hurricane scene.

Figure 5. Four screenshots from the two video footage used in the study.

3.4.2 Audio Corpus

The audio corpus that participants had to use to synthesize the sound effects for the shattering scene consists of four audio recordings of glass shattering events. The corpus that is used to synthesize the soundscape for the hurricane scene consists of four audio recordings of windy acoustic environments. All eight audio files have been segmented to audio-units with durations of 242 milliseconds. The selection of the audio files used to prepare the two corpus was predominately determined by the theme of the video footage. However these two videos were selected to allow testing the mapping in two very different auditory contexts. For example the shattering scene requires a corpus that consists of sounds that are relatively dissonant, non-periodic, and abrupt such as impact/percussive sounds. The second hurricane scene requires a corpus that contain moderately harmonic, slightly periodic and continuous sounds.

4. RESULTS

The first question aimed at assessing participant satisfaction of the sounds created using Morpheme, see **Table 3**. The participants' average response shows that they were neutral regarding this question. Participants' responses show that there was a strong correlation between the user input (i.e. sketch) and the outcome sound, and that it was easy to understand the mapping. Although the degree of correlation was not as strong at all times. Participants' responses indicate that Morpheme's sketching interface helped them articulate their sound design ideas in visual terms, and that they felt they had control over the sound synthesis parameters. However the responses also indicate that more precise control of the audio parameters would be desired. Participants felt equally in control using either corpus (i.e. wind and impacts) while there was indication that there was a stronger preference in working with the impacts corpus. Finally, participants agreed that Morpheme 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. An analysis of the data gathered by the open ended questions was performed manually. Every time a new theme was encountered in the answers, it was used to form a new *category*. Then the frequency of these categories was recorded to identify which the most prominent issues and desired technical features. The usability improvements identified are summarized in **Table 4**

5. DISCUSSION

Based on the results presented above, it can be concluded that overall Morpheme achieves a satisfactory level of performance. The subjective level of control of the sound parameters through sketching, and the participants' level of satisfaction with the sounds they designed was average. These results might be attributed to three factors. The first factor is the user's unfamiliarity with sketching as a model of interaction with sound synthesis parameters. The second factor might be their unfamiliarity with the way concatenative synthesis works. This view is further supported from the average responses (M=3 SD=1) to the question 'I felt confused in several occasions about how my drawing affected the audio output'. This is also reflected in some of the user comments, for example:

"Unpredictable results at times",

"It wasn't always easy to be precise",

"It was complicated at times to identify the correlation between the pitch and the type of sounds played".

As it was mentioned earlier in Section 3.2 the information that was provided to the participants prior to the experimental task was mainly about how to use the interface. Minimal information was provided about the synthesis method. This decision was made primarily to avoid the development of positive biases towards the system due to enthusiasm about the way the system synthesises sound. The third factor might be related to the usability issues identified.

Overall, the perceived correlation between the visual and sound features were satisfactory. Participants' responses showed that Morpheme is easy to use, offers an interesting approach to interacting with sound synthesis and supported that the interface helped them think about sound in visual terms. Furthermore, the majority of participants thought that Morpheme would be a useful

Suggested user interface improvements	N
Image processing tools for refinement of the sketch	1
Timestamps navigation of the timeline	2
Edit the position of graphics based on timestamps	1
Larger canvas	8
Canvas zoom-in function	6
Temporal looping function based on user defined loop points	1
Undo function	1
Latency between graphics and audio timeline	5
Non-linear sketch exploration	1
Enable layering of multiple sounds/sketches and ability to shift between layers	1

N=Number of participants

Table 4. Participants' answers to the question: What changes to the User Interface would you suggest to improve it?

	Questions	Mean	STD
1	I am satisfied with the sound I designed using this mapping.	3	0.85
2	I felt there was a strong correlation between the sketch and the sound that was synthesised by the system.	4.18	0.38
3	I felt I understood how attributes of the sketch were associated to attributes of the sound.	4.54	0.65
4	I felt I could articulate my creative intentions using this mapping.	3.9	0.51
5	I felt I had control over the synthesis parameters while using the system.	4.18	0.57
6	I am satisfied with the level and precision of the control I had over the audio parameters while using the system.	3	0.85
7	I felt confused in several occasions about how my drawing affected the audio output.	3	1.04
8	Overall, I am satisfied with Morpheme's Graphical User Interface.	4	0.42
9	I believe that Morpheme offers an interesting approach to interacting with sound synthesis.	4.81	0.38
10	I believe that Morpheme would be a useful addition to the audio tools I currently use.	4.45	0.65
11	I felt Morpheme helped me think about sound in visual terms.	4.27	0.86
12	I felt equally in control while using the two sound corpora.	3.54	0.65
13	I felt frustrated about certain aspects of the interface/interaction.	2.9	0.79
15	I felt that Morpheme was complicated and difficult to use.	1.9	0.5

Table 3. Statistics of the Likert type questions that were answered by participants (1= strongly disagree, 5 = strongly agree).

addition to the audio tools they currently use. Participants responses were not conclusive as to whether the corpora that was used affected their perceived level of control over the system as participants response was (M= 3.5 SD=0.6), while seven out of eleven participants seem to prefer working with the impacts corpus, three preferred the wind corpus and one neither. One of the differences between the impacts and the wind corpora is that the former is much larger. Based on the findings from the evaluation it appears that a larger corpus can result in both positive and a negative effects. Some of the negative effects became evident from some of the participants comments discussed above such as more unpredictable results, because the probabilities of getting a sequence of audio-units with very distinct timbre is higher when there is a large nonhomogeneous corpus (e.g. impacts corpus used for the evaluation) than when a small and homogeneous corpus (such as the wind corpus) is used. Furthermore, it is worth noting that participants' were moderately satisfied with the sounds which they designed using the system (M=3 SD=0.8).

Many usability issues were also revealed, mainly related to the lack of standard controls found in other image processing applications (e.g. photoshop) such as zooming in and out, resize canvas and undo function. Further, participants also pointed out the lack of other functions that tend to be standard functionality in time-based media production applications such as setting loop and cue points on the timeline, having a precise transport panel and a sequencer were sounds can be layered. Moreover, several participants complained about latency between the timeline and the output sound. Latency depends on two factors: the size of the audio corpus (i.e. how many audio units are stored in the corpus) and how many comparisons the algorithm has to perform until it finds the audio-unit that its features best match the target. Another factor that might cause the perception of latency is that in the present version of morpheme, the current position of the analysis window is indicated by a slider that does not reflect well the actual position of the window, see top image in **Figure 6**. The problem is that the window is 9 pixels wide while the current cursor used to represent the position of the analysis window suggest that the window is smaller. A better solution would be to use a cursor as shown in **Figure 6** bottom.

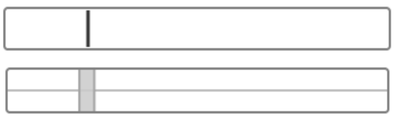


Figure 6. The top figure shows the current visual feedback for the representation of the position of the analysis window. The bottom figure shows a more precise visual feedback.

6. CONCLUSIONS

The evaluation of Morpheme showed that the performance of Morpheme was satisfactory and participants seemed to recognise the creative potential of the tool. From the analysis of the results, we could distinguish between two types of issues. The first type were issues related to the user interface. Most of the usability and functionality features that the participants noted could be relatively easily addressed with the implementation of standard controls found in other time-based applications, or in more advanced drawing packages. The second type were issues related to the type of sound synthesis used by the application (i.e. target based automatic selection synthesis using low and high level descriptions). Some of issues involved the unexpected transition between audio-units that sounded very different, which gave participants the impression of lack of control. In order to create sounds that are plausible variations of the original audio used in the corpus a degree of awareness not only of the micro but also of the meso and the macro levels of the sound is required. The issues identified through this evaluation will form the basis for future development of the *Moprheme* interface.

Acknowledgments

The authors would like to acknowledge the IRCAM's IMTR research centre for sharing the *CataRT* system.

7. REFERENCES

[1] S. B. Diemo Schwarz , Grégory Beller , Bruno Verbrugghe, "Real-Time Corpus-Based Concatenative Synthesis with CataRT," in *Digital Audio FX*, 2006, pp. 279–282.

[2] D. Schwarz and B. Hackbarth, "Navigating variation: composing for audio mosaicing," in *International Computer Music Conference*, 2012.

[3] M. Savary, D. Schwarz, and D. Pellerin, "Dirty tangible interfaces: expressive control of computers with true grit," *CHI'13 Ext.* , pp. 2991–2994, 2013.

[4] J. Comajuncosas, "Nuvolet: 3d gesture-driven collaborative audio mosaicing," *New Interfaces Music. Expr.*, no. June, pp. 252–255, 2011.

[5] E. Tomás and M. Kaltenbrunner, "Tangible Scores: Shaping the Inherent Instrument Score," *Proc. Int. Conf. New Interfaces Music. Expr.*, pp. 609–614, 2014.

[6] G. Levin, "The table is the score: An augmented-reality interface for real-time, tangible, spectrographic performance," in *proc. of DAFX - Digital Audio Effects*, 2006, pp. 151–154.

[7] M. Farbood, H. Kaufman, H. Line, and K. Jennings, "Composing with Hyperscore: An Intuitive Interface for Visualizing Musical Structure," in *International Computer Music Conference*, 2007, pp. 111–117.

[8] P. Nelson, "The UPIC system as an instrument of learning," *Organised Sound*, vol. 2, no. 01, 1997.

[9] J. Garcia, T. Tsandilas, C. Agon, and W. E. Mackay, "InkSplorer : Exploring Musical Ideas on Paper and Computer," *Proc. Int. Conf. New Interfaces Music. Expr.*, pp. 361–366, 2011.

[10] J. Thiebaut, "Sketching music: representation and composition," Queen Mary London, 2010.