

CABOTO: A Graphic-Based Interactive System for Composing and Performing Electronic Music

Riccardo Marogna
Institute of Sonology
Royal Conservatoire in The Hague
Juliana Van Stolberglaan 1
2595 CA The Hague, Netherlands
riccardomorgana@gmail.com

ABSTRACT

CABOTO is an interactive system for live performance and composition. A graphic score sketched on paper is read by a computer vision system. The graphic elements are scanned following a *symbolic-raw* hybrid approach, that is, they are recognized and classified according to their shapes but also scanned as waveforms and optical signals. All this information is mapped into the synthesis engine, which implements different kind of synthesis techniques for different shapes. In CABOTO the score is viewed as a cartographic map explored by some navigators. These navigators traverse the score in a semi-autonomous way, scanning the graphic elements found along their paths. The system tries to challenge the boundaries between the concepts of composition, score, performance, instrument, since the musical result will depend both on the composed score and the way the navigators will traverse it during the live performance.

Author Keywords

Graphic score, optical sound, sound synthesis

1. INTRODUCTION

In a previous work [13] I developed a graphic notation system for improvisation, called *Graphograms*. In that system, a graphic vocabulary was organized in a graph-like structure, and the musicians could choose, under certain rules, their own path through it. The idea was to give them enough freedom to express their ideas, keeping an overall control of the structure. From these experiments in improvisation came the idea to explore a similar graphic-based approach for composing and performing electronic sounds. Both improvisation and electronic music composition share a similar issue: within these scenarios, traditional notation systems are maybe not the most useful tools for representing the sonic material and the musical gestures. In a more deep sense, as noted by Trevor Wishart [21], traditional Western notation is based on a time/pitch lattice logic, which strongly influences the way music is composed. The main idea behind CABOTO was to develop a graphic-based notation system which could be defined in a continuous domain, as opposed to the lattice, and to find a way to scan these graphic shapes and map them into sounds. The system

was originally intended to be an offline tool for composing. However, the project evolved towards a performative scenario, where the graphic score becomes an interface for real-time synthesis of electronic sounds. The score scanning is not entirely controlled by the performer: a set of semi-autonomous *navigators* traverse the score. The system can be defined as an *inherent score-based system* [11] like the tangible scores developed by Tomás and Kaltenbrunner [19]. With respect to previous works on the same topic, in developing CABOTO the focus has been on four original features: the use of sketching on a real canvas, the introduction of an hybrid approach in interpreting the score, a *polymorphic* mapping, and the concept of the score as a map. The system also tries to exploit the morphophoric intrinsic characteristic of the graphic shapes as an immediate and intuitive cue for the performer, which can look at the score as a palette of *sonic elements* available for the live performance.



Figure 1: An example of a graphic score used in CABOTO.

2. HISTORY: SCANNING GRAPHICS

The idea of synthesizing sound from graphics has a long history, which tracks back to the early experiments by pioneers in Soviet Russia during the 30's [17]. In 1930, Arseny Avraamov produced the first hand-drawn motion picture soundtracks, realized by means of shooting still images of sound waves sketched by hand. During the same year, Evgeny Sholpo developed the Variophone, which made use of rotating discs of paper with the desired shapes. In the meanwhile, similar researches were conducted in Germany by Rudolf Pfenninger and Oskar Fischinger. Later on, in Canada, Norman McLaren started his experiments in sketching sound on film [6], while Daphne Oram explored optical sound and developed her *Oramics* instrument [4]. During the '70s these explorations moved into the digital do-



Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

NIME'18, June 3-6, 2018, Blacksburg, Virginia, USA.

main. A well-known computer-based interface for composing with drawings is the UPIC system conceived by Xenakis at CEMAMU in 1977 [10]. In that system, the user could draw on a graphic tablet, and the system had a great degree of customization, using the sketched material as waveforms, control signals, tendency masks. In more recent years, there have been some projects inspired by Xenakis work, such as the HighC software [2] and Music Sketcher, a project developed by Thiebaut et al. [18]. Golan Levin's work [9] focused on an audio/video interface, which was intended to allow the user to express audiovisual ideas in a "free-form, non-diagrammatic context". Though the resulting sound was intended to be more a kind of *sonification* rather than the output of an instrument or a composition, in his work Levin has discussed several interesting design issues, such as the representation of time and the quest for an intuitive but rich expressive interface. An interesting project is Toshio Iwai's *Music Insects* [5], an interactive sequencer developed in 1991, where the notes, represented by colored pixels, were triggered by "insects" which were moving on a virtual canvas. The idea of multiple agents that scan the graphic score has been an inspiring source for CABOTO, and it can be found in other previous works, such as the one proposed by Zadel and Scavone [22]. They developed a software for live performance which makes use of a virtual canvas in which the user can draw strokes. These strokes define paths along which some playheads, called "particles" may travel. Their movements and positions drive the sound playback, and an interesting feature is that the drawing gesture is recorded by the system, and the particles mimics the recorded motion. Other authors have explored the possibility of scanning sketches as a tool for composing in traditional notation systems [20] [3].

3. DESCRIPTION OF THE SYSTEM

A graphic score (Fig.1) sketched on paper using traditional drawing tools is read by a computer vision system. The graphic elements are then scanned following a symbolic-raw hybrid approach, that is, they are interpreted by a symbolic classifier (according to a vocabulary) but also as waveforms and optical signals. The score is viewed according to a cartographic map metaphor and the development in time of the composition depends on how we traverse the score. Some navigators are defined, which traverse the map according to real-time generated paths and scan a certain area of the canvas. The performer has some kind of macro-control on how to develop the composition, but the navigators are programmed for exhibiting a semi-autonomous behavior. The compositional process is therefore split in two phases: the sketching of the graphic score, which can be performed both offline or in real time, and the generation of the trajectories for reading it (and thus the synthesis of the sonic result), which is performed in real time during the performance.

4. THE SCORE AS A MAP

Athanasopoulos et al. [1] have recently published a comparative study on the visual representation of sound in different cultural environments. An interesting result of this study is that the Cartesian representation of sound events, where time is represented on the x axis, is a cultural influence probably derived from literacy. In developing CABOTO, the issue of time has been a crucial one. In the first prototypes, in which the system was intended as a composing tool rather than an instrument for live performance, time was represented on the x axis, as in traditional Western notation. This led to a conventional representation of the composition, which was quite intuitive on one side, while on

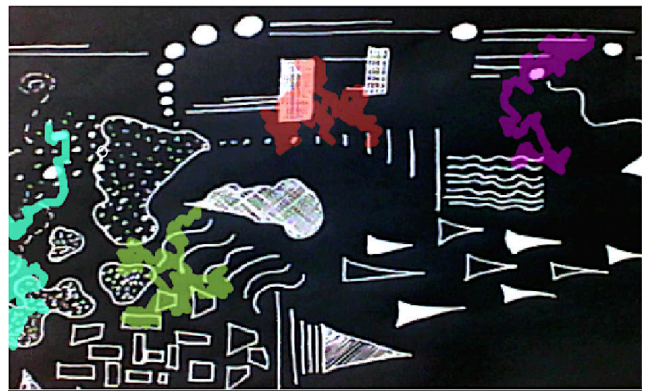


Figure 2: The traces of four navigators scanning the graphic score.

the other side it led to predictability, and invited the user to think about the composition process in a time-oriented way. These considerations led to the shift from the time-based score to the concept of the score as a map [14]. According to the map metaphor, the two-dimensional canvas is viewed as the representation of some kind of *terra incognita* which is explored by some navigators (Figure 2). There are different kind of maps, and different ways of reading them. Thus, we can define different kind of scanners, or navigators, which traverse the map collecting data which are then used in the sound synthesis engine. The way we traverse the score, the path we choose, affects the resulting information we gather from the score itself. One or more paths (or an algorithms that generates paths) can be defined in order to explore the score. The performer can guide the navigators forcing them to certain areas of the score-map, or constrain them to generate certain paths.

5. DEFINING A GRAPHIC VOCABULARY

The graphic notation developed for composing the score (Figure 1) is the result of personal aesthetic choices. In this abstract vocabulary, geometry plays a leading role. Simple geometric shapes such as points, lines, planes form the basic elements for the development of the graphic sketch. These elements are combined according to relations that can be expressed in the terms of physics: mass, density, rarefaction, tension, release. This vocabulary draws inspiration from various sources. One is the work of Wassily Kandinsky [7]: in his writings he tried to develop a theory of shapes and colors, and the study upon elementary shapes that he proposed is quite interesting. Other important sources of inspiration have been the works of John Cage, Earle Brown, Cornelius Cardew, Roman Haubenstock-Ramati and Anestis Logothetis.

6. SCANNING THE SCORE

Each navigator traversing the score scans a certain area centered at its current position. When a graphic element enters the navigator's scope, it is processed and it will result in a sound output. The graphic material is interpreted using three different scanning algorithms: a symbolic classifier, a waveform scanner and an optical scanner. These scanners will be presented in detail in the next sections.

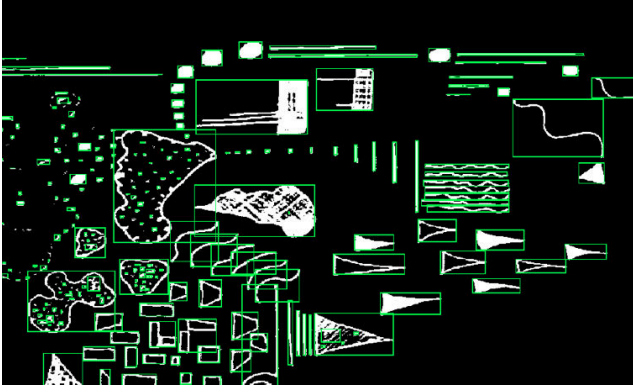


Figure 3: Blobs recognition applied to the example score.

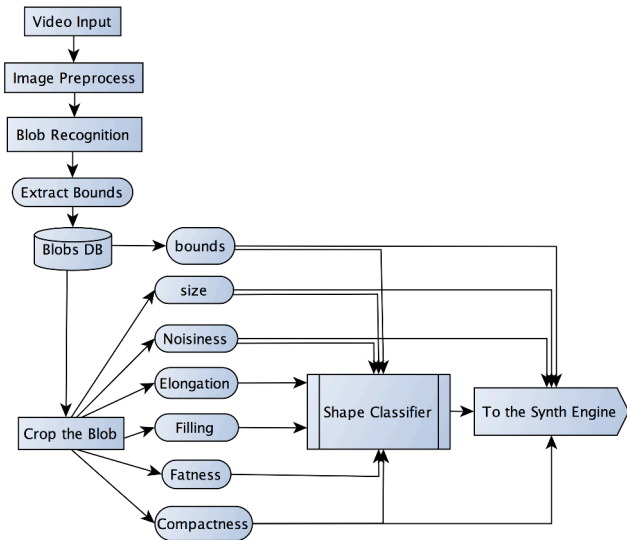


Figure 4: A general scheme of the shapes recognition, features extraction and classification algorithm.

6.1 Image preprocessing, features extraction and classification

During a preprocessing phase, all the blobs - that is, the connected components in the score - are detected, along with their boundaries in the Cartesian plane. The algorithm then computes a set of geometric features: size, dimensions ratio, orientation, filling, compactness, *fatness* and *noisiness*. The filling is a measure of the total luminance with respect to the blob area. The compactness is the ratio between the area and the perimeter of the shape, thus a filled circle has the highest compactness value. *Fatness* is a parameter that measures the average thickness of the shape along its main orientation, in order to tell curved lines from plane-like shapes. The *noisiness* of the blob is defined by the average number of zero-crossings of the first derivative along a set of paths that traverse the shape. Thus, a compact blob which is mostly filled or mostly empty will have a very low noisiness value, while a complex line will exhibit high noisiness. All these features are then used as parameters in the synthesis engine, and they are also used for classifying the shape. According to its features and a set of thresholds, each blob is classified into 7 categories or *classes* (Figure 6). The classification algorithm is depicted in Figure 5. It can be noted that this kind of classification algorithm is an untrained one, therefore it could be

objected that is a quite naive kind of classifier. Nevertheless, this choice is a deliberate one. In a previous version, a more sophisticated classifier was developed, which made use of a trained pattern recognition algorithm. This led to an over-classification of shapes, which tend to become a sort of dictionary or a taxonomy of graphic elements. A symbolic mapping implies an interpretation. In this sense, classifying is a way of quantizing the collected data, and thus, in a certain sense, it's an operation which leads to a reduction of information. Moreover, since different synthesis techniques are defined for different classes, we may have discontinuities in the sound result when moving between adjacent classes of shapes. These are the reasons why the classifier has been designed in a simple and general way, while introducing two other scanning algorithms for keeping the richness of the hand-drawn sketch.

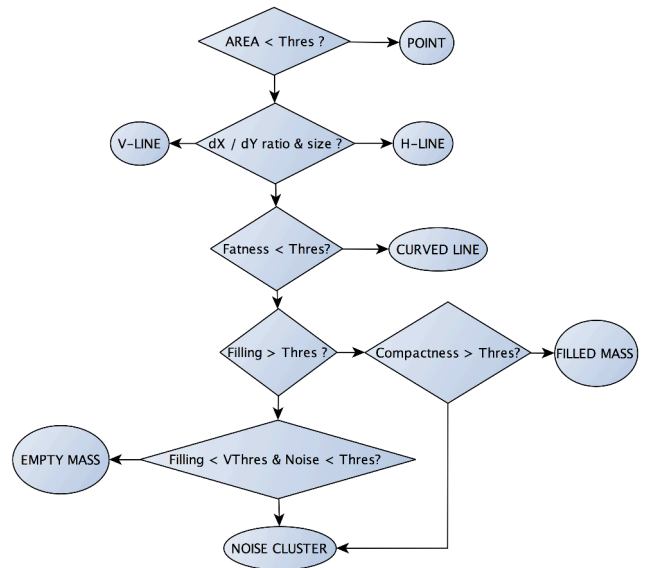


Figure 5: Diagram showing the classification procedure.

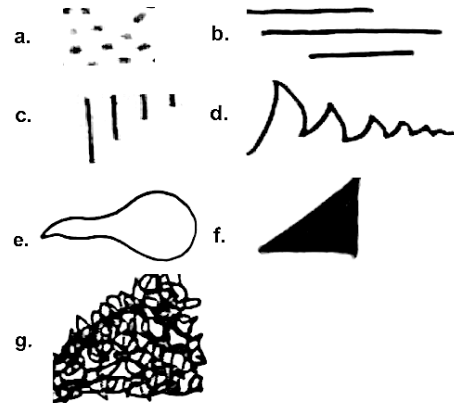


Figure 6: The classes of shapes recognized by the symbolic classifier: a) point, b) horizontal straight line, c) vertical straight line, d) curved line, e) empty mass, f) compact filled mass, g) noise cluster.

6.2 The Waveform Scanner

Another technique implemented in CABOTO is the *waveform scanner*. The blob is cropped and its edges are scanned

along its main orientation axis. The optical signal is extracted as a measure of the distance between the outer edge of the shape and the median line with respect to the blob size. Once the scanner reaches the bound of the blob (with respect of its main axis), it wraps around the shape and goes backward scanning the opposite edge. The output signal is sent to the synthesis engine as an audio stream, and is then used as an envelope, modulator, control signal or directly as an audio signal, according to the synthesis algorithm involved for the specific class of the current shape.

6.3 The Optical Scanner

An optical scanner is associated with each navigator traversing the score. The scanner crops a view in a chosen color channel (if available) and extract the overall mass, that is, a measure of the luminance. The area covered by the scanner can be controlled in real time, thus varying the resolution and gain of the resulting signal. The output of the optical scanner is a raw signal, that is, it's not derived from some sort of interpretation according to a vocabulary, but from a scanning operation upon the values stored in the image matrix, and it brings richness and unpredictability to the sound synthesis. Moreover, since it depends strongly on the instantaneous position of the navigator, has an immediate correlation with the visual feedback that can be seen on the visualized score. This allows the performer to have a certain degree of control on the optical signal output. An interesting outcome of the optical scanning is that, since it can act on a pixel resolution level, it is highly affected by the imperfections of the hand drawn sketch and the canvas.

7. MAPPING

In a famous experiment by Ramachandran and Hubbard [16], derived from Köhler [8], people were asked to assign names to two different geometric shapes. The provided names were “Bouba” and “Kiki”, and the shapes were a curved, smooth shape and a more sharp-angled one. The results of this experiment suggested that the association between shape and sound is not connected to cultural biases, but to a human brain feature. We can note a curious link between the results of these studies and the mathematical properties of waveforms. Consider the graphical representation of a sound pressure wave, that is, the pressure vs time Cartesian plot (or voltage vs time). If we listen to the synthesized sound corresponding to that shape by reading the wave as a wavetable (in the digital domain) or playing it with an optical device similar to the ones used in the analog film technique, we can verify that a more sharp-like kind of waveform will sound harsher, since its spectrum will contain more components, more partials. On the other side, a sinusoidal-like shape will have few or even just one spectral component (the fundamental), resulting in a smoother sound output. These considerations have been taken into account in designing the mapping strategy and the sound synthesis processes. It is important to note, however, that this mapping is still arbitrary, and reflects personal aesthetic choices. For the rendering of the different shape classes, different processes and synthesizers have been designed, each one characterized by a set of control parameters. This results in a *polymorphic mapping*, that is, different mapping strategies for different kind of sonic events. For instance, the relative position of the navigator with respect to the sound object boundaries is mapped and used for the *noise cluster*, but is ignored in the case of the *point* class. Part of the mapping is presented in Figure 7. For some classes, multiple sound processes have been defined, which are different *realizations* of the same shape/sound class. In this

case, the actual sound process used for a certain shape is chosen randomly at runtime.

	points (PBind)	h line (Synth)	v line (PBind)	noise cluster 1 (PBind)	noise cluster 2 (PBind)
Ymin,max	grain rate	sin freq, max cutoff freq	wave rate / sin freq	max grain rate	max grain rate
dY	-	freq rand deviation range	impulse duration	rate range	rate range
Xmin, max	grain position, rate deviation	pulse duration	-	grain pos	grain pos max
dX	-	pulse trigger freq	-		grain pos range
Opti signal	-	cutoff freq	LP cutoff	gain envelope	gain envelope
Waveform	-	audio signal 2, mod signal	audio signal 1	mod signal	audio signal
Xe	-	-	-	grain pos	grain pos
Ye	-	-	-		min grain

Figure 7: Mapping between extracted parameters and synthesis parameters, for some classes realizations. X_e , Y_e denote the navigator position. X_{min} , X_{max} , Y_{min} , Y_{max} the shape bounding box.

8. ADJUSTING THE SAILS

The navigators trajectories are generated in real time according to four different modes: *forced*, *random*, *jar of flies*, *loop*. The *forced* mode allows the performer to manually send a navigator to a certain position in the score, using a cursor on the score view interface (Fig. 8). In *random* mode, the navigators are moving autonomously, performing a random walk. A more interesting motion is defined by the *jar of flies* algorithm. This is a random walk in which the step increment is inversely proportional to the optical signal value detected at the current position. This means that a navigator will move slowly when it is in a densely populated area of the score (that is, with more elements), while it will run faster when nothing is detected. This simple technique results in a sort of “organic” motion, which has some interesting effect on the development of the sound output. Finally, a *loop* mode is available, which generates a trajectory modulating the X and Y coordinates of the navigator with periodical signals. Since the rate and amplitude of these signals can be set independently for the two axis, it is then possible to have different kind of motions, from simple loops along one axis to more complex trajectories. Some of these modes can be mixed or superimposed. For example, a navigator can perform a random walk while looping in a certain interval across the X axis.

9. IMPLEMENTATION

The system is designed according to a modular logic, with different pieces of software integrated through Open Sound Control (Figure 9). The image processing module has been developed in the Max/MSP programming environment, using Jitter and the cv.jit library developed by Jean-Marc Pelletier [15]. The image processing is quite CPU intensive, therefore some routines have been written in Java and C++ for optimization. The sound synthesis engine has been developed in the Supercollider language, which provides a powerful framework for generating complex sound events in the form of processes controlled by a set of macro-parameters. The sound is projected in the performance space through a 4-channel audio system, and the output from each navigator is mapped to one of the four channels.

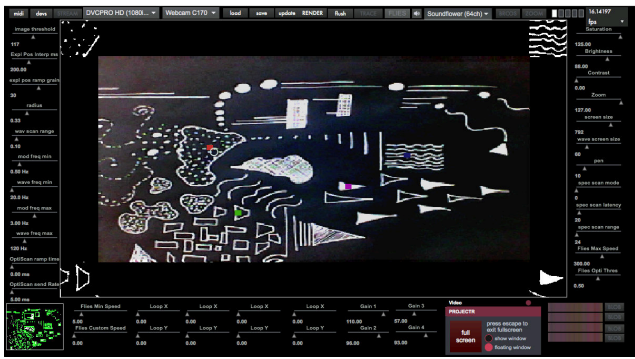


Figure 8: The CABOTO console.

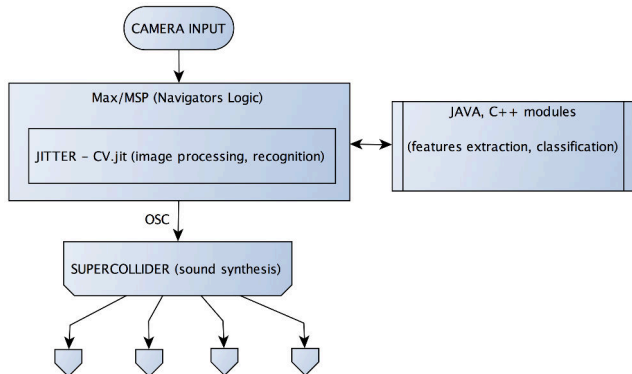


Figure 9: Implementation diagram of the CABOTO system.

10. LIVE PERFORMANCE

As previously noted in Section 1, The system was originally conceived as a tool for composing. However, the project evolved towards the design of an instrument for live performance. This evolution is connected to the fact that, as a musician and improviser, I felt the need for a system for live performance and improvisation. The live setup includes a light table for the canvas, a camera, a laptop, an audio interface and a midi controller. Moreover, a video output is provided for screen projection, which shows the score to the audience, along with the current scopes of the navigators and the trajectories (Figure 10). During the performance it's possible to sketch or modify the score: in order to avoid the hand interference, the image can be grabbed with a one-shot button, once the drawing gesture has been completed. Another option is to disable the video streaming according to a motion detection algorithm. Nevertheless, I found more interesting to keep the video streaming on and let the drawing action interfere with the score scanning, thus resulting in glitches, noise and unexpected sonic output. In designing the live setup, some decisions had to be made regarding the parameters to be controlled. Since I'm dealing with multiple navigators and the drawing action, I decided to keep the control over few macro-parameters, such as the output gain of each navigator (which also enables/disables the navigator itself), the trajectories generation mode and speed, and the score image settings (brightness, contrast, saturation, zoom, blob recognition thresholds). A video documentation of a live performance with the instrument can be found in [12].

11. CONCLUSIONS

A novel system for performing electronic music through graphic notation has been presented, which focuses on three

features: the use of sketching by hand on paper, the introduction of an hybrid approach in interpreting the score, a polymorphic mapping, and the concept of the score as a map. The system has to be considered as a work in progress and many improvements are currently under development. The sonic palette and parameters control need to be extended and developed further. In particular, new strategies will be introduced for generating the navigators trajectories. Moreover, in the current version each navigator can deal with only one blob at a time, thus if more than one shape is detected in the navigator scope, only the bigger one is synthesized. This limitation is going to be addressed in future updates. Much effort has been put into code optimization, since the image processing algorithms are quite CPU demanding. Also, further explorations will focus on developing the visual feedback which is presented to the audience during the live performance. In future developments, CABOTO will be used for live performance, both in solo and in collaborative scenarios with improvising musicians, and as an interactive installation.

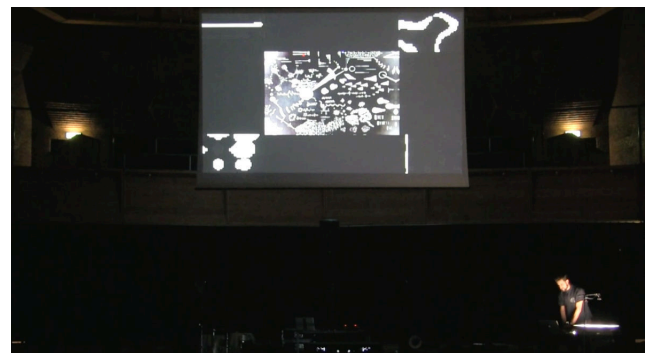


Figure 10: Live performance with CABOTO. On the bottom right, the light table with the camera.

12. ACKNOWLEDGMENTS

CABOTO is part of my Research Project for the Master in Sonology at the Royal Conservatoire in The Hague. I would like to thank all the staff members and colleagues at the Institute of Sonology for their advice and support, in particular: Prof Kees Tazelaar and Prof. Richard Barrett.

13. REFERENCES

- [1] G. Athanasopoulos, S.-L. Tan, and N. Moran. Influence of literacy on representation of time in musical stimuli: an exploratory cross-cultural study in the UK, Japan, and Papua New Guinea. *Psychology of Music*, 44(5):1126–1144, 2016.
- [2] T. Baudel. HighC, draw your music. <https://highc.org> (accessed on November, 29th 2017).
- [3] J. Garcia, P. Leroux, and J. Bresson. pOM: Linking pen gestures to computer-aided composition processes. In *40th International Computer Music Conference (ICMC) joint with the 11th Sound & Music Computing conference (SMC)*, 2014.
- [4] J. Hutton. *Daphne Oram: innovator, writer and composer*, volume 8. Cambridge University Press, 2003.
- [5] T. Iwai. Piano-as image media. *Leonardo*, 34:183–183, 2001.
- [6] W. E. Jordan. Norman McLaren: His career and techniques. *The Quarterly of Film Radio and*

- Television*, 8(1):1–14, 1953.
- [7] W. Kandinsky. *Point and Line to Plane*. Dover Publications, 1947.
 - [8] W. Köhler. *Gestalt psychology*. H. Liverights, New York, 1929.
 - [9] G. Levin. *Painterly Interfaces for Audiovisual Performance*. M.S. Thesis, MIT Media Laboratory, 2000.
 - [10] H. Lohner. The UPIC system: A user’s report. *Computer Music Journal*, 10(4):42, 1986.
 - [11] E. Maestri and P. Antoniadis. Notation as Instrument: from Representation to Enaction. In *Proc. First International Conference on Technologies for Music Notation and Representation TENOR 2015*, Paris, France, May 2015. IRCAM - IReMus.
 - [12] R. Marogna. CABOTO - Live at Koninklijk Conservatorium, Arnold Schoenbergzaal, March 21st 2018. <http://riccardomarogna.com/caboto> (accessed on April, 6th 2018).
 - [13] R. Marogna. Graphograms. <http://riccardomarogna.com/graphograms> (accessed on November, 29th 2017).
 - [14] D. Miller. Are scores maps? A cartographic response to Goodman. In *Proc. of the Int. Conference on Technologies for Music Notation and Representation - TENOR2017*, pages 57–67, A Coruña, Spain, 2017. Universidade A Coruña.
 - [15] J.-M. Pelletier. CV-JIT <http://jmpelletier.com/cvjit> (accessed on December, 11th 2017).
 - [16] V. S. Ramachandran and E. M. Hubbard. Synaesthesia - a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12):3–34, Mar. 2001.
 - [17] A. Smirnov. *Sound in Z - Experiments in Sound and Electronic Music in Early 20th Century Russia*. König, 2013.
 - [18] J.-B. Thiebaut, P. G. Healey, and N. Bryan-Kinns. Drawing electroacoustic music. In *Proceedings ICMC*, 2008.
 - [19] E. Tomás and M. Kaltenbrunner. Tangible scores: Shaping the inherent instrument score. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 609–614, London, United Kingdom, June 2014. Goldsmiths, University of London.
 - [20] T. Tsandilas, C. Letondal, and W. E. Mackay. Musink: composing music through augmented drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 819–828. ACM, 2009.
 - [21] T. Wishart and S. Emmerson. *On Sonic Art*, volume 12. Psychology Press, 1996.
 - [22] M. Zadel and G. Scavone. Different strokes: A prototype software system for laptop performance and improvisation. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, NIME ’06, pages 168–171, Paris, France, France, 2006. IRCAM, Centre Pompidou.