ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/331782629

Shaping Microsound Using Physical Gestures

Conference Paper · June 2015

CITATION
1

READS
9

2 authors, including:

Gordan Kreković
Visage Technologies

17 PUBLICATIONS **17** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Intuitive control of sound synthesis View project

Project Visual programming languages for music creation View project

All content following this page was uploaded by Gordan Kreković on 27 March 2019.

The user has requested enhancement of the downloaded file.

xCoAx 2015
Computation
Communication
Aesthetics
and X

Glasgow
Scotland
2015.xCoAx.org

Shaping Microsound Using Physical Gestures

Gordan Kreković
Faculty of Electrical Engineering and Computing,
Zagreb, Croatia
gordan.krekovic@fer.hr

Antonio Pošćić
Faculty of Electrical Engineering and Computing,
Zagreb, Croatia
antonio.poscic@fer.hr

Keywords: sound synthesis, parameter mapping, fuzzy logic, dynamic stochastic synthesis.

This paper presents a system for controlling the structure of synthesized sounds at the waveform level using physical gestures. The purpose of the system is to allow intuitive, natural, and immediate interaction with a sound synthesis model based on a non-standard synthesis technique. Instead of manipulating numerical parameters which are, in case of non-standard synthesis, typically abstract and without acoustical meaning, musicians can shake a mobile device in order to shape the structure of synthesized waveforms. The system receives raw data captured from accelerators, extracts relevant statistical features, and maps them into parameters of a dynamic stochastic synthesizer. Mapping is based on fuzzy logic in order to ensure a non-linear and non-injective relation defined within explicit mapping rules. Experimentation proves that the system provides natural, immediate, and expressive control which is convenient both in the composition process and in live performances.

1 Introduction

Sound synthesis using analog and digital electronic devices allows composers to create novel sonorities that characterize their compositions uniquely. Controlling the timbre and its changes over time is an important compositional aspect which provides coherence between musical form, structure, and material (Manousakis, 2009). Stockhausen emphasized that "every sound is the result of a compositional act" (Stockhausen, 1963), while Di Scipio wrote that "synthesis can often be thought of as micro-level composition" (Di Scipio, 1995) referring to the idea that sound synthesis allows composing timbers instead of just employing them in higher-level musical structures (Brün, 2004).

Sound synthesis techniques particularly oriented to the micro-level composition and sound microstructure are non-standard methods (Thomson, 2004). Instead of relying on theoretical acoustical models, reproduction of actual sounds, or psycho-acoustic phenomena, non-standard synthesis methods are based on mathematical models and compositional abstraction (Holtzman, 1979). Such an approach allows composers to describe waveforms, their organization and transformation, without imposing their acoustical consequences. Thereby, many compositional aspects are reduced to controlling the sound synthesis process and creating sounds at the waveform level.

Non-standard techniques, idiomatic to digital sound synthesis, attracted the attention of researchers and composers, especially in the 1970s. Even though most of the non-standard techniques produce sounds by generating waveforms in the time domain, several principally distinct approaches emerged: synthesis based on rules (Berg, 1979; Berg, Rowe and Theriault, 1980, Brün and Chandra, 2001; Holtzman, 1979), stochastic approach (Xenakis, 1992), fractal interpolation techniques (Yadegari, 1991; Monro, 1995; Dashow, 1996), and other approaches (Valsamakis and Miranda, 2005; Collins, 2008).

Since non-standard synthesis techniques are not focused on acoustical features of the synthesized sound, their controllable parameters usually do not bear acoustical meaning. The parameters serve as abstract numerical inputs of mathematical models for waveform generation. In order to achieve desired waveforms, composers need to understand all the details of the applied synthesis model and its capabilities. While this is not a limiting factor for composers who developed the synthesis models themselves, the lack of intuitiveness may negatively affect the efficiency and inspiration of those who do not use their own models. The process of shaping micro-sound should be closer to the way how composers imagine sound structures. A more intuitive way of controlling parameters of non-standard synthesis models would allow composers to think within the musical domain during their creative process. Additionally, such an approach would be more convenient for applications in which immediate control is needed, such as for live performances or interactive installations.

As a solution for intuitive control of non-standard synthesis techniques, we propose a system for detecting physical gestures and mapping them into sound synthesis parameters. Physical gestures as a means for controlling the process of sound generation are already widely used in the area of dance-music intermodalities and applications related to enhancing musical content by physical actions (Friberg, 2005; Heile, 2006). An advantage of physical gestures for composing at the micro-level would be the intuitive and immediate relation of natural movement with waveforms synthesized by non-standard models. Instead of manipulating numerical parameters, composers could achieve their ideas and develop unique expressivity by experimenting with physical movements. The system proposed in this paper extracts selected features from the physical gestures and maps them into synthesis parameters. The purpose of this mapping is to achieve natural relations between gestures and the structure of synthesized waveforms. Transitively, physical gestures may be also related to the acoustical features of the synthesized sound, but only to the extent to which synthesis parameters are related to acoustical features.

The system for shaping microsound using physical gestures described in this paper employs the dynamic stochastic synthesis as an underlying synthesis model. We selected this synthesis technique because it is purely parametric unlike some other non-standard methods which require setting up rules or initial states.

2 Dynamic Stochastic Synthesis

Before presenting the overall system, here is a short overview of the employed synthesis model. Dynamic stochastic synthesis was devised by Iannis Xenakis as a result of his ambition to achieve unified and simultaneous engagement on different time-scales within the composition, from the overall structure of the composition to its microstructure and tone quality.

Dynamic stochastic synthesis generates samples by interpolating a set of breakpoints which change their amplitudes and positions in time stochastically. A breakpoint position is represented relatively to the preceding breakpoint and it is commonly called "breakpoint duration". Initial amplitudes and durations are usually chosen randomly or taken from a trigonometric function.

At every repetition of the waveform, these values are varied independently of each other using random walk. That means that both the amplitude and the duration of a certain breakpoint are changed by adding random steps to the values in the previous cycle as shown in Figure 1. A succession of random steps applied on all breakpoints causes the continuous variation of the waveform. The amount and character of the variation depend on a selected probability distribution and its parameters. Both amplitude and duration random walks are limited each with two reflecting barriers which bounce excessive values back into the predefined range.

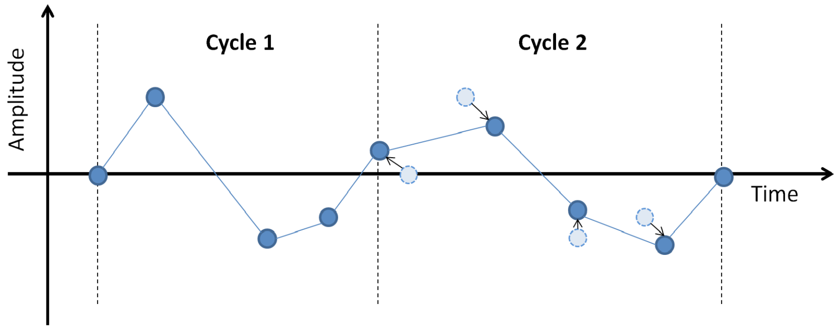


Fig 1. Breakpoints change their positions from one repetition to another. Light blue circles in the second represent positions from the first cycle, while darker circles represent new positions.

These barriers prevent breakpoints from straying too far from their initial positions and therefore enable control over amplitude and frequency ranges of the overall waveform.

Parameterization of the synthesis model is achieved through: (1) the number of breakpoints in a waveform, (2) barriers of the amplitude random walk, (3) probability distribution of the amplitude random walk and its parameters, (4) barriers of the duration random walk, and (5) probability distribution of the duration random walk and its parameters. The amplitude barriers provide control over the amplitude range of the generated waveform, while the duration barriers define minimal and maximal number of samples between two breakpoints. If changes in amplitude and duration between successive repetitions are small, the synthesized sound is relatively simple, but it can have interesting modulation effects. On the other hand, as changes become more prominent, the sound becomes more complex and noisier.

Detailed explanations of the synthesis model can be found in (Serra, 1993) and (Luque, 2009), while several other researchers proposed extensions of the original algorithm (Hoffman, 2000; Brown, 2005; Young, 2010). Since the first implementation by Iannis Xenakis did not provide any means of controlling the synthesis process, some authors suggested interface designs for direct parameter control (Hoffman, 2000; Bokesoy and Pape, 2003; Brown, 2005). An interesting solution was also a mobile application which obtained parameters from multi-touch gestures and accelerometers (Collins, 2011). In all of the mentioned approaches, the values from controllers or sensors were directly mapped into synthesis parameters which remained transparent to composers. In order to hide numerical parameters, in our previous research we proposed intuitive control by an input audio signal (Kreković and Brkić, 2012) and using MIDI messages (Kreković and Petrinović, 2013).

In this paper we present a novel approach focused on using physical movements for controlling dynamic stochastic synthesis and thereby for the intuitive shaping of synthesized waveforms. The following chapters describe the system design, experiments, and results yielded within this research.

3 System Overview

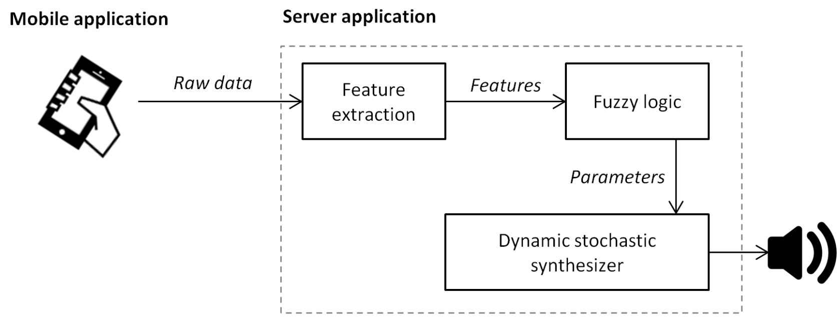
As has been mentioned previously, the main goal of this interactive system is to provide a simple and intuitive method for shaping waveforms by shaking a mobile device. Therefore, the central problem of the research is establishing natural mappings between the shaking gestures and the synthesis parameters.

The first step was to choose features of the shaking gesture which could be extracted from the raw physical movements. Since the intention was to establish a relation between the features and the waveform structure, we searched for lower-level features which could not bear meaning, symbols, or metaphors. The information used for shaping microsound should be contained at the phenomenological level of the physical movement and not on the symbolic level. To cover cinematic, frequential, and spatial phenomenological aspects of physical movements we selected the following features as relevant for our study: intensity, frequency, and shaking direction.

In order to extract the aforementioned features using a computer system, the prerequisite is to capture raw physical movements and represent them as a stream of numbers. That functionality is available in a mobile application which serves as a controller. As the user shakes the mobile device, the mobile application captures data from accelerometers and sends them to the server side. The communication between the application and the server side is based on the Open Sound Control (OSC) protocol which is a widely-used and standardized protocol for networking sound synthesizers, computers, and other multimedia devices. This way, we can use any existing mobile application that can read data from accelerometers and form appropriate OSC messages. Besides the compatibility with many existing iOS and Android applications for smartphones, the implementation of the OSC protocol also opens opportunities for running client applications on different types of devices such as smart watches and other wearable devices with accelerometers. There are countless possibilities when the use of such devices in performances is concerned.

The server-side application receives the raw data from accelerometers, extracts relevant features, maps those features into sound synthesis parameters, and finally produces an audio signal using dynamic stochastic synthesis. The mapping between features of the shaking gestures and synthesis parameters is a non-linear and non-injective mapping based on rules which can be elicited from knowledge of a human expert. To implement such a mapping we opted for an expert system based on fuzzy logic. The overall system architecture with corresponding data flow is

Fig 2. Overall system architecture and data flow.



shown in Figure 2.

3.1 Movement Analysis

Raw data from accelerometers represents instantaneous accelerations of the mobile device and do not quantify physical gestures directly. However, there are higher-level features extracted from the raw data that can better describe the nature of movements. Since the dynamic stochastic synthesis produces rich and complex sounds with organic quality, we opted for shaking movement as a gesture which can, to some degree, metaphorically represent the waveform structure and its acoustical qualities.

The first feature, which represents the cinematic phenomenological aspect of the physical gesture, was the shaking intensity. The intensity is calculated as the root mean square (RMS) of the acceleration changes for all three axes:

$$intensity = \sqrt{\frac{1}{3}\left[\left(a_x[n]-a_x[n-1]\right)^2+\left(a_y[n]-a_y[n-1]\right)^2+\left(a_z[n]-a_z[n-1]\right)^2\right]}$$

where $a_x[n]$, $a_y[n]$, and $a_z[n]$ represent discrete values of acceleration respectively in the n-th step at the axes x, y, and z. In order to smooth the spikes, a running average filter is applied to the calculated root mean square.

The second feature we selected was the measure of how fast the user shakes the device. We called this feature “shaking frequency”. This measure is calculated based on the number of zero-crossings. Each time that the motion of the device changes direction, the sign for acceleration on some axes changes as well. Therefore, the number of crossings through the zero value can be approximately correlated with the frequency.

The purpose of the third and final feature is to quantify how complex the shaking movement is. For simplicity’s sake, we call this measure “shaking direction”. If the user shakes the device just along one axis (e.g. up and down), this feature has a low value, but if the user makes loops and changes directions very often, the feature will have higher values. The measure is based on the maximal absolute difference between the acceleration changes on two axes in the same moment. A running average filter is again used to smooth spikes.

3.2 Fuzzy Mapping

The higher-level features are mapped into sound synthesis parameters according to desired relations to the synthesized waveforms. The intensity is intended to correlate with the amplitude and the structural complexity of the generated waveform. More vigorous movements of the device should cause higher amplitudes of the synthesized signal and more prominent changes of breakpoint positions. The shaking frequency is intended to be related to the frequency range and frequency drifts of the synthesized waveform. It should, consequently, affect the impressions of pitch and timbral flux. Faster movements of the device should result with smaller duration limits, greater frequency drifts, and shorter waveform cycles. Finally, the shaking direction is intended to control the dynamicity of the amplitude and the frequency changes of the waveform. Simpler movements of the device should cause steadier waveforms and simpler timbres, while loops and sudden changes of the shaking direction should cause faster developments of the waveform and thereby more complex timbres.

An expert system based on fuzzy logic was chosen as the most convenient solution for mapping gestural features into synthesis parameters. Fuzzy logic is a form of probabilistic logic which supports the concepts of partial truth and linguistic variables (Zadeh, 1965). This is suitable for quantifying imprecise information and making decisions based on incomplete data (Kosko, 1993).

Mappings between gestural features and synthesis parameters are described by fuzzy rules with linguistic variables. An example of a linguistic variable is the "intensity", whilst its linguistic terms are "low", "medium", and "large". Inputs in a fuzzy logic system are usually numeric, so it is necessary to convert these numeric values into linguistic terms. An input value can partially satisfy several linguistic variables at the same time. For example, if the feature has a value of 0.3, the "intensity" is somewhere between "low" and "medium".

Fuzzy rules are specified in the form of IF-THEN statements:

```
IF (x1 IS S1) AND/OR ..., (xn IS Sn) THEN y IS T
```

where x_i represents input fuzzy variables, y is the output variable, while S_n and T stand for input and output linguistic terms. The first step of applying the fuzzy model is to convert input variables into fuzzy logic variables. Then, output variables are calculated by evaluating the rules and the output values are converted to numeric form.

Fuzzy logic enables nonlinear many-to-many mappings between gestural features and synthesis parameters, while the rules based on linguistic variables can be easily understood and specified by composers. Because of these properties, systems based on fuzzy logic have been previously used in the musical domain for coding musical gestures (Orio and De Prio, 1998), analyzing the emotional expression in music performance (Friberg 2004), mapping between visual and aural information (Cádiz, 2006), sound synthesis (Miranda and Maia, 2005; Cádiz and Kendall 2005), and several other applications.

The fuzzy logic model specifies input and output variables, membership functions, fuzzification and deffuzification methods, and mapping rules for the expert system based on fuzzy logic. In our implementation, the fuzzy logic model can be specified using Fuzzy Control Language (FCL). This language is standardized by the International Electrotechnical Commission standard IEC 61131-7. The fuzzy rules have an intuitive IF-THEN form which allows musicians to modify and write new rules by themselves.

The fuzzy logic model used for this research accepts three features extracted from the raw data received from the mobile application. The fuzzy logic model has 5 outputs which represent values of sound synthesis parameters. All the membership functions used in the fuzzy model have a triangular form. To defuzzify output variables, the fuzzy logic model uses a technique based on the center of gravity which is the typical approach for models with real-valued output variables.

The rule set for calculating audio features consists of 30 rules which were manually written and adjusted after several iterations of subjective testing by the authors. Here are some examples of the rules:

```
IF frequency IS little THEN durationUpperLimit IS small;
```

```
IF direction IS prominent THEN amplitudeVariation IS large;
IF intensity IS moderate THEN amplitudeLimit IS moderate;
```

3.3 Implementation

The server-side was implemented using Pure Data, a visual programming environment for music and multimedia projects. The expert system based on fuzzy logic was implemented as a Pure Data external component which served as an interface between Pure Data and the jFuzzyLogic library (Cingalogani and Alcalá Fernández, 2012). Since jFuzzyLogic was written in Java, wrapper functions that rely on JNI calls have been implemented to serve as glue code between the main functions of the external written in C and the Java library.

4 Results

Experimentation with the system has shown that the intended mapping between physical gestures and the synthesis parameters has been achieved in accordance with initial requirements. Since we implemented a priori knowledge about synthesis parameters within the fuzzy model, physical gestures are also intuitively related to acoustical results. Stronger shaking causes more complex and louder sound, faster shaking produces higher average pitch with an increase in frequency drifts, while changes in the direction of shaking also affect the synthesized timbre. The overall impression of the sound is transparently and immediately related to physical gestures. Figure 3 shows data captured during

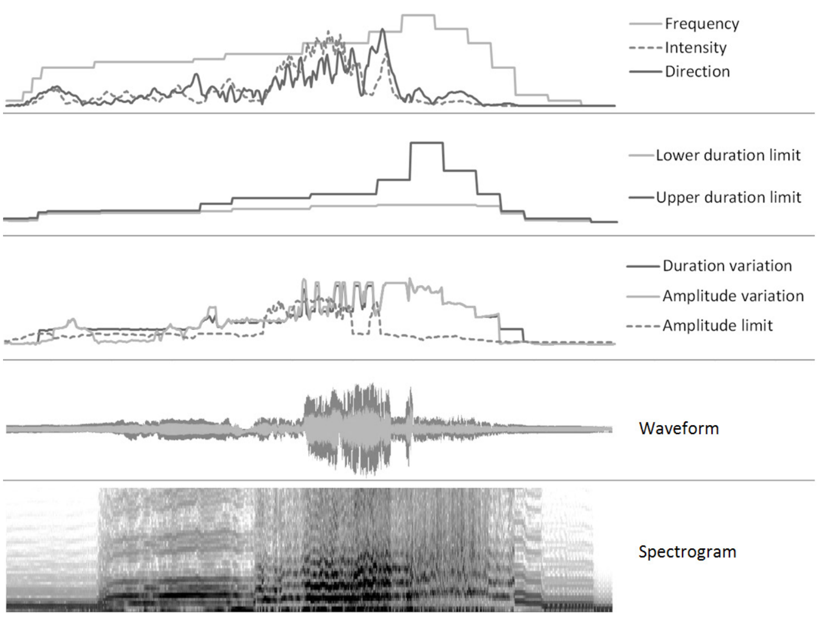


Fig 3. Data captured during a period of 11 seconds. The top chart shows features extracted from physical movements, while the second and the third charts show values of synthesis parameters produced by the expert system based on fuzzy logic. Pictured at the bottom are the waveform and the spectrogram of the resulting sound.

one experiment and proves the established relations between selected features of the shaking movement, sound synthesis parameters, and the acoustical qualities of the synthesized sound.

The expressivity of this interface is satisfying. By combining different shaking intensities, frequencies, and directions, various timbral results can be achieved, from simple steady sounds to buzzy and noisy timbres which are characteristic of dynamic stochastic synthesis. However, mapping three movement features into five sound synthesis parameters meant deliberate and necessary limitations of the expressivity when compared to direct parameter control. Additionally, the selected features of shaking movements are not completely independent. For instance, it is difficult to increase the shaking intensity without increasing the shaking frequency. The consequence of such a dependency is that sounds with both low pitch and complex timbral texture cannot be easily achieved. However, the expressivity of the synthesized sound corresponds to the expressivity of the shaking movement, so we believe that users will not be able to notice those missing aspects of expressiveness, unless they have a lot of experience with dynamic stochastic synthesis and strict expectations before they start using the system.

5 Conclusions and Future Direction

As a solution for intuitive and immediate sound shaping at the micro-level, we proposed a system for mapping physical gestures into parameters of dynamic stochastic synthesis. The benefit of the proposed approach in comparison with direct parameter control is that it is straightforward and does not require deep understanding of the underlying sound synthesis model. This system is also convenient for live performances in which transparent mapping between movements and sound can be exploited in many ways. Unlike other similar systems for controlling synthesis parameters, this one is particularly focused on a non-standard synthesis method and therefore can be observed as an approach for shaping microsound using physical gestures.

From this point on, future research can continue in two different directions. The first one is achieving even closer connections between movements and the microsound by developing a new non-standard synthesis technique which would directly rely on nuances of physical movements instead on stochastic processes. Such a synthesis method would be interesting for dancers and choreographers who could explore the links between microstructures and acoustical features of synthesized sounds and dance movements.

The other research direction would be completely different. Instead of subordinating a synthesis model to the nature of movement, the system for gesture detection could be extended to understand a much larger vocabulary of complex gestures. That way, gestures could be used as symbols and metaphors for triggering various modes of sound synthesis. As a result, higher expressivity could be achieved with interesting results in the context of dance-music intermodalities.

To conclude, the results of this research are generally encouraging with regards to our intention to develop a system for controlling the sound microstructure with physical movements. The proposed approach can be employed to control dynamic stochastic synthesis more easily and effectively, it can be used in live performances, and it can serve as a base for future research.

References

Berg, Paul. 1979. “PILE – A Language for Sound Synthesis”. Computer Music Journal 3(1), pp 30-37

Berg, Paul, Rowe, Robert, and Theriault, David. 1980. “SSP and Sound Description”. Computer Music Journal 4(1), pp 25-35

Bokesoy, Sinan and Pape, Gerard. 2003. “Stochos: software for Real-time Synthesis of Stochastic Music”, Computer Music Journal 27(3), pp 33-43

Brown, Andrew. 2005. “Extending Dynamic Stochastic Synthesis”. International Computer Music Conference, Barcelona, Spain

Brün, Herbert. 2004. *When Music Resists Meaning, chapter From Musical Ideas to Computers and Back.* Wesleyan University Press

Brün, Herbert and Chandra, Arun. 2001. A Manual for Sawdust. Accessed December 30, 2014. http://academic.evergreen.edu/a/arunc/brun/sawdust/sawdust.htm

Cádiz, Rodrigo.2006.“A Fuzzy-Logic Mapper for Audiovisual Media”. Computer Music Journal 30(1), 67–82.

Cádiz, Rodrigo and Kendall, Gray. 2005. “A Particle-Based Fuzzy Logic Approach to Sound Synthesis”. Paper presented at the Conference on Interdisciplinary Musicology, Montreal, Canada

Cingolani, Pablo and Alcalá Fernández, Jesús. 2012. “jFuzzyLogic: A Robust and Flexible Fuzzy-Logic Inference System Language Implementation”. Paper presented at the 2012 IEEE International Conference on Fuzzy Systems, Brisbane, Australia, pp 1-8

Collins, Nick. 2008. “Errant Sound Synthesis”. Paper presented at the International Computer Music Conference, Belfast, UK

Collins, Nick. 2011. “Implementing Stochastic Synthesis for SupperCollider and iPhone”. Paper presented at the Xenakis International Symposium, London

Dashow, James. 1996. “Fractal Interpolation”. Computer Music Journal 20(1), pp 8-10

Di Scipio, Agostino. 1995. “Inseparable Models of Materials and of Musical Design in Electroacoustic and Computer Music”. Journal of New Music Research 24(1), pp 34-50

Friberg, Anders. 2004. “A Fuzzy Analyzer of Emotional Expression in Music Performance and Body Motion”. Paper presented at the Music and Music Science, Stockholm, Sweden

Friberg, Anders. 2005. “Home Conducting – Control the Overall Musical Expression with Gestures”. Paper presented at the International Computer Music Conference, Barcelona, Spain

Heile, Bjorn. 2006. “Recent Approaches to Experimental Music Theatre and Contemporary Opera”, Music and Letters 87(1), pp 72-81

Hoffmann, Peter. 2000. “The New GENDYN Program”. Computer Music Journal 24(2), pp 31-38

Holtzman, Steven R. 1979. “A Description of an Automated Digital Sound Synthesis Instrument”. Computer Music Journal 3(2), pp 53-61

Kosko, Bart. 1993. *Fuzzy Thinking. The new science of fuzzy logic.* New York: Hyperion

Kreković, Gordan and Brkić, Igor. 2012. “Controlling Dynamic Stochastic Synthesis with an Audio Signal”. Paper presented at the International Computer Music Conference, Ljubljana, Slovenia

Kreković, Gordan and Petrinović, Davor. 2013. “A Versatile Toolkit for Controlling Dynamic Stochastic Synthesis”. Paper presented at the Sound and Music Computing Conference, Stockholm, Sweden

Luque, Sergio. 2009. “The Stochastic Synthesis of Iannis Xenakis”. Leonardo Music Journal, 19, pp 77-84

Manousakis, Stelios. 2009. “Non-standard Sound Synthesis with L-systems”. Leonardo Music Journal 19, pp 85-94

Monro, Gordon. 1995. “Fractal Interpolation Waveforms”. Computer Music Journal 19(1), pp 88-98

Orio, Nicola and De Piro, Carlo. 1998. “Controlled Refractions: Two Levels Coding of Musical Gestures for Interactive Live Performances”. Paper presented at the International Computer Music Conference, Ann Arbor, Michigan, USA

Serra, Marie-Hélène. 1993. “Stochastic Composition and Stochastic Timbre: GENDY3 by Iannis Xenakis”. Perspectives of New Music 31(1), pp 236-257

Stockhausen, Karlheinz. 1963. *Texte zur elektronischen und instrumentalen Musik.* Verlag M. DuMont Schauberg

Thomson, Phil. 2004. “Atoms and Errors: Towards a History and Aesthetics of Microsound”. Organized Sound 9(2), pp 207-218

Valsamakis, Nikolas and Miranda, Eduardo R. 2005. “Extended Wave-form Segment Synthesis, a Non-standard Synthesis Model for Composition”. Paper presented at the Sound and Music Computing Conference, Salerno, Italy

Xenakis, Iannis. 1992. *Formalized Music.* Stuyvesant, NY: Pendragon Press.

Yagaderi, Shahrokh David. 1991. "Using Self-Similarity for Sound-Music Synthesis". Proceedings of the International Computer Music Conference, Monetral, Canada

Young, Jonathan. 2010. "Rethinking Synthesis: Extending and Exploring Gendyn". BA thesis, University of Sussex: Department of Informatics.

Zadeh, Lotfi A. 1965. "Fuzzy sets". Information and Control 8, pp338–353

View publication stats