Putting the Pieces Together

Investigating Concatenative Sound Synthesis

Joshua Penrose

Graduate Independent Study

Dr. Gerald Philips

15 December 2007

Introduction

Concatenative Sound Synthesis (CSS), also referred to as "micro-sampling" and "micromontage", is a novel, computer-based process used in the organization and construction of sounds. While the necessary technology advances have already been achieved, there continue to be significant practical and legal bottlenecks within activity of this field. This paper endeavors to survey the legal issues, technological underpinnings, historical development, as well as the aesthetic ramifications of this growing area of investigation.

Granular Synthesis

The most basic technological precedent comes directly out of the developments of a digital signal process called granular synthesis. In granular synthesis, a sound file is segmented into very short durational units – as short as hundreds of a millisecond. These segmented units are called "grains". Thus accomplished, the grains can be treated in any number of ways for a desired sonic effect. A common application is to loop the grains in a way that "stretches" the sound file without altering the perceived pitch. For example, if one wanted to make a sample twice as long as the original, simply slowing down the playback would work (as is common with older technologies like the phonograph and magnetic tape), except that the resulting pitch would be one octave lower. With granular synthesis, the file can be segmented into grains. If the grains are small enough (e.g. 100 – 200 milliseconds), each grain can be "looped", in this example, played two times each consecutively, and the perceived pitch would remain the same as in the original sound file. Of course this process is not perfect. By interrupting the flow of the recorded sound

roughly 5 - 10 times per second, unintended frequencies and harmonics from subtle clipping are added in the resulting product. These unintended additions are called "artifacts" - byproducts generated from the procedure.

CSS Overview

In traditional granular synthesis, however, the sound file generally remains conceptualized linearly – each grain is a member belonging to the whole parent file. Within the area of concatenative sound synthesis, the grains are conceptualized as independent units; their relationship to the parent sound file (in terms of temporal ordering) is not retained. This subtle difference in conceptualization grows out of three significant technological developments.

First, in conjunction with the segmentation process, each grain is analyzed and subsequently assigned descriptive "tags". These descriptors are chosen from a fixed set as a result of the analysis process, and are one of the central organizational elements of CSS, known as "feature-as-control"[1]. The descriptors can be generated from the signal, perceptual, spectral, and harmonic contents of each individual grain. This development aids in the conceptualization of the grain as a discrete unit, in that its identity is no longer limited to its temporal position in relationship to a larger whole. Instead, the identity of the grain has significantly more depth and individuality, due to the many descriptor tags that may be applied. In the case of Diemo Schwartz's *CataRT* CSS software application,

[1]Adam Lindsay, "Introduction," *Journal of New Music Research* Vol. 35, No. 1 (2006): 1, http://www.jstor.org.

these tags are derived from the 230 descriptors native to the MPEG-7 file format, in addition to further descriptors calculated in the software during the analysis process.[2]

The second development is the availability of large collections of audio files. Not only are the available digital repositories of audio files vast (and growing), but also large storage devices for these files are becoming more and more economically feasible at the consumer level. These databases and storage units are another necessary development used in CSS. The larger the audio file database, the more likely the contents will provide a heterogeneous offering from which the creative agent may select. This relatively new technology, coupled with the process of assigning content-generated descriptors to individual grains, allows for an unprecedented collection (or in CSS terminology, "corpus") of raw material.

The third element, one that is more particular to CSS, is the algorithms used to organize the myriad of descriptive tags, with the user interfaces, which together make up the distinct CSS software applications. Some applications, such as Bob L. Sturm's *MATCONCAT*, uses a "target" sample, with its own content-based feature descriptors. In Sturm's piece "Concatenative Variations of a Passage by Mahler", Sturm uses a percussion *crescendo* from Mahler's *Second Symphony* as the target file for each of the 11 movements. In each of the movements, however, a different corpus of sources is used in assembling the grains from which the movement is generated. In this approach, the result is a kind of "automated micromontage", very much analogous to the Photomosaics of Robert Silvers.

---

[2] Diemo Schwartz, "Real-Time Corpus-Based Concatenative Synthesis with CataRT," *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx-06)* (September, 2006): 4.1, http://recherche.ircam.fr/equipes/analyse-synthese/schwarz.

However, in other software applications such as Schwartz's *CataRT*, a 2-dimensional X/Y field is supplied as a graphical user interface (GUI). According to how the user chooses to organize the grains (by which descriptive tags are employed), each grain appears in the GUI with relation to the different descriptive tags. With the click of a button, the grains within the corpus can be instantaneously re-organized by any number of chosen collections of tags, and "played" via the GUI. Schwartz states that the reason for this software design is to enable sonic exploration. In his paper accompanying the software, he states:

> Explorative real-time synthesis from heterogeneous sound databases allows a sound composer to exploit the richness of detail of recorded sound while retaining efficient control of the acoustic result by using perceptually meaningful descriptors to specify a target in the multi-dimensional descriptor space. If the selection happens in real-time, this allows [one] to browse and explore a corpus of sounds interactively.[3]

Historical Precedent

Where did all of this work come from? What were some of the forerunners? Many researchers and practitioners attribute John Oswald as a pioneer.[4] Beginning in the 1970's, Oswald began using a cut-up technique in his work. In an essay presented to the Wired Society Electro-Acoustic Conference in 1985, he coined the term "Plunderphonics", referencing the fact that these pieces were taken from a single artist in most cases, or from a few at most.[5] In 1988, he released an EP entitled

---

[3] Diemo Schwartz, "Real-Time Corpus-Based Concatenative Synthesis with CataRT," 1.

[4] Adam T. Lindsay, "Introduction," 2.

[5] John Oswald, "Plunderphonics, or Audio Piracy as a Compositional Prerogative," *Plunderphonics*, http://www.plunderphonics.com.

"Plunderphonics", which he sent to the press and several radio stations. The result of this was increased notice of his work. However, the project came to a halt in 1989 when CBS Broadcasting sued Oswald for copyright infringement (on behalf of Michael Jackson). The case was settled out of court, and Oswald was ordered to hand over the master, and destroy all remaining copies[6]. The sources from this album were taken from a wide variety of recorded music: The Beatles, Metallica, Michael Jackson, Dolly Parton, Beethoven, Stockhausen, Ligeti, etc. It is also interesting to point out that none of the Plunderphonics albums were sold, or were intended to create any kind of monetary profit. Nonetheless, CBS argued that Oswald's created sound objectsnegatively affected the image and existing musical objects of Michael Jackson's body of work. This point will be revisited later in the discussion of the aesthetic and socio-economic ramifications.

Without the aid of any computer or automated algorithmic process, Oswald assembled hundreds of small samples and spliced them together in novel ways via the method using magnetic tape. This process is not so much different from other forerunners of CSS, all of which predate Oswald's work. Pierre Schaeffer and his circle in France during the late 1940's and 1950's is one example. These composers used a similar process in creating music that came to be known as *Musique Concrète*. Other examples exist from the same post-war period, but from different parts of the world. Karlheinz Stockhausen's *Étude des 1000 collants* (1952) used cropped samples of hammered piano strings. John Cage, in creating *Williams Mix* (1953), divided his source

---

[6] Bob L. Sturm, "Concatenative Sound Synthesis and Intellectual Property: An Analysis of the Legal Issues Surrounding the Synthesis of Novel Sounds From Copyright-Protected Work," *Journal of New Music Research* Vol. 35, No. 1 (2006): 29, http://repositories.cdlib.org/escholarship.

sounds (approximately 600) into content-based categories from which the samples were selected according to chance procedures.[7]

Contemporary Situation

However, it was Oswald who first engaged the cultural connotative issue of identity and ownership in a context of ever increasing sound objects. His world was markedly different from the world of the aforementioned post-war experiments. In his 1985 Manifesto "Plunderphonics, or Audio Piracy as a Compositional Prerogative", he states, "Although people in general are making more noise than ever before, fewer people are making more of the total noise; specifically in music, those with megawatt PA's, triple platinum sales, and heavy rotation."[8] The world described by Oswald here, although over 20 years old, still closely resembles the world we live in today.

Recorded sound is ubiquitous in our world. It is reproduced in both background and foreground of our everyday experience. This sound interacts with our anatomical processes, which translate the vibrations into chunks of memory in our brains. This in turn actually enables us to hear these sounds, this process of creating memories. These memory patterns subsequently affect and inform (or reinforce) how we perceive and experience future sound in a subconscious, yet visceral way.[9] These sounds are indeed powerful.

---

[7] Diemo Schwartz, "Concatenative Sound Synthesis: The Early Years," *Journal of New Music Research* Vol. 35, No. 1 (2006): 10-11, http://www.jstor.org.

[8] Oswald.

[9] Bob Snyder, *Music and Memory: An Introduction* (Cambridge: The MIT Press, 2000), 1-18.

As both marketing tools and objects for sale in and of themselves, the engineers and distributors of these sounds have not underestimated them. The waveforms we hear today have undergone several evolutionary stages over the last several decades. The most recent developments in the art of mixing and mastering have fueled a kind of loudness "arms race". Compression and limiting, processes which reduce the peaks in a wave, thereby enabling the entire wave to be amplified within a particular dynamic range, exist in all kinds of proprietary forms, most of which exist only within the budgets of large studios and radio stations. If a radio station chooses not to invoke a proprietary compression process, they risk not being perceived as "loud", or as "full" as the next station on the dial.

Legalities

This is an uncomfortable situation for the consumer. These waveforms, which have become denser - packed with more sonic data, are unavoidable. They are everywhere. They enter our bodies where they have a persistent, residual effect. And yet our access to these vibrations, the rights that one has over their use, is ambiguous. The holders of copyright have an economic stake in mitigating any ambiguity (the money that the entertainment industry spends on cyclical anti-piracy campaigns underscores this point). The industry that profits from the creation and distribution of commercially viable sonic artifacts promotes a strict hands-off approach when it comes to consuming their wares. However, if human beings in this context are rationalized as mere economic actors, then not only is one's very person altered by these waveforms through constructed memory patterns, but also one is further objectified by serial passive consumption.

Theodor Adorno's insights into *homo oeconomicus* are illustrative of this point: "If the exchange form is the standard social structure, its rationality constitutes people; what they are for themselves, what they seem to be to themselves, is secondary."[10] This conceptualization of a human being is unsupportable. Certainly there are a myriad of appropriate responses to such a situation. Oswald's solution is stated metaphorically at the conclusion of his 1985 essay:

> Difficult to ignore, pointlessly redundant to imitate, how does one not become a passive recipient? Proposing their game plan to apprehend the Titanic one it had been located at the bottom of the Atlantic, oceanographer Bob Ballard of the Deep Emergence Laboratory suggested 'you pound the hell out of it with every imaging system you have.'[11]

A strategy of constructive engagement, he argued is a valid solution to passive consumption. While Oswald made this statement before his dealings with CBS Broadcasting, it is important to note that he did not discontinue his work in this medium.

Bob L. Sturm' paper "Concatenative Sound Synthesis and Intellectual Property: An Analysis of the Legal Issues Surrounding the Synthesis of Novel Sounds From Copyright- Protected Work" surveys the legal precedent established regarding music sampling, and as it may or may not apply to CSS. Sturm, who is also a composer, draws out several points of difference between traditional sampling and CSS by investigating how the legal system would deal with his CSS-based work "Concatenative Variations of a Passage by Mahler". Although there are significant barriers constructed around traditional sampling in commercial music through legal precedent, CSS as a process transcends this dialectic. Perhaps one of the ironies of development of the big business

---

[10] Theador W. Adorno, "Subject and Object" in *The Essential Frankfurt School Reader*, ed. Andrew Arato and Eike Gebhardt (New York: Continuum, 1988), 501.

[11] Oswald.

that has grown out of copyright protection, is that copyright was initially granted in order to stimulate research, and the creation of new works and technologies. In order to stimulate this kind of activity, the U.S. government granted these temporary protections. Those engaging in these kinds of activities historically required more time to cover their investments in time and capital. The writers of the constitution could hardly have anticipated the current buying and selling of copyright as stock in a portfolio. Sturm argues that this historical intent of copyright protection is still upheld in the legal precedent. He concludes,

> Through the doctrines of de minimis and fair use, copyright-protected work may be used without fear of punishment in researching and developing CSS. The creation of rich databases from copy-right protected work is defensible since their compilation serves a completely different function than the originals.[12]

CSS and Cultural Salience

Beyond legal defensibility, another element that makes CSS a favorable domain for creative endeavor is the sheer ubiquity of commercial music. Whether a proponent of one particular genre over another, or perhaps even antagonistic towards commercial music altogether, the experience of it is a unifying factor. Regardless of the medium, it would be difficult to imagine a day when a person does not hear commercial music in some kind of broadcast advertisement in either a public space, during travel, or in one's home (and no genre, whether "popular" or "classical" is exempt). We can't help but be familiar with these sounds. And it is not only the "tunes", with which we are familiar, but also the elements that are more difficult to describe. Again, Oswald: "The precarious

[12] Sturm, 32.

commodity in music today is no longer the tune. A fan can recognize a hit from a ten-millisecond burst, faster than a Fairlight can whistle Dixie. Notes with their rhythm and pitch values are trivial components in the corporate harmonization of cacophony."[13] These elements that are more elusive, these timbres, are what make up the most basic formants that are common in the contemporary context of music making.

Western art music has a long history of quotation. Quotation offers the opportunity to add depth, or an enlarged multi-dimensionality to an experience. However, any quotation presupposes shared knowledge. It assumes that the listener can recognize the quote, and trace it to its source.[14] The current situation with regards to classical music and classical music education is, at best, anachronistic. A shared knowledge of the masterpieces of the western tradition cannot be taken for granted. To appreciate the music of the common practice era with any depth, one must learn the vocabulary of the common practice era. This doesn't happen by default in the contemporary context. By appropriating timbres that are held in common among contemporary listeners, CSS offers a way into re-engaging the listener in new ways with the material with which he/she already is already familiar.

(re)Formations

In a recent work entitled *(re)Formations* (2007), I engaged with the sampling of copyrighted work as raw material. It is written as a structured improvisation for three performers, each manipulating software instruments I designed for laptop computers, and

---

[13] Oswald.
[14] Lindsay Vickery, "Nine Aspects of ~~Beauty~~ Appropriation," (Edith Cowan University, 2004), 1.

manipulated by USB gamepad controllers. The piece consists of three movements, each consisting of a different corpus of sources.

The first movement, *(re)Constituted*, is sourced from an amalgamation of popular music samples, spoken word, mechanical sounds, and percussion instrument samples. At the time of the creation of this piece, I had not yet discovered the algorithmically driven computer software applications that are becoming common within the CSS field. Therefore, I resorted to manual extraction and manipulation of each grain from within my corpus. Without any target descriptors, I chose the treatment and concatenation of the grains intuitively.

One of the performers in the first movement, *(re)Constited)*, played only from a one minute long concatenation of vowel sounds, each taken from a collection of popular songs. I chose the 10 most downloaded singles from the iTunes online store on March 3, 2007, and sampled the free 30-second preview clip of each single. The performer, via the software instrument, can then "scrub", or move the starting location back and forth within the one-minute long concatenation of vowels. The performer also has the flexibility to choose, the length, attack and decay of their gestures, as well as a series of various signal routes to further color the material. A similar scheme is provided for the other two performers, but with spoken word and mechanical sounds, respectively.

The second movement, *(re)Constituted*, is generated entirely from sounds sampled during the live performance of controlled feedback via a specially designed instrument. The third movement, *(re)Generated*, is all sourced from 24 distinct recordings of the fourth movement of Beethoven's 5th Symphony. In this movement, *(re)Generated),* the graphical score in combination with the software instrument, moves

through a series of 29 moments of the original work. Each moment is temporally disjunct with relation to the source work. However, within each moment, all three performers explore material from the same moment of the original piece.

All of the pre-compositional procedures were done manually, and had a significant impact on the direction of the composition. The incredible amount of time spent on the extracting and treatment of the grains significantly limited how much time was spent exploring the source material in preparation for the ensemble improvisations.

Further developments

(re)Formations also engages with the notion of manipulating and performing pre-constructed databases of grains in real-time as improvisational performance. This adds an additional interesting juxtaposition between the intention of the sourced material and the realization of its newly appropriated form. Particularly interesting to the author is the use of recorded concert music; music intended for the concert hall, abstracted and objectified into a recorded unit of sale, then subsequently mined for raw materials used in a live performance. While Oswald and his predecessors were making new sound objects from pre-existing sound objects, the introduction of new technology and the development of gestural capture as novel controllers for organized sound allows for even further work in the area of performance applications for CSS.

Conclusion

The technological advances in the field of CSS, particularly those of content-based descriptive tags, large heterogeneous databases of digital sound files, and software

applications for their organization, have opened new vistas for musical exploration. The existing body of works incorporating methods of micro-sampling have only scratched the surface of possibilities. While the ramifications of the sampling of copyrighted material may be perceived as threatening, the precedent established thus far is only tangentially related to the field of CSS research and experimentation. Furthermore, the precedent only reinforces the spirit of copyright protection, which is indeed the stimulation of research and experimentation. Given this condition, the ubiquitous timbral memories available as a unifying element and cultural phenomenon makes CSS a field ripe for creative activity.

Bibliography

Adorno, Theodor W. "Subject and Object" in *The Essential Frankfurt School Reader*. Edited by Andrew Arato and Eike Gebhardt. New York: Continuum, 1988.

Oswald, John. "Plunderphonics, or Audio Piracy as a Compositional Prerogative." *Plunderphonics*. http://www.plunderphonics.com.

Lindsay, Adam T. "Introduction." *Journal of New Music Research* Vol. 35, No. 1 (2006): 1-2. http://www.jstor.org.

Schwartz, Diemo. "Concatenative Sound Synthesis: The Early Years." *Journal of New Music Research* Vol. 35, No. 1 (2006): 3-22. http://www.jstor.org.

Schwartz, Diemo. "Real-Time Corpus-Based Concatenative Synthesis with CataRT." *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx-06)* (September, 2006). http://recherche.ircam.fr/equipes/analyse-synthese/schwarz.

Snyder, Bob. *Music and Memory: An Introduction.* Cambridge: The MIT Press, 2000.

Sturm, Bob L. "Concatenative Sound Synthesis and Intellectual Property: An Analysis of the Legal Issues Surrounding the Synthesis of Novel Sounds From Copyright-Protected Work." *Journal of New Music Research* Vol. 35, No. 1 (2006): 23-33. http://repositories.cdlib.org/escholarship.

Vickery, Lindsay. "Nine Aspects of ~~Beauty~~ Appropriation." Edith Cowan University, 2004.