Self-similar Synthesis
On the Border Between Sound and Music

by

Shahrokh David Yadegari

B.S., Electrical Engineering
Purdue University, West Lafayette, Indiana
1982

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1992

© Massachusetts Institute of Technology 1992
All Rights Reserved

Signature of Author ....
Media Arts and Sciences Section,
School of Architecture and Planning
August 25, 1992

Certified by ....
Tod Machover
Associate Professor of Music and Media, MIT Media Laboratory
Thesis Supervisor

Accepted by ....
Stephen A. Benton
Chairman, Departmental Committee on Graduate Students

Self-similar Synthesis
On the Border Between Sound and Music

by

Shahrokh David Yadegari

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning
on August 25, 1992, in partial fulfillment of the
requirements for the degree of
Master of Science
at the
Massachusetts Institute of Technology

Abstract

A study of the application of self-similarity to music synthesis was conducted with special emphasis on the relationship of form and material in art. Tonal and serial form in music was put in perspective in relation to self-similarity. The relationship between form and content was presented both in ambiguous communication systems such as music, and in mathematical systems in relationship with Gödel's incompleteness theorem[30]. These communication systems were related to the main topic of Schoenberg's ideas of form[40], which is "comprehensibility", and to the uncommunicatability of Kierkegaard's "faith"[21].

Auditory qualities were defined as "sound" and "music" using a definition for musical communication over a self-similar channel whose plexus is the relationship between form and content. The term "musical timbre" was introduced in contrast to the timbre of sound, and a uniformity among the different time scales of musical perception (i.e., form, rhythm, and pitch) was established. Schoenberg's theory of harmony was studied and the physical continuum of consonances and dissonances was extended to the relationship between sound and music (i.e. physical and psychological effects of music).

Self-similarity, self-referentiality, and chaos were briefly explained. A simple but intuitive, explanation of a class of self-similar signals were represented. The results of an analysis of some pieces in this context was presented.

It was established that serialism is a powerful basis for computer music, and the use of self-similarity is a logical step toward the evolution of music. A synthesis method based on self-similarity was devised and implemented. No distinction is made between sound and music, or form and content in this paradigm. A few techniques for using this system were described and the results were presented as audio examples on an accompanying digital audio cassette.

Thesis Supervisor: Tod Machover
Title: Associate Professor of Music and Media, MIT Media Laboratory

This work was supported in part by Yamaha Corporation.

Self-similar Synthesis
On the Border Between Sound and Music

by

Shahrokh David Yadegari

August 25, 1992

Reader: ______________________________

Rosalind W. Picard
Assistant Professor of Media Technology, MIT Media Laboratory

Reader: ______________________________

Xavier Rodet
Professor of Computer Science, Université Paris VI
Maître de Recherches at IRCAM

Acknowledgments

This work is as much Tod Machover's as it is mine. I had thought about this thesis before I arrived at MIT; however, the leap of faith was made by Tod and not by me; for that, I know, I will always be indebted to him. His stability, diligence, broad range of knowledge, support, and care for the most conceptual parts as well as the finest details, consistently encouraged and surprised me. From what I learned through my interaction with him during the past two years, Tod became a professor to me in the true sense of the word and I do not have adequate means to thank him for it. And finally, I would like to thank him for being a true musician.

I would like to thank my readers Rosalind Picard and Xavier Rodet who were always there for my simple-minded and abstract questions regarding precise matters of science. Xavier's help in many conceptual issues already started in 1987.

Special thanks go to Dan Ellis for creating the audio signal processing tools, without which the synthesis system would not have gotten off the ground. He has also been a rich source of technical help and insight for me. I would like to thank Barry Vercoe for his insightful help and comments during the proposal stages of the work, and for creating Csound.

The Media Lab and especially Music&Cognition has been a wonderful environment for thinking, working, and playing. I would like to thank Alan, Andy, Betty Lou, Bill, Casimir, Dave, Fumi, Janet, Jeff, Jin, Joe, Judy, Kathryn, Linda, Mary Ann, Mike H, Mike T, Neil, and Robert. I will especially remember Bill Gardner, Dan Ellis, Dave Rosenthal, Janet Cahn and Robert Rowe for debugging talks, proofreading and much moral and technical support. Andy Hong's art of sound was always a savior. If you are going to listen to the accompanying tape, you should thank Janet Cahn who saved you from listening to my voice. Stan Sclaroff provided solid code for the least square fit algorithm which I did not even have to look at.

David Nahamoo, who always had logical answers to my metalogical questions, is the one who started this project many years go. Esmaeel Tehrani, Hossein Omoumi, Sally Sargent, Khosro Pishévar, and Isabella Khan formed my understanding of music. Working with Miller Puckette had many influences on the initial design of the system. Discussions with Gerhard Eckel were always challenging, informative, and enjoyable. The structure hierarchy was designed in summer of 1989 in Vienna with Gerhard who taught me that music and visual art are related. Cort Lippe's consistent passion for music always amazed me. Many thanks go to Lev Koblyakov for many hours of sharing his rich knowledge of music with me.

Janet, Judith, Julia, Karen, Laura, Mok, Patrick, Stephan, and Thomas built my very pleasant reality in the past two years.

Finally and most importantly, I want to thank my family and long known dear friends Althea, Bijan, Fardad, Farhad, Fariba, Faryar, Ferial, Hayedé, Mammad, Saied, Shahriar, Shahyad, Shidokht, Sorour, and Yadollah for being a part of my life.

"For Bella"

Contents

List of Figures

List of Tables

Chapter 1

Introduction

I am told that many readers will first read the abstract of the thesis and if they are interested, they will read the introduction and conclusion; finally, if those parts have an inviting taste, the readers will proceed to read the rest of the thesis. Interestingly enough, these are the two chapters that are usually written last (or at least that is true in this case). At first, just by the fact that these two chapters were written at the same time, I had a difficult time separating their materials. However, once they were written, I could clearly separate them. Actually, to my official "readers", this introduction acts as a conclusion as well, since they have already seen the rest of the material.

Other than being a requirement for my graduation, this thesis attempts to communicate something to its reader to create a relationship between the reader and the content of the thesis, or, in other words set up the context for it. However, it assumes many relationships already. For example, it assumes that the reader has the thesis physically in his or her hand and can read it[1]. Perhaps, it does not even need to assume that much either. The thesis may be available electronically, or the ideas of the thesis may actually be transmitted through the mind of a third person. In this view, the thesis assumes some kind of relationship which acts as introduction to this introduction. If we think of the problem in classical information theory, we can say that the thesis wants to transmit some information to the reader. Even though the content of the thesis has been solidified through the process of archiving at the Massachusetts Institute

[1] Even though this thesis has many graphs and an accompanying audio tape, for the sake of simplicity, let us only talk about the words of this thesis.

of Technology, the amount of information which every reader obtains from it is different. Let us assume that this thesis is only available in English. To a person who does not have any knowledge of English, this thesis provides no information. However, perhaps the potential still exists since the person can study the language, and then read the thesis. To an entity which has no relation whatsoever to the thesis, there is not even the potential of transmission of information. We may conclude that the greater the relationship between the thesis and the reader, the greater the potential for transmission of information. However, if this thesis was created in a single instance of time, and the author had not gone through any changes himself, this thesis would have offered no information to the author either, who has perhaps the greatest relationship to the thesis, since there would have been nothing new in the thesis for the author to learn.

Let us make the situation a bit simpler. If this thesis, in the most rigorous way, proved a fact generally known as true, (e.g., the sun would be seen in the sky tomorrow assuming there would be no clouds), it would offer no information to its readers since they would know that fact already. Therefore, if the thesis says something that is known as a true statement to all beings, it cannot transmit any information to them[2]. On the other hand, if the thesis stated a fact generally known as false (e.g. the sun would blow up in a year), the first thing the reader would doubt would be the assumptions and reasoning of the thesis; since the reader already knows that the conclusions are wrong, the thesis still would not provide any information.

Let us assume that the thesis has a single message. If the transmitter (the thesis) and the receiver (the reader) both clearly agree or disagree on the truth value of the message, there can be no transmission of information. Therefore, if communication is transmission of information, *we can only communicate through ambiguity.* This is a paradoxical situation, since we usually attribute communication with clarity. Once we accept such a paradox as our starting point of communication, we cannot be completely sure of the truth value of any knowledge which we have received of the world. This problem is explained by Weaver as follows[44, page 96]:

> One essential complication is illustrated by the remark that if Mr. X is suspected not to understand what Mr. Y says, then it is theoretically not possible, by having Mr.

[2]The question of whether information exists when there is no perceiver for it is up for discussion; however, we believe, there is little content in that question.

> Y do nothing but talk further with Mr. X, completely to clarify this situation in any finite time. If Mr. Y says "Do you now understand me?" and Mr. X says "Certainly, I do," this is not necessarily a certification that understanding has been achieved.

In Shannon's discrete theory of communication, the amount of accepted information of every event depends on other knowledge. If we hear a sentence and we know the person who sends the sentence to us, we can judge the truth value of the sentence by what we know of that person. Thus, the truth value of the sentence which is the content of the message is dependent on its context which is what we know of that person. However, what we know of that person is the result of a series of "judgments" about that person's past history, to which this new sentence will add itself. However, due to the reasoning presented above, we can never be sure of the complete truth of our judgment[3].

If we recreate this scenario in our own mind, there is no need for any information to have any truth value. Truth values are attached to our sensations for the sake of communication, even if the communication is to oneself. For example, if we heard the bark of a tiger (which sounded hungry) and if we were sure that it came from a speaker, we attach a "false" truth value to the statement: "there was a tiger in the room". However, if we heard the barking and, from its acoustical elements, deduced that the sound was transmitted from the throat of a tiger, and on top of that, we physically saw the tiger, rather than thinking about truth values we would try to get out of that room. Therefore, we act according to a certain coherency among our senses governed by a faculty which we may call "common sense". It has been our experience that the idea of "common sense" or intelligence in general is treated as something which is not related to our physical self. It is known to be a faculty which understands meanings. However, it is not clear where the combination of our senses go through a transformation which suddenly change our physical sensations to meaning. When we communicate with others, we create collective entities (i.e. societies) which themselves possess a certain level of intelligence. These societies will in turn be able to understand and act *independently* of the individuals in the same way that we are able to act independently of the cells composing our bodies. If we try to explain such situations in a linear and logical manner we run into many paradoxes. For example, we assume that we are free, yet we have to abide by the laws of society. We accept a certain selection

[3]This is the largest flaw in being judgmental about the world, and above all, about the people around us.

process in nature which suggests that only the fittest will survive, yet we can see much altruistic behavior in nature which helps the underdog. Perhaps the biggest paradox of all is the physical experience of life and death. These experiences are simply sensations; however, once we assume that we have a faculty called intelligence which can understand these situations, we run into self-referential paradoxes.

In this thesis we have approached the problem of communication and comprehension from a different angle. This project started as an art project. The engineering of the system went through a scientific research process, and while writing the thesis some philosophical and psychological conclusions were made. The subjective meaning of music in the mind of the author was used as an assumption of the work. This may seem as a very unscientific approach. However, if we replace the word "music" with "faith", such work can be thought of as philosophy which borrows from Kierkegaard and Omar Khayyam[11]. Kierkegaard says[21, page 71]:

> On the one side, it has the expression for the highest egotism (to do the terrible act, do it for one's own sake), on the other side, the expression of the most absolute devotion, to do it for God's sake. Faith itself cannot be mediated into the universal, for thereby it is canceled. Faith is this paradox, and the single individual simply cannot make himself understandable to anyone.

At the same time it is rather difficult to put science and philosophy apart, as Chomsky writes[4, page 2]:

> In discussing the intellectual tradition in which I believe contemporary work finds its natural place, I do not make a sharp distinction between philosophy and science. The distinction, justifiable or not, is a fairly recent one. In dealing with the topics that concerns us here, traditional thinkers did not regard themselves as "philosophers" as distinct from "scientists." Descartes, for example, was one of the leading scientists of his day.

If this thesis is stating the truth, I do not know this in its every detail, and I know that I will never know. I am also sure that there are wrong statements in the thesis; however I do not know where they are yet, otherwise I would have corrected them. If the thesis is taken as a mathematical system, by the fact that there exists a wrong statement in the system, we announce the system as a whole wrong and in need for correction. We can never know if the correction needed is only for that single wrong statement, or if the system as a whole has to be re-implemented, redefining its assumptions and operations. We believe that we should look at this thesis as a mixture of true and false statements.

False statements can easily be hidden within true statements such as: "The statement '$1 + 1 = 3$' is wrong". We believe projecting such layering of true and false statements, as well as the continuum between truth and falsity, upon physical matter, can create a uniform relationship among our different levels of perception through which we can simultaneously understand our individuality as well as our universality. When communication happens, a universe is created by the ensemble of the communicators *and the communicative entity*; or in other words, the communicative entity is created according to the relationship between the communicating parties.

We believe that music is a form of communication where such issues can be studied through the relationship of form and content. The technical part of this thesis consists of a synthesis method which provides uniform control over the micro and macro-structures of sound. Thus, the definitions of the structures of synthesis not only define the small-scale structures (which can become the material to the perceiver) but also the large-scale structures (which can become the form). A synthesis language, with an eye toward a graphical interface, was developed to support the definition of such structures. Some results of the synthesis method are presented and analyzed, and the synthesis method itself is explained toward the end of the thesis.

We will study tonality and atonality in the context of Arnold Schoenberg's ideas and theories. We believe that the idea of tonal form deriving from the internal structures of harmonic sound is the central theme of his theory, by which he established a physical relationship between consonances and dissonances. We shall extend Schoenberg's idea, which apparently was meant to address normal to large-scale levels of music perception, to the structures of sound itself. We shall also propose that the relationship between consonances and dissonances can be extended to a highly perceptual level, which we call the sound and music relationship. We shall attempt to establish definitions for sound and music in a context where music is modeled as transmission of information, reaching the conclusion that form and material has to be treated in the same way, especially in computer music where we have the freedom of creating any type of sound. We shall suggest that, contrary to some cognitive psychologists' and composers' beliefs that serialism is not in accord with our cognitive system, serialism is natural and necessary for the evolution of electronic and computer music. Some of the works and ideas of Stockhausen, especially those concerning the uniformity of perception, will be briefly analyzed.

When we assume the unity of form and material, we are also assuming the existence of self-similar or self-affine structures. A short explanation of self-similarity and chaos, which is where the physical manifestations of self-similarity were first observed, will be given. The idea of self-referentiality, which we believe to be the underlying concept behind self-similarity, is mentioned in connection with Gödel's incompleteness theorem, and two cases of self-referentiality in literature.

We shall also study a class of signals called $1/f$ noise which have been seen in different instances of nature, including music. The purpose of this study is to create an intuitive feeling about what $1/f$ noise is and what its characteristics are, which implies a sense of (perhaps statistical) self-similarity in the signal it characterizes.

Overview

Chapter 1 introduces the problem of comprehension and puts the rest of the thesis in perspective.

Chapter 2, Sound or Music, is a study of Schoenberg's theory of tonality. The main purpose of this chapter is to establish a physical continuum between the physical and psychological effects of music, which we call sound and music. In the context of the problems presented in this chapter, the object is to establish a relationship between sensations and meanings in music. We shall also establish the fact that this continuum is non-linear, and can be modeled by a self-similar structure. The most important idea to understand from this chapter is that all forms come from the inner necessity of the material, or as Kandinsky says[3, page 152]: "*The form is the outer expression of the inner content.*" We will also show the unity of form and material in the context of some of the works of Stockhausen. We have tried to show that self-similarity is the natural necessity and outcome of the unity of form and material. We make no assumption about the knowledge of the reader concerning self-similarity in this chapter, and hope that the concept will intuitively emerge from the arguments. However, one can read chapter 3 before reading this chapter, if one is interested to read this chapter with some knowledge of self-similarity.

Chapter 3, What is Self-similarity?, is a portrait of self-similarity and its underlying concept, self-referentiality.

Chapter 4, Self-similarity in Sound and Music, is a technical presentation of a few cases of self-similarity in music. Specifically, we have tried to make the problem of $1/f$ noise more intuitive. Even though in this chapter very little technical knowledge is assumed, and no formulas have to be understood, this chapter may be skipped by those who do not like to look at formulas. This chapter very lightly suggests that it is possible to study music (i.e. meaning) without making any judgment on the "intelligence" (e.g. memory or musical training) of the listener.

Chapter 5, Self-similar Synthesis, is the most original part of this thesis. In this chapter, we shall put the problem of composition with computers in context, explaining that the process of composition has to define not only the organization of the piece but also of the material. We shall define a synthesis technique based on the principles of self-similarity and present some of the results we have obtained. Many audio examples accompany this chapter.

Chapter 6 is the conclusion.

Appendix A has the results of a simple analysis we did on 57 different pieces. The analysis is related to chapter 4 and $1/f$ noise.

Appendix B provides simple descriptions of the examples on the accompanying audio tape. Much care has been taken for the sound quality of the audio examples, and we suggests that the examples be listened to on an audio system with good low and high frequency response.

Appendix C is an explanation of the principles used in composing *Morphosis* (1992), which is a piece composed by the author using the synthesis technique described in this thesis.

Chapter 2

Sound or Music

2.1 Introduction

In this chapter we will discuss the physical and psychological aspects of music, which in a sense is a way of distinguishing between its objective and subjective qualities. First we shall try to establish an awareness of the existence of such qualities. Further we will show that even though one is derived from the other, they are perceived in very different manners by our mind, thus separating the two concepts. And finally we shall try to unify the two concepts in a musical context, and show that in some perspectives their difference is only a matter of degree and not of kind. Thus, we will try to establish a continuum between the two. The chapter has three sections (excluding the introduction and conclusion). The first two sections define poles, and the third section attempts to unite the two in a continuum. This structure is also repeated in every subsection, where the same formula applies.

2.2 Music: Logical or Physical?

Music is a form of expression. For any music there has to be a listener, even if the listener is the musician herself or god; otherwise the act becomes only a gymnastic exercise of the body (for performance) or the mind (for composition). Music can have meaning in many different forms and layers. If we view music as a coherent assembly of proportions in time, we can go as far as describing the movements of the heavenly objects (stars, planets, molecules, and atoms) as a piece of music. Nevertheless, we think of music as an art form, and as with any other art form, music is a very subjective matter. This assumption implies that every person can have a different idea of what music is, and probably every person's idea of music is different from the other's.

In any basic communication system, information is interchanged between two (or more[1]) entities. Before any communication can be achieved, there has to exist a channel[2] upon which the information is transmitted. In the case of music, the coherence between how the two entities feel about music can create a channel. For example, peoples of the same culture may have similar ideas about music and that similarity in their minds can create a channel on which they can communicate musical ideas. If we take this issue to the deepest formal level possible, we arrive at the physical (e.g., genetic) similarities of the two entities (e.g., if they sense the auditory information in the same way). A dog is able to hear frequencies which are inaudible to humans, and music composed by a dog using those frequencies cannot even be heard by us. The physical similarity creates a relationship, and therefore a channel, between the two entities in a specific direction (which for the sake of clarity we call "vertical"[3]). This vertical relationship is an outcome of many years of evolution.

On the other hand the subjective meaning of music can also create a relationship, and therefore a channel, between the two entities. In the mind of an idealistic musician, music is a universal language; this means that in music one could convey a feeling to another regardless of culture, race, or even species. In this case the channel is more ephemeral, since intrinsically there is very little physical history which supports this channel except other ephemeral and ideal feelings such as honesty, truth, beauty, love, or god. However, we humans attach a rather special quality to this case, since it is only through quality of work that one can pass the boundaries of history and culture, and convey a musical idea. For example, the many hours of internal and solitary work and struggle of an instrumentalist are readily apparent to any

[1]If we assume more than two entities we will have to think about the three body problem, which is still a hard problem for human beings to think about.

[2]The words "channel" and "linear" have very precise technical definitions. In this chapter we have used these words in contexts in which it is difficult to be scientifically precise. These words should be taken in their technical sense, but not with a scientific precision.

[3]The words "vertical" and "horizontal" are used to show two orthogonal axes. In this section, their orientation in space may not have any meaning; however, later they will be used for the time/frequency relationships where they have more literal meanings.

ear that chooses to listen[4] regardless of their differences in culture, style, or taste. This is the moment that one feels that the music flows, and interestingly enough in such situations the complexity of music becomes hidden.

This relationship or channel between the two entities on the aesthetic level acts on an orthogonal axis (horizontal) to the one previously explained. However, as soon as information (e.g., a musical idea) is passed through this horizontal axis (e.g., the aesthetic channel), the channel becomes vertical since the communication proves the existence of the channel and becomes part of the history and therefore creates room for evolution of that channel vertically. On the other hand, two entities that are culturally so close to each other, to the extent that they can be called identical copies, have very little to communicate to each other, even though they have a channel with enormous capacity for communication. Anytime one of them tries to be original, he needs to step away from the culture and therefore decrease the capacity of the channel. Anyone who has tried fusion of music in different cultures or even in different styles knows that this is a very difficult task, and can only be done through quality and hard work. In this case it is the vertical axis which is "sacrificed" to a more ephemeral channel. Any system of communication can be perceived in this manner, in which the channel becomes a plexus of orthogonal axes, where one axis can transform to another depending on how information is transmitted through the plexus.

Every musician knows the moment of total synchrony in feelings with another musician in a musical activity. This feeling can be created when listening to a performance, or, more powerfully, while one is performing. This synchrony is an unstable and paradoxical situation. Let us examine a simple and powerful instance of this situation in a case of improvisation between two performers. While performing, the sound that the performers create is not only a function of the musical structure they start with, but also a function of the instantaneous communication between them. If they both are thinking and feeling exactly the same, they have a very strong vertical channel, yet as far as the performance, one of them is superfluous since they are exactly identical. As their minds and feelings wander away from each other, they create a new entity, which is the instantaneous music being created according to the balance of their being related yet apart from each other. If they wander away too far from each other

[4] Such a choice means that one has to be able to go into the state of "not thinking" when listening.

this entity disappears and they will be playing two solo pieces at the same time. Here we can think of a continuum characterizing the state of their playing. One end of this continuum is when they are exactly the same, and the other end of the continuum is when they have nothing whatsoever in common.

This continuum is not a simple linear line. First let us examine its boundary conditions. In order to reach the ends of the continuum, we have to push the concept logically and formally to its fullest extent. If we push their state of thinking and feeling so it is the same as their very physical beings, they actually become the same entity and there will not be any way of distinguishing them from each other. While being at this point of complete sameness, it will be impossible for them to move apart from each other, since, because of their sameness, one will imitate the other. If we push their state of being completely apart, they become random noise to each other, and in that case they will never be able to establish any channel between them and therefore their state of communication will never move from that end of the continuum either.

Formally speaking, this is also true for any point on this medium. As soon as we analyze the state of communication, we can factor out their common factors, thus singling out their differences and creating a local continuum. In this way, we define their state of communication as the boundaries of that local continuum. Since according to the reasoning presented above none of those points can move from their position, their communication has to stay in that mode forever. We can look at this point on a different angle as well. Any new development between the two performers has to go through a paradoxical test of a selection process. A truly original theme cannot be introduced since, due to its originality, it will not have any relation to communication and could therefore stop the performance. The original theme could be ignored by the receiver, in which case the communication has not moved from its state. The receiver could try to understand the new development, but due to the originality of the idea the receiver cannot establish a channel with the idea, and it becomes impossible to understand the new theme. Therefore, in this context, a communicable original idea is not really an original idea, and an important part of the act becomes the balance between originality and comprehensibility. If the balance is natural and uniform, it is the balance itself which becomes original and not the idea.

2.2.1 Physical and Psychological Effects of Music

If the existence of an entity covered the complete continuum of time and space, it would be imperceptible to us. "Existence" itself is an entity (concept) which abides by this law. Any time we assume an existence subjectively, we also assume the negation of that existence objectively. If we assume that we have a consciousness and a mental faculty, we should also assume that we are nothing but a collection of matter in a single point of space. Physically speaking, we hear sound through our ears, and thus, the immediate aspect of music becomes a "thing" in the vibrations which reach our ear. Stockhausen refers to this experience as "receiving vibrations"[47, page 31]:

> When I speak about receiving vibrations, I am referring to the simple banality that everybody experiences a constant bombardment of rays from the cosmos.

Even though music connects itself to many of our activities, and in that case becomes something else, in its purest form it is a piece of sound — a collection of vibrations. On this point of view, in *Concerning the Spiritual in Art*, Kandinsky explains the effect of colors in the field of vision as follows[3, page 152]:

> If you let your eye astray over a palette of colors, you experience two things. In the first place you receive *a purely physical effect*, namely the eye itself is enchanted by the beauty and other qualities of color.

And further he states:

> But to a more sensitive soul the effect of colors is deeper and intensely moving. And so we come to the second result of looking at colors: *their psychological effect*. They produce a correspondent spiritual vibration, and it is only as a step toward this spiritual vibration that the physical impression is of importance.
>
> Whether the psychological effect of color is direct, as these last few lines imply, or whether it is the outcome of association, is open to question.

The psychological effect is probably due to association as well as a direct result of the physical, and also due to the innate self-referentiality of the interaction between these two effects. The relationship between these effects is not as linear and simple as it may seem at first glance. The psychological effects are understood and realized by our mental faculties; however, our mental faculties are nothing but the collection of our psychological perceptions of the world built around our physical senses. Kandinsky explains the construction of such knowledge as follows[3, page 153]:

> This is the experience of the child discovering the world; every object is new to him. He sees light, wishes to hold it, burns his finger and feels henceforth a proper respect for flame. But later he learns that light has a friendly side as well, that it drives away the darkness, makes the day longer, is essential to warmth and cooking, and affords a cheerful spectacle. From the accumulation of these experiences comes a knowledge of light, indelibly fixed in his mind.

While thinking about the association between different psychological effects, we notice that they can manifest themselves on many different levels. For example, the effects of a single tone from one instrument may associate itself with the tone of another instrument, or the feeling of a piece of music may associate itself with a view of a landscape or with the imageries created in the mind by a piece of poetry. Some psychological effects are created according to the collection of other psychological effects. In this case, the lower psychological effects are acting as physical effects. Note that this is an important point of departure. An effect in our mind is a psychological effect because we humans define it that way and communicate it to each other in such symbols as the word "psychological" in our language; while looking at ourselves as a collection of matter, these psychological effects are nothing but the state of arrangement of the physical matter. As we go higher in the hierarchy of perception and the association between these effects, we are actually descending deeper into the primitive qualities of our physical being. The more we move our consciousness to higher levels of our psychological mind, the more we understand about the state of our physical being. A work of art has very few boundaries, if any. When an artist feels and thinks about his art, his whole existence is in relationship with the work. Different forms of expression use different physical material and effect different physical senses, and therefore may seem to have different psychological effects. However, all the different forms of art in their "highest" psychological levels are perhaps affecting a single fundamental relation in our very "lowest" physical beings. Kandinsky writes[3, page 346]:

> All the arts derive from the same and unique root. Consequently, all the arts are identical.

And further, he discusses the similarities of music and painting:

> It is very simple at first glance. Music expresses itself by sounds, painting by colors, etc. facts that are generally recognized. But the difference does not end here. Music, for example, organizes its means (sounds) within time, and painting its means (colors) upon a plane. Time and plane must be exactly "measured" and sound and color must be exactly "limited." These "limits" are the preconditions of "balance" and hence of composition.

He also discusses how one can see, hear, smell, touch, and taste a painting and further says:

> Do not deceive yourself; do not think that you "receive" painting by the eye alone. No, unknown to you, you receive it by your five senses.

In this context, we could think of hearing a piece of sound as a purely physical experience, and listening to music as the psychological effect which this experience creates. The physical experience is probably very similar among living beings of the same species. However, as we try to understand the deeper psychological effects, we arrive at issues which are inherently subjective and cannot be objectified in principle. That is to say that by objectifying these issues we neither create nor gain anything. On the other hand, should we stay honest and true (a purely subjective matter) to the material of our study (which is music, and that means being musical), we could objectify any matter that serves the process of music. This is true because with every objectifying step we open many subjective doors useful for creation. Following this thread of thought we can objectify ourselves, ignoring all spiritual concepts and even life itself, and look at ourselves as simple matter.[5] Thus, music becomes nothing but a piece of sound.

2.2.2 Differentiating between Music and Sound

Perhaps the first thing that comes to our mind when asked to differentiate between sound and music is that the physical vibration generated by an instrument is the sound and the

[5]Spiritual concepts, especially those concerning life and survival, are important to us. Inquiring into the reason for that importance is philosophy, and that itself is a spiritual concept in the life of the one who inquires about that issue.

structure applied to the sound by the musician is the music. It is usually very easy to tell a good musician by a single note coming from her instrument, and there the differentiating parameter is the sound quality. This view can be argued against in cases like the sound coming from technologically more advanced instruments like the piano (from which one can get a good sound rather easily). However, the fact of the matter is that a musician spends a considerable amount of time and concentration to get a good sound out of the instrument. And this is not a fact that is learned once, but rather is a continuing effort in almost every new piece or performance. We can view that activity as an effort to match the music to the sound of the instrument or match the sound of the instrument to the music. Can we not say that the sound of a single note of a good musician contains music?

We do not need to stop our contemplation of this scenario with the trained musician. A good instrument maker is judged by the sound of the instrument he makes. Many years of training and craftsmanship goes into choosing and shaping the raw material for making the instrument. Can an instrument maker make an instrument that sounds good without thinking about music? If not, should we not call his efforts for creating an instrument which sounds good, part of the music played on that instrument? On the other hand, we can study a musically gifted child who will do something meaningful anytime he takes that instrument in his hand. Due to the isolation of the child's mind from preconceived notions, the music which comes out of that improvisation is largely dependent on the sound that the instrument makes. The more the logical part of the child learns about music, the more distant he gets from the sound. Sound is the most tangible form of music. A good musician is the one who uses this distance as a tool to bring these two opposites together. Once that process is experienced and learned, the farther the distance between the poles are, the more powerful their union becomes.

Music has the interesting property that it can be heard repeatedly. Every music lover has favorite pieces which he or she can hear over and over, and perhaps be drawn to the piece even more with every listening. It is very difficult to decide if such pieces are examples of how one thinks music should sound like, or if it is actually the sound of these pieces which form one's musical perception.

When music is notated, a set of instructions and codes are chosen which, in conjunction with the musical culture, try to communicate a musical structure to the performer. The music heard is not only present in what is notated but also in the subtle conscious or unconscious musical gestures of the musician. A Glenn Gould fan can repeatedly listen to his recording of Bach's French Suites, and find his humming with the music most beautiful. However, if we have a player piano play the piece accurately without any deviation in timing or dynamics, after understanding the structure of the piece, we will start ignoring the auditory information as static sound of the piano.

It seems fair to say that the basic structure of traditional orchestral music is laid out before orchestration; however, the act of orchestration itself (which creates the sound of the music) requires a deep understanding of the preliminary musical structures. The composer writes the music while having the sound of the music in mind. In contrast, orchestration becomes the process of creating sound textures while having the musical structures in mind.

Computers provide us with very powerful and precise control over the physical sound spectrum in time and frequency. Composers are now able to convey musical information through the evolution of timbres (for examples of use of the continuum of timbre see Machover[27] and Saariaho[39]). Such musical structures are conceived with an intention toward creating an abstract sound that the composer might imagine, and the physical sound is created according to the finally-evolved musical structures. Thus, separating the functions of sound and music in today's compositions can be troublesome, both for the listener and the composer.

2.2.3 The Dichotomy

We may be able to start putting some of the described qualities about sound and music into a form of dichotomy. For example, we can say that music is progressive and dynamic, while sound is instantaneous and static, or that music is alive while sound is lifeless. Dichotomies are created by applying a central duality principle to a subject. It is worth noting that most dualities can be created by a simple negation operation. Thus, we can separate sound from music by saying that what is sound is not part of the music, or in other words, what is music is not in the sound. However, in real life experiences and especially in the creative process, we try to break these boundaries, to create life from death, and beauty out of ugliness. Such thoughts make our path for the search of simplicity and beauty, complex and sometimes ugly. However, beauty is not a matter of right or wrong, or true or false. The path is as much a creation path as it is a search path. In our search we can reach for the most simple results by the most complex paths, and we can create the most complex results by taking the simple paths. It is only through balance, and understanding the continuum between the poles, that we can achieve stability and communication. Transcendence of the banal and ordinary is only the first step of creation; making sure that the creation is going to last the nature's selective process is an issue of survival. Originality is a source of transcendence while use of techniques and traditions create more support for survival.

One of the techniques of assuring a relative balance in music is tonality. The balance is created by a uniform coherency between musical structures and what they organize, which is the harmonic sound. In the next section we will examine tonality and its antithesis and try to establish a relationship between the two.

2.3 Techniques – Ways to Create Form

Techniques assure a certain balance in any activity, but they cannot create. When the systematic principles of a work of art are understood, those principles become common knowledge as techniques. Techniques create a channel between the composer and the listener. The composer uses the techniques to create the piece, while the listener, using the knowledge about the technique, looks for cues and feels a satisfaction when his expectations are met. From approximately 1650 to 1910, tonality was one of the most powerful techniques for creating form and coherency in music. In this section we will try to establish a relationship between self-similarity and the development of principles of tonality and, its counterpart, serialism. We shall show that the principles of tonality — a technique for organizing sound — come from the structures of the harmonic sound. Further we shall suggest that the technique of serialism should be regarded as creation of sound from musical structures.

2.3.1 Tonality

Charles Rosen defines tonality as follows[38, page 23]:

> There are so many conflicting accounts of tonality that it will be useful to restate its premises, axiomatically rather than historically for brevity's sake. Tonality is a hier-

> archical arrangement of the triads based on the natural harmonics of overtone series of a note.

Harmony has been one of the very basic principles of Western music in the last few centuries. The roots of harmony go far beyond music, and its most beautiful manifestation is found in the rules governing the movement of stars and planets in space. Before Schoenberg formulated his theory of harmony, tonality was the underlying principle in the theory of form in Western music. Tonality can be thought of as organizing the harmony of structures to a single point of gravity. This idea is itself embedded in harmony, and that is the same way that all harmonics of a tone are integer multiples of the fundamental. Following those principles, many forms and techniques for composition in tonal music have evolved.

Schoenberg says[41, page 19]:

> The material of music is the tone; what it affects first, the ear.

He further says about tonality[page 27]:

> Tonality is a formal possibility that emerges from the nature of the tonal material, a possibility of attaining a certain completeness or closure (*Geschlossenheit*) by means of a certain uniformity. To realize this possibility it is necessary to use in the course of a piece only those sounds (*Kläng*) and successions of sounds, and these only in a suitable arrangement, whose relations to the fundamental tone of the key, to the tonic of the piece, can be grasped without difficulty.

The "*certain completeness of closure by means of certain uniformity*" which Schoenberg talks about is in fact the uniformity of material and organization, or in other words, sound and music. He further says[page 29]:

> It [tonality] is one of the techniques that contribute most to the assurance of order in musical works — that order, consistent with the material, which so greatly facilitates the untroubled enjoyment of essential beauties in the music.

Again here he is talking about "*that order, consistent with the material*" which is the order of tonal music consistent with its material, or the harmonic sound.

Many schools of classical music, for educational purposes, treat harmony and counterpoint as two different and separate elements. It is perhaps a wrong approach to try to find out if harmony evolved according to the mixture of two or more melodies, or if the scales, upon which the melodies are based, were decided according to harmony. The well-tempered scale is obviously a compromise between harmonicity and position independence from the frame of reference. Rosen says[38, page 25]:

> Equal temperament absolves us from considering at length whether or not tonality is a 'natural' or a 'conventional' language. It is quite evidently based on the physical properties of a tone, and it equally evidently deforms and even 'denatures' these properties in the interests of creating a regular language of more complex and richer expressive capacities.

If in fact all the scale values would have been chosen according to the physics of a tone, then the values of the elements of the scale would have to be changed any time we change our frame of reference (the tonal center). In other words, the scale gives us a constraint on the continuum of frequency, which is created according to the vertical requirement of cohesion in physics of the tone (sound). However, if we fully abide to this constraint, the horizontal plane (the plane for melodies) becomes so constrained that melodies with the same intervals will sound different in different positions in reference to the tonal center, and further, movement of the tonal center will require repositioning of the elements of the scale.

When a singer[6] wants to learn a melody, there are two orthogonal requirements which have to be learned, namely time and frequency. For now we refrain from involving the sound parameters and ignore such important factors as timbre and dynamics. The simplest case is when the basic structure of the melody is exactly in the well-tempered (or some other) scale and notes all have equal durations. Then, the act of learning the melody is to remember the sequence of the scale values which have to be sung. However, that only makes up for very expressionless melodies, since the structure of the melody does not carry itself into the duration of every note. A singer can create the feeling of this carry over by changing the intonation or duration of notes. In the simplest case, the amount of deviation has a linear relationship with

[6]We use a singer here rather than an instrumentalist, as many believe that one has learned a melody only when one can sing the melody.

the structure of the melody. However, in reality, this relationship is not linear. An intonation which matches the structure of the melody is a requirement in the vertical plane, and the order of the progression of the notes in time is a requirement in the horizontal plane. When the singer hears the melody for the first time, she gets an impression of the central idea. In formal terms, this central point is the basic structure of the melody, and perhaps on the plane of poetics we can call that the "emotional meaning" of the melody. Once the first impression is learned, the singer builds a relationship with the melody, and with every repeated listening or performance of the melody the vertical and horizontal requirements change to better accommodate the basic structure. A simple change in an element of one of the planes may require changes in the values of the elements in the same plane as well as the orthogonal plane. A very subtle change in the intonation of one note may require changes to many other notes as well as changes to the duration of the notes in time. In this scenario we also have to account for the relationship of the mood of the singer and the "emotional meaning" of the melody. Obviously that relationship is not by any means linear either. Through this evolution, the melody "comes alive", and it finds its own character which specifies the vertical (scale) and the horizontal (durations and operations in time such as vibrato) requirements. Thus, even though the melody was primarily defined by the scale and durations of the notes, once it is subjected to our thought and emotion, it sets its own terms for scale and durations.

Schoenberg says[41, page 23]:

> Intuition and inference (*Kombination*) assisted in translating the most important characteristic of the tone, the overtone series, from the vertical (as we imagine the position of all simultaneous sounds) into the horizontal, into separate, successive tones.

This process is true not only for a melody and a singer or a composer and a piece, but also for a society and a musical culture. This evolutionary point of view is perhaps a much better way of looking at the development of such principles as scale, voice leading, and chord progression which shape the principles of tonality[7].

The harmony and voice leading rules imply an interrelated network of constraint for prolonging the structures of a tone and at the same time it is the structures of a tone which leads

[7] Or is it the other way around? Do these rules come from tonality, or is it these rules that shape tonality itself?

us to realize this plexus of constraints. This technique also creates a paradigm for the interplay of content and form. A melody has to agree with its harmonic context. A piece may be composed of two different themes whose harmonic (and basic structures) are far apart from each other. The first theme sets up its own harmonic context, and through the grammar of harmonic modulations we can accommodate the second theme. However, if these two themes are too far apart from each other, the integrity between the content (the themes) and the form (the harmonic context) of the piece is broken and the relations will not be comprehensible as a unit. Again if the two themes are too close to each other, they cannot stand for themselves and they become variations of each other. Such circular thoughts are part of the process of evolution. The paths for combinations are endless, and here we need a musical intuition to prune the paths. On the evolution of principles of harmony Schoenberg says[41, page 26]:

> It is much more correct to say that the development of harmony was not only essentially influenced by melodic principles, that the development of possibility of voice leading was not only essentially influenced by harmonic principles, but that in many ways each was actually determined by the other.

As more and more we try to apply the operations implied by our intuition and inference to the tone, in our mind we derive a different entity (music) from it. The more these operations are applied to the tone, the more distant the new entity is going to be from the tone. However, the closer these operations are to the nature of the tone, the more they will emphasize the structures in the tone itself, and therefore, the closer the entity becomes to what characterizes a tone in our mind.

Chords are instantaneous entities, and melodies are progressive. A chord progression in a sense is a form of melody in itself. On a higher level, key changes, which use pivot chords for their connections, create another sense of melody. All these progressive elements in different layers are heard by the sensitive ear, and in order to have a closure for the piece as a whole, all these melodies have to be related to each other (i.e., be in harmony). The harmony in the structure of a chord is a harmony in sound, and the harmony of the melodies in different layers of time is a harmony in music. The relationship of such logical entities as music, movements, and melodies has to abide by the same rules that govern the relationship of their parts, namely sound, tones, and chords. Schoenberg's theory of harmony is largely based upon this idea on the level of tones and chords, and the continuum which exists between consonances and dissonances.

Consonance and Dissonance

On the relationship between melodies and chords, Schoenberg says[41, page 26]:

> If the scale is imitation of the tone on the horizontal plane, that is, note after note, then chords are imitation on the vertical, notes sounded together. If the scale is analysis, then the chord is synthesis of the tone.

And he further says[41, page 26]:

> The triad is without a doubt similar to the tone, but it is no more similar to its model than, say, Assyrian reliefs are to their human models.

By recognizing that all the simple elements of scale (i.e. the scale tones) as well as the compound elements (i.e., the chord and melodies) are all derived from the same principle, and realizing the relationship between analysis (i.e., breaking down an object to its parts) and perception of impression (i.e., the holistic view), he establishes a continuum between the consonances and dissonances[41, page 20]:

> That is to say, here the musical ear does indeed abandon the attempt at exact analysis, but it still takes note of the impression. The more remote overtones are recorded by the subconscious, and when they ascend into the conscious they are analyzed and their relation to the total sound is determined. But this relation is, to repeat, as follows: the more immediate overtones contribute *more*, the more remote contribute *less*. Hence, the distinction between them is only a matter of degree, not of kind.

There is a very subtle and important point in this analogy. What this relationship is implying is a relationship between content and form. Before we can grasp this point we need to fully understand the function of tonality and its implications.

Function of Tonality

Tonality is a technique for assuring a certain integrity in a composition; its major goal is to make comprehension easier. However, we pay a great price for this service, and the price is being constrained to a single type of relationship in the composition — namely the integer harmonic relationship. In a tonal context, the relationship of the tonic to every note and all operations resulting to those notes have to be clearly comprehended. The tonal structure of chords is used as a reinforcement of the physical structure of the tonic. The operations which are applied to notes to build chords or melodies, as well as the operations applied to chords and chord progressions, are themselves completely in accord with the structure of the tone. All these operations and state of relationships act as agents of context in relation to the tonic. Thus, every chord implies a certain context. For example, a stable chord implies a certain resolution in the musical idea, and a dissonant chord implies tension. Therefore, formally speaking, the relationship between the overtone series of notes cannot be used as content of musical meaning, since if it is in accordance with the context then it becomes part of the context and cannot be distinguished. On the other hand if this content — the relationship of the overtone series of the notes creating the chord — would not agree with the contextual requirements, we run into a contradiction of form and content. Schoenberg says [40, page 217]:

> Formerly the harmony had served not only as a source of beauty, but, more important, as a means of distinguishing the features of form.

and on the functions of tonality, he says[40, page 277]:

> Though the development of tonality was by leaps and bounds, though it has not signified the identical thing at all times, its function has, nevertheless, been one and the same. It has always been the referring of all results to a centre, to a fundamental tone, to an emanation point of tonality, which rendered important service to the composer in matters of form. All the tonal successions, chords, and chord-successions in a piece achieve a unified meaning through their definite relation to a tonal centre and also through their mutual ties.
>
> That is the unifying function of tonality.

Schoenberg repeatedly emphasized that the function of form is for comprehensibility[40, page 316]:

> I have, above all, repeatedly pointed out the *purpose of all forms*: a layout which guarantees comprehensibility.

Notice the emphasis on "purpose of all forms". This is a central idea in Schoenberg's theory. If we break the principles of tonality, according to this idea, all we risk is comprehensibility and not any musical content, and it is following this belief that he says[40, page 216]:

> What distinguishes dissonances from consonances is not a greater or lesser degree of beauty, but a greater or lesser degree of *comprehensibility*.

What distinguishes dissonances from consonances is the way the overtone series of their parts — the two or more combined tones — match each other. After all, what does combining two tones mean? He says[40, page 270]:

> The question is more important than it seems at first; nevertheless to my knowledge it has not previously been raised. Although all imaginable and far reaching problems have been considered, no one has yet asked: How, after all, can two tones be joined one with another?
>
> My answer is that such a juxtaposition of tones, if a connection is to be brought about from which a piece of music may be the result, is only possible because a relation already exists between the tones themselves.
>
> Logically, we can only join things that are related, directly or indirectly. In a piece of music I cannot establish a relation between a tone and, let us say, an eraser; simply because no musical relation exists.

Notice how far he has pushed his ideas when he is considering a piece of music resulting only from connection of two tones; and at the same time he has to make such seemingly simple-minded examples as the relationship of a tone and an eraser to communicate his idea. No doubt we can in turn use Schoenberg's reasoning to imply that the fact that a tone and an eraser are brought up in a single sentence shows they do have a relationship with each other. Then, the question is if they have a musical relationship or not, and if there is any border between what is called "musical relationship" and other kind of relationship. Again, we can use Schoenberg's own reasoning about consonances and dissonances and establish a relationship between musical relationships and other types of relationships, and say that this is matter of degree and not of kind.

However, the importance of what Schoenberg is saying is not in the "truth" of his statement, but in what it communicates to us, which is a relationship between what characterizes music in our minds and nature. After explaining that the major and chromatic scales both are derived from the nature of the tone itself, he goes on to say that our music making is just simply an imitation of nature[40, page 272]:

> And here is the answer to our question regarding the possibility of interconnection of the tones. It is founded on the fact that in the sounding tones and its nearest relative, the union and the companionship of the tones is continuously demonstrated to our ear, so that we do nothing more than imitate nature when we make use of these relations.

In other words, in the language of this essay, music is nothing but a piece of sound. If we apply this reasoning to every aspect of our mind and our intelligence, we reach a very obvious conclusion: our mind and our intelligence are simply an imitation of nature, and therefore they, and whatever results from them — including this sentence — are part of the nature.

Let us reiterate what the continuum of consonances and dissonances mean. Before, the relationship of the harmonics were only used as form; now we can denounce that type of form and use this relation as part of the content in the music. Therefore, in this way we can communicate musical ideas (relationships), which were communicated horizontally and progressively, vertically and instantaneously. These situations had come about in music before Schoenberg formulated his theories. In preparation for explanation of his "twelve-tone method", he writes[40, page 216]:

> Richard Wagner's harmony had promoted a change in the logic and constructive power of harmony. One of its sequences was the so-called *impressionistic* use of harmonies, especially practiced by Debussy. His harmonies, without constructive meaning, often served the colouristic purpose of expressing moods and pictures.

And he further writes:

> One no longer expected preparations of Wagner's dissonances or resolutions of Strauss' discords; one was not disturbed by Debussy's non-functional harmonies, or by harsh counterpoints of later composers.

Once Schoenberg formalized the functions of such chords, he went further and declared that tonality was not an eternal law of music. His "twelve-tone method" and serialism were methods which were devised to assure form in music which did not depend on tonality.

2.3.2 Serialism

At first glance, it may seem that music today does not need any order. That one can just put any number of notes together and make music. This is a common misconception among the public (that we have encountered) about the music of 20th century. The freedom of music from tonality did not bring anything new to music; on the contrary, it took a very prominent history of form away from it. In the paradigm of serialism, achievement of order, while satisfying comprehension requirements (a subjective matter), is a much more difficult task. Tonality, by its rich history of developed complexity, provides the composer with a framework in which a certain amount of comprehensibility is guaranteed. Once we take this framework away, the composer has to create his own framework for assuring order in the musical communication. For that, Schoenberg devised the "twelve-tone method", about which he says[40, page 207]:

> The weightiest assumption behind twelve-tone composition is this thesis:
>
> Whatever sounds together (harmonies, chords, the result of part-writing) plays its part in expression and in presentation of the musical idea in just the same way as does all that sounds successively (motive, shape, phrase, sentence, melody, etc.) and it is equally subject to the law of comprehensibility.

Let us, for the sake of comprehensibility, assume that an object is comprehensible[8] when a meaning is attached to it, and therefore, it is required to have a certain degree of self-sufficiency. This implies that now simultaneous notes (which are called chords in the tonal context) are used as content with musical meaning, and they do not imply any context resulting from the tonal relationship of the simultaneously sounding tones. Please note that here we are talking about musical content in what was previously considered as instantaneous sound. Therefore, the content of simultaneous sounds does not have to abide by any eternal prefixed rule; now the content of a simultaneous sound (which is the relationship of all the overtones to each other

[8]Please note, if we question the meaning of comprehensibility, which after all is what Schoenberg is emphasizing, then neither our explanation nor his comment have any meaning.

in conjunction with the development of those elements in the duration of the sound) can be anything that serves the underlying musical idea.

And it is following this idea that Schoenberg introduced the idea of *Klangfarbenmelodien*, which is progression of tone colors independent of pitch or harmony. On this idea Schoenberg writes[41, page 421]:

> I think the tone becomes perceptible by virtue of tone color, of which one dimension is pitch. Tone color is, thus, the main topic, pitch a subdivision. Pitch is nothing else but tone color measured in one direction. Now, if it is possible to create patterns out of tone colors that are differentiated according to pitch, patterns we call "melodies", progressions whose coherence (*Zusammenhang*) evokes an effect analogous to thought process, then it must also be possible to make such progressions out of the tone colors of the other dimension, out of that which we call simply "tone color", progressions whose relations with one another work with a kind of logic entirely equivalent to that logic which satisfies us in the melody of pitches.

And he ends his "*Theory of harmony*" by the following passage:

> Tone-color Melodies! How acute the senses that would be able to perceive them! How high the development of spirit that could find pleasure in such subtle things!
>
> In such domain, who dares ask for theory![9]

Let us examine these two ideas, *Klangfarbenmelodien* and "simultaneous sounds which are subject to the laws of comprehensibility."[10] Please note how the role of pitch and tone color changes in the *Klangfarbenmelodien* concept. In traditional tonal music, the pitch structures are conceived and then the musical idea is orchestrated, which creates the sound of the music. However, as Schoenberg states, pitch is nothing but timbre reduced to a one dimensional instantaneous value. And therefore, the tonal system is dependent and capable of producing only a single type of (musical) timbre — the natural harmonic timbre. Now pitch has become a secondary issue, and one is still capable of communicating a musical idea without any dependency

[9] Would Schoenberg say the same thing, if he had computers to help him create and control new timbres? And would he still feel the same way, if he had heard the contemporary computer music of today?

[10] These two ideas are really portraits of the same concept; however, since they have been used in different contexts in the music of 20th century, we will discuss them as separate entities.

on it. This is communication based upon a type of progression which we previously understood as sound.

The same analogy applies to the "simultaneous sounds which are subject to the laws of comprehensibility." Here the simultaneous sounds create a single timbre which has to be understood. Again, please note, how the roles have reversed; in tonal context a unity was assumed, and pitches (melodies) or chords, and then timbres, were used to portray that unity. Any deviation from this unity was only to build a stronger context for affirmation of the assumed unity. However, in an atonal context the unity is created only when all the parts are combined together; it is a physical unity rather than a logical pre-assumed unity. As one of the steps which has to be taken for new music, Schoenberg says[40, page 137]:

> The path to be trodden here seems to me the following: not to look for harmonies, since there are no rules for those, no values, no laws of construction, no assessment. Rather, to write parts. From the way these sound, harmonies will later be abstracted.

How far can we move away from this presumption? What Schoenberg attacked was tonality of pitch, and he created a method to substitute the function of form in his music. So, why not apply the same idea to *all* parameters of music? This principle was what many of the composers following Schoenberg's footstep used as agents of form for their music. They applied the serial idea to parameters such as duration (rhythm), loudness, and timbre (orchestration). Rhythm defines a constraint plane in the horizontal dimension. If we apply serialism to form and rhythm and finally to inner structures of the sounds, we create aperiodic waveforms. The most aperiodic sound is white noise. Thus, we can define a continuum between tones and noise. On periodicity and noise, Stockhausen says[47, page 93]:

> So the continuum between sound of fixed pitch and noise is nothing more than that between a more and a less stable periodicity: the noisiest noise being the most aperiodic. This discovery of a continuum between sound and noise, the fourth criterion of electronic music, was extremely important, because once such a continuum becomes available, you can control it, you can compose it, you can organize it.

John Cage took a different route. He used organized chance to control the process of sound and not the sound itself, and in this way freed music from his own personal intentions. His approach is to move from thoughts about order to no thoughts about order. On choices of what to do with sounds he says[2, page 10]:

> Or, as before, one may give up the desire to control sound, clear his mind of music, and set about discovering means to let sounds be themselves rather than vehicles for man-made theories or expressions of human sentiments.

How far can we push such ideas as serialism and organized chance? Understanding John Cage's philosophy about music requires a certain approach to life, and for now, we will refrain from any linear reasoning to interpret what he suggests. Serialism implies a complete breakdown of the channel of communication between the composer and the listener, since if we fully abide by the idea, we are left with nothing in common between the composer and the listener. Fifteen years after Schoenberg completed his "*Theory of Harmony*" he writes[40, page 259]:

> Tonality's origin is found — and rightly so — in the laws of sound. But there are other laws that music obeys, apart from these and the laws that resulted from the combination of time and sound: namely, those governing the working of our minds.

Why can we not apply the same argument against tonality to any other formal concept in music? If tonality is a uniform structure in music and sound, and if in fact, as Schoenberg seems to imply, the real content of music is our thought, why should we not find the same elements, which free pitch from tonality, in the "*rules governing the working of our minds*"? In other words why can we not free "*rules governing the working of our minds*" from the rules governing the working of our mind? In fact we can, and in this way we will free music from communication and we will reach a subjective idea of music. Every person can have his own idea of music; however, the music cannot be communicated at all, perhaps not even to ourselves.

2.3.3 Tonality of Atonality

When Schoenberg started to compose with his twelve tone method, he only serialized pitch and not other parameters. He says[40, page 87]:

> Coherence in classic compositions is based — broadly speaking — on the unifying qualities of such structural factors as rhythms, motifs, phrases, and the constant ref-

> erence of all melodic and harmonic features to the centre of gravitation — the tonic. Renouncement of the unifying power of the tonic still leaves all the others in operation.

In his atonal works, Schoenberg also avoided any chord which implied a tonal context like any combination of major or minor thirds[40, page 263]. Does that not sound like a contradiction? A complete reversal of tonality is itself a type of tonality. He recognizes this issue, and he wrote that his conscious avoidance of such circumstances was only due to the fact that he felt that the veil of the classical tonal culture was still too heavy. He felt that listeners still could not hear tonal chords, which in tonal context require a specific progression, only for their colors. The question of why Schoenberg did not apply his method to all parameters himself, and why he avoided tonal chords, is an important question. He recognized that for a musical idea to be understood a relationship has to exist between its parts. Schoenberg never liked the term atonal; however, this is the term that has been since used to characterize his music. He says[40, page 283]:

> 'Atonal can only signify something that does not correspond to the nature of tone.' And further: 'A piece of music will necessarily always be tonal in so far as a relation exists from tone to tone, whereby tones, placed next to or above one another, result in a perceptible succession. The tonality might then be neither felt nor possible of proof, these relations might be obscure and difficult to comprehend, yes, even, incomprehensible. But to call any relation of tones atonal is as little justified as to designate a relation of colours aspectral or acomplementary. Such an antithesis does not exist.'

All these issues go back to what concerned Schoenberg the most — comprehensibility. What lies in the music is not only *"what lies in the music"* but also the mentality that creates it and the way it is communicated[11]. The music, the composer, the musician, and the listener are all part of the musical idea, and in the same way that Schoenberg says "we can only join things that are related," they themselves — music, composer, musician, and listener — have to be related to each other. By breaking every kind of tonality in pitch, rhythm, harmony, thought, emotions and even common sense, we may create new ideas in music; however with every new

[11] The mentality that creates the music and the way it is communicated are all apparent in what lies in the music.

step in that direction we break a channel of communication. If there were no such a thing as *time*, we would have to just sit and do nothing since it seems that with every step toward progress, we regress in a different direction in what we are trying to achieve. Fortunately, we live in a temporal world, and falsities of today can be truths of tomorrow, and it is only through this understanding that an artist, or for that matter any being, can feel that he or she can be free to think and still stay hopeful. Schoenberg was (and still is) misunderstood, and about the labels put on his music he says[40, page 283]:

> If audiences and musicians would ask about these more important things and attempt to receive answers by listening, if further they would leave the idle talk and strife rather to the school-masters, who also must have something to do and wish to make a living, I, who have the hope that in a few decades audiences will recognize the *tonality* of this music today called *atonal*, would not then be compelled to attempt to point out any other difference than a *gradual* one between the tonality of yesterday and the tonality of today. Indeed, tonal is perhaps nothing else than what is understood *today* and atonal what will be understood in the *future*.

Indeed, Schoenberg's work was a gradual movement in music. He formulated what was already being practiced. However, the act of his consciousness of '*how*' these impressionistic entities were used and how they could be formulated was perhaps a revolution, since now our point of view is different. In a sense, we can tell that by the fact that he brought his practice into a theory and explained it in a linear fashion, he changed *truth*. In a less stronger term, he broke an accepted truth, with a seemingly strong knowledge of its theory and practice, only to combine his internal inspiration — his internal truth — with it, and through a concise, diligent, and patient expression of himself, he returned his truth to the world outside of himself. What he made us conscious of is now a technique which we can apply to many aspects of music (and other forms of art and thought) to create new sounds and music. He explains his first inspirations about his method as follows[40, page 49]:

> I was inspired by poems of Stefan George, the German poet, to compose music to some of his poems, and surprisingly, without any expectation on my part, these songs showed a style quite different from everything I had written before. And this was only

> the first step on a new path, but one beset with thorns. It was the first step towards a style which has since been called the style of 'atonality'. Among progressive musicians it aroused great enthusiasm. New sounds were produced, a new kind of melody appeared, a new approach to expression of moods and characters was discovered. In fact, it called into existence a change of such an extent that many people instead of realizing its evolutionary element, called it a revolution.

Now we are confronted with a sense of ambivalence. First we are not quite sure of the nature of what has happened: is it an evolution or a revolution? Was something created, or did it evolve? Secondly we are not sure what is tonal and what is not tonal; it seems to be just a point of view. By the fact that a piece of music is a piece of music it has a tonality in its sense of existence. When we listen to it, first it is not being played, then it is played and then we go back to it not being played. Schoenberg and Stockhausen had also gone as far as saying that such music does not have a start or an end, calling the atonal sequence "endless melody", and therefore breaking the tonality of its existence. However, is this not a property of sound? Depending on our point of view a piece of sound can become music. A timbre does not have a start or an end. Atonal music is a type of sound on a very high level; it defines a new musical timbre.

J. S. Bach created (formulated, or helped the evolution of) a form for music based upon the structures of the harmonic tone and a uniform connection between the music based on it — tonal music. What Bach did to music, Schoenberg did to sound[12]. As pointed out before, Schoenberg formulated a connection between form and content — music and sound — and he became aware of this fact by understanding the relationship of the tonal form and the harmonic sound. Schoenberg did not only emancipate pitch, he emancipated the structures of sound. Notice that in the last quote, Schoenberg says *"New sounds were created"*.

Schoenberg also says that this method creates impressionistic music that has to be listened to differently. He implies a very primitive way of listening to this music, and that is how we listen to sound, impressionistically. We receive the vibrations and get a feeling from them; there is very little analysis. At the same time, Schoenberg asks that every simultaneous sound be

[12]Schoenberg believed that there were similarities between historical situations, but he says: *"I am no Bach"*[40, page 119]. Schoenberg was inclined to call Bach the first twelve tone composer[page 117].

subject to the laws of comprehensibility as far as the musical idea is concerned. This point of view means that when we listen to music as a whole we are listening to sound, and when we try to comprehend the sound by the progression of its elements we are listening to music. As mentioned before, the tonal form is only capable of creating harmonic musical timbre, while with serialism we are free to create any type of (musical) timbre we please. Comprehension of serial music is not easy. Webern was so optimistic about atonality that he thought people would be humming atonal melodies in the street by the 1950s. Serialism has been attacked for its problems of comprehension, which is precisely what Schoenberg was most concerned about. It is my belief, that such attacks are short-sighted in their view of what serialism is. In today's music, it is rather difficult to separate the functions of sound and music. Especially in computer music, composers are able to convey musical information through control of sound parameters. Lerdhal calls the holistic effect of parts of Boulez's *Le Marteau sans Maître* (1954) pure sound[24], when he says:

> *Le Marteau* does not feel complex in the way, for example, that Beethoven or Schoenberg do. Vast numbers of nonredundant events fly by, but the effect is of a smooth sheen of pretty sounds.

In our analysis, this is no shortcoming. Creating serial music by using acoustical instruments is like building a house with a single type of material (e.g., building electrical circuits and water pipes out of bricks). To label serialism as a system which is not in accord with our cognition, metaphorically and literally, implies that our cognition is based on integer and not real numbers: let us stop using real numbers! The work of the past century concerning serialism has been fundamental for electronic music, where serialism will be able to show its real fruits. Serialism is a natural concept for music whose potentials will not be understood until we have a natural theory for composition with computers. We would like to reiterate the fact that serialism, and perhaps any technique, used systematically without any musical intuition, can only create sound. Thus, serialism can be a foundation for the sound of computers played by human musicians.

Now that we are able to compose even to the finest structures of sound, and at the same time, by using algorithmic composition, create large-scale sounds using musical structures, our point of view toward material and organization changes; they become intertwined with one another. The unity of form and content is not only a convenient paradigm, but is a necessary step, technically and — far more important — aesthetically, for the future evolution of music.

2.4 Unity of Material and Organization

In this section we will look at the relationship between sound and music in large-scale structures while reiterating some of the characteristics of what we have called sound. We shall suggest the term "musical timbre" to characterize the similar elements in musics which sound the same. We shall establish a need for a certain scale-independent uniformity in a piece of music which will also imply a certain uniformity in our perception. We shall suggest that such uniformity in music and perception suggests the unity of form and material through self-similar structures. We shall also suggest that the unity of material and organization (which we believe is the concept underlying serialism) seems to be a natural base for a theory of composition for electronic and computer music.

2.4.1 Sound — Recapitulation

Let us review what we have talked about so far in this chapter. In most of our analysis, we have focused on the relationship of the normal level of hearing to the micro levels (for an in-depth discussion of the different levels of musical perception see Koblyakov[22]). By now, we should have an awareness of such qualities as sound and music on any level of the musical communication process. That means that at any point that we focus our attention where there are structures below or above the focus point, we should be able to understand what we hear in terms of the sound and music relationship. In the communication process, every focus point by itself can be looked at as a point of trade-off between the channel of the communication — which is mostly dependent on the past — and the information which is transmitted over that channel. Where this trade-off between channel and information, or sound and music, or material and organization, becomes inherently ambiguous, we can use the ambiguity for communicating a musical idea. Once we remove the ambiguity by committing to a definition of our focus point, then the rest of the structures in relation to the focus point become clear as far as this communication process is concerned. That is to say that we become aware of the plexus which every focus point defines while acting as content, while the plexus, acting as context defines the focus point.

By now, we may have built an intuition about how this plexus is created for a tonal piece of music. This plexus is a somewhat subjective entity which is created by the relationships which exist in the structures of a tone, resulting in special forms and operations in time or frequency and creating a special type of musical timbre. Through the passage of time, not only the form has been affected by our consciousness of the structures of the tone, but also this form has helped us to better recognize the structures themselves. This effect can also be seen in the development of (almost all) instruments whose evolution not only changes how the form is used, but is deeply affected by the requirements of the form (e.g., the relationship of piano and piano reduction). We may also be able to see the sound/music relationship on higher levels; for example, we may agree that we can tell apart the music of two composers, or two different eras, by the sound and not by the music. One needs no academic music training to be able to learn to recognize a composer's style. It seems very plausible to say that we can recognize two different styles, in exactly the same way that we can recognize the timbres of two different instruments; the only difference is that one is the timbre of the sound and the other the timbre of music. Here, we would like to define the idea of a "musical timbre"[13] as the quality which makes two pieces of music different to us independent of any logical (conscious) analysis. This may seem vague; however, it is no more ambiguous than the definition (or the lack of definition) of sound timbre, which is whatever is left in the characteristics of the sound after we account for pitch, loudness, and duration[10, page 63]. We believe that the case where we are not able to tell the difference between two composers by their musical timbre, but by conscious analysis of their music, is similar to being able to tell apart the sound of two instruments only by conscious analysis of their partials.

Any time that we define an acoustical entity as timbre, we also have to define its instrument. For example, the timbre of piano is played by the piano, and the timbre of tonal music is played by the tonal form; or the sound of Mozart's music is played by his style, or the timbre of the music of Pierre Boulez is played by his compositional style[14]. This is not to say that a single

[13]This idea was first introduced to me by Marc-André Dalbavie during late-night discussions when I stopped him from working at IRCAM.

[14]The term "*Le Son Boulez*" is familiar among the composers and scientists of IRCAM (Institut de Recherche

composer has only one type of musical timbre. However, again, we come into the idea of unity in a composer's language, and one can usually feel the evolution of the musical timbre in the progression of the composer's pieces in her lifetime.

Music has its own evolution, and it is no surprise that usually the music of the composers who live in the same era sounds very similar. Their music, or in the other words its emotional content, may be completely different; however, due to social and cultural issues, what they hear and what they learn is perhaps similar. Therefore, they come up with instruments for their music which are very close to each other; that is one way that the musical language of an era comes about. The same analogy about sound and music applies to this level as well; however, there is a certain distinction on this level. The timbre of the music of different eras is played by a society of humans, and not individuals any more. The implication is that music separates itself from the personal freedom of the single individual, becoming an entity in itself.

The evolution of the material and organization of tonal music is the fruit of many centuries of work of musicians. Many composers of the late 19th century had digressed away from the formal requirement of tonality, not by conscious choice, but out of the necessity of feelings. Once Schoenberg realized why and how this path should be taken, the composers who wanted to be adventurous and revolutionary were suddenly confronted with a dilemma. The revolutionary who was ready to break barriers and tradition, came face to face with a space which had no barriers. Schoenberg formally broke all barriers of music on all levels by recognizing that the logical difference which had been assigned to consonances and dissonances was actually a physical continuum.

Schoenberg was not an anarchist. While discovering these principles, he also realized that the practice of music is very far from dogmatic theory. He understood the implications of blindly applying a newly founded theory to art would be useless. The only parameter he attacked was pitch, and even that only relative terms. He attacked the long-term relationship of pitches in form (long term being three or more pitches), and created a technique in which pitches are only related to one another, different from the tonal form where all pitches are only related to a single pitch. When asked about the further subdivision of the octave, something that has already evolved in monophonic music cultures, he first said[41, page 424]:

et Coordination Acoustique/Musique), the computer music research institute in Paris.

> However that may be, attempts to compose in quarter or third tones, as are being undertaken here and there, seem senseless, as long as there are too few instruments available that can play them.

In the second edition of his *Theory of Harmony*, Schoenberg reconsiders the question and adds a footnote, mainly to show that music cannot change by theory alone and that change has to come from musical necessity. It is unfortunate that Schoenberg did not know about computer music, otherwise he would understand that not only could there be instruments capable of playing all tones with the greatest precision, but that one can also control them with unlimited temporal accuracy. He says[41, page 26]:

> Perhaps here, once again, laws and scales will be erected and accorded an aesthetic timelessness. To the man of vision, even that will not be the end. He recognizes that any material can be suitable for art — if it is well enough defined that one can shape it in accordance with its supposed nature, yet not so well defined that the imagination has no unexplored territory left in which to roam, in which to establish mystical connection with the universe.

Did Schoenberg know that he himself proposed one of the greatest laws, which is lawlessness? The material for computer music is a strange beast; it has no intrinsic constraint, which means that it has no shape and no form; it is not only *not well enough defined*, it is not defined at all. In the other words, computer music (or music conceived in that spirit) has no material, and according to Schoenberg's argument, no form. Can we conclude that we cannot make music with computers? This is a paradox. From freedom we reach the point of no choice at all. However, we can live with this paradox by a paradoxical way of looking at the music of computers, which is to assume that form and material are the same parameter. It is paradoxical since when we listen to the music we feel the form, and we hear the material as well; however, the unity implies that if we go deeper into the structures of what we perceived as material we should find the structures of the higher level form again (or a form related to it), and if we look into that form we would find the same material again. This is so since they are both defined according to the same parameter. Stockhausen, who is one of the pioneers of electronic music, says[47, page 111]:

Harmony and melody are no longer abstract systems to be filled with any given sounds we may choose as material. There is a very subtle relationship nowadays between form and material. I would even go as far as to say that form and material have to be considered as one and the same. I think it is perhaps the most important fact to come out in the twentieth century, that in several fields material and form are no longer regarded as separate, in the sense that I take this material and I put it into that form. Rather, a given material determines its own best form according to its inner nature. The old dialectic based on the antinomy — or dichotomy — of form and matter had really vanished since we have begun to produce electronic music, and have come to understand the nature and relativity of sound.

2.4.2 Homogeneity of Music

All things from the lowest to the loftiest, from the smallest to the greatest, exist within you as equal things. In one atom are found all the elements of the earth. One drop of water contains all the secrets of the oceans. In one motion of the mind are found all the motions of all the laws of existence.

Khalil Gibran[13, page 46]

A piece of music is a single piece of music. This fact may sound like a simple truism, but it is not. How can we have a single physical entity? Without getting into deep philosophy or physics, we have to agree that everything is composed of its parts. Even though we consider music as a logical entity, it has to abide by this rule as well. However, would an artist admit that rules govern her most intimate aesthetical thought and emotions? If there exists such a rule, then it has to be a universal rule, not only true for that specific space-time and that specific piece of music, but also for all places, moments, and art. Once conceived[15], a piece of music breathes on its own, sets its own terms[38, page 7], and will be its own living entity. Schoenberg says[40, page 144]:

Thence it became clear to me that the work of art is like every other complete organism. It is so homogeneous in its composition that in every little detail it reveals

[15]For an interesting discussion on conception, as opposed to composition, of music refer to[40, page 166].

> its truest, inmost essence. When one cuts into any part of the human body, the same thing always comes out — blood[16]. When one hears a verse of a poem, a measure of a composition, one is in a position to comprehend the whole. Even so, a word, a glance, a gesture, the gait, even the colour of the hair, are sufficient to reveal the personality of a human being.

It is not only romanticism which unifies form and content; it is an issue intrinsic to our intelligence and the way we perceive the world.

A composition has a message which, however, is not a clear one. If the message is too clear, the listener gets bored before the piece is finished; if it is too complicated, it becomes difficult to grasp, and again is not interesting. If the piece is composed of different parts, by the end of the first part the listener should get a feeling of introduction which is coherent with the structure of the piece as a whole, not only on the first hearing of the piece, but on every listening. No matter at which level the piece is listened to, the introduction has to feel like the introduction. On the second listening, the listener grasps more structure in two directions. He hears the more detailed ornaments better, while a longer-term structure manifests itself. All these manifestations have to be in accord or, in other words, related to each other. The listener should be able to assign a relationship not only to the process in which these different layers of structure manifest themselves, but also, once manifested, to the feeling which these structures portray, while the feelings and the process which fleshes out the feelings have to be in turn related to each other as well. Again, all these relationships, which can become quite entangled if we try to follow them in every macro and micro structure, have to be connected to each other by a single relationship – a single sentiment. Schoenberg says[40, page 290]:

> Anyway, whatever one's views about the pleasure that can lie in conducting each part in polyphony independently, melodiously and meaningfully, there is a higher level, and it is at this level that one finds the question which needs answering in order to arrive at the postulate: 'Whatever happens in a piece of music is nothing but the endless reshaping of a basic shape.' Or, in other words, there is nothing in a piece of music but what comes from the theme, springs from it and can be traced back to it; to put it still

[16] Schoenberg would have been even more excited if he had known about DNA.

> more severely, nothing but the theme itself. Or, all the shapes appearing in a piece of music are *foreseen* in the 'theme'. (I say a piece of music is a picture-book consisting of a series of shapes, which for all their variety still (a) always cohere with one another, (b) are presented as variations (*in keeping with the idea*) of a basic shape, the various characters and forms arising from the fact that variation is carried out in a number of different ways; the method of presentation used can either 'unfold' or 'develop'.)

If a composition is rich enough it can be listened to more than once. While we may *think* that we *know* everything about the piece, the physical sensation of the sound will always surprise us. The introduction of a piece in the second listening has to follow the end of the piece after the first listening; therefore, the end of the piece has to act as a prelude to the beginning. When we assume such self-sufficiency in every part in every scale of perception, which says that every part has a message of its own, and at the same time we assume that the ensemble of all parts has a message which is related to the message of the parts composing the ensemble, we are assuming a sense of self-similarity or self-affinity.

One might suspect that: "This is a very simple minded way of looking at what is in music and does not take into account the composer's emotional complexity or the hard labor of the realization", however, we need to understand what self-similarity and its implications are. Self-similarity is a a very simple idea. However, its different ways of appearing in the physical world, and our thought and emotions are extremely complex.

When a composer is inspired, he imagines the whole piece at once. The inspiration seems to come from nowhere. Even though many of the elements of its creation (or evolution) process are dependent on the past, what characterizes it as original comes from nowhere. The inspiration seems to be self-sufficient, and by re-applying its own idea to itself, the inspiration grows. There are perhaps many contradicting accounts on this issue. Some composers may see a whole work in an instance and some may find the true self of the work during the compositional process. However, we believe that there is a point in time, which may not even become conscious to the composer, that the composition detaches itself from the composer and defines all its parts by itself. About inspiration, Schoenberg says[40, page 107]:

> This comes about because in my case the productive process has its own way; what I sense is not a melody, a motive, a bar, but merely a whole work. Its sections: the

movements; their sections: the themes; their sections: the motives and bars — all that is detail, arrived as the work is progressively realized. The fact that the details are realized with the strictest, most conscientious care, that everything is logical, purposeful and organically deft, without the visionary images, thereby losing fullness, number, clarity, beauty, originality, or pregnancy — that is merely a question of intellectual energy, which may only be taken amiss by those who themselves possess it and believe themselves entitled to despise it.

Briefly recapitulating:

The inspiration, the vision, the whole, breaks down during its representation into details whose constructed realization reunites them into the whole.

How does the "*constructed realization*" come about? In the mind of the composer, once she is finished with the mental work or when she is finished with the score? Or is it in the mind of the musician who reads the score and creates the sound? Or does it happen through the feedback of playing and listening at the same time? Or does the reconstruction happen in the mind of the listener who uses nothing but ears? Music has to be able to communicate itself, even if it is just to oneself. Therefore, should this *whole* not imply a coherency between all these wholes, in the mind of composer, musician, and listener. The path that the composer takes to realize an idea may be different from the path that a musician takes to learn the piece for playing. However, there is a certain feeling that remains the same in the mind of the composer and the experience of the musician, and that feeling is what makes that piece different from another piece. Again, this unity, this feeling, is not only a horizontal unity between the mind of the creator and the listener, but also a vertical unity in different levels of the perception of the piece. This last issue is very important in the practice of electronic music today. This perceptual relationship was perhaps the most basic principle which Stockhausen used for his electronic and acoustic compositions.

2.4.3 Unity of Perception

When we think of music in its linear form, we can isolate the different parameters in time and frequency. Even though used in a technical way, the term frequency can have many implications. For example, the way we usually think about an event which has a frequency of 0.1 Hz is not in the frequency domain, especially when we are thinking about music. The frequency domain in music is usually referred to as the way we perceive tones and their combinations, whose spectrum lies in the range of 20-20 KHz. This practice is strongly backed by the fact that we are not able to hear physical frequencies of less than 20 Hz[10, page 21].

Pitch or Beat

Stockhausen says[47, page 92]:

> What we perceive as rhythm from a certain perspective, is perceived at a faster time of perception as pitch, with its melodic implications.

If we take a stick and hit an object with it at a rate of once every second, we hear the sound of that object very consciously once a second. If the succession of the impulses of sound are precise within a certain amount of accuracy, we feel a sensation which we call sensing a beat. Now, if we speed up the rate of impulses from 1 Hz to 300 Hz, the sensation of what we hear changes in a very drastic way. In this case we hear a pitch at 300 Hz and will not feel a sensation of beat. However, if we speed up the impulses gradually, depending on the timbre of the object we experience different sensations. In general, the beat first changes to a texture in the fuzzy boundaries, and then it becomes a pitch. Structures of pitch and rhythm, which are classically two very different concepts, can be related to each other just by changing the scale in which they are being perceived. When we listen to very fast rhythms (such as African, Indian, or Persian drumming, minimal music of Steve Reich, or simply a roll of a drum) we do not consciously hear every impulse; we listen to the texture that these rhythms create. Many art rock musicians (such as Brian Eno) view their music as textures, which means that even if the underlying musical structures which are employed to create every layer sound simple (and in fact they usually are complex and only sound simple), and repetitive, the combination of the sounds together is a texture which is pleasing and interesting to the ear as a whole.

We can establish a relationship between pitch and sound, since we associate with both of them a feeling of instantaneousness. We can also establish the same relationship between rhythm and music for their progressive elements in time. However, we can change these relationships around. For example, in the context of tonal music, the basis of what is felt as consonances or dissonances is in the relationship of the pitches of what constitutes the chords, and much of the analysis which is based on the linear form usually views music as a sequences of pitches[38, page 29]. Therefore, in that regard we should relate pitch and music to each other. The connection between rhythm and sound is rather more difficult to grasp. If indeed we listen to rhythms as textures, we are listening to them as instantaneous entities. If we assume that music is information and sound acts as a medium, the sense of beat, by the assurance of being static, acts as a medium for a musical idea being transmitted as a form of melody on top of the beat. On a larger time scale, perhaps the feeling of form is not as much of a conscious entity, as the feeling of pitch or beat are.

If we are able to perceive a single musical idea in many different scales of perception as melody, rhythm, or form, and if indeed, it is the single musical unit which manifests itself as these apparently different perceptual values, what happens if a musical idea defines structures which lie between these perceptual boundaries? Does it not make more sense to believe that there exists a physical continuum between these sensations and (to put it in Schoenberg's term) that their difference is only a matter of *degree* and not of *kind*? To put these perceptual actions into separate categories implies that listening to music is a *logical* act, while the logic of it has no physical basis. To be more specific, our senses detect a certain coherency in different scales of time, and all of them are sensed at the same time. If these senses are not connected to each other through our physical apparatus, there has to exist a layer which *suddenly* changes all these sensations to "meaning" and creates a whole out of them. To assume that such intelligence can exist without any physical basis is inconceivable. Music is an imitation of sound in nature. Listening to music, as well as any other "intelligent" act we do, is a physical action and should not be explained by metaphysics. Our intelligence is nothing but a sensation, which itself comes from the physical connection of our five senses in time.

Uniform Time

Let us not digress too much from the subject at hand which is, after all, music and its practice of composition. The composer's inspiration is a timeless entity and its manifestation in time is only for the sake of communication. As discussed before, the purpose of all forms is comprehensibility. Whatever the psychological implications may be, the treatment of time as a unified entity is a much more natural view of composition than separating the different parameters in time.

Stockhausen says[47, page 46]:

> I think that the most important innovations in musical form come about from building on the relationships of the three time regions: form, which is everything that happens between, say, eight seconds and half an hour; rhythm and metre, which is everything that happens between one-sixteenth of a second and eight seconds; and melody, which is everything that is organized between one-sixteenth and one-fourthousandth of a second, between 16 and 4000 cycles per second. It is almost technically possible to stretch a single sound lasting one second, to a length of half an hour, so that you have an overall form which has the characteristics structure of the original sound. On the other hand, if you are able to compress an entire Beethoven symphony into half a second, then you have a new sound, and its inner structures has been composed by Beethoven. Naturally it has a very particular quality compared to sound resulting from the compression of another Beethoven symphony. Not to mention a Schoenberg symphony, because there are many more aperiodicities in Schoenberg; that would be more of a noise, whereas the Beethoven would be a vowel, because it is more periodic in its structure.

Stockhausen was well aware of the rich relationships between sound and music and used them in many of his pieces. In ".....how time passes...."[46], he discusses a system of composing "phase-durations" according to structures of pitch composition. He establishes the relations between beats in the same ways the overtones of harmonic sounds are related to each other. He also recognizes the fact that rhythms are perceived as textures, and from the idea of tone-colors devises a system for composing rhythm timbres which he calls "formant-rhythms". He used these ideas to compose *Zeitmasse* (1955-56), *Gruppen für drei Orchester* (1955-57), *Klavierstück XI* (1956), and *Carreé* (1959-60). In *Gruppen*, three orchestras surround the audience, with each orchestra having its own conductor, each playing in a different tempo. We can analyze this situation in the context of what has been mentioned in this chapter; it is as if every orchestra is a single instrument whose sound (timbre) is created by the musical structures played by the musicians of the orchestra using the sounds of their individual instruments.

Once we recognize the continuum of our perception in time, by controlling it we can use the continuum as a compositional tool. Stockhausen used this continuum as his basic medium of communication for the piece *Kontakte* (1959-60). About this composition he writes[47, page

95]:

> There is a very crucial moment in my composition KONTAKTE for electronic sounds, beginning just before 17' 0,5" in the printed score. A translation of the title might be 'Contacts', and the contacts are also between different forms and speeds in different layers. The moment begins with a tone of about 169 cycles per second, approximately F below middle C. Many of the various sounds in KONTAKTE have been composed by determining specific rhythms and speeding them up several hundred times or more, thereby obtaining distinctive timbres. What is interesting about this moment is that if I were to play little bits of the passage one after another, like notes on the piano, nobody would be able to hear the transition that takes place from one field of time perception to another. The fact that I make the transition continuously changes our whole attitude towards our acoustic environment. Every sound becomes a very mysterious thing, it has its own time.

In this way, traditional meaning of parameters like rhythm and melodies become intertwined with the sound timbre qualities. In fact, we believe that the uniformity of the continuum of time connects the two concepts of the musical timbre and sound timbre. However, note that when this connection (the continuum itself) is used and made clearly apparent as a part of the composition, the traditional parameters (e.g., rhythm and melody) go through a circular transformation; meaning that for example, in listening to a process which is decelerating, rhythmic forms emerge out of timbral sounds while the contents of what creates the rhythmic from itself is a new, yet related timbre; therefore, timbral form also emerges out of rhythmic sounds. By slowing down or speeding up sounds, we are physically listening to the different scales of the signal, and for the signal to have a certain meaning by having a continuous uniform relationship among its different scales we are assuming a self-similar or self-affine structure in the sound and music. This view changes not only the way we compose music, but also how we listen to it analytically. Stockhausen says[47, page 95]:

> The ranges of perception are ranges of time, and the time is subdivided by us, by the construction of our bodies and by our organs of perception. And since these modern means have become available, to change the time of perception continuously, from one

range to another, from a rhythm into a pitch, or a tone or noise into a formal structure, the composer can now work within a unified time domain. And that completely changes the traditional concept of how to compose and think music, because previously they were all in separate boxes: harmony and melody in one box, rhythm and metre in another, then periods, phrasing, larger formal entities in another, while in the timbre field we had only names of instruments, no unity of reference at all. (I sometimes think we are fortunate in having such a poor language to describe sounds, much poorer than the visual field. That's why, in the visual field, almost all perception has been rationalized and no longer has any magic.)

Self-similarity

The coherencies which exist in music have to agree with each other in any scale and dimension in which they are being perceived. The auditory experience which comes from a performance in the way that the sound of the instrument (a single instrument or an orchestra) matches the music, gives complete freedom to the listener to choose the scale of audition. However, it enslaves him or her by providing the same message at every level. The listener is free to tune in at any scale of perception. However, the composition has a single feeling to it. In fact this feeling may change in every performance. However those different feelings are in turn related to each other by the integrity of the piece. A score of a composition is the coding of a musical idea in some accepted dimensions as parameters for the sake of communication. Schoenberg says[40, page 220]:

> THE TWO-OR-MORE-DIMENSIONAL SPACE IN WHICH MUSICAL IDEAS ARE PRESENTED IS A UNIT.[17] Though the elements of these ideas appear separate and independent to the eye and the ear, they reveal their true meaning only through their co-operation, even as no single word alone can express a thought without relation to other words. All that happens at any point of this musical space has more than a local effect. It functions not only in its own plane, but also in all other directions and planes, and is not without influence even at remote points. For instance, the effect

[17] The capitalization of this sentence is Schoenberg's.

> of progressive rhythmical subdivision, through what I call 'the tendency of the shortest notes' to multiply themselves, can be observed in every classic composition.

Such a definition perhaps takes a dimensionless concept such as the musical idea and projects it onto a plexus of dimensions for communication. The fact that every part is part of a whole and abides by global law while at the same time, as Schoenberg says, *"all that happens at any point of this musical space has more than a local effect"*, can be modeled with self-similar structures and self-referentiality. It is true that such a model tries to capture a sense of aesthetics and romantic feeling about music; however, there is no need to fear since we will never reach a true self-similar shape since they only exist in infinity. The idea of self-similarity can also capture the uniformity of time and perception. Stockhausen has noticed this fact as well, and one of his acoustical piece, *Mantra* (1970), may be called a fractal piece. About it he writes[47, page 57]:

> I can give an example of a more recent concept of sequential form, my composition MANTRA for two pianos and electronic modulation. In this work I use a 13-note formula, and nothing but this formula throughout the whole duration of the composition. The formula is expanded and compressed in its pitch and time intervals, but it is always the same formula. Each note of the original statement of the formula has certain characteristics: a periodic repetition, an accent at the end of the note, an ornament, and so on, these characteristics are seeds of later development. The structure of the whole composition is an enlargement in time of that one small formula to more than 60 minutes, and the sections of the composition correspond to the notes of the original formula, and their characteristics. The form is sequential, but with an overall development.

2.5 Summary and Conclusion

Let us briefly review the content of this chapter. We first established an awareness of physical and psychological effects and connected those effects to the concepts of like sound and music. We established a dichotomy between the two, thus separating them from each other as poles, and suggested that for music to be coherent and meaningful it has to bring these two poles together in a natural way. We also tried to establish an awareness of a finely detailed plexus of communication whose axes could be transformed one to the other. In section 2.3, we discussed Schoenberg's theory of harmony and the function of form in general, which according to Schoenberg is comprehensibility. We explained that Schoenberg established a physical continuum between consonances and dissonances by recognizing that tonality's origin can be found in the physics of its material, which is harmonic sound. This is a theory of the relationship between form and content in tonal form. We briefly explained serialism and explained that it is in accord with electronic music, where it can be used as a technique for creating high-level sounds. In section 2.3.3, we discussed the tonality of atonality, which in our opinion concerns the inner musical necessities and aesthetics of music. In the sense of defining atonality as an act which is musical and exists outside the system of form, it becomes a social and political issue of questioning authority in our societies. Tonality of atonality has to be worked on in every moment of the aesthetical process and deserves far more attention than can be provided in the context of this thesis. Finally we explained the idea of uniformity in musical time, and unity in our perception. We showed, however, that this unity implies a sense of self-similarity, while self-similarity provides us with a convenient and consistent tool to model the relationships of sound and music.

The serialist composers extended the physical continuum between consonances and dissonances to a continuum between tone and noises. We suggest in this thesis that this is actually a continuum between sound and music, or in other words a continuum between out physical and logical beings.

Schoenberg reduced his concept of music to the relationship between two tones and stopped himself from the manipulation of the structure of harmonic sound, even though as we have suggested, he freed its structures. Perhaps, he found such thought silly and strictly theoretical without any musical foundation; in that respect he remained "tonal". Stockhausen took a step further by reducing his musical entity to a single sinusoidal function. He writes[47, page 88]:

> Until around 1950 the idea of music as sound was largely ignored. That composing with sounds could also involve the composition of sounds themselves, was no longer self-evident. It was revived as a result, we might say, of a historical development. The Viennese School of Schoenberg, Berg, and Webern had reduced their musical themes and motifs to entities of only two sounds, to intervals. Webern in particular, Anton

> von Webern. And when I started to compose music, I was certainly a child of the first half of the century, continuing and expanding what the composers of the first half had prepared. It took a little leap forward to reach the idea of composing, or synthesizing, the individual sound.

Electronic music was one story; computer music is a different one. We not only have the capability to generate any relationship in the structure of sound, and not only can we control them with practically unlimited precision, but we can also define logical processes which take over such controls as well. We have total control and no material; therefore the sound or the music using computers is all form in every scale, and it is the relationship among these forms in different scales which constitutes the composition. The material is nothing, and therefore the form has to be infinitely detailed. Self-similarity is the form for nothing.

Chapter 3

What is Self-similarity?

3.1 Introduction

It is best to understand self-similarity in its geometrical sense. However, before we discuss it in this way, let us examine M. C. Escher's *square-limit* which has been reproduced in Figure 3-1[1]. The drawing is coded in a graphical language by first defining a very simple shape and then a set of operations to be applied to it. The progression of a fourth of this drawing is illustrated in Figure 3-2. When we look at the center of the piece in figure 3-1, we are more conscious of the lines and areas which create the shapes; in other words, we create a mental representation of how the shapes look to us. As we move toward the outer edges, the shapes start to turn into textures; thus, the same shape and the same procedures are used in two levels of our vision perception. The procedure which creates this drawing is a recursive process. It could be made as big as one would wish. However, what we see on the page is actually just a snapshot of the forth level of recursion, and actually what is coded in this document is not the exact drawing but just the procedure. Therefore if one had access to the machine readable format of this document, one could change the number of levels of recursion and create a picture with more or less detail.

The self-similarity of this drawing is a bit difficult to grasp. If the drawing was made so that the shapes were built around the edges and the recursion process filled the center of the page, we could take any carefully picked segment of the picture form its center and magnify

[1]This drawing was coded in Post Script by John Pratt.

it, and we would come up with the same picture. (Many of Escher's engraving and drawing have this property; a very clear example is *Path of life II* by M. C. Escher). In this drawing (figure 3-1) there are actually no defined edges. If we cut a carefully chosen square from the middle of the drawing and then stretch every other part toward the center of the drawing so that the cut square would disappear, we would again come up with the same picture, except that some of the gray scales would be different, in this case, we call this picture self-affine.

Schroeder opens his recent book called "*Fractals, Chaos, Power Laws*" with the following paragraph[43, page xii]:

> The unifying concept underlying fractals, chaos, and power laws is self-similarity. Self-similarity, or invariance against changes in scale or size, is an attribute of many laws of nature and innumerable phenomena in the world around us. Self-similarity is, in fact one of the decisive symmetries that shape our universe and our efforts to comprehend it.

Invariancy against change of scale is called *self-similarity*, and if there are more than one scale factor involved we call that *self-affine*.

In this chapter, we will try to create an impressionistic view of what self-similarity is, and touch upon a few of the cases which create its history. Self-similarity is created when a self-referential entity is observed. Chaos provides a physical proof of the tangible importance of the idea of self-similarity. Self-referentiality is deep at the heart of Gödel's proof, whose real implications for mathematics and logic, we believe, is not yet fully understood.

3.2 What is Chaos?

Until recently signals were categorized as either being deterministic or random. If a deterministic signal was an oscillating signal and had an infinite amount of energy it was supposed to be periodic. The discovery of chaotic systems meant that this assumption no longer holds. When Lorenz detected chaos, he called it: "Deterministic Nonperiodic Flow"[26]. Chaos was an observed phenomenon which went against the usual scientific intuition; obviously intuition is a highly subjective matter and one should create ones own perception of this statement. Lorenz studied the phenomenon of convection in fluids. However, his equations can be mapped to a

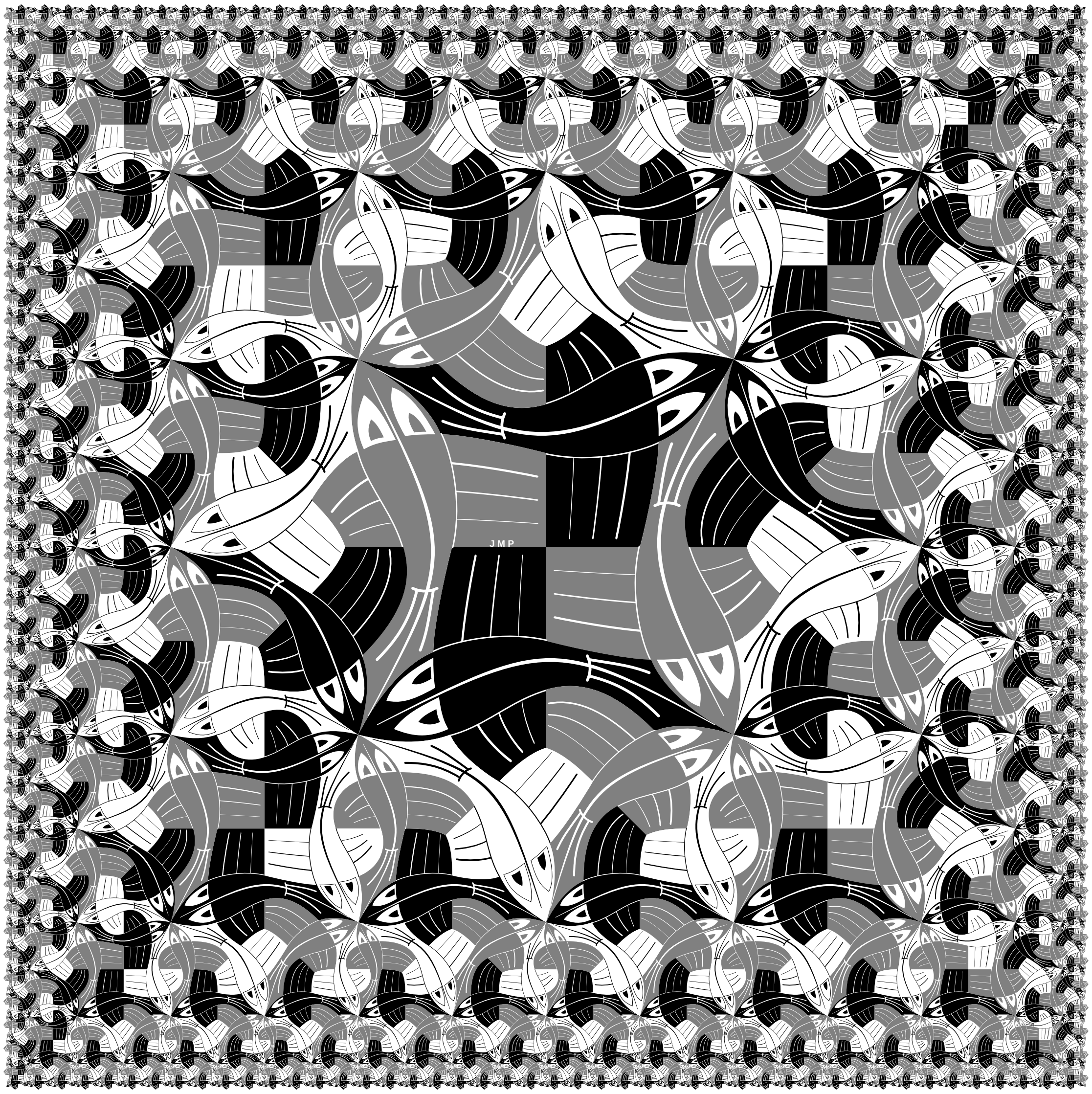
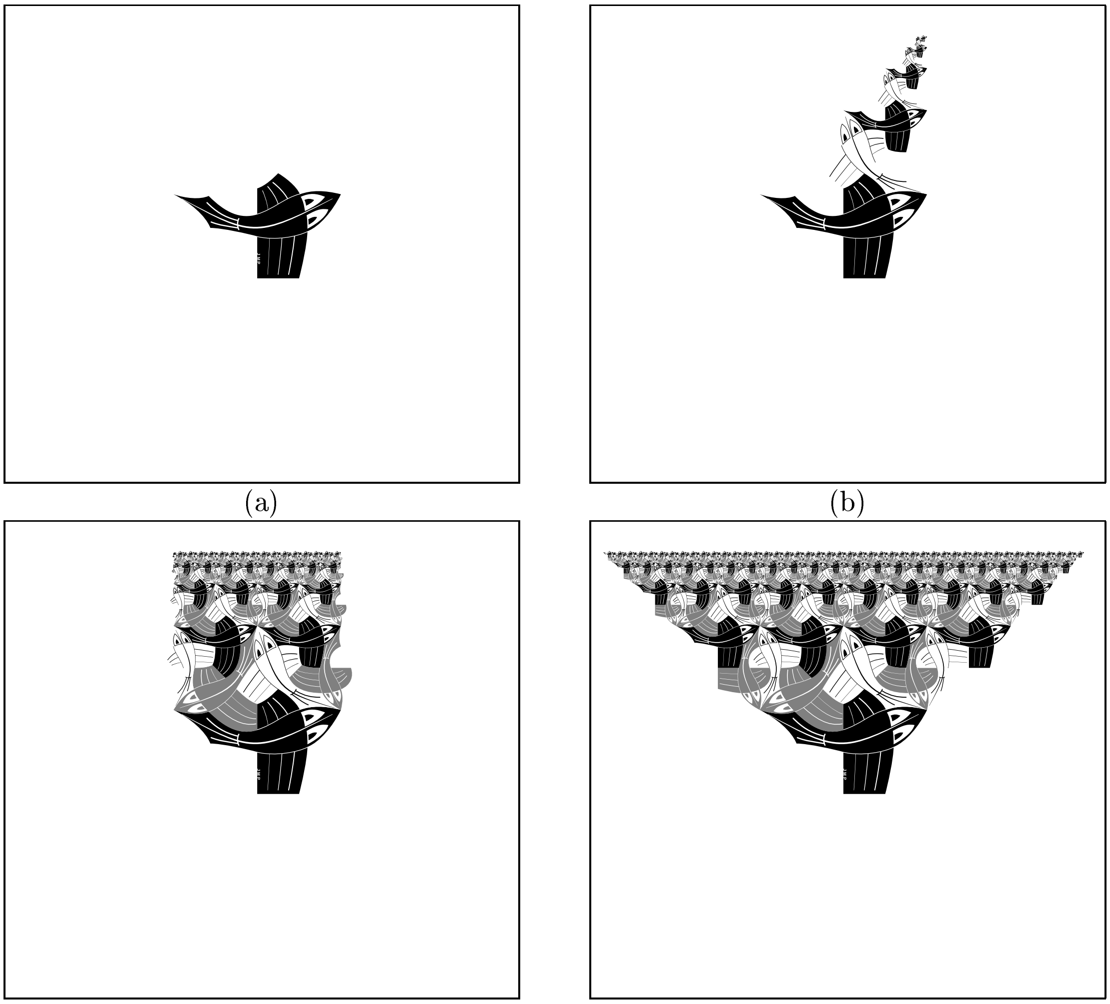


Figure 3-1: Square Limit by M. C. Escher



(a) (b) (c) (d)

Figure 3-2: Square Limit progression

very simple mechanical system. Imagine a water-wheel with many buckets connected to it (see Figure 3-3). All the buckets have holes in the bottom so that the water can run out. A steady flow of water is supplied from the top. If the wheel is started with a small push, the buckets on the top are filled and by the time they reach the bottom, they are mostly empty. Therefore, one side of the wheel becomes heavier than the other. If we increase the flow of the water the wheel starts to turn faster. Once we have passed a certain threshold, the system can start to act chaotic. The wheel can turn so fast that by the buckets which reach the bottom of the wheel are not completely empty, and the buckets that pass under the flow do not have enough time to fill up, and the wheel starts to get slower, then it gets slow enough that the original situation causes it to speed up again. This oscillation becomes damped to the point that the wheel starts to turn the other way around; this means that the oscillation of getting faster and slower damps out at the point that if the wheel was turning to the right, the left buckets would be heavier and the wheel starts to turn in the other direction. What would happen if we let such a system "cool down" without changing any parameters? This is where the scientific intuition used to provide different answers than nature. One may think that the system will eventually pick up a pattern, however long this pattern may be, and keep repeating that pattern. Lorenz showed that this system will never repeat itself, which means that even though the behavior of the system is called deterministic (i.e., three differential equations model the system), the resulting behavior is nonperiodic. Lorenz explained such behavior by showing that the phase-space of this system contains a space which is created from volumeless surfaces with infinitely detailed structure.

3.3 Relationship Between Chaos and Self-similarity

A phase-space is an N dimensional space whose every point fully characterizes the state of a system. The phase space of the water-wheel system can be characterized by three variables[2]. If we plot every state of the system in time according to these three variables, we come up with a trajectory which characterizes the behavior of the system over time. Now, if we would be able to predict how this trajectory moves in the phase space, we would be able to predict

[2]The three variables are the angular velocity of the wheel, and the first sine and cosine coefficients of first harmonics of the fourier series of the amount of water in the buckets.

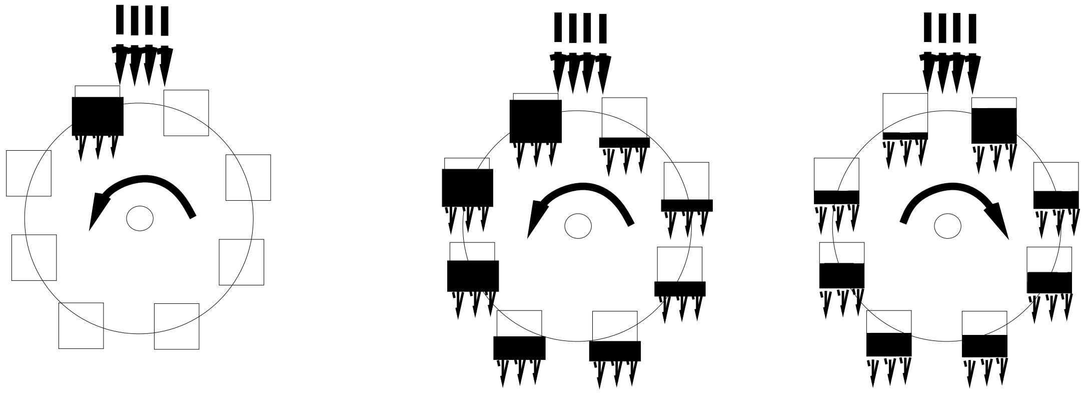


Figure 3-3: Water-wheel imitating the convection system of Lorenz[14].

the behavior of the system. If the trajectory was a line moving on a simple 2 dimensional surface and the system was linear, by having three samples of this trajectory we would be able to predict the behavior of the system. However, there are places in the phase space where the trajectory seems to trace a very thin volume, and the volume is created by infinite stretching and folding of a surface. The shape underlying this strange "surface" is the cantor set, which is self-similar. This implies that the trajectory is moving on a shape with infinite amount of detail, meaning that a different direction could be taken according to infinitesimal differences in initial condition. In a linear system, a small error in initial condition could only cause an error proportional to the original error. However in a system like this, an error (i.e., our inability to measure conditions with infinite precision) could cause completely different directions to be predicted for the trajectory. It is important to note that there is no noise introduced into the system, and this interesting behavior can be seen on the computer by trying to predict the trajectory of the system by using the three differential equations characterizing the system. By changing the integration interval and initial conditions, we obtain completely different results, while we are usually used to obtaining more accurate results when we integrate over smaller segments of time.

3.4 Fractional Dimensions

The Euclidean geometry implies integer values for the dimension of geometrical shapes. There is perhaps no physical object known to our consciousness in the world whose shape conforms to the Euclidean paradigm. Every object known to us can be broken down into smaller objects until we reach the principle of quantum mechanics where, with the present status of physics, we have to treat matter in a shapeless form. When we try to measure the length of a straight Euclidean line, its length does not depend on the length of the ruler we use to measure it with. For example, if we try to measure a 10 cm line with a 5 cm ruler we have to cover the line with 2 copies of the ruler, and if we measure the line with a 1 cm ruler we will have to cover the line 10 times with the new ruler. This relationship can be written as r^d, where r is the ratio of the rulers' length and r^d is the ratio of the number of times that we have to cover the line. In this case $r = 5$ and $d = 1$. The value of d is called the Hausdorff-Besicovitch dimension (which, from now on, we simply call dimension). Therefore, we say that this line has a dimension of 1.

Imagine a square whose every side is 9 cm. If we try to cover this square with smaller squares whose side is 3 cm, we will need 9 copies of our measuring square. If we use a measuring square whose side is 1 cm, which means we are choosing an $r = 3$, we will need 81 copies of this square, which means our ratio of the number of covering squares is $81/9 = 9$. Therefore, $3^d = 9$, which implies that $d = 2$, or in other words the surface of the square has a dimension of 2.

When we apply this idea to a self-similar cure we get fractional values for d. This situation arises since a self-similar object has infinite amount of detail and no matter how small our measuring unit is, we will be ignoring some details whose lengths may actually not converge. The Koch snowflake (Figure 3-4) is a very famous self-similar shape, or in other words, fractal[3]. The process of the construction of the curve is illustrated in Figure 3-4. Let us assume that the first level of the cure is an equilateral triangle, whose every side is 3 cm. If our measuring stick is 3 cm, we will need 3 copies of the stick to cover the whole shape, and in this case we are ignoring all the other details which result from the other levels of progression. However, if we use a measuring stick of 1 cm ($r = 3$), we can cover an extra level of detail and we will need 12 copies

[3]The name fractal was coined by Mandelbrot to bring together many mathematical shapes and ideas which prior to that were called with names such as monsters, wobbly, twisted, or crooked because of their infinite amount of details[29].

of our measuring stick, which implies that $r^d = 12/3$ or $3^d = 4$, or $d = log(4)/log(3) \approx 1.262$.

As described before, the phase space of Lorenz equations can also be created with such a self-similar procedure whose dimension is 2.06[31, page 126]. One way to think about a fractional dimension is to think, for example, that the Koch curve covers a space more than a straight line and less than a surface. It is also possible to have self-similar shapes whose dimension is integer like the Hilbert non-intersecting curve whose dimension is 2[43, page 10]. The progression of the Hilbert curve is illustrated in figure 3-5[4]. In this case we are covering a two-dimensional surface with a topologically one-dimensional line.

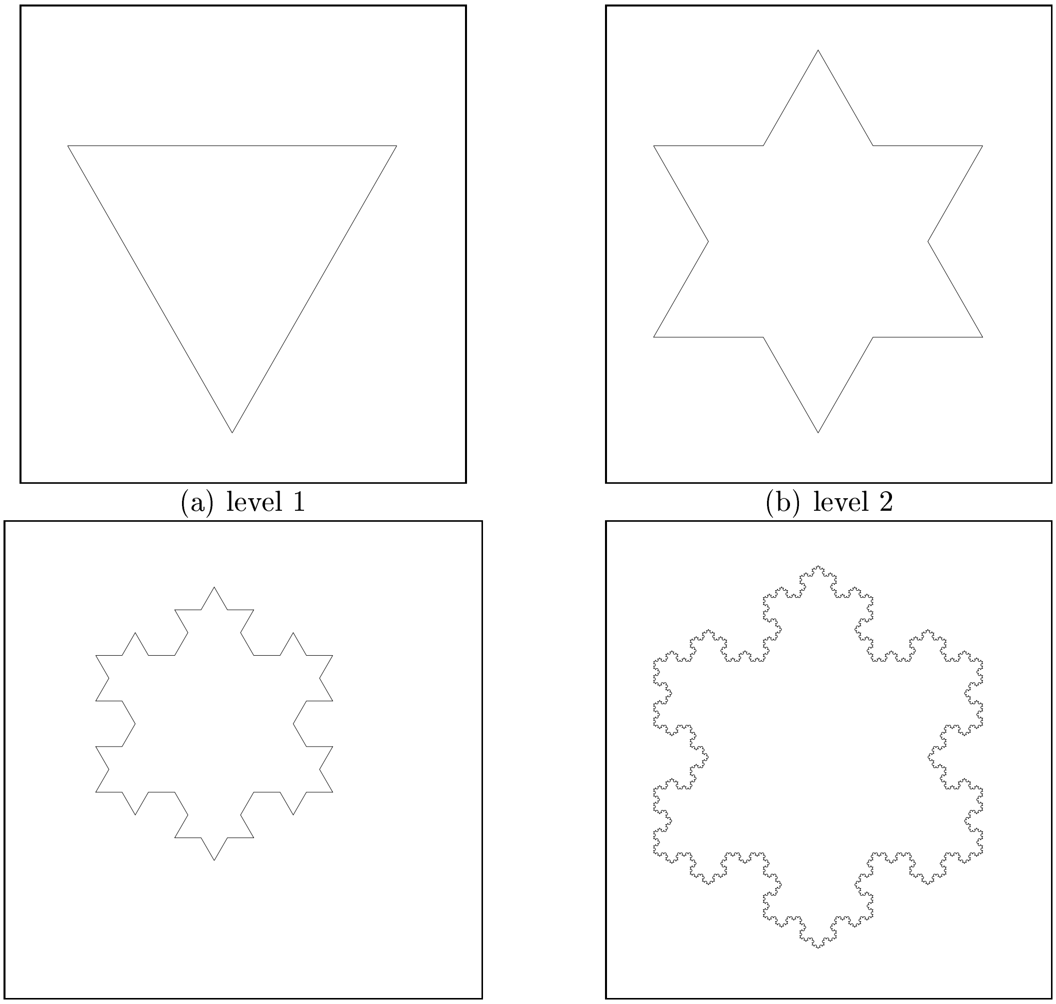
The concept of dimensions is discussed since we believe that it is an important idea for understanding the idea of what a continuum is. For example, if we assume a very simple idea of thinking about music as a two-dimensional (time and frequency) entity, a melody can be one-dimensional, in which case it behaves like a simple line, or it can cover the whole spectrum as a white noise by having a dimension of 2. This is one way to model the continuum of tone and noise which Stockhausen sets as a criteria for electronic music[47, page 109].

3.5 Self-referentiality

Self-similarity should be thought of as a portrait of a self-referential entity. A self-referential entity refers to itself before it exists, and this process is essential in its existence. For example, if we think about the self-similar shapes discussed in this chapter, all we have seen from them is simply a snapshot of a certain level, their true and complete selves existing only in infinity. Self-referentiality in science is a new idea. Cantor's set theory is probably responsible for its recent developments. The many paradoxes which Cantor's set theory created were first thought to be pathological cases. Notably, Henri Poincaré called Cantorism "a sickness from which mathematics would have to recover", while Hilbert thought that Cantor had created a new paradise in mathematics[7, page 1]. However, once Gödel published his paper, "*On Formally Undecidable Propositions Of Principia Mathematica And Related Systems*" in 1931[15], self-referentiality was taken very seriously.

The basic idea behind Gödel's paper is that no formal system can be complete and consistent

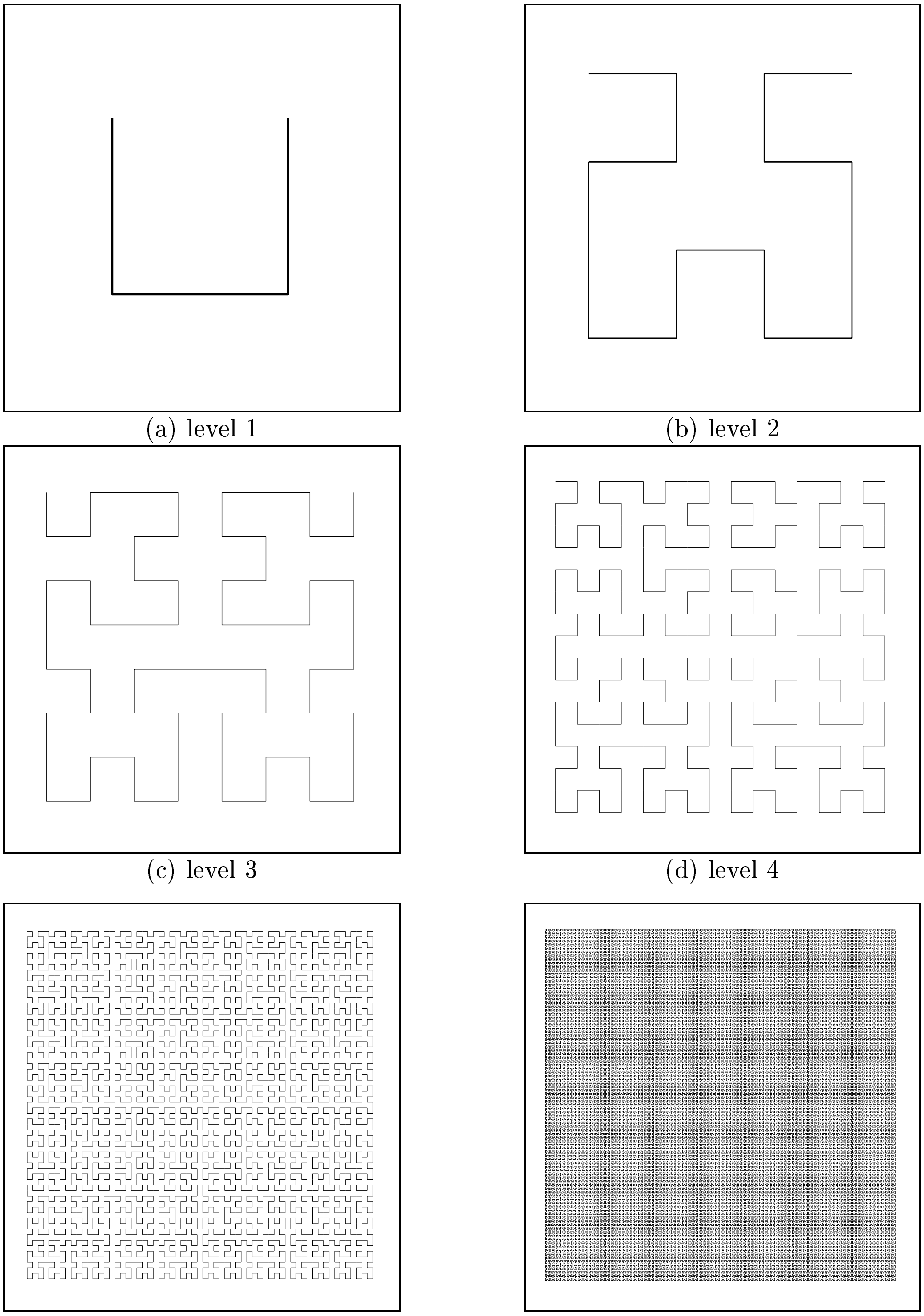
[4]This illustration was created by Jin Choi.



(a) level 1 (b) level 2

(c) level 3 (d) level 6

Figure 3-4: The Koch Snowflake



(a) level 1 (b) level 2

(c) level 3 (d) level 4

(d) level 5 (d) level 6

Figure 3-5: The asymptotically self-similar Hilbert curve. Notice the effect of aliasing that is created when the detail exceeds the capacity of a printer.

at the same time, or in other words, no formal system, no matter how rigorous, can cover the whole truth. There are perhaps many interpretation of Gödel's work, and it is generally understood that a full comprehension of the paper has not yet been reached. Gödel's paper is a completely rigorous mathematical work. However, the basic idea is very simple and intuitive[18, page 17]. He was inspired by the Richard paradox, which is a self-referential paradox in number theory, showing that any meta-mathematical statement which is about a formalized calculus can indeed be formalized within the system itself[30, page 66]. In this way a system can create undecidable propositions. There is really no need to think about mathematics to understand Gödel's work; all one needs to do is to try to decide if the following statement is true:

> This statement is wrong.

This statement can neither be true or false. Once one applies a truth value to this sentence, the sentence itself reverses its truth value. This situation arises since the statement refers to itself ("this statement") before it is completed. Such statements are deep within the system of our thought and senses. For example, the idea of seeing is not an issue unrelated to what we see. As a child what we see creates the idea of seeing; any new visual information can change our concept of vision. If we take this idea on the path of evolution, we may ask: "Was something seen first before an eye was evolved, or is it the other way around?" There is really no substance in such questions, except that they make us aware of the self-referential issues in evolution. Dawkin treats the paradoxical issue of survival in being selfish or altruistic to our own or other species in "*The Selfish Gene*"[8], and for that he almost takes the consciousness away from living beings to the gene level.

These types of questions inevitably take us on the path of philosophy. Self-referentiality is one of the strongest elements in the philosophy of Zen Buddhism and Taoism[23]. Self-refentiality is especially found in the poetry of many of the eastern cultures. The 20th-century western literature and philosophy of the absurd is mainly concerned with questions of authority and power, which, once questioned, become self-referential entities. Many of the works of Kierkegaard deal with issues like paradoxes and ironies of life. One of his most influential works "*Fear and Trembling*", deals with the paradox of faith. He says[21, page 55]:

> Faith is namely this paradox that the single individual is higher than the universal — yet, please note, in such a way that the movement repeats itself.

What Kierkegaard meant as repetition, is actually understood by us now as recursion. He further says:

> Faith is precisely the paradox that the single individual as the single individual is higher than the universal, is justified before it, not as inferior to it but as superior — yet in such a way, please note, that it is the single individual who, after being subordinate as the single individual to the universal, now by means of the universal becomes the single individual who as the single individual is superior, that the single individual as the single individual stands in an absolute relation to the absolute.

Kafka's work which now is hailed as a masterpiece of 20th century modern literature is also deeply based upon self-referentiality. The following is one of his short paradoxes called "*On Parables*" which has many levels of self-referentiality[19]:

> Many complain that the words of the wise are always merely parables and of no use in daily life, which is the only life we have. When the sage says: 'Go over', he does not mean that we should cross to some actual place, which we could do anyhow if the labor were worth it; he means some fabulous yonder, something unknown to us, something too that he cannot designate more precisely, and therefore cannot help us here in the very least. All these parables really set out to say merely that the incomprehensible is incomprehensible, and we know that already. But the cares we have to struggle with every day: that is a different matter. Concerning this a man once said: Why such reluctance? If you only followed the parables you yourselves would become parables and with that rid of all your daily cares.
>
> Another said: I bet that is also a parable.
>
> The first said: You have won.
>
> The second said: But unfortunately only in parable.
>
> The first said: No, in reality: in parable you have lost.

And here is the shortest self-referential statement we have arrived at:

> Nothing exists[5].

[5] Start with: "nothing" exists.

Chapter 4

Self-similarity in Sound and Music

4.1 Introduction

In this chapter we will explain some of the previously discovered cases of self-similarity in sound and music. We will also present our results in recreating the cases mentioned in the literature.

4.2 The Shepard Tone

The partials of harmonic sounds are related to each other by an arithmetical relationship. The partials of a Shepard Tone are related to each other by a geometrical relationship. Shepard used such signals to prove his hypothesis of the circularity of pitch perception[45].

Schroeder[43, page 96] shows that the auditory paradox created by the Shepard Tone, which is generated according to a Weierstrass function, has become possible due to the self-similarity of the signal. A Weierstrass function is constructed as follows:

$$w(t) = \sum_{k=0}^{\infty} \alpha^k \cos(\beta^k t) \tag{4.1}$$

where α is real and β is odd. Weierstrass showed that under certain conditions of α and β, this function is everywhere continuous but nowhere differentiable. For creating a Shepard tone we can drop the α^k term since we are only going to be dealing with a finite number of partials.

Therefore, we have:

$$w(t) = \sum_{k=0}^{M} \cos(\beta^k t) \tag{4.2}$$

where M is the number of partials and β is the geometrical relationship between two adjacent partials. Although Shepard applies a formant-like envelope to the frequency domain representation of the signal, this is done for smoothing the perceptual transition and sustaining the paradox effect. The paradox is created from the fact that the ear attempts to extract a one-dimensional signal (the variable being pitch) out of a multidimensional signal (timbre). We can think of pitch as a value which identifies a relationship between the partials of a signal in a one-dimensional way. If we view the frequency domain representation of the signal, then time scaling according to the same geometrical relationship β does not change the "body" of the signal but only its boundary conditions; therefore we hear the same pitch and not a pitch scaled according to the scale factor. Scaling the function $w(t)$ in time by a factor of β gives $w(\beta t)$: substituting into equation 4.2, we get:

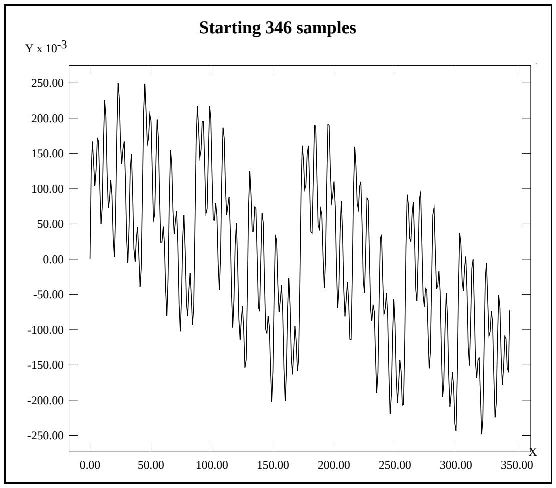
$$w(\beta t) = \sum_{k=0}^{M} \cos(\beta^{k+1} t) = \sum_{k=1}^{M+1} \cos(\beta^k t) \tag{4.3}$$

which is the same as $w(t)$ except for the boundary conditions of $k = 0$ and $k = M + 1$.

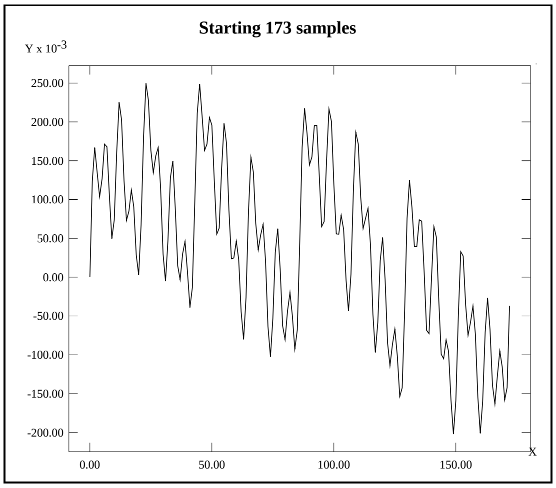
4.2.1 Recreated Results

One could argue that if we start with $k = 0$, we are actually creating partials which are lower than the audibility range (less than 20 Hz). And by rescaling the signal (playing it faster) we are only changing the audible high frequency spectrum. For that reason we start the partials from 32 Hz. Audio example 1 is an example of a Weierstrass function with $\beta = 2$ and $k = 5, 6, ..., 12$, therefore, the sound is composed of geometrically related partials from 32 to 4096. The example was created [1] at a sampling rate of 22050 Hz. It is first played at a sampling rate of (22050 Hz) and then at double that rate (44100 Hz). Even though one notices that the center of mass of the energy has increased in the frequency spectrum, one does not get the feeling that the pitch has

[1] All the audio examples for this chapter were created using Csound[49].



(a)



(b)

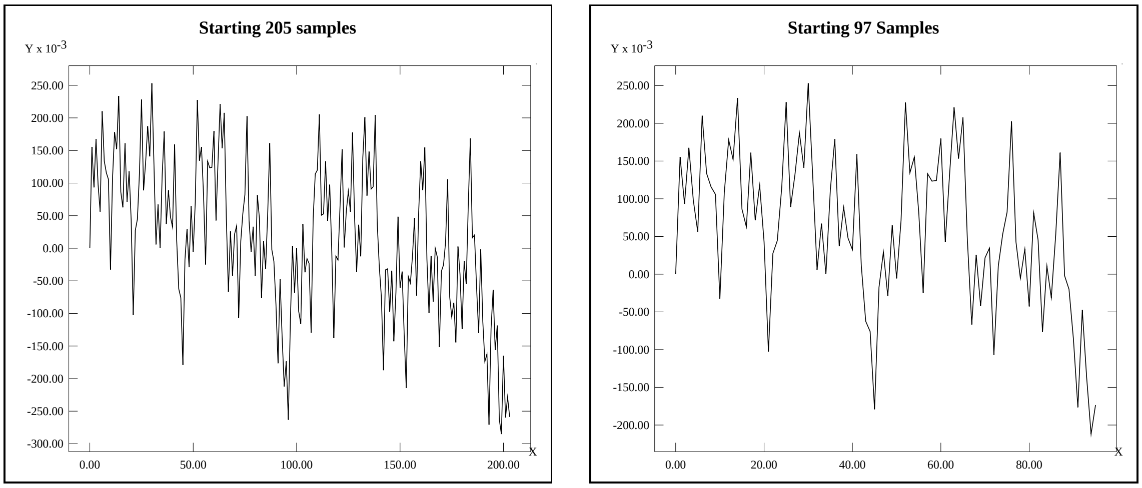
Figure 4-1: A simple Weierstrass function with $\beta = 2$ and $k = 5, 6, ..., 12$. (a) shows the starting 346 samples at sampling rate of 22050 Hz and (b) shows exactly half of that signal (the starting 173 samples). Notice that the two shapes differ only in high frequency details. This similarity can be seen in higher or lower time scales as well.

moved one octave higher. The self-similarity of this signal can be seen in figure 4-1. In figure 4-1-a the first 346 samples (which at a sampling rate of 22050 Hz is about 16 milliseconds), is plotted against exactly half ($\beta = 2$) of those samples in figure 4-1-b.

By carefully choosing β, we can create a signal whose perceived pitch will descend by a semitone when the signal is played at twice the speed. Therefore the relationship between the new β and scaling value (which is 2 since we are playing at twice the sampling rate), should be the twelfth root of two (the frequency multiplier for a semitone). Therefore:

$$\beta = 2\sqrt[12]{2} \simeq 2.1189 \tag{4.4}$$

Audio example 2 is an example of a Weierstrass function with $Beta = 2.1189$ and $k = 5, 6, ..., 12$. Again, the sound is played at the original sampling rate (22050 Hz) and then played at twice the sampling rate (44100 Hz). The self-similarity of this signal in the time domain can be seen in figure 4-2. Listen to this and the previous example at first with no attempt to find the pitch, and you will simply hear the movement of the mass of frequencies. Then, listen to the examples while concentrating on finding a pitch, and notice that the paradox effects gets stronger.



(a) (b)

Figure 4-2: A Weierstrass function with $\beta = 2.1189$ and $k = 5, 6, ..., 12$. (a) and (b) are the first 205 and 97 (205/2.1189) samples.

4.2.2 Conclusions and Speculations

One may ask why such an illusion, be it for self-similarity or not, is interesting. Helmholtz once said[34, page 218]:

> The study of what are called illusions of the senses is, however, a very prominent part of the psychology of the senses; for it is just those cases which are not in accordance with reality which are particularly instructive for discussing the laws of those processes by which normal perception originates.

Our ears are very familiar with harmonic sounds and we are familiar with their properties. We use the arithmetical relationship between the partials of harmonic sounds (i.e. the pitch) as a channel for communicating musical thought. Harmonic sounds are one of the simplest type of sounds, whose spectrum we have been able to control by our acoustical instruments. With computers not only are we able to create sounds which do not have any correspondence to the natural physical word, but we can also control their spectrum in frequency and in time in almost any way. It is a rather different way of looking at the problem of composition. Before electronic music, a composer had a series of constraints dictating the type of sound and its control. These constraints are imposed on the composer in the physical domain of sound.

However once we move into the domain of musical structures such constraints vanish. Here, it is only the psychological issues, such as culture, aesthetic, or style, which may impose a constraint on the mind of the composer. In other words, the composer is free to assume any structure he or she pleases on top of the time scale of the physical constraints of sound.

However, there are now practically no constraints in the relationships which exist in the domain of sound, and that in itself is probably the only constraint that the composer is faced with. The composer is now able to compose down to the smallest micro-structures of the sound. The Shepard tone is a sound whose internal constraints are not that of the harmonic relationship. It seems apparent that the internal constraints of such tones create an integrity in the sound which can be used as a tool for the communication of musical thought. Risset used different flavors of the Shepard tone for his composition *Trois Moments Newtonians* (1979)[35]. Risset[36] also explains the work of Ken Knowlton regiarding this issue, showing how the same principles could be applied in the time domain to create rhythms which seem to become faster and faster while there is actually no change in speed. Such a rhythm can be created by superimposing several beats which have geometrical relationships to each other, and then slowly fading in the slow ones while fading out the fast ones. If, in fact, the reason for the illusion of the Shepard tone is the self-similar structure in the sound, we may be able to conclude that we can detect and relate to self-similarity in the auditory domain.

4.3 The Well-tempered Scale

Imagine that we have recorded a melody on tape. If we play the tape twice as fast as it was recorded, the melody is transposed up an octave, and if we play the tape one and a half times faster, the melody is transposed a fifth above. In almost any tonal scale other than the well-tempered scale, not all the new notes resulting from transposition by time scaling would fall exactly over the scale values. In other words, the melodies in the well-tempered scale are invariant against time scaling with a similarity factor of $\sqrt[12]{2}$, meaning that if we transpose any melody according to any of the frequency factors of the scale, we come up with a melody whose notes are all in scale. Schroeder[43, page 99] explains the different power laws which govern this property of the well-tempered scale, and he also explains that if we had all the notes of a piano (which was tuned exactly according to the well-tempered scale), sounded simultaneously, we would hear a self-similar Weierstrass function with $\beta = \sqrt[12]{2}$ and its harmonics.

4.4 The Ubiquitous $1/f$ Noise

A work of art has to be complete and at the same time it should be devoid of any extra part; meaning that a complete piece needs no part added to it while nothing can be taken out of it. In the case of music composed with traditional notational systems, no notes can be taken out, and there is no room for any new notes to be added. This means that every single note should have a meaning and a function. Every note contributes not only to the instantaneous color of sound (i.e. creating its own individuality and meaning), but also it satisfies a context built by the previous note and sets up a new context for the notes which proceed it (i.e. satisfying its function). Satisfying immediate functions means that successive notes have to be "correlated" with each other. A complete correlation in the time scale of notes dictates very boring melodies. It is important to note that in a longer scale of time the phrase "it satisfies a context built by previous notes", does not mean that there has to be a conformation to the immediate context. It may be that a conscious breakdown of context is needed to satisfy a higher level goal (context) in a higher time scale, and that might be what creates the element of surprise. This breakdown of lower level context can also be controlled by higher level organized chance operations.

Now we can simply replace the word "note" with "melody" in the previous paragraph, and move to a higher plane with the same type of requirements. When we apply this idea to all levels of time in music we reach a rather obvious fact: that a piece of music has to have structures on all levels of our perception. However, these structures themselves have to be related in some way to each other. Again the same rule which we described for the successive elements (e.g., notes, melodies, etc.) applies to the entities which these correlations create. If we visualize music laid out in the conventional time-frequency plane of spectrograms, then the relationship among successive events is a relationship along the horizontal axes, while the relationship between correlations in different time scales of perception is in the vertical direction. In other words a piece of music has to have some "correlation" in all its time scales while the values of the correlations are in turn correlated within themselves. Having this correlation and at the same time not being boring, a piece of music creates a plexus for every note (event) which has to strive for its individuality while conforming to its context.

Different techniques in signal processing provide us with ways to become more concrete about qualities such as "correlation", as long as we are precise about what comprises our signal. For example, there are different algorithms for pitch detection using fast Fourier transform or analysis of the auto-correlation function by using the sound pressure level as the signal. Such analyses take a physical signal (e.g. sound pressure level) and try to come up with a perceptual value (pitch). It seems plausible that applying the same type of analysis, which finds some type of correlation in the physical signal, to the newly found perceptual values would result in some tangible understanding of a higher level entity. There are two questions which we have to keep in mind. (1) Is there any clear-cut boundary between perceptual and physical events? (2) Are the physical and (many) perceptual levels of our mind governed by the same principles, and therefore can they be analyzed in a similar fashion?

In this section we will briefly touch upon these two questions by analyzing "pitch signals". A pitch signal is composed of a single line melody which is extracted from a piece for the duration of the whole piece. Please note, we make no claim to the fact that the extracted is "the" melody of the piece; we define a procedure and extract "a" melody from the piece. Voss and Clarke[50] conducted some such studies and concluded that what they assumed to be a pitch signal of almost all music behaves like $1/f$ noise. Before explaining their results, we will try to achieve an intuition about how a $1/f$ noise behaves and what are its properties. The text is written in a way that, with the help of graphs, the formulas may be ignored. One of the goals of this section is to show how musical signals such as pitch can be analyzed in the same way that we analyze sound.

4.4.1 What Is $1/f$ Noise

Before understanding how $1/f$ noise behaves, we have to intuitively understand what a power spectrum is. We will try to achieve that by visually looking at the effect of changing some parameters on random signals. Imagine that the signal we are using for our experiment is a pitch signal. This means that the values of the signal are pitches chosen (randomly or deterministicly) in time for a single melody for a specified time.

Power Spectrum

One way to look at a signal is in the discrete time domain, which puts a series of values consecutively in time. In this way we can tell something about the behavior of the signal at every moment in time, and can also make some simple statements about its long-term behavior. However, it is rather difficult to say anything about how the long-term behavior is related to the short-term development of the signal. Another way to look at a signal is to view its spectral density (i.e., the Fourier transform of the signal). The Fourier transform views the signal as a whole. It swaps the dimension of time with the dimension of frequency. One can think of the Fourier transform as a combination of slow and fast oscillations with different amplitude. A very strong and slow component in the frequency domain implies that there is a high correlation between the large-scale pieces of the signal in time (macro-structures), while a very strong and fast oscillation implies correlation in the micro-structures. Therefore, if our signal $f(t)$ represents values in every single moment of time, its Fourier transform $F(\omega)$ represents the strength of every oscillation in a holistic way in that chunk of time. These two signals are related to each other by the following formula[48]:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt. \tag{4.5}$$

One can think of the time domain function as how one listens to a melody and the frequency domain function as how one listens to a chord. Even though the situation in musical communication is not as simple as that (i.e. the time scales in which we listen to melodies and chords are different), this metaphor can give us a starting point in understanding this analysis.

In the Fourier transform, oscillations are characterized with sinusoid functions. Auditorily speaking, these functions are the purest sounds one can create (i.e. they are "clean as a whistle"). The average value of any smooth oscillation, fast or slow, strong or weak, is zero. If we use the square of the values in time we can study the power of these oscillations in the same way we studied the original signal (i.e. take its Fourier transform). *Parseval's theorem for energy signals* states that:

$$\int_{-\infty}^{\infty} \mid f(t) \mid^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mid F(\omega) \mid^2 d\omega. \tag{4.6}$$

The Fourier transform analysis assumes the life of a signal from $-\infty$ to ∞. For that reason when an analysis is carried out for a finite amount of time, it is either assumed that the signal is periodic or that it has a finite amount of energy. A true power spectrum of a signal has to consider the signal from $-\infty$ to ∞. However, we are not always able to observe a signal that way or derive precise functions for it. We can define $F_T(\omega)$ which is the fourier transform of the signal in period T, and define the power spectrum as the following:

$$S_f(\omega) = \lim_{T\to\infty} \frac{1}{T} \mid F_T(\omega) \mid^2. \tag{4.7}$$

The power spectrum itself is the Fourier transform of the auto-correlation function. Auto-correlation function represents the relationship of long and short-term correlation within the signal itself.

$$< f(t)f(t+\tau) >= \frac{1}{2\pi} \int_0^\infty S_f(\omega) e^{j\omega t} d\omega. \tag{4.8}$$

In this experiment, it is this last relationship which is of immediate interest to us. The power spectrum is a function in the frequency domain, which means that we can examine the long-term behavior of fast and slow oscillations. We will be looking at power spectrums approximately in the range of 0.001 to 5 Hz, which corresponds to oscillations which happen from 0.2 to 1000 seconds. Thus, a high value in the low spectral region, close to 0.001 Hz, means a high correlation in a very long time scale (i.e. in macro-structures) and a high value in the high region of the spectrum close to 5 Hz implies high correlations in the micro-structures[2]. A relationship between the different sections of the power spectrum implies a relationship between the auto-correlation of the signal in the time domain to which those frequency sections are referring. In the following section we will examine the effects of changing some parameters of a random signal on its power spectrum.

[2] Customarily, micro-structures in music refer to structures which happen in the sound domain in frequencies above 20 or even 100 Hz. However, we are using this as a relativistic term in reference to the structures in the region of our inspection.

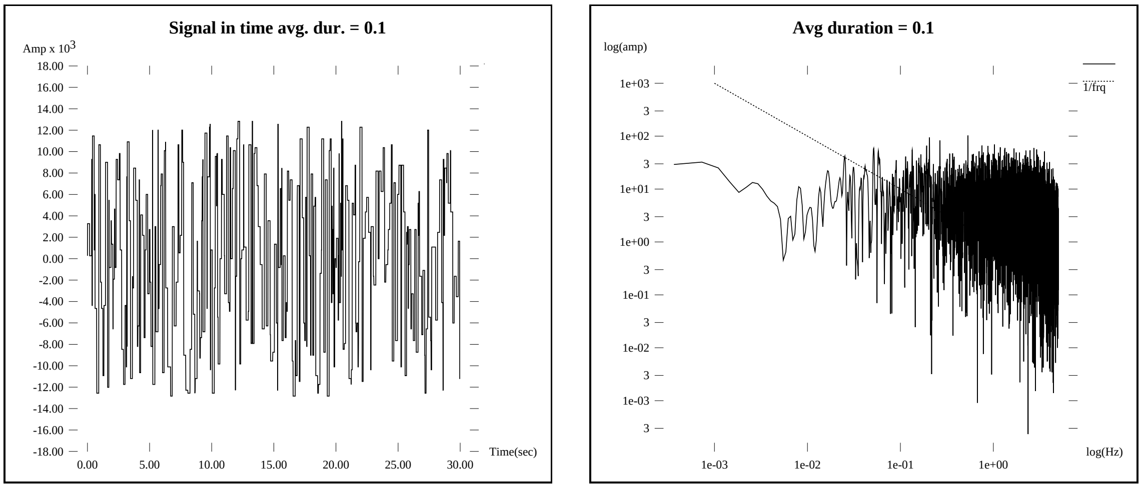
Effect of Changing the Average Duration

In this section we will examine random pitch signals. The values have been chosen from a logarithmic scale of frequencies with various quantization levels. Later we will use the same method for analyzing some pieces according to their MIDI encryptions. The pitch signals are stored as sound files. The frequency of the middle C, or the MIDI note number 60, is used as a reference point. We can have up to 273 quantization levels per semitone. The value of the pitches are restricted to 20 to 2100 Hz. Unless noted, in all the signals the pitches are quantized to frequency values of the well-tempered scale. Once the random signal is generated the average value of the signal in time is subtracted from all the samples.

Figure 4-3 shows the power spectrum and the first 30 seconds of a random signal with average note duration of 0.1 second. We can see that the power spectrum for this random signal is flat, which means that there are as many fast oscillations (structures) as there are slow oscillations. The power spectrum is shown on a log-log scale and for having a reference, the line which represents the $1/f$ spectrum is plotted on top of all the plots in this section. Figure 4-4 shows the power spectrum and the first 10 seconds of random signals with average note durations of 0.5, 2, and 200 seconds (for all the signals 1000 seconds of the random signal was generated.) Notice how these signals start to show a "slope" on the high frequency spectrum. This slope indicates some temporal correlation in that region. Obviously a constant value is more correlated than a random signal; therefore with a higher value of average note duration, the signal becomes more correlated. In fact we can characterize these functions as a $1/f^{\beta}$ spectrum, while in the case of the flat spectrum $\beta = 0$ for all the regions and for the other cases $\beta = 2$ in the region of correlation and $\beta = 0$ in other regions.

Long-term correlation

If a signal is truly random we will never observe any long term correlation (i.e., no power concentration in the low frequency region). However, some operations can create such correlations. Obviously the simplest one is to add such a structure to the random signal, which is not currently what we are inspecting. A bad quantization method can also create correlations in the low frequency region in a random signal. In this case an artificial DC power is added to our signal and that creates a correlation in the low frequency region. In this case one can say that:



(a) (b)

Figure 4-3: The time domain and power spectrum of white noise with average duration of 0.1 seconds. (a) illustrates the first 30 seconds of the signal and (b) is the power spectrum of such a signal in log-log scale. The line representing the $1/f$ is also drawn for reference. Notice that the power spectrum for this signal is flat for the area of our inspection which is between 0.001 and 5 Hz.

"Correlation is in the eye of the beholder"; meaning that it is a correlation in the process of our measurement and not in the signal itself. Figure 4-5 shows the effect of truncating values for quantization rather than rounding them to the nearest integer value. Notice that as the number of quantization levels gets smaller the low frequency power gets larger.

One other way to create low frequency power is to add deterministic structures on top of the random values in micro-structures. Figure 4-6 shows the power spectrum and the first 30 seconds of a random signal with average note duration of .5 seconds with a simple vibrato added to every note. The vibrato's period is determined by the duration of the note.

Relationship between Long and Short-term Correlations

As Voss and Clarke[50] point out many fluctuating signals can be characterized by a single correlation time τ_c. In which case, for time scales much smaller than τ_c (which means for frequencies much larger than $1/\tau_c$) the signal is correlated and the power spectrum's slope is close to that of the $1/f^2$ line, and in the regions much bigger than τ_c ($f \ll 1/\tau_c$) the spectrum is similar to that of white noise. However, a signal which behaves like $1/f$ noise cannot be

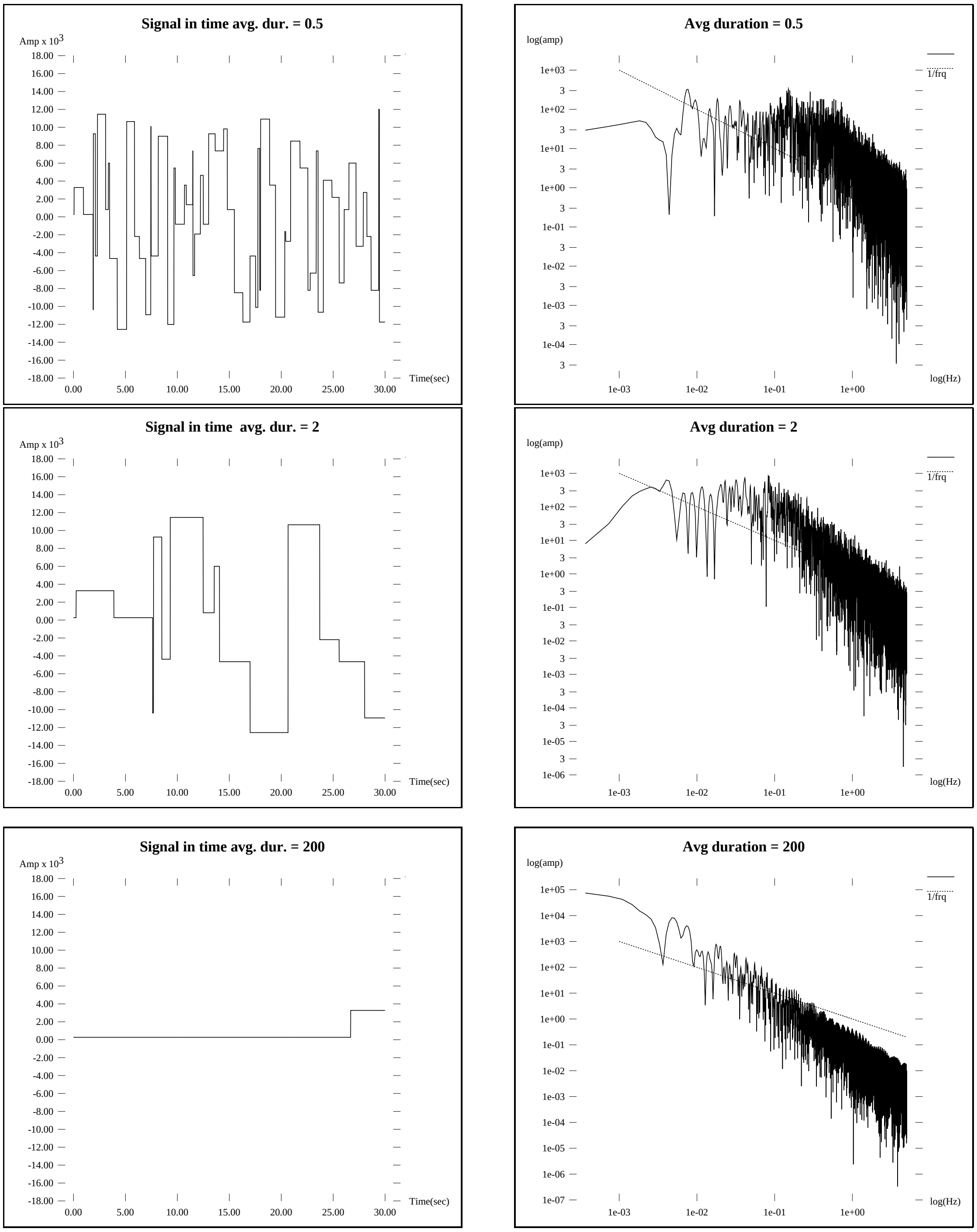


Figure 4-4: The time domain and (log-log scale) power spectrum of random signals with average duration of 0.5, 2, and 200 are illustrated. The line representing the $1/f$ line is also drawn for reference. Notice that the power spectrums show a slope steeper than the $1/f$ line in the area of correlation while the rest of the spectrum stays flat.

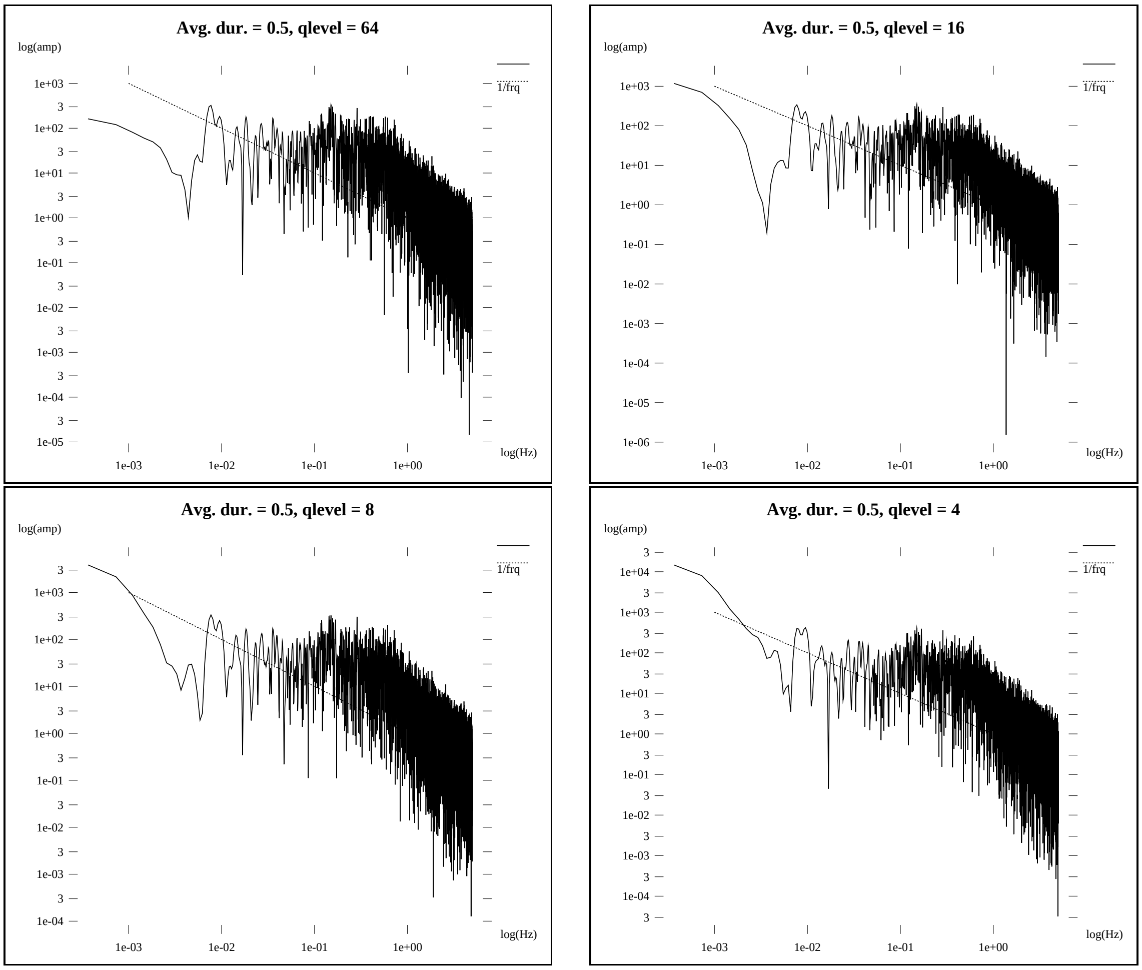
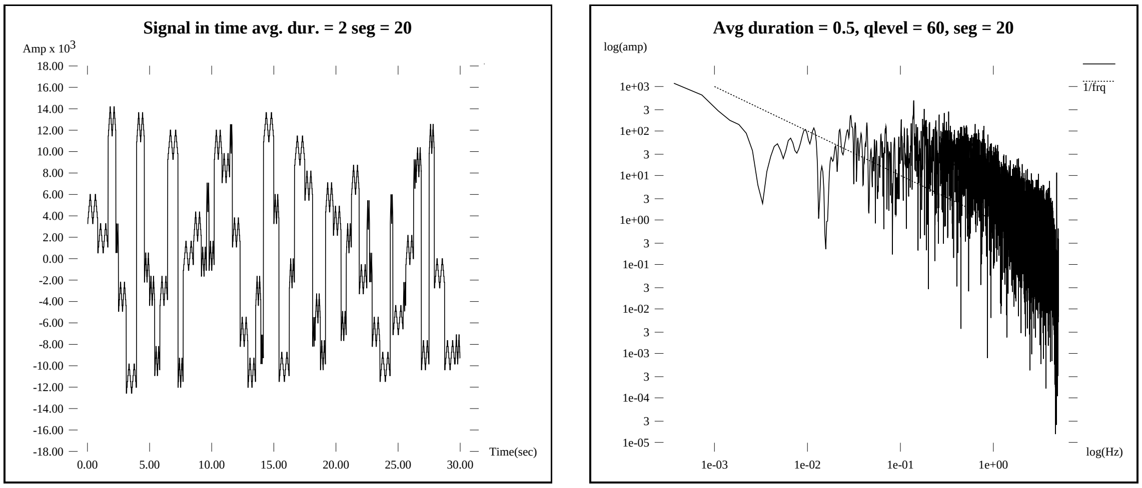


Figure 4-5: In this figure the effect of bad quantization of a random signal is illustrated. The values of a random signal are truncated to the quantized level rather than being rounded to the nearest level. This figures should be compared to the first illustration of figure 4-4. Notice how the low frequency power increases as we use fewer quantization levels.



(a) (b)

Figure 4-6: The first 30 seconds and the power spectrum of a random signal with a simple deterministic shape added to it is illustrated. The deterministic shaped is scaled to the duration of the note. Notice that such a process shows up as low frequency (i.e. long term correlation) on the power spectrum.

characterized by a single correlation time. In fact a spectrum with a $1/f$ slope implies a scale-invariant correlation between long-term and short-term correlation in the region in which the spectrum is exhibiting the $1/f$ slope.

4.4.2 Self-similarity of $1/f$ Noise

The scale invariancy of the signal can be explained by the simple scaling rule of Fourier transforms.

$$
\begin{aligned}
S_f(f) &= 1/f && (4.9)\\
\mathcal{R}_f(\tau) &= \mathcal{F}^{-1}(S_f(f)) = \mathcal{F}^{-1}(1/f) && (4.10)\\
\mathcal{R}_f(\alpha\tau) &= \mathcal{F}^{-1}(\frac{1}{\alpha}S_f(f/\alpha)) \\
&= \mathcal{F}^{-1}(\frac{1}{\alpha} \times \frac{\alpha}{f}) \\
&= \mathcal{F}^{-1}(1/f) && (4.11)
\end{aligned}
$$

From equations 4.10 and 4.11 we can conclude:

$$\mathcal{R}_f(\tau) = \mathcal{R}_f(\alpha\tau), \tag{4.12}$$

which means that our auto-correlation function is scale independent, or in the other words the auto-correlation function is a fractal. We should note that most observed $1/f$ signals are random signals. Mandelbrot[28] suggests that these signals should be treated as nonstationary random signals to get around the infinite invariance problems. Thus, the autocorrelation and spectrum of $1/f$ noise would be time-dependent. The problem with infinite invariance is that a true self-similar signal has an infinite amount of energy in its high spectral region. In the case of the $1/f$ signal the integral:

$$\int_F^\infty 1/f^\beta df \qquad where F > 0, \tag{4.13}$$

is finite when $\beta > 1$, and infinite when $\beta < 1$. This border is where we make the distinction between random and deterministic signals. Notice that the equation 4.12 holds for any α, and that fact should be interpreted as the statistical behavior of the signal. There is no one-to-one relationship between a power spectrum and a signal. Many different signals in time may have the same power spectrum. If we were dealing with a deterministic signal and not a random process, one way we could explain equation 4.12 is that the auto-correlation function has to be a DC function. However, the fluctuations of the observed phenomena which exhibit a $1/f$ power spectrum are far more erratic than unit functions. (For rigorous mathematical treatment of $1/f$ noise see Keshner[20], Flandrin[12], Wornell[51].)

4.4.3 Observed $1/f$ Noises

When a process is assumed to be random and treated as such, the accuracy and scale of its power spectrum depends on the accuracy and stability of the equipment and the method of observation of the signal. Keshner[20] lists many observed fluctuations which behave like $1/f$ noise. These phenomena range from the voltage or currents of vacuum tubes, diodes, and transistors; the resistance of carbon microphones and semiconductors; the frequency of quartz crystal oscillators; the voltage across nerve membranes, to average seasonal temperature, annual amount of rainfall, rate of traffic flow, economic data; and finally, as Voss and Clarke claim, in pitch and loudness of music.

One would imagine that if these phenomena were observed for a very long period of time or with very high precision, one would find regions in the power spectrum which either act as white noise or as deterministic processes. Currently science has a difficult time understanding the $1/f$ noise since it is neither a deterministic periodic (or quasi-periodic) nor a random signal. Some experimenters have measured the $1/f$ noise in MOSFET's down to $10^{-6.3}$ Hz, or 1 cycle in 3 weeks. Other experimenters have computed the weather data using geological techniques to 10^{-10} Hz, or 1 cycle in 300 years. Yet still in neither of these cases was any change observed in the power spectrum. Keshner points to two cases (the resistance of fluctuations of thin-films, and of tin film at the temperature of the superconducting transition and in the voltage fluctuations across nerve membranes) where changes were observed.

4.4.4 $1/f$ in Music

Voss and Clarke conducted some studies on some selected musical compositions. In the first experiment the audio signal was run through a bandpass filter of 0.1-10KHz. The output of the filter was squared to obtain a power function, and that signal was run through a low pass filter with the cutoff at 20Hz. The data from this filter was plotted and it was reported that almost all kinds of music (ranging from a recording of Bach's First Brandenburg Concerto to arbitrary selections of signals from different types of radio stations) behaved like $1/f$ noise. With this experiment they concluded that the "audio power fluctuations" of music, which they called loudness, varies according to $1/f$ noise. We would like to point out that the structures observed were actually the rhythmical structures in the fast regions (about 0.25 to 8 seconds) and the formal structures in the slow regions (greater than 8 seconds). One way to interpret this data is that it describes the uniformity of the loudness between these two regions.

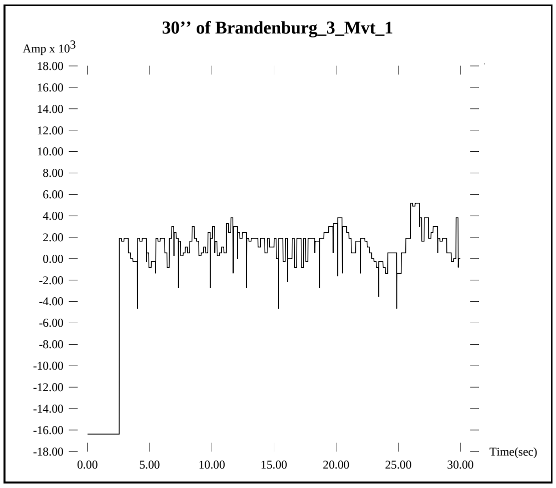
Voss and Clarke also studied the "instantaneous pitch" fluctuation of music. The "instantaneous pitch" was measured by counting the number of zero crossings of the audio signal in specific periods of time. Thus, a new signal $Z(t)$ was extracted from the audio signal $V(t)$, which they assumed, in this case, follows the melody of the music. $Z(t)$ was passed through a low pass filter at 20 Hz and then its power spectrum was measured. Again they found that $Z(t)$

for many different kinds of music and radio stations behaved as $1/f$ noise. In this study they also produced some sounds using white, $1/f$, and $1/f^2$ noises. For every one of the samples the same process was used to control the pitch as well as the duration of every note. The pitches were rounded off to different musical scales such as pentatonic, major, or 12 tone chromatic. These examples were played to several hundreds of listeners, and it was reported that listeners classified the "compositions" according to: white noise, too random, $1/f^2$ noise too correlated, and $1/f$ closest to what listeners expected of music.

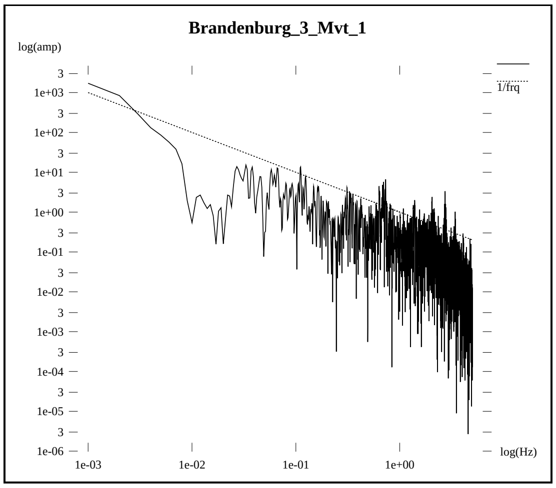
They argued that even though low-level Markov models, or deterministic constraints imposed on white noise, can create some local correlations, they fail to provide a long-term correlation. They suggested that $1/f$ noise is the natural way of adding long-term correlations to stochastic compositions.

4.4.5 Recreated Results

In this section we will present the result of our analysis of the pieces we had access to. Rather than looking at the audio signal, we took a different route for our analysis. We used the data from 57 pieces which were coded in MIDI file format. We extracted a top voice from these pieces. The top voice is defined as the highest sounding pitch at any moment. Silences were eliminated by extending the last highest pitch. The data was stored as described in section 4.4.1. The tempo was set by the first tempo marking and all other tempo changes during the piece were ignored. The DC value of the pitch signal was subtracted from all samples and the power spectrum of the resulting signal was computed. We would like to emphasize the fact that we are not saying that such a signal is "the" melody of the piece; however, we are assuming that with the defined procedure we will obtain "a" melody which has some musical integrity. Audio example 3 is the resynthesis of the first 30 seconds of the pitch signal extracted from the J. S. Bach's 3rd Brandenburg concerto. As it can be clearly heard, there are still problems in the extraction method which, due to not having enough time, we did not solve. Figure 4-7 shows the first 30 seconds of the extracted pitch signal and the power spectrum computed for the duration of piece. The problems of the extraction method can be seen as the vertical spikes in the figure. As it can be seen, the power spectrum of this signal is best fitted by the $1/f$ line. Appendix A contains the result of all the pieces whose power spectrum were systematically



(a)

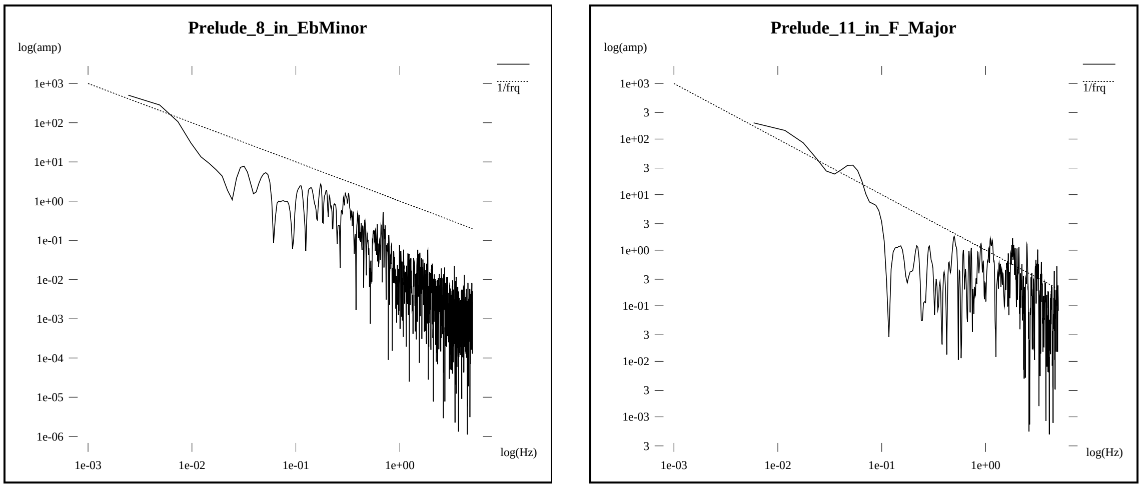


(b)

Figure 4-7: (a) is the first 30 seconds of the "top voice" signal extracted from J. S. Bach's 3rd Brandenburg Concerto. (b) is the the power spectrum of the "top voice" signal for the duration of the piece. Notice that the line representing the $1/f$ line fits the slope of the power spectrum.

computed.

Almost all pieces behaved very closely to the $1/f$ noise. It is worth noting that we were able to find the fault of our extraction method by looking at the resulting power spectrums, and that shows that the power spectrum does carry useful analysis information. For example, the power spectrum of Prelude 11 from the first book of the Well-tempered Clavier (see figure 4-8-b) was the most odd looking spectrum. When we listened to the extracted signal we found that the many trills of the dotted quarters (which are scattered throughout the piece) mixed with the bottom voice created a "noisy" melody which accounts for the flat section of the spectrum between .1 to 5 Hz. The slope of the power spectrum is a good measure of how much material is coded in the melody. For example, the spectrum of Prelude 8 (see figure 4-8-a) showed a slope steeper than other pieces, which should mean that the melody of the extracted signal should be more correlated than the others. When we looked at the score for that piece, we noticed that much of the melody is coded in other voices rather than the top voice, and the highest pitch is kept for long periods of time; in one case (measures 32 to 34) the highest note is kept sounding for 3 full measures.



(a) (b)

Figure 4-8: The spectrum of two of the odd cases of the analysis is shown. Figure (a) is the power spectrum of the "top voice" signal of the 8th prelude from J. S. Bach's Well-tempered Clavier Part I. Notice that the slope of the spectrum is sharper than the $1/f$ line and that can be explained by the static melody of the top voice in that piece. Figure (b) is the power spectrum of the 11th prelude from J. S. Bach's Well-tempered Clavier Part I. Notice that the spectrum is flat in the 0.1-5 Hz region. This effect was caused by the way we extracted the top voice. The interaction between the half note trills and our down-sampling of the MIDI data created a noisy melody which is characterized by a flat spectrum.

4.4.6 Conclusions

This section has tried to touch upon a few different issues concerning $1/f$ noise. In general we view signals as either random or deterministic. If a signal is not periodic and has an infinite amount of energy and all its maximum and minimum values are in a finite range, with our present state of signal processing we must treat the signal as random. However, if the signal has finite energy (and a finite number of discontinuities) we will be able to mathematically, rather than statistically, derive and specifically apply the Fourier transform theorem (Dirichlet conditions[48, page 84]) to the signal. A $1/f$ signal lives on the border of these dichotomy. The high frequency energy of a $1/f^{(1+\epsilon)}$ spectrum is finite, while the high frequency region of a $1/f$ spectrum is infinite. The power spectrum of random processes is usually also divided into two sections, a high frequency region with a slope steeper than $1/f$ and close to $1/f^2$, and the low frequency region which is flat. The flat low frequency region implies that there is no long-term correlation in the signal, while the steep high frequency slope implies a short-term correlation. Keshner[20] points out:

> The presence of $1/f$ noise in MOSFET's, down to the lowest frequency allowed by the limited observation time, suggests that the division into just two subsystems is inappropriate.

The $1/f$ noise is an evolutionary signal, meaning that its whole past history effects is present and future state. This implies a certain type of memory in a $1/f$ process. Dodge[9] finds fractals and $1/f$ noise to be an interesting paradigm for computer-aided composition. He also suggests that the "memory" of $1/f$ noise can account for its success.

The study of music as a $1/f$ noise has a certain value, in that it treats a musical signal as a physical signal. The uniformity that a $1/f$ model of music suggests exists on all levels of our perception down to about 5 Hz. There are no psychological issues to be considered. This is not to undermine the psychological implication of music, but rather to suggest that if we would like to make comments about music in a scientifically rigorous paradigm, it is possible, as we really should, to ignore all psychological issues (the most important of all of them being the assumption of "intelligence"). The study of music as $1/f$ noise assumes no intelligent entity except the music itself.

Chapter 5

Self-similar Synthesis

5.1 Introduction

We can think of musical sound as an entity which lives on the continuum between silence and white noise. Then, a composition becomes a procedure which defines a path along this continuum. Compositions usually start with silence, at some point get closer to white noise, and eventually return to silence. Due to the nature of infinity, a continuum can never be traversed by humans unless it is "quantized". Pitch scales quantize the continuum of frequency, while rhythms do the same for the continuum of time. In tonal form the quantization methods, as well as all the formal operations, are derived from the structures of the harmonic sound. In this paradigm – tonal form — one can only create a single type of musical timbre. Schoenberg's theory of harmony[41] implies a new perspective on music and sound. According to this theory, music is capable of conveying any type of relationship, and not only that of harmonic sound, as discussed in the chapter 2 of this thesis.

Computers seem to be useful to the world of music in a few different ways. They make ideal mechanistic instruments and instrumentalists for the precision of their sound creation capabilities. The computation power of computers makes them ideal for algorithmic composition. Given the correct paradigm, computers are also capable of managing huge databases of information, and provide our imagination with enough primitives to build logical interfaces to the stored data. However, there is a certain dilemma in composing with computers, namely the extent of the freedom they provide.

When writing for acoustical instruments, the composer already uses a quantization of the continuum of timbres, which is defined by the available instruments. In electronic music, the act of composition is stretched to the micro structures of sound. The field of sound has no constraints, and therefore no shape. There are no defined timbres or scales. The point is not that we cannot define such things; currently different synthesis methods are capable of creating distinct sounds for computers. However, "not having a defined sound" is inherently part of the spirit of using computers for music, and perhaps in general is a major part of the spirit of modern music. In this domain, a musical idea has to define not only the organization but also the material of composition. Thus, material and organization become intimately interconnected (refer to chapter 2, Stockhausen[47] and Koblyakov[22]).

If we would like to take advantage of the freedom that computers provide us, we have to come up with paradigms of composition which treat material and organization in the same way, not only emotionally and spiritually, but also very precisely and logically. In this work, we have experimented with the principle of self-similarity which is very close to Lindenmayer's L-system[25] as a synthesis method[1] The question of why self-similarity can be useful in music is discussed in chapter 2. Self-similarity provides us with a simple paradigm to view material and organization as a single parameter, and therefore, view sound and music as the same. Self-similarity also provides us with tools to control the perceptual continuum which exists between pitch, rhythm, and form. In this chapter we will explain the synthesis method we have devised and present some of its results.

5.2 Synthesis Method

5.2.1 The Synthesis Paradigm

In this method the user defines a hierarchy of structures to create the sound. The hierarchy can contain recursive elements. This structure is defined by a series of factor arrays. "Time" is the factor which defines the segmentation of time into different *cells*. All parameters are developed by applying the current level factors to higher level values. Thus, a segment of sound becomes

[1] We would like to point out that this work started before we had any formal knowledge of the L-system, fractals or chaos.

a multi-layer collection of *cells* organized in time, while a series of parameters are active for the duration of every *cell*. It is useful to explain the synthesis method with a simple example. In this example, we will explain how the time segmentation and development of a single parameter (frequency), is achieved. Imagine that we define a structure with equal time segmentation (0.5, 0.5), and frequency factors of 1 and 2. We will assume that we want to synthesize 2 seconds of sound with an initial frequency value of 100 (the word "initial" does not mean that the sound is going to start with that frequency or even have a partial at that frequency; it simply means that this is the value with which the parameter development starts). First we divide the time according to the time segmentation factors. Then we multiply the initial value by the two factors and assign new values to each segment. If we recursively apply this process to each segment, we obtain a multi-layer series of frequency values (Table 5.1). These values can be used for a variety of methods of synthesis of sound (e.g. waveshaping, granular, FOF, or MIDI pitch sequences) or graphics.

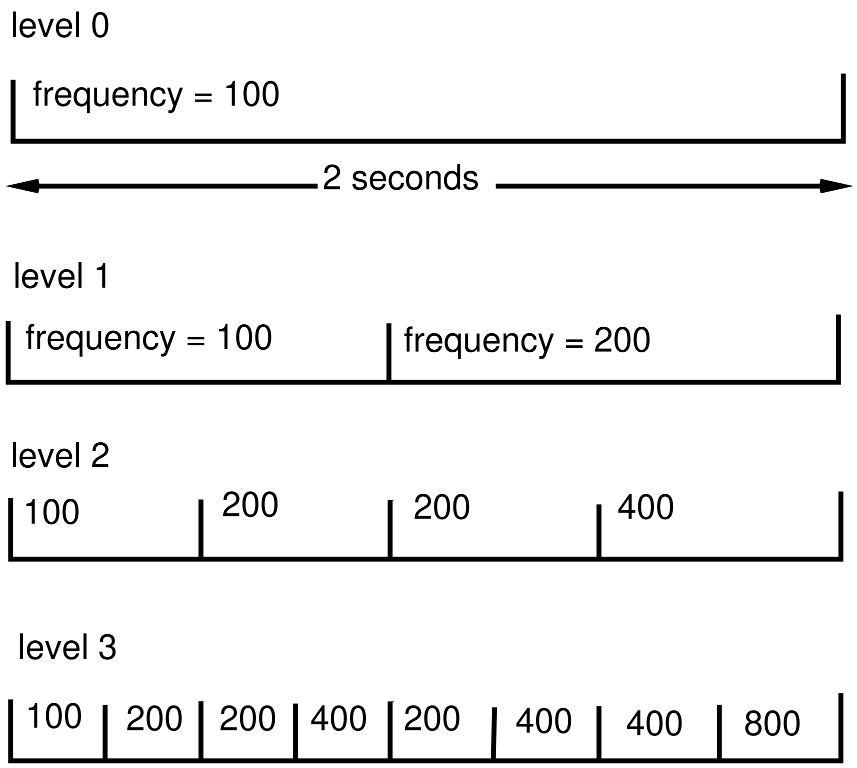
5.2.2 The Synthesis Language

A language was developed for specification of the synthesis hierarchy[2]. For every layer of the parameter definition, one defines a *seed*[3], which itself is a collection of *structures*, and pointers to objects for production of the end result. These latter objects are responsible for mapping the developed parameters to the desired output (e.g., soundfiles, scores for other systems, or graphical pictures). *Structures* are a collection of *points. Points* are a collection of *factors* and *options* and a pointer to a *seed*, which defines their lower content. Some of the normally used factors are "time", "frequency", "amplitude", and "channel values" (for multi-channel synthesis.) The program first starts with the *seed* called "mainseed", which has a *point* as its initial starting value. Then, according to the factors found in the *points* in the *structure* of "mainseed", it re-writes the initial "main" as a series of *seeds.* This procedure is repeated recursively until the duration of a *cell* is smaller than the "stop recursion" value. At every level for every *seed* an output production service routine is called, with the *seed* value (which is represented as a *point*) as its argument. The *factors* for *points* can either be double

[2]Mammad Zadeh developed the initial parser.

[3]The name, seed, was suggested by Gerhard Eckel during discussions on the subject in summer of 1989.

Table 5.1: Parameter development used in the synthesis method is illustrated in this table. The time segmentation of (0.5,0.5) implies an equal binary segmentation of time.



	Time Segmentation	Frequency Factor
Segment 1	0.5	1
Segment 2	0.5	2

precision values, or expressions. One is able to access all the values of the higher levels by using expressions. A single value used as a factor, for example α, without an expression is a shorthand notation for the expression:

$$x_{l+1} = \alpha x_l$$

where x_l represents the value of the *factor* x at level l.

The production objects can have a single *table* and a single *window* attached to them. In the *sound* production object, the *table* is used as a lookup table with increments defined by the frequency *factor*, while the *window* is used as an amplitude window for the duration of the *cell*. Every *point* can as well have a *table* and a *window* which override those in the production objects. The language itself is rather simple to understand once one understand the connection between different objects. Its syntax is very close to structure declaration of the C language, and in fact, every score is passed through the C language preprocessor, so that comments and C style macros can be used in the score. Rather than explaining every detail of the language, we will go through a few examples, and shall explain the scores and the synthesis method in more detail while discussing the results.

5.3 Examples and Results

5.3.1 Two Simple Examples

The first operations that may come to mind using self-similarity involve a fractal as a set of pitch sequences. We will first present two very simple examples which, we think, will make the method more clear. The score for audio example 4 is printed in table 5.2; the values used in this example are similar to the example explained in table 5.1.

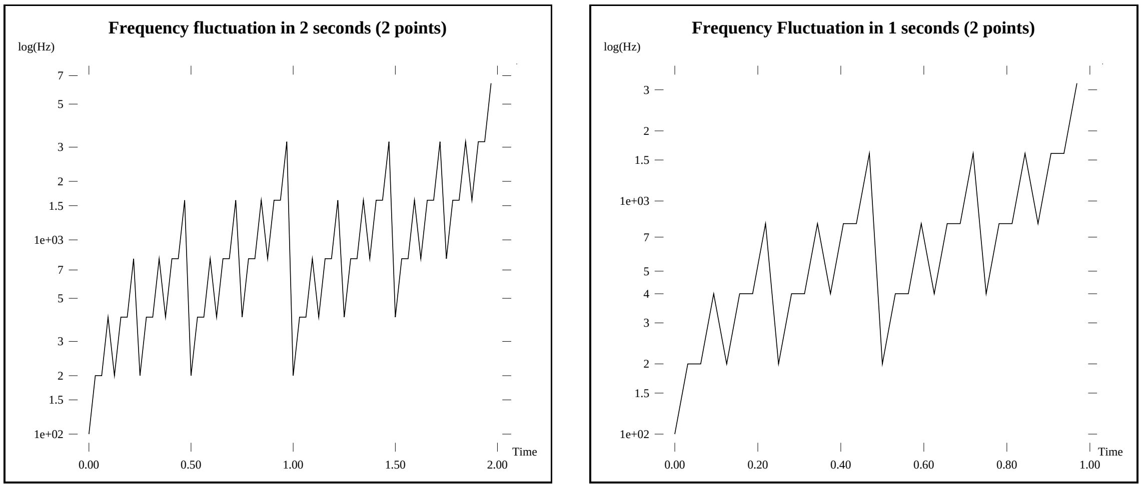
At the end of the score is the definition for the "mainseed"; it defines an initial value of "init", which is a *point*, a structure ("twopoint") and a production object ("snd"). The *structure* "twopoint" is composed of two points "a1" and "a2", which define an equal binary time segmentation (0.5,0.5) and frequency multipliers of 1 and 2. The option "lastlevel" means that we will only use the last level *cells* of the developed parameters. Both of these *points* refer to the mainseed; therefore, we have a single level recursive hierarchy. The object "snd" defines

```
point init {time: 2; freq: 100; amp: .1; seed: mainseed;}

sound snd {
        time: 10.0; srate: 22050; file: "2p.1.snd"; window: "nowin";
        stop_rec: .05;
}

point a1 {time: 0.5; freq: 1; amp: 1; seed: mainseed; options: lastlevel;}

point a2 {time: 0.5; freq: 2; amp: 1; seed: mainseed; options: lastlevel;}

struct twopoint {a1; a2;}

seed mainseed {value: init; struct: twopoint; seedobj: snd;}
```

Table 5.2: The score for audio example 4.

the sound production values. Since the system has no way of knowing how long the synthesized sound is going to be, we have to specifically define the allocation of the sound buffer, and that is specified by the "time: 10.0" entry. The "srate" entry defines the sampling rate, and "file" specifies the file name to which the produced sound will be written. As discussed before, "window" defines an amplitude window whose length is adjusted to the length of the *cell*; for now we can ignore this entry since we are not applying a window in the process. (Actually, we are applying the window "nowin", which is just a constant value of 1.) A sinusoid table is used by default for a table lookup, and the "stop_rec" (which stands for "stop recursion") specifies a time threshold for the last level of parameter development. We will stop the parameter development process, once we reach a *cell* whose duration is less than the value of the "stop_rec". The *point* "init", which is the initial value of the "mainseed", specifies that we are asking for 2 seconds of sounds to be synthesized while the initial values for the development of the frequency and amplitude parameters are 100, and 0.1 respectively. This score will produce a sinusoid whose frequency is ascending fractally. The frequency fluctuation of this example as well as the frequency fluctuation of half of its duration is illustrated in figure 5-1. The similarity of the two graphs can be seen as four broken lines ascending in 1 or 2 seconds.

In the next example we will show the use of expressions and make the self-similarity of the frequency fluctuation clearer by a trinary segmentation of time. The score for audio example



(a) (b)

Figure 5-1: The frequency fluctuation of the audio example 4 is illustrated. (a) shows the frequency fluctuation in 2 seconds and (b) shows the frequency fluctuation in 1 second. The basic shape of both graphs are similar to each other.

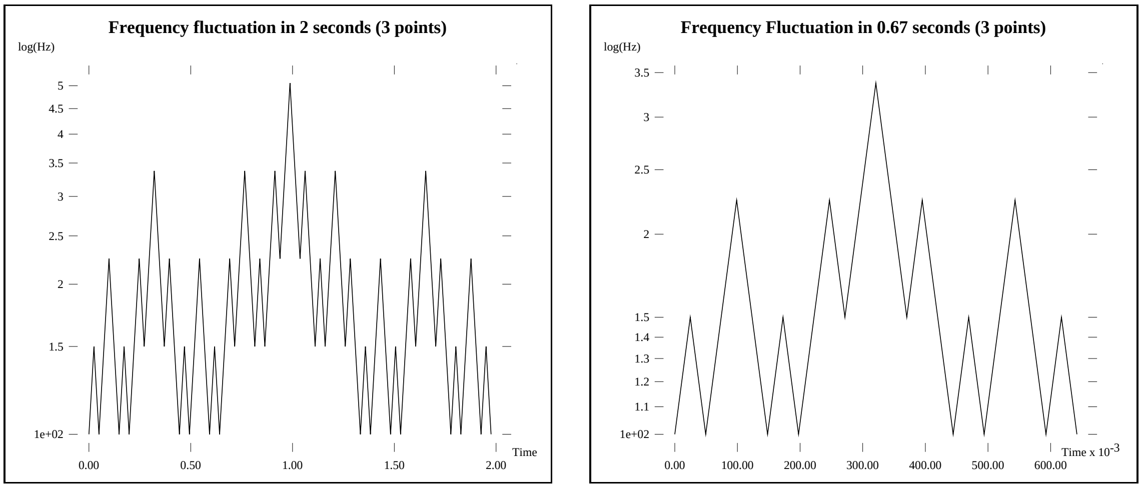
5 is printed in table 5.3. The basic shape of the hierarchy of this score is the same as the first example, except that the *structure* of this example is composed of three points (using the point “a1” twice). The *time* factor is defined as an expression, and the frequency factors used are 1 and 1.5. Expressions have to be quoted with backquotes (‘). In expressions there is no assumption about how the higher level values are treated, and any operation has to be explicitly specified. In this case the variable “time” in the expression ‘time / 3‘ implies that the duration of the *cell* in the current level is one third of the duration of the higher level *cell*. If we had simply used a value of 0.333 for the *time* entry in “a1”, then we had to use a value of 0.334 for the time entry in “a2” to make sure that the duration of all of our *cells* in every level adds up to be the same. It is legal to use time segmentation factors which do not add up to 1. However, that should be used with the knowledge of how the global time is managed in the system, so that undesired side effects would be avoided. The system parses the recursion tree depth first, and advances the global time anytime it reaches the “stop_rec” value in the last level. Therefore, if we used values which did not add up to 1, as we get deeper into the recursion tree the addition of the *cell* duration becomes smaller, and finally we will create a sound shorter than what we had initially asked for. All of the values of all the higher level *factors* can be

```
point init {time: 2; freq: 100; amp: .3; seed: mainseed;}

sound snd {
        time: 10.0; srate: 22050; file: "3p.1.snd"; window: "nowin";
        stop_rec: .05;
}
point a1 {
        time: ‘time / 3‘; freq: 1;   amp: 1; seed: mainseed;
        options: lastlevel;
}
point a2 {
        time: ‘time / 3‘; freq: 1.5; amp: 1; seed: mainseed;
        options: lastlevel;
}

struct threepoint {a1; a2; a1;}

seed mainseed {value: init; struct: threepoint; seedobj: snd;}
```

Table 5.3: The score for audio example 5.

used in an expression. For example ‘freq + 10‘ means that the value of the frequency *factor* in the current level is equal to the value of the frequency *factor* in higher level plus 10. *Factors* can be indexed as arrays to access values of *factors* in the levels not immediately preceding the current level. For example, ‘freq[1] + 10‘ means that the value of the frequency in the current level is equal to the value of the frequency factor in two levels above. Notice that ‘freq + 10‘ is a shorthand for ‘freq[0] + 10‘. Currently two global variables are recognized: “rec_level” is the value of the current recursion level and “cur_time” is the value of the currently advanced global time. Let us get back to our examples. Figure 5-2 illustrates the frequency fluctuation of the audio example 5. In this case the self-similarity of the frequency fluctuation is rather apparent.

The next step is to use all the values of the factors in all levels for synthesis. The result will be as if we had synthesized a signal for every level of the parameter development (as described above) and had added all the signals together. The system takes this action by default, unless the “lastlevel” option is set. The score for audio example 6 is printed in table 5.4, which is the same as the previous example without the “lastlevel” option. Notice that we have used



(a) (b)

Figure 5-2: The frequency fluctuation of the audio example 5 is illustrated. (a) shows the frequency fluctuation in 2 seconds and (b) shows the frequency fluctuation in 0.667 second. The basic shape of both graphs is a triangle.

```
point init {time: 2.08286; freq: 100; amp: .7; seed: mainseed;}

sound snd {
        time: 10.0; srate: 22050; file: "3p.2.snd"; window: "nowin";
        stop_rec: .05;
}

point a1 {time: ‘time / 3‘; freq: 1;   amp: .6; seed: mainseed;}

point a2 {time: ‘time / 3‘; freq: 1.5; amp: .6; seed: mainseed;}

struct threepoint {a1; a2; a1;}

seed mainseed {value: init; struct: threepoint; seedobj: snd;}
```

Table 5.4: The score for audio example 6.

```
point init {time: 2.04336; freq: 100; amp: .5; seed: mainseed;}

sound snd {
        time: 10.0; srate: 22050; file: "2p.2.snd"; window: "nowin";
        stop_rec: .05; loop: 3;
}

point a1 {time: 0.5; freq: 1; amp: .5; seed: mainseed;}

point a2 {time: 0.5; freq: 2; amp: .5; seed: mainseed;}

struct two8 {a1; a2;}

seed mainseed {value: init; struct: two8; seedobj: snd;}
```

Table 5.5: The score for audio example 7.

amplitude factors of 0.6, so that the higher frequency partials would have lower amplitude. We have also used a value of 2.08286 for the duration of sound, so that the number of samples can be divided by 3 up to the point that we stop the parameter development. This is an important issue in this example since we are not using any amplitude window for the cells. Had we used a value of 2.0, we would have produced clicks due to the round-off error of calculating the number of samples of the duration of every cell. Finally, we can hear the additive version of our first example as audio example 7, whose score is printed in table 5.5. In this example we have used the "loop" option of the sound object and looped the result 3 times.

5.3.2 Self-contained Examples

The score for audio example 8 is printed in Table 5.6. The time segmentation in this example is 20 to 1, and the different partials are added to the sound from top to bottom. The spectrogram of the whole duration and three seconds of the sound, which is 60×0.05, is illustrated in figure 5-3[4]. As it can be seen, the same structure is manifested in both spectrogram. Almost any picked segment according to the similarity factors of this sound manifests the same structure. For example the segments 3.0-5.85 ($5.85 = 3 + 3 \times 0.95$) is a scaled down version of the segment

[4]All the spectrograms for this theses as well as the soundfile interface tools were written by Dan Ellis. These tools were indispensable to development of this project.

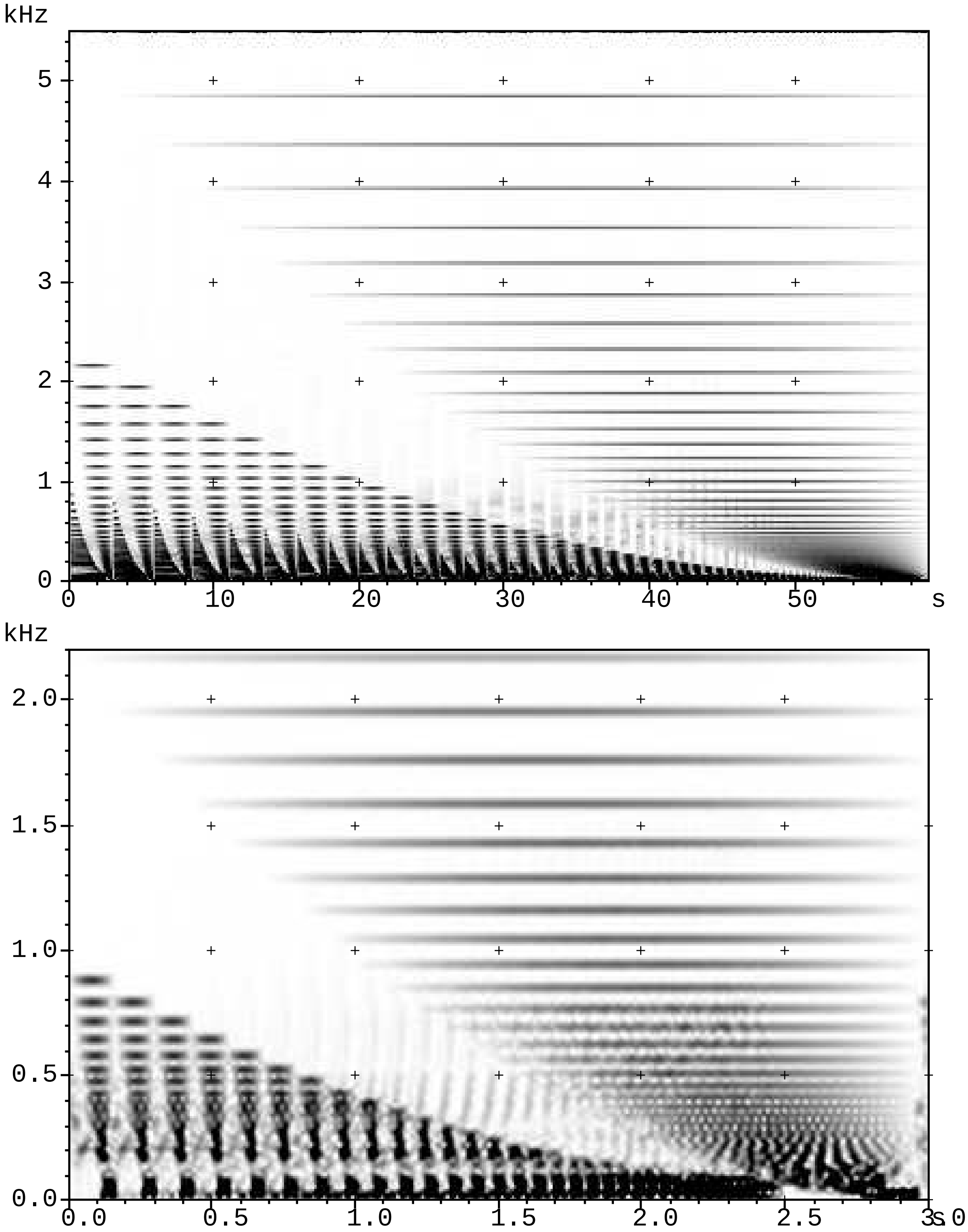


Figure 5-3: The spectrogram of the first 60 and the first 3 seconds of audio example 8 is illustrated. The spectrogram of the first 3 seconds is rescaled by a factor of 0.4. As it can be seen, the same structure is manifested in this sound in two levels of our auditory perception.

```
point init {time: 60; freq: 5400; amp: .01; seed: mainseed;}

sound snd {time: 60; srate: 11025; file: "sound"; stop_rec: .02;}

point a1 {time: 0.95; freq: .9; amp: 1.01; seed: mainseed;}

point a2 {time: 0.05; freq: .4; amp: 1.5;  seed: mainseed;}

struct s1 {a2; a1;}

seed mainseed {value: init; struct: s1; seedobj: snd;}
```

Table 5.6: The score for audio example 8.

```
point init {time: 5; freq: 5400; amp: .01; seed: mainseed;}

sound snd {time: 60; srate: 11025; file: "short"; stop_rec: .02;}

point a1 {time: 0.95; freq: .9; amp: 1.01; seed: mainseed;}

point a2 {time: 0.05; freq: .4; amp: 1.5;  seed: mainseed;}

struct s1 {a2; a1;}

seed mainseed {value: init; struct: s1; seedobj: snd;}
```

Table 5.7: The score for audio example 9.

0.0-3.0. This similarity can be seen as an exponentially decaying shape in the lower spectrum of the sound. The sound starts with this shape and at the same time that the listener is becoming aware of this decaying shape, the larger picture of the sound emerges, which is the similarity of the ending segments 0-60, 3-60, 5.85-60, etc.

The time segmentations of 0.05 to 0.95, or frequency factors of 0.4 and 0.9, may look arbitrary. In fact, in the process of the development of the system, the examples which we have called self-contained started as experimentations and the numbers were tuned with every listening. In this paradigm, one is able to work with smaller versions of the sound for development and tuning, and in this way save time in the synthesis process. Table 5.7 is the score for audio example 9, which is a short version of the previous example.

```
point init {time: 40; freq: 4000; amp: .01; seed: mainseed;}

sound snd {time: 60.0; srate: 22050; file: "water"; stop_rec: .1;}

point a1 {time: 0.18; freq: `freq - 280`; amp: 1.3;  seed: mainseed;}

point a2 {time: 0.02; freq: `freq - 160`; amp: 1.4;  seed: mainseed;}

point a3 {time: 0.8;  freq: `freq - 360`; amp: 1.01; seed: mainseed;}

struct s1 {a1; a2; a3;}

seed mainseed {value: init; struct: s1; seedobj: snd;}
```

Table 5.8: The score for audio example 10.

All the frequency partials in the previous examples were geometrically related to each other. We can create harmonically related partials by using expressions for frequency factors. The score for the audio example 10 is printed in table 5.8 and figure 5-4 illustrates the spectrogram for this audio example.

5.3.3 Layered Examples

Audio example 11 was created by layering many transposed copies of a single shape. The score for this example is printed in table 5.9. The *structure* for this example has 3 points; the first and the last points both have the "silent" option on; therefore, it is only the middle point "a2" which creates any sound. The "window" used for this example is the final 0.3 seconds of the spoken word "light" without the letter 'l'. Therefore, the "window" starts with a voiced sound and ends with a noisy fricative, and this structure is magnified to 20 seconds in the duration of the example. Notice the use of the "interpol" option. By default, the system does not interpolate any of the values either when applying amplitude windows or when looking up tables. The option "finterpol" means to interpolate during a table look up, and "ainterpol" to interpolate when applying amplitude windows, and "interpol" to interpolate in both cases. The spectrogram for this example can be seen in figure 5-5. This sound was used as the opening sound of *Morphosis* (1992), which is a piece composed by the author using this system and is

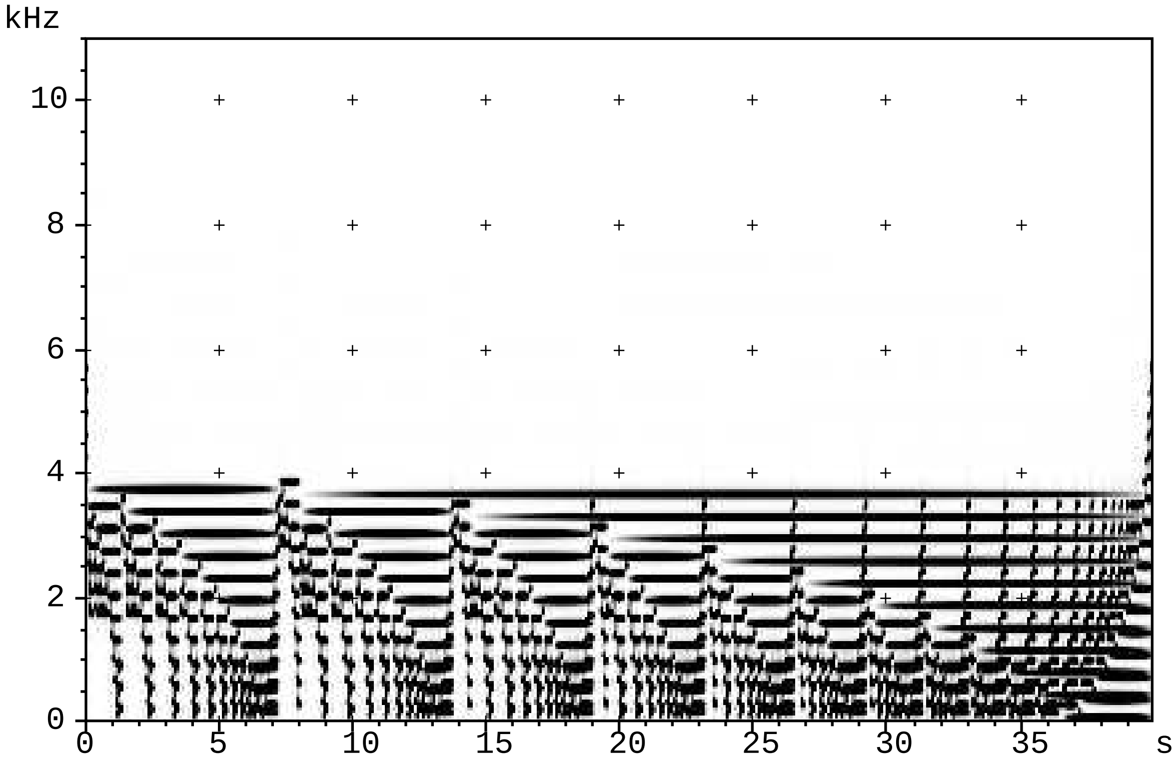


Figure 5-4: The spectrogram of audio example 10 is illustrated.

```
point init {time: 20; freq: 1; amp: .7; seed: mainseed;}

sound snd {
        time: 25; srate: 44100; file: "sound";
        stop_rec: 5; window: "voice";
}

point a1 {time: 0.015; freq: 1; amp: 0; seed: mainseed; options: silent;}

point a2 {
        time: 0.98; freq:  ‘freq + 150‘; amp: 1.005;
        seed: mainseed; options: interpol;
}

point a3 {time: 0.005; freq: 1; amp: 0; seed: mainseed; options: silent;}

struct s1 {a1; a2; a3;}

seed mainseed {value: init; struct: s1; seedobj: snd;}
```

Table 5.9: The score for audio example 11.

partly described in appendix C. Audio example 12 has been created by applying the same type of procedure to a longer melody of a cello sound.

The score for audio example 13 is printed in Table 5.10. This example adds many layers of looped sound of a piano note. The entry "`table: "piano/d2":25000-157300;`" picks 3 seconds of a sampled piano sound. The numbers specified in the table entry are sample numbers, and this option is provided for precise definition of tables. The point "a2" is "silent". The point "a1" segments the time by a factor of 0.95 while the frequency factor of it is 1.052632 which is 1/0.95. Thus, as the segments get shorter the frequency value gets larger by the same factor. In this way, every layer becomes 20 (60/3) notes. This example also shows how we can create stereo outputs. The number of channels are specified in the "snd" object by the "`nchnls: 2;`" entry. The factors "ch1" and "ch2" in the point "a1" are applied to channel 1 and channel 2 respectively. The factors for "ch1" and "ch2" are specified as expressions by using the "if" function. Three arguments are passed to "if"; the first is a condition, and the value of the "if" function is either the second or the third argument depending on the truth value of the first argument. Therefore, in this case the values of both "ch1" and "ch2" are 0.5 if we are in the

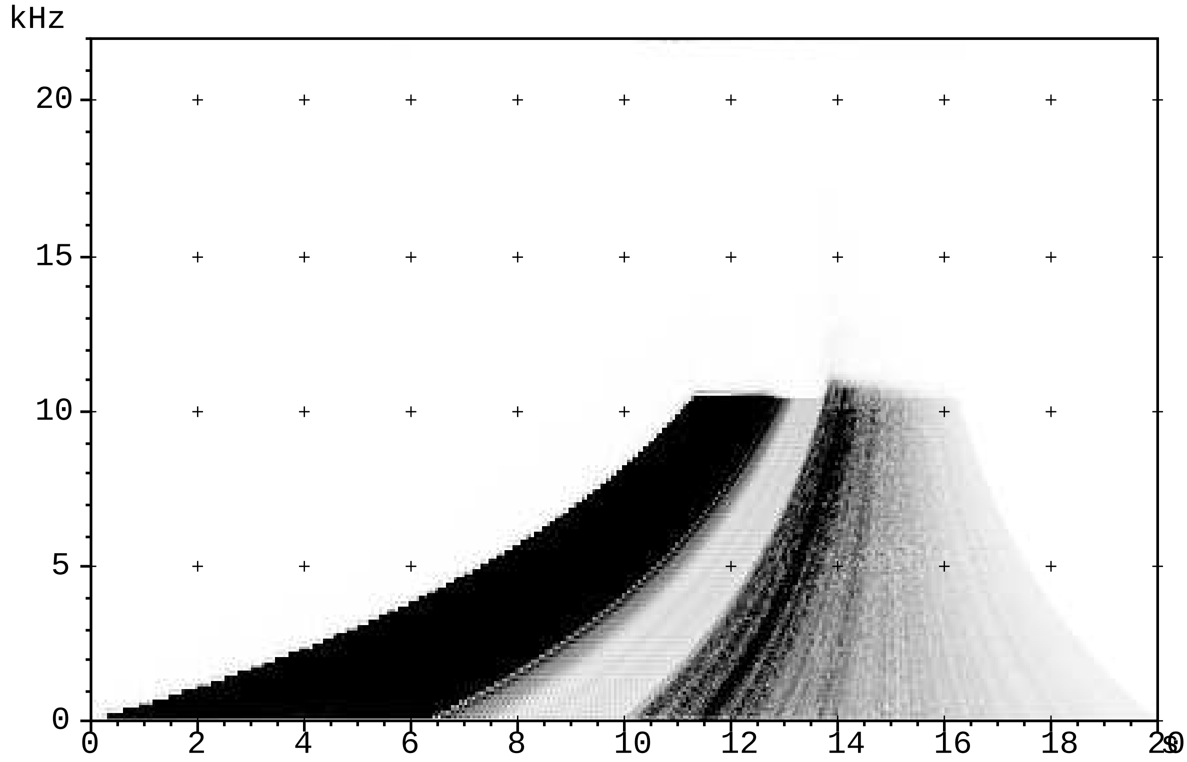


Figure 5-5: The spectrogram of sound example 11 is illustrated.

```
point init {time: 60; freq: 1; amp: .08; seed: mainseed;}

sound snd {
        time: 60; srate: 44100; file: "piano"; stop_rec: 10;
        window: "nowin"; table: "piano/d2":25000-157300; nchnls: 2;
}

point a1 {
        time: 0.95; freq: 1.052632; amp: .98; seed: mainseed;
        options: fcycle finterpol;
        ch1: ‘if (rec_level == 1, .5, rec_level % 2)‘;
        ch2: ‘if (rec_level == 1, .5, (rec_level + 1) % 2)‘;
}

point a2 {time: 0.05; freq: 1; amp: 1; seed: mainseed; options: silent;}

struct s1 {a1; a2;}

seed mainseed {value: init; struct: s1; seedobj: snd;}
```

Table 5.10: The score for audio example 13.

first level (rec_level == 1); otherwise their values is either 1 or 0 depending on the level. Thus, except for the first level, every other level of the sound is assigned to either channel one or two. A similar version of this sound was used for the ending of *Morphosis*.

5.3.4 Rhythm Examples

In this section we will present two examples for creating pieces with strong rhythmical chatacter. The score for audio example 14 is printed in table 5.11. This example has two sections, and its structure is reflected in the score as three *seeds*. The "mainseed" has the information about how the two sections are organized. The window for this example is the first 4722 samples of a powertom drum. The sound of the sampled segment is played before the synthesized segment in the audio example. The value 4722 was picked carefully so that the last amplitude value of the window would be 0. Using such a window creates a deep drum sound for long *cells* and high pitched sound for short *cells*. Notice that in this example we are using a sampled sound as an amplitude window.

```
point init {time: 90; freq: .5; amp: .23; seed: mainseed;}

sound snd {
        time: 100.0; srate: 44100; file: "rhythm1"; stop_rec: .1;
        window: "drums/powertom10":4722;
}
point part1 {time: `time * 2 / 3`; freq: 1; amp: 1; seed: tense;}
point part2 {time: `time / 3`;     freq: 1; amp: 1; seed: resolve;}
struct parts {part1; part2;}
seed mainseed { value: init; struct: parts; seedobj: snd;}

/* tense seed */
point p1_1 {time: 0.5; freq: 1;   amp: .8;  seed: tense;}
point p1_2 {time: 0.5; freq: 1.5; amp: 1.1; seed: tense;}
struct s1 {p1_1; p1_2;}
seed tense {value: init; struct: s1; seedobj: snd;}

/* the resolve seed, the inverse of the tense seed */
point p2_1 {time: 0.5; freq: 1.5; amp: 1.1; seed: resolve;}
point p2_2 {time: 0.5; freq: 1;   amp: .6;  seed: resolve;}
struct s2 {p2_1; p2_2;}
seed resolve {value: init; struct: s2; seedobj: snd;}
```

Table 5.11: The score for audio example 14.

This example useusess a very subtle feature of the synthesis program. By default, anytime a *cell* is ready to be synthesized, the frequency value is adjusted so that an integer number of cycles would fit in the duration of the *cell*. At first this method was used to reduce the noise due to the fractal modulation of the amplitude, ensuring that the amplitude factors would change when the amplitude of the signal is zero (if the table is cropped carefully). The system turns off this processing if the option "fcycle" (which stands for fractional cycle) is set. The first section of this example is the *tense* seed whose structure contains two points. The initial frequency value is 0.5, and since we have not set the "fcycle" option, it gets translated to 0. A frequency of zero is equivalent to using the value of the last sample used from the "table" in any level. A sinusoid function is used as a "table" for this example, and since the frequency factor of the first point of the first part is 1, at the beginning of the sound all the frequency values for all the levels are 0 and no sound is generated for 3 seconds. As the frequency value is modulated by the second point, whose frequency factor is 1.5, different layers start to generate sound. This process can actually be heard clearly in the audio example. If we were to graph time versus the number of layers present in the sound, we would come up with a shape similar to figure 5-1. The amplitude factor of the second point is also higher than the first, and the shape of the amplitude of the first part of the sound is also similar to the shape of figure 5-1. The second part of the example is basically the inverse of the first part. As it is coded in the two points "part1" and "part2", the first part lasts for 2/3 of 90 seconds which is 60 seconds and the second part lasts for 1/3 of 90 seconds which is 30 seconds. The amplitude factor of the second point in the second part is 0.6 as opposed to the amplitude factor of the first point in the first part which is 0.8. This difference causes a faster drop in amplitude in the second part of the sound.

Audio example 15 uses the same principles as the last example, except that its hierarchies have two levels of recursions, and different windows are assigned to different points. Since we are using sound samples as amplitude windows, the character of the window is heard as the timbre for that segment. Therefore, by assigning different windows to different points, we are actually assigning different instruments to them. This example uses three different windows, which are samples from: a powertom (which was used in the last example), the sound of breath, and a ride cymbal.

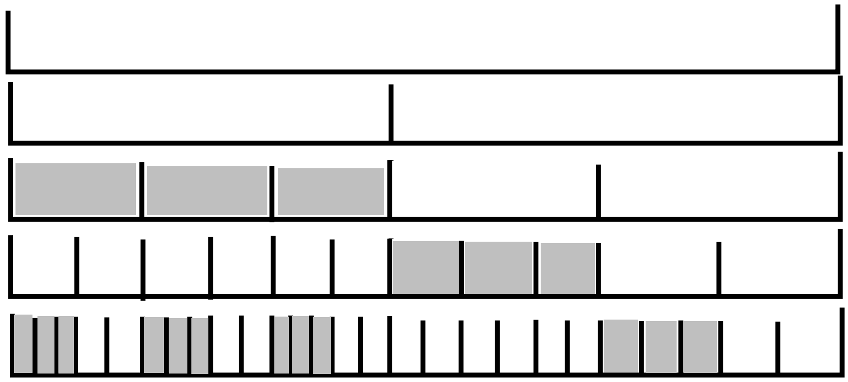


Figure 5-6: The time segmentation of the first 4 levels for audio example 15 is illustrated. This example has a two-level hierarchy. The first level goes through a binary segmentation, and the first part of the second level goes through a trinary segmentation. The time segments which have gone through a trinary segmentation are shaded.

The first section of the score for this example is printed in table 5.12. This example has three sections and we will briefly explain its first section. The default window for the score, which is the breath sound, is defined in the "snd" object. The first section is specified by the point "part1", whose seed is "tense". The seed "tense" is composed of two points with equal time segmentations. The seed for the first point is "tense2" which is composed of three points whose seeds are "tense". Therefore, the hierarchy for the first section of this example ("part1"), is a binary segmentation whose first part has a trinary segmentation, and the second segment a binary segmentation. The segmentation of time for the first 4 levels is illustrated in figure 5-6. Any time segment which has gone through a trinary segmentation has been shaded in the figure. The effect of this hierarchy can be heard as a compound rhythm in the first section, which gradually moves toward a simple rhythm and connects itself to the second section at 60 seconds. The second section is a two-level simple binary segmentation, and the basic structure of the third section is the inverse of the first section.

5.4 Future Development

The language described in this chapter is not ideally meant to be manipulated or looked at by humans. The language was designed as an intermediary protocol for storage of the hierarchies. A major part of the future development of this system is creating a graphical interface to the

```
point init  {time: 100; freq: .5; amp: .4; seed: mainseed;}

sound snd {
        time: 120.0; srate: 44100; file: "sound";
        stop_rec: .1; window: "breath";
}
point part1 {time: 0.6;  freq: 1; amp: 1;   seed: tense;}
point part2 {time: 0.15; freq: 2; amp: 1;   seed: sustain;}
point part3 {time: 0.3;  freq: 1; amp: .75; seed: resolve;}
struct parts {part1; part2; part3;}
seed mainseed {value: init; struct: parts; seedobj: snd;}

/* tense seed */
point p1_1 {
        time: 0.5; freq: 1; amp: .8; seed: tense2;
        window: "drums/powertom10":4722;
}
point  p1_2 {time: 0.5; freq: 1.5; amp: 1.07; seed: tense;}
struct p1_s {p1_1; p1_2;}
seed tense  {value: init; struct: p1_s; seedobj: snd;}

/* tense2 */
point px1_1 {
        time: `time / 3`; freq: 1; amp: .6; seed: tense;
        window: "drums/powertom10":4722;
}
point px1_2 {
        time: `time / 3`; freq: `freq * 4 / 3`; amp: 1; seed: tense;
}
struct px1_s {px1_1; px1_2; px1_2;}
seed tense2 {value: init; struct: px1_s; seedobj: snd;}
```

Table 5.12: The score for audio example 15.

synthesis language. Since this system treats the parameters of all levels of the sound in the same manner, the graphical interface has to be able to represent the hierarchical structures in sound as well as in the music domain. Once such an interface is created, it will become possible to create a library of sound and musical structures which could be used by other scores.

There are other features in the method which we have used for creation of *Morphosis*; however, due to their experimental nature they have not been explained here. The basic idea behind these features is to define some linear operations which will be applied to the different synthesis parameters for the duration of the *cell*. For example, the frequency factor in all the presented examples stayed constant for the duration of every *cell*. One can imagine a frequency envelope which could be applied to the frequency value of every *cell*. The parameters for the frequency envelope would themselves go through the system's development process.

Currently all the development processes in the system are deterministic. Even though adding random elements may have seemed to be an interesting addition to the features of the system, we believed that they would create paths of development which would be hard to understand. However, once the current state of the system is better understood, the system could be used for organizing chance operations, and perhaps adding some flavor of a $1/f$ process would in fact enrich the system.

It is easy to create self-similar structures; however, not every self-similar structure is musically interesting. In fact, most of the presented examples have been arrived at after many hours of searching and tuning. At first, the behavior of the system seemed very erratic the reason being that it performs massive amounts of related operations on the initial structures. Most parameters can take on different roles at the same time. For example, consider the parameter for time segmentation. When we apply a "window" to every *cell*, the shape of this "window" is scaled to fit the duration of the *cell*. Thus, the frequency in which the window is played in every *cell* is inversely proportional to the duration of the *cell*. Therefore, the time segmentation factor defines a plexus of time-frequency relationships. The frequency factor can also act as two different agents as well. When we define a frequency factor in the low frequency region (e.g., 0.1 to 2 Hz) depending on the shape of our lookup "table", this factor can actually behave as an amplitude window. For example, the shape of a sinusoid at the frequency of 0.25 Hz and phase of zero can act as a fade-in structure in a *cell* whose duration is 1 second. Thus, small changes to the initial conditions could result in drastic perceptual differences. This situation can best be thought of as the "Butterfly Effect" which is described by Glieck as[14, page 8]:

> In weather, for example, this translates into what is half-jokingly known as the Butterfly Effect — the notion that a butterfly stirring the air today in Peking can transform storm systems next month in New York.

Sensitivity to initial conditions is a characteristic of chaotic systems.

We arrived at the different categories of the presented sounds rather intuitively. Some basic principles have become clear to us. For example, equal time segmentation, in conjunction with a "window" which contains a percussive sound, creates a rhythmical form. If the "window" is a simple shape, the rhythmical structures are heard as the characteristics of the timbre of the sound; in this case, the form is usually determined by the frequency and amplitude factors. Layering different transposed copies of related shapes is probably one of the simplest and most finely controllable structures which we can create. By controlling the concentration of the material (in the simplest case, the number of shapes added together) we can create sounds with archetypical climactic form. This idea was used in the first 45 seconds of *Morphosis*. We believe that the musical possibilities of the system in its current shape have not yet been exhausted, and a major part of the future work will be to use and understand the behavior of the system.

An important future goal is to create a notation system which is completely intuitive to the composer. Obviously, we must assume some knowledge of electronic and computer music. However, the main effort is to draw the line between what should be the task of science and what should be the task of music. For example, a composer does not need to know the different types of metal which are used for piano strings. However, using the behavior of such characteristics in a piece could create wonderful subtle effects. Asking a composer to program in standard computer science languages is similar to asking him to make his own pen before transcribing the music on paper. The language we have defined in this thesis is meant to be used as a format for storing different types of structures. The interface to the composer would be a programmable notation system which provides a way of notating music and sound in the same manner[52]. Different composers have tried to create such systems. In dealing with the continuum of pitched sounds to noise Machover writes[27]:

> An efficient notation that includes complex timbral transformations is still to be

> found. I believe that those systems that incorporate the most elements from common practice notation will be the most successful! (I use, for example, a simple system of note-heads that indicate gradual transition from pitch to complete noise: normal note-head, note-head in parentheses, cross in parentheses, cross alone. This seems to be clear to most players.)

Notice that timbral changes are changes in sound, and the notation system before the 20th century had never been used for notating sound. Schoenberg was aware that the traditional notation system had to be changed to support his new ideas, and he made an attempt on creating one[40, page 354]. Even though, this notation system provides a more uniform quantization of the pitch continuum, it does not address the problem of timbre. Perhaps if Schoenberg had not stopped himself from breaking the harmonic structures of the individual tones in music, he also would have provided us with such a timbre notation system. It may be interesting to note that my initial inspiration to conduct the research that led to the work described in this thesis was, in fact, the desire to invent a totally new, formally intelligent and interactive, notation system for computer music.

Chapter 6

Summary, Conclusions and Speculations

> I discovered the secret of the sea in the meditation upon the dewdrop.
>
> Khalil Gibran[13, page 11]

Communication, specifically in the musical domain, is the main topic of this thesis. I believe that music exists not only in the structures which organize sound, not only in the micro-structures of the sounds being organized, but also in every decision that the musician makes in everyday life. For a musician there is little difference, if any, between music and truth. When we try to communicate, we have to compromise our truth. Even though the concept is universal, truth is a highly personal and local entity and it will stay that way. Communication is an art. It is, however, the art of stating the truth with lies, in a way that sets up a significant relationship between the parties involved. Music happens when we communicate with no compromise.

6.1 Technical Issues

In this project we applied the evolution of text processing in computers to the practice of computer-aided composition. The basic idea was to create an abstraction layer between the compositional and the computational process. Very few synthesis methods have been able to create such a layer which gives the composer the ability to think globally about music. With its efficiency for creating sound, FM synthesis[5] was probably a strong factor in the commercialization of synthesizers; however, it falls short of providing any tool which is musically intuitive. In my opinion, that is why our ears are able to recognize the sound of FM rather easily and get bored by it. I believe that CHANT, which uses the FOF synthesis method developed by Rodet[37], has been the most successful technique for providing a musical synthesis method for the composer (See Barrier[1] and Harvey[17]).

We defined a synthesis method which made no distinction between the micro and macro-structures of music. The synthesis parameters are defined as a hierarchy of structures which can contain recursive elements. The system can create self-similar or self-affine sounds from a number of different point of views (e.g., pitch fluctuation, amplitude fluctuation, the shape of the spectrogram, and the way different layers of sound or music are faded in). A simple language was developed for specification of the hierarchy. The system proved to be able to create extremely complex results with very simple structures; however, not every result was musically interesting to us. Most of the research work with the system was to search for structures which resulted in musically interesting sounds. These structures showed a certain versatility that, through very little change, could create new sounds which were different from the original results but still remained interesting to our ears. Thus, the relationship in the structures defined a certain class of sound in the system which could be tailored for a specific purpose. A piece was composed using the system which shows that it is possible to create sounds with specific intentions.

The concept of self-similarity was used because since self-referentiality was formally introduced to me by "*Gödel, Escher, Bach*"[18], I have been rediscovering it in many unexpected contexts. When we combined some very simple computer science ideas such as programmability, hierarchy, and functionality to what we knew of computer music, self-similarity had developed itself in the design of the system by unifying the different perceptual levels in the model. In our search, a sense of duality was discovered in the traditional way that two concepts were treated: one in the treatment of sound and music, and the other in the technical treatment of random and deterministic signals.

$1/f$ noise was studied and a simple analysis of many pieces which we had access to in MIDI format was conducted. A signal was extracted from these pieces and almost all the spectrum of the extracted signals showed a slope close to that of $1/f$ noise, which means that the signal is neither random nor too correlated. $1/f$ noise falls on the border between the signals which we treat as random (for which we use statistical methods to study) and deterministic signals (for which we use very precise functions). This class of signals creates some technical problems by the fact that a signal with a $1/f$ spectrum shows a scale invariant auto-correlation, which in turn means that there are certain correlations among all levels of the signal (i.e. micro or macro-structures).

6.2 Musical Issues

Schoenberg's theory of harmony was studied from a very abstract point of view. We concluded that almost all of his conceptions were based upon the relationship between form and content. He recognized that tonal form was the expression of the inner content of its material which is harmonic sound. By this recognition, he established a physical continuum between consonances and dissonances. Logically and aesthetically, this discovery had a revolutionary effect. The revolution was the breakdown of tonal form in music, which also coincided in time with the 20th century breakdown of the traditional form in painting, poetry, and mathematics. Schoenberg's greatest concern was clarity and comprehensibility. By denouncing tonal form as an eternal law of music, he basically denounced all pre-established forms. Tonal form provides a convenient way of communication, where some protocols are already agreed upon between the composer and the listener. However, in Schoenberg's theory, the composition has to define not only its content but also its form, or in other words, it has to define not only what it wants to communicate to the listener, but also how it is going to communicate it. These are not two different tasks; the form and the content are intertwined in the musical idea, and the way they show up in the composition can be thought of as the sound and the music. We established the idea of a plexus in musical communications which can be thought of as the manifestation of the non-linear relationship of form/content, sound/music, channel/information, or comprehensibility/originality.

By breaking the logical barrier between consonances and dissonances in a physical way, Schoenberg freed the structures of sound. However, perhaps he himself was not aware of the full implications of his ideas. We repeat one of Schoenberg's quotes for its importance[40, page 137]:

> The path to be trodden here seems to me the following: not to look for harmonies, since there are no rules for those, no values, no laws of construction, no assessment. Rather, to write parts. From the way these sound, harmonies will later be abstracted.

Aesthetically, the idea is simple. Art is not up for judgment unless it is done for its own sake. However, when we think about that idea and take it to its formal end, we reach some very complex issues. Where does harmony of communication come from if it is not worked on? Will god write the harmonies for us? If so, that is faith, and, as we have very briefly mentioned, faith is a paradox that cannot be communicated[21]; concerning these issues we showed that there is a certain compromise between originality and comprehensibility. From this quote, we conclude, that art, which is really a way of life, is not a job and the artist cannot be concerned with the assessment of his or her work; not because it is not important, but because it does not help in any way and perhaps can never be "known".

Technically, Schoenberg's ideas opened so many doors in music that the problem was not how to find an original idea but rather how to make such originality aesthetically accessible. Schoenberg stopped himself from manipulating the structures of sound since he thought there were no instruments that could play what his imagination would have created[41, page 424]. He broke the preestablished forms up to the boundaries of note intervals in the well-tempered scale. Many composers who followed his path (such as Cage, Boulez, and Stockhausen) devised their own language of form. Stockhausen went a step further in understanding the relationships between material and organization, and introduced the idea of synthesis or the composition of sound.

There is perhaps very little argument about the fact that a real work of art has a certain homogeneity. The idea of a musical theme defining the music as well as the sound, or in other words unity of form and content, made Stockhausen aware of the unity in different levels of our perception. As we have shown, the requirements of homogeneity in a balanced (random vs. correlated) piece of music and the unity of form and content are the requirements of unity in our different perceptual levels. Every one of these factors points to the concept of self-similarity. The homogeneity of music implies that every part of the piece sends the same information; that should be true not only for the smaller sections which follow one another but also for the larger sections as well which are composed of those smaller sections. The relationship of the macro-structures and micro-structures are in fact the relationships between the material and organization of the piece. The different levels of our perception, which in music are represented by the feeling of form, rhythm, and pitch, are connected to each other with self-similar structures. The sensation of pitch comes from a rhythmical organization of vibrations; rhythms are created from the repetition of simple forms of pitches with related variations. The feeling of form comes from a certain coherency in the rhythmical structures of pitch. And finally for the form to have any meaning, for example in the tonal form, the feeling of form connects itself to a large-scale feeling of pitch. This is one possible view of how form works in tonality.

Pitch is timbre reduced to a single dimension according to harmonic relationships. Serialism simply implies that the unifying concept relating our different levels of perception does not have to be the harmonic relationship. By this fact, serialism implies that not only the composition has to define how this relationship is used, but also that it has to define the relationship in the first place. Before electronic music existed, it was difficult to conceive of such an idea since we had very few instruments that could create inharmonic sounds that could be precisely controlled. Every piece of music has to be adapted for its instruments, while at the same time, it is the sound of the instrument which defines what kind of music should be played on it. Having a computer in our hands which could create any sound with any type of relationship, and being able to control them with any precision we pleased, implied a reconsideration of the relationship of the content as form, and the form as the relationships in the content.

6.3 Future Work

As explained before, Stockhausen used the unity of form and material extensively. *Kontakte (1959-60)* was one the first of his purely electronic pieces and it uses the unity in different perceptual layers as a principal theme. (*Kontakte* is composed of an electronic part and an instrumental part; however, the electronic part can be listened to by itself as a complete piece.) One is very surprised to find out with what kind of primitive instruments, compared to today's digital computers, these pieces were created. Stockhausen writes[47, page 131]:

> In some sections of KONTAKTE I had to splice it all by hand, which is an unbelievable labour. Imagine, I worked on the last section of KONTAKTE, beginning around 23' 00" or 24' 00", together with Gottfried Michael Koenig in Studio 11 on the third floor of Cologne Radio, for *three months.* And when it was completely ready, I spliced it together with the previous sections, listened, turned pale, left the studio and was totally depressed for a whole day. And I came back next morning and announced to Koenig that we had to do it *all over again.* I mean, he almost fainted.

Compare their instruments with the speed of today's central processing units or the versatility of modern operating systems. On the contrary to the belief that computers are not still good enough for music, I believe we have to concentrate on creating software bases for computers suited for music which keep up with the fast pace of changing hardware, rather than building special purpose hardware[1]. We can compare the works of the early serialist composer to the work of computer scientists who coded assemblers by entering the bits of the binary object codes by keys on the front panel of the old computers. These computers, which would fill up a room 30 years ago, today can be installed in the door of our microwave ovens.

In the synthesis method described in this thesis, we believe that we have captured the serial ideas of Stockhausen, perhaps unconsciously, since we were not aware of these composition methods when we started this project. We also believe that this thesis shows that serialism and self-similarity are intertwined and that they are natural and necessary for the future development of computer music. The use of self-similarity has provided a system which can create very complex results by using very simple structures. It also provides us with many tools to not only assure the uniformity of form and content but also to use the unity of the perceptual layers as a musical tool. Self-similarity and chaos are among the most fascinating findings of our century, and there is still a great deal for us to learn about them.

6.4 Perceptual Issues

I believe that the work of every composer of the 20th century who succeeded in being a profound thinker can also be a base for software abstraction. Through trial and error we will find the

[1]Miller Puckette's MAX is an excellent example of such efforts[33].

natural paths, or more correctly, the paths will find their own natural flow. What Schoenberg started is not an easy path and has very strong implications in our lives; many may not agree. Electronic and instrumental music of the 20th century in general is not easily accessible, and for every piece that survives, hundred of others will die. As we mentioned before, serialism has been attacked for being difficult to be understood.

Tonal form is a very strong form, and I am yet to find a human who has really listened to Bach and has not been affected by the sounds, even without having any knowledge of the intellectual energy that has been put into the music. In fact, as the ideas of John Cage imply, to hear music all we need to do is to listen. The idea of serialism is not to go against nature; rather, the idea is respect musical relationships and, thus, provide the grounds for music to evolve. The tonality of atonality, which is the communication of originality, has to be understood. Serialism is an issue regarding communication and our relationship with nature and the people around us. Paul Griffiths writes[16]:

> That electronic music is, as I have already suggested several times, a mirror music, a music which offers new perspectives in the world of the mind, new perspectives in our understanding of music and of ourselves. One may ask why perspectives are being discovered so slowly, why the outstanding works of electronic music are so few. But one may consider the history of the piano, which was invented around 1700, but which had to wait three-quarters of a century before composers found and used its characteristic properties. Perhaps the wait in the race of electronic music will be shorter.

So let us briefly reflect on ourselves (i.e., be self-referential[2]), concerning the issues discussed in this thesis.

Gödel proved that we have no way of reaching the whole truth by any formal means, no matter how rigorously we have defined our system. No laws in physics are accepted unless they are proven by experiments. However, laws have to be theorized at first; which theory can theorize theorizing? The history of physics has shown repeatedly that anytime we have found a theory which became a law, another theory has superseded it. Then, what is a law? Do we know of any law that has actually not been broken? If we could find a single law that assures us

[2]This sentence is not just a figure of speech, self-referentiality is as simple and profound as this situation.

of being a law, it implies that we would have a complete understanding of the future indicating that the law would not be broken. At that moment all of our freedom, identity, and "existence" is taken away from us, since we become defined as a deterministic process. Such questions are no longer in the realm of science or philosophy. Rather they are concerned with social and political situations which we have to deal with not only on the individual level, but also on a global scale. Once we talk of human relations, it is naive to assume that logic alone could go very far. In this case truth becomes a matter of probability rather than what is usually known as "hard truth".

The awareness of the physical similarities of consonances and dissonances brings a sense of "justice" to musical form where one pitch is not more important than any other. Consonant chords are a minority in comparison to the countless number of dissonant chords. The continuum between the consonances and dissonances is the same continuum which exists between our physical and psychological constructs. Both of them are the manifestations of the evolution of relationships perceived by our senses. This means that our psychological constructs are simply the state of matter from which we are formed.

The consonant chords are based on integer power relationships, while real number relationships create dissonant chords. There are more real numbers between 0 and 1 than there are integer numbers. Cantor spent a good part of his life trying to find out how many real numbers exist between 0 and 1. We find all these continua (namely consonance/dissonance, sound/music, physical/psychological, channel/information, Cantor's 0/1, Gödels work which we interpret as the continuum of truth and falsities) to be similar in the sense that they all connect symbolic entities of meaning to the physical world. All these ideas tell us that we, and whatever we do, is part of the nature. In this view, communication is not a symbolic act as the idea of exchanging information may imply, but rather an interaction of matter in the physical world. On the contrary to general belief, it is neither surprising nor magical that we find the most abstract constructs of mathematics in nature (e.g. finding of Cantor set in Chaos — refer to chapter 3)[3]; we are part of nature, and what results from our mind (be it music, mathematics or idle thought) is also part of nature. It is magic that we are able to communicate at all with

[3]It is also no surprise that Gödel's incompleteness theorem connects itself with the computer science halting problem and the non-computability of Kolmogorov complexity[6, page 162]. Kolmogorov complexity is the extension, or actually a superset, of the classical information theory.

each other, and perhaps the reason that we can is that we are physically the same as that with which we communicate.

The uniformity of time and perception, the idea of a composition being a unit in and of itself, the idea of the existence of music as a conceptual entity, and much romantic spiritual thought about the unity of mind all suggest the existence of self-similar structures in our musical communication. Such an issue takes on a different color in electronic and computer music. In instrumental music, no matter how far we push the use of non-conventional instruments, there are still physical limitations and constraints, and the composition takes on its form around those constraints. Computers can implement the specifications of sound for a composition to the smallest detail, in any physical relationships that are precisely defined. The constraint of computers lie in a different domain. It seems to us that they lie in the domain of communication, where we need to understand what ambiguity means when possible, and in fact, have to specifically define that ambiguity. Whether this path is good or bad we do not know; it is a path to be tried. The path seems natural and consistent with some of the body of thought in philosophy, mathematics, and discoveries in our physical world. After all it is a path rich in poetic possibilities.

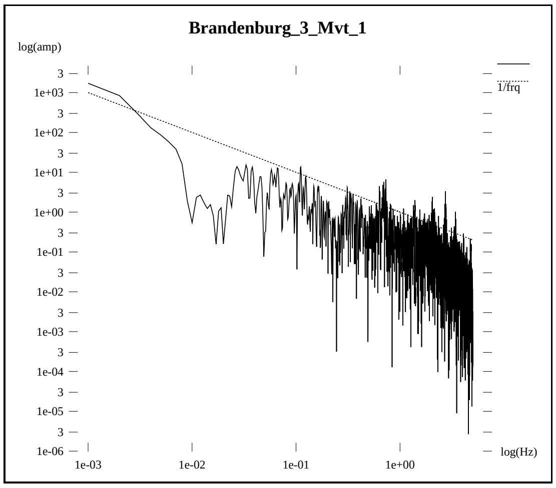
Appendix A

The Slope of Correlated Music

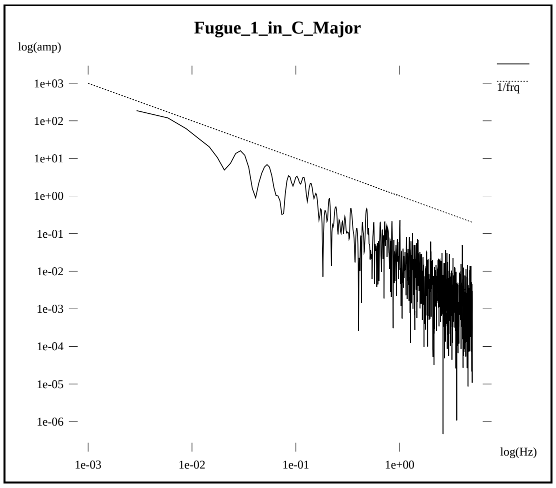
The result of our analysis of 57 different pieces are presented in this appendix. The third movement of the Brandenburg concerto and all of preludes and fugues form the Well-Tempered Clavier Part-I, along with a few other pieces, were analyzed. This analysis was an attempt to recreate the results of Voss and Clarke's[50] in their study of music as $1/f$ noise, which is explained in chapter 4. In our study we extracted a simple signal (which we called the "top voice") from the MIDI encoding of these pieces. Our methods of extraction and analysis are also explained in chapter 4. A line was interpolated from the data using least squares error[1] to find an approximate value for the slope of the power spectrums[32]. The approximated slope for all the power spectrums is shown. The line which corresponds to the $1/f$ (which has a slope of -1) spectrum is also printed on all graphs.

The higher the absolute value of the slope, the more correlated the melodies sound. We make no general claim that music in general works like $1/f$ noise, except that these pieces show a uniform relationship between their small and large-scale structures. These pieces are similar to each other and are all from a specific era of Western music. Unfortunately we did not have access to other pieces in MIDI format.

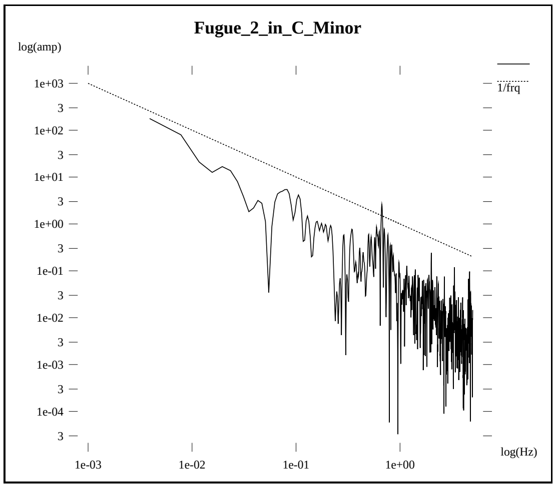
[1] Stan Sclaroff provided the code for the algorithm.



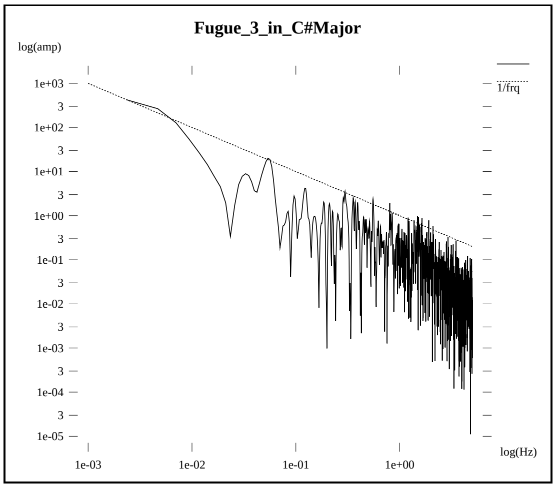
$(slope \approx -1.234)$



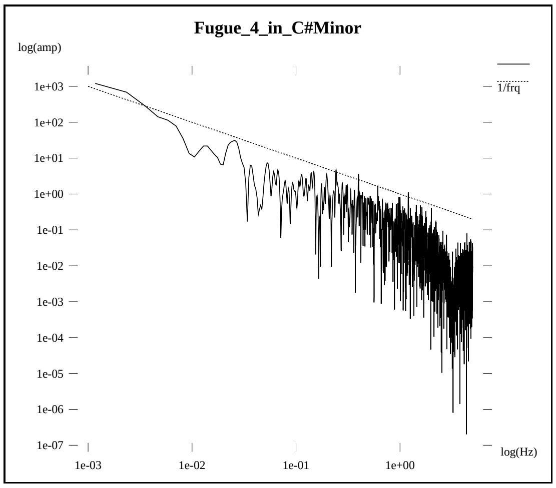
$(slope \approx -1.746)$



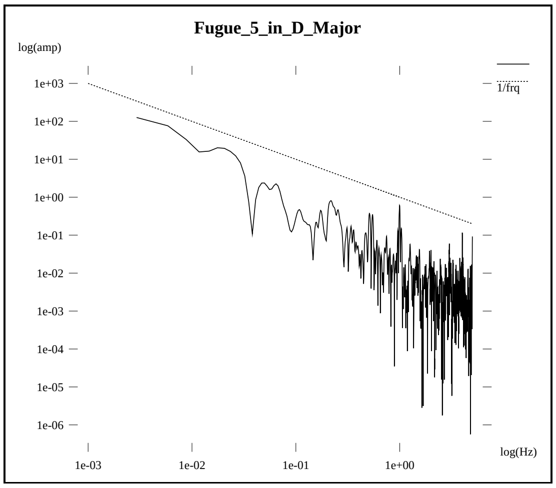
$(slope \approx -1.512)$



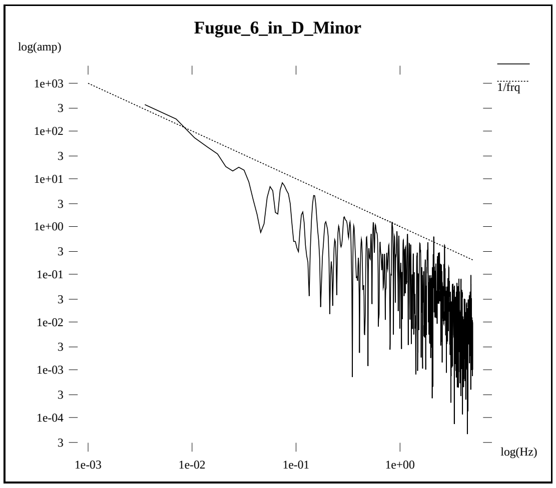
$(slope \approx -1.234)$



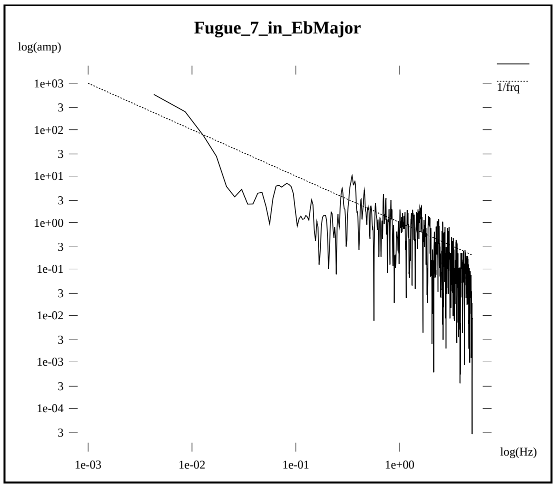
$(slope \approx -1.707)$



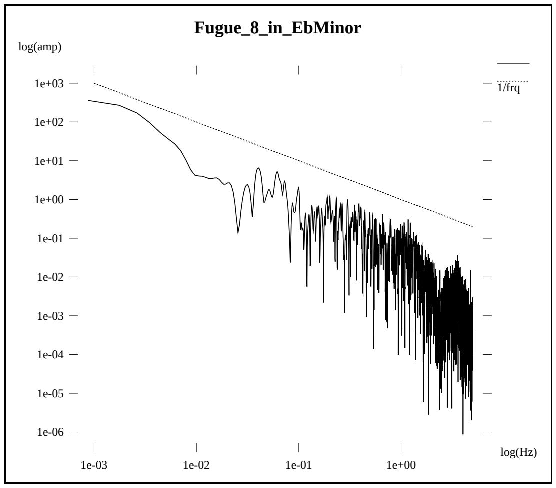
$(slope \approx -1.398)$



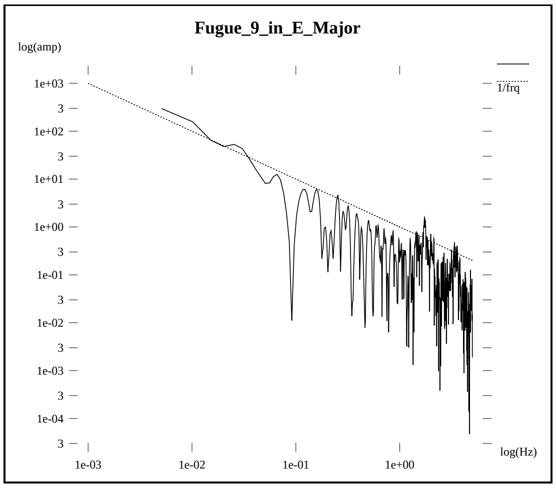
$(slope \approx -1.384)$



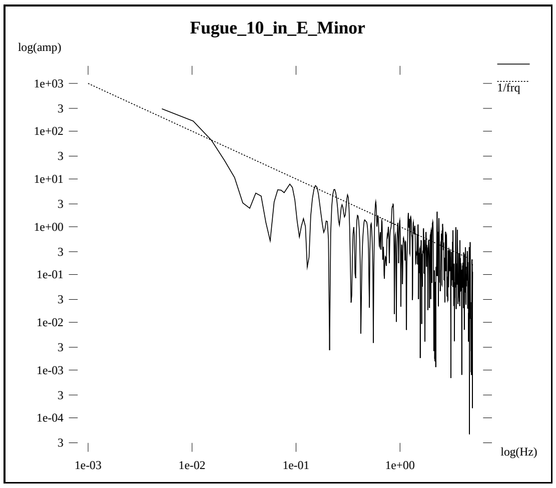
$(slope \approx -1.245)$



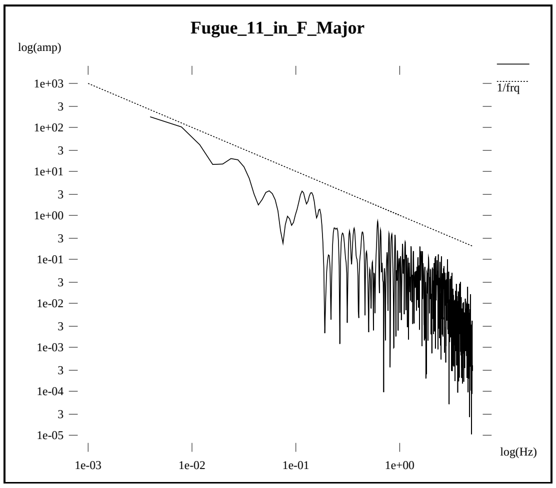
$(slope \approx -1.843)$



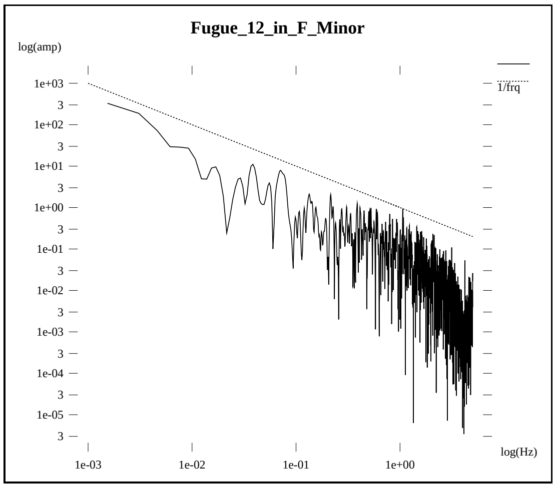
$(slope \approx -1.162)$



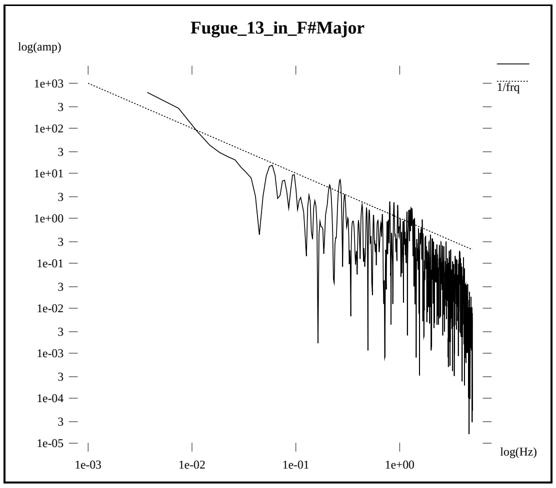
$(slope \approx -1.004)$



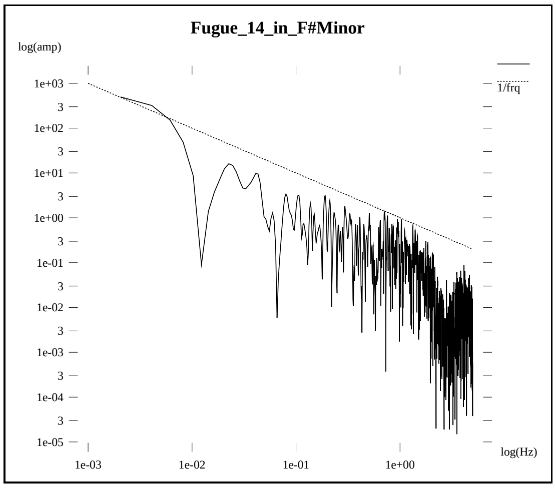
$(slope \approx -1.539)$



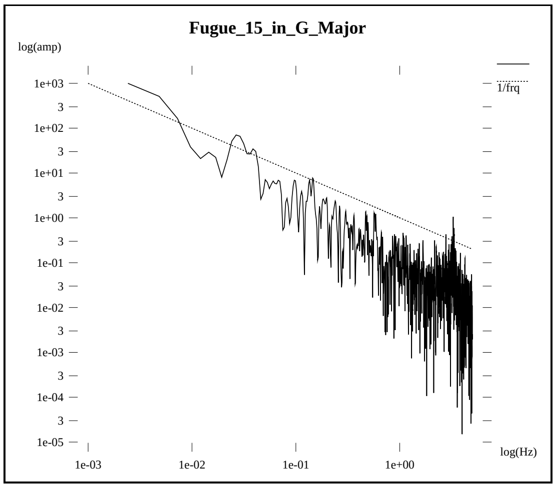
$(slope \approx -1.690)$



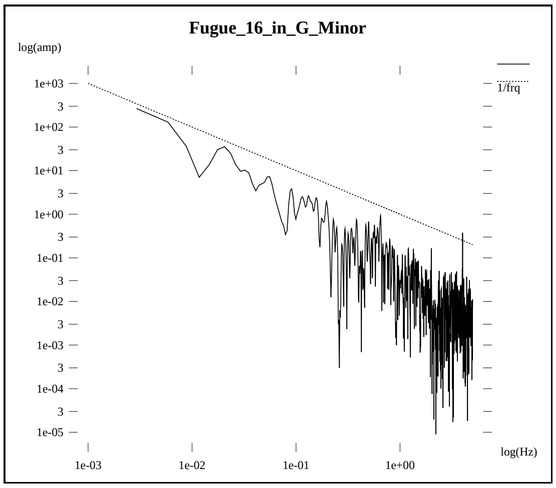
$(slope \approx -1.554)$



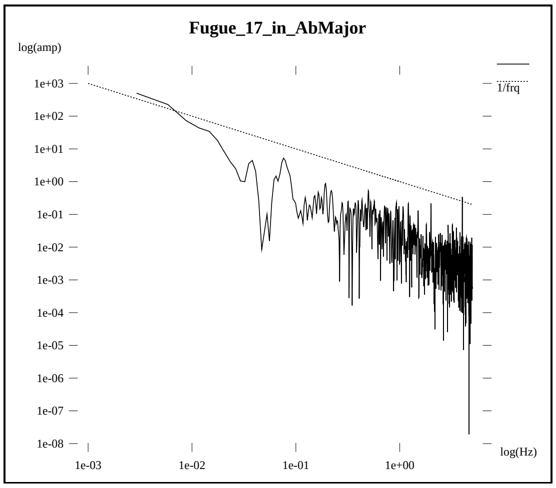
$(slope \approx -1.541)$



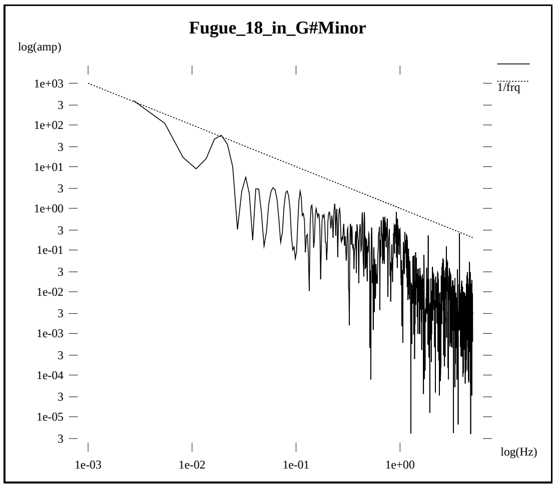
$(slope \approx -1.308)$



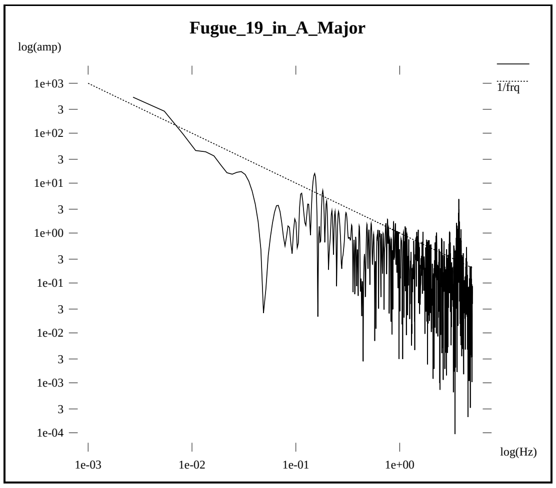
$(slope \approx -1.488)$



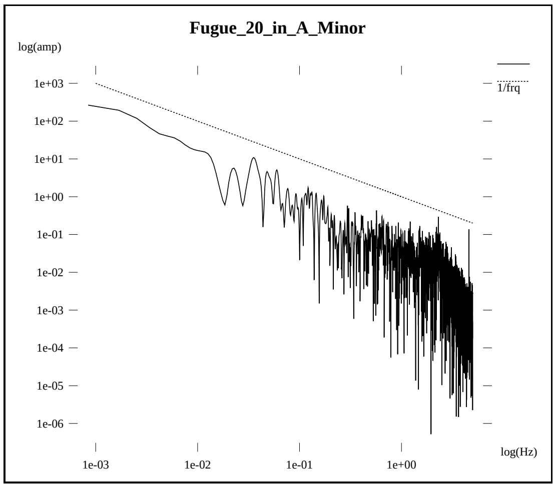
$(slope \approx -1.349)$



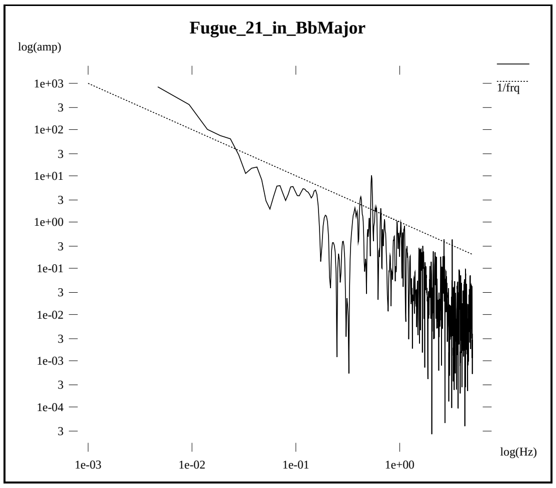
$(slope \approx -1.516)$



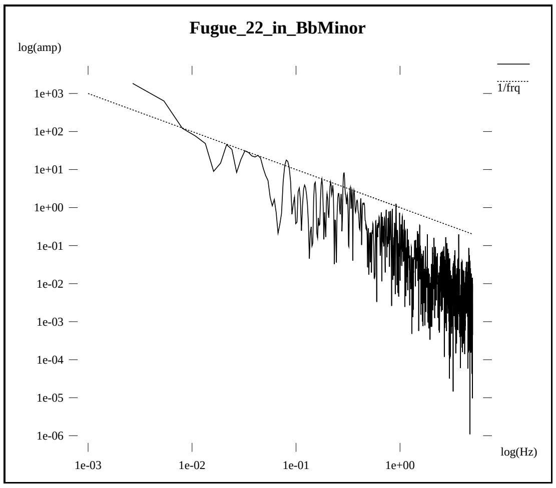
$(slope \approx -0.873)$



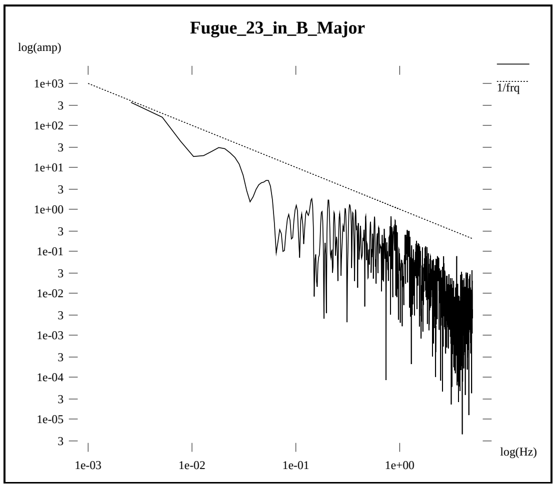
$(slope \approx -1.605)$



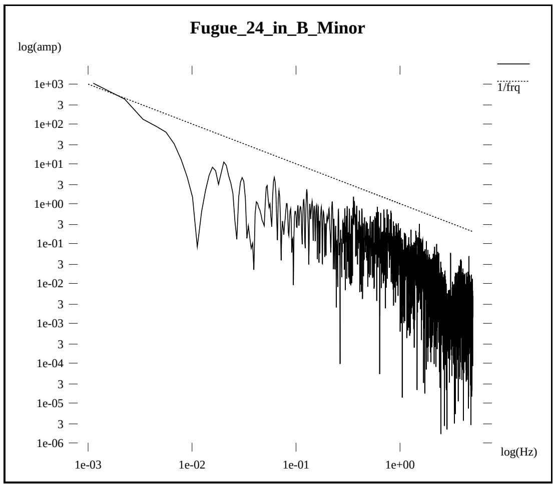
$(slope \approx -1.577)$



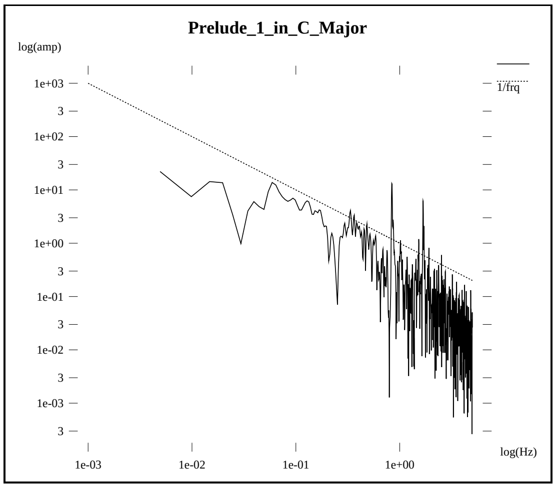
$(slope \approx -1.759)$



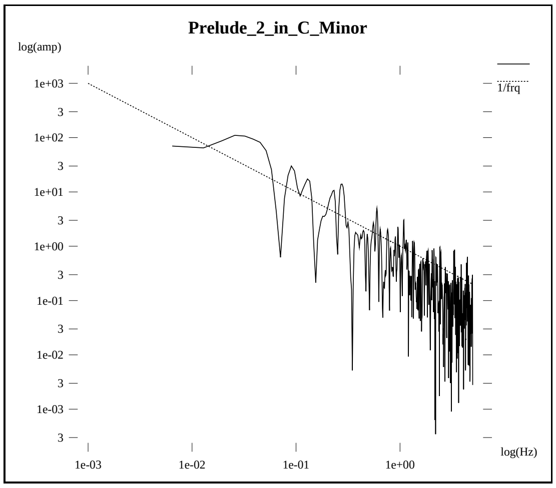
$(slope \approx -1.496)$



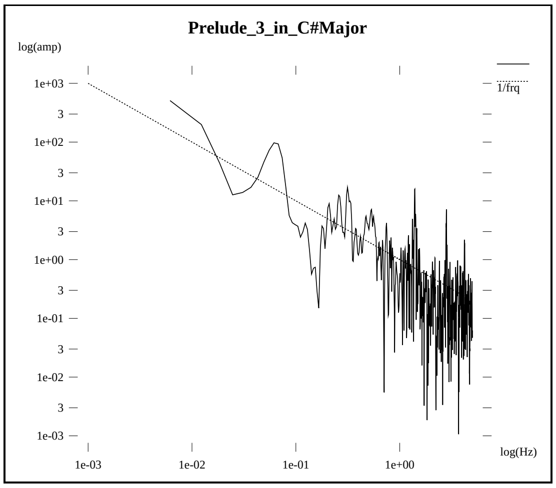
($slope \approx -1.584$)



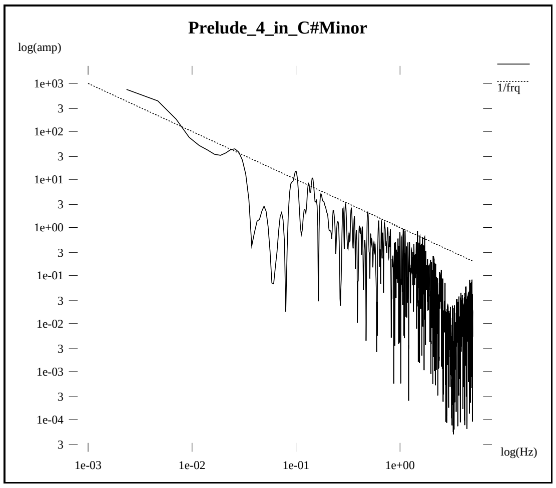
($slope \approx -1.481$)



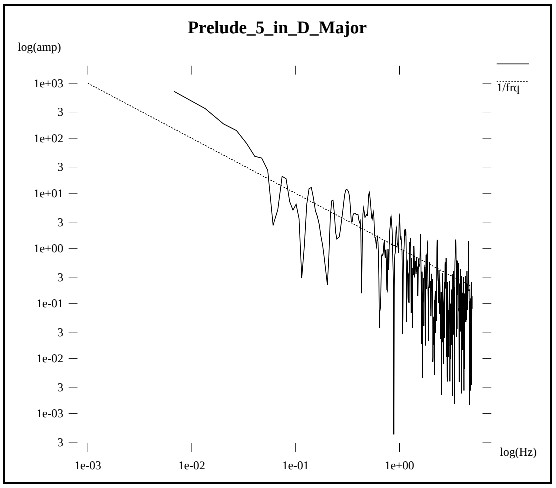
($slope \approx -1.300$)



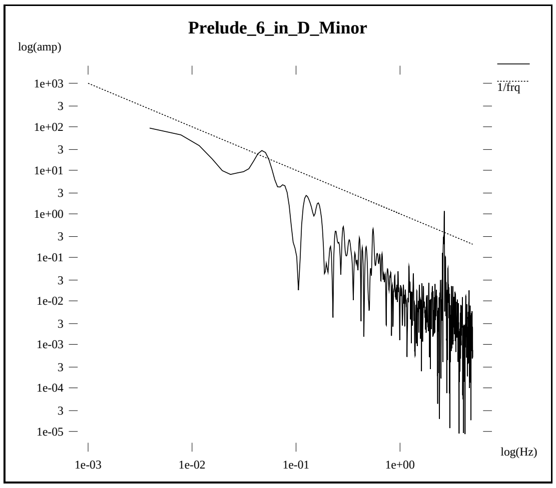
($slope \approx -1.113$)



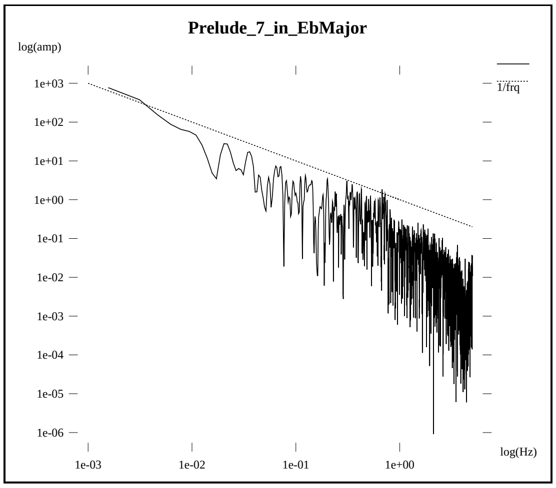
($slope \approx -1.675$)



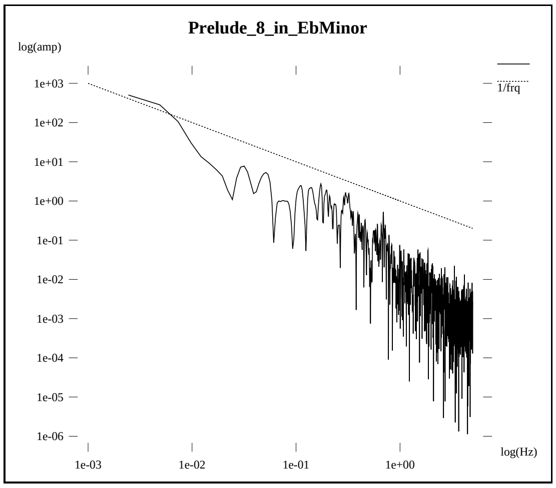
($slope \approx -1.337$)



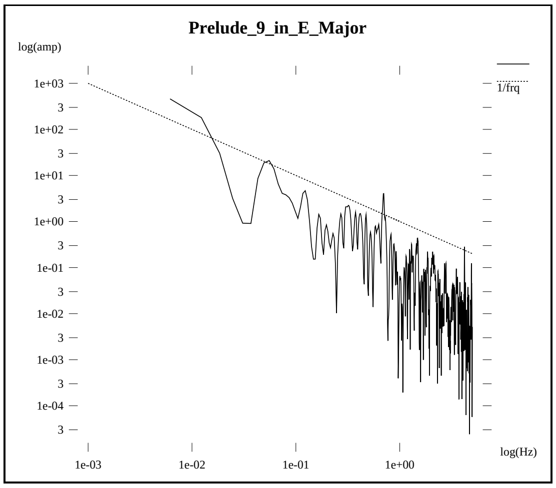
$(slope \approx -1.570)$



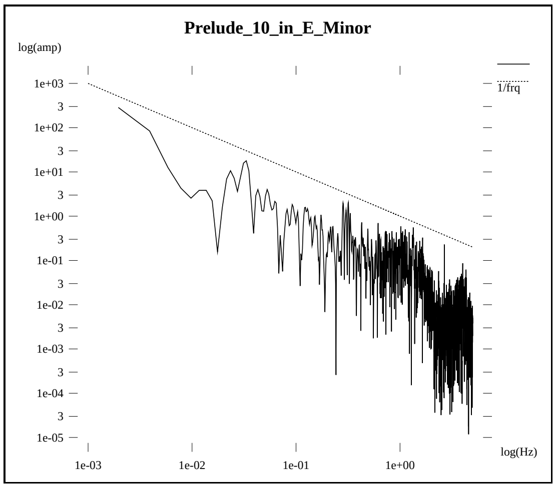
$(slope \approx -1.777)$



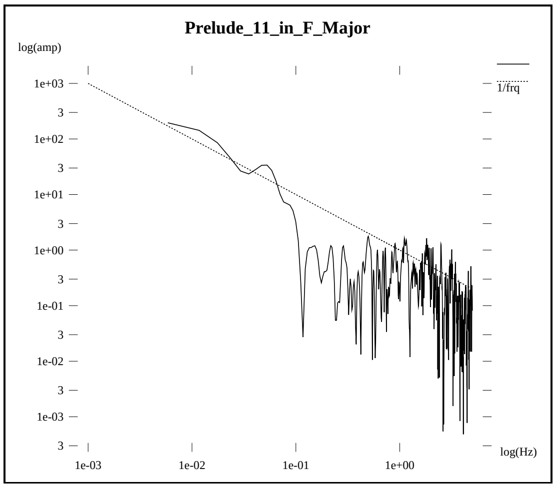
$(slope \approx -1.907)$



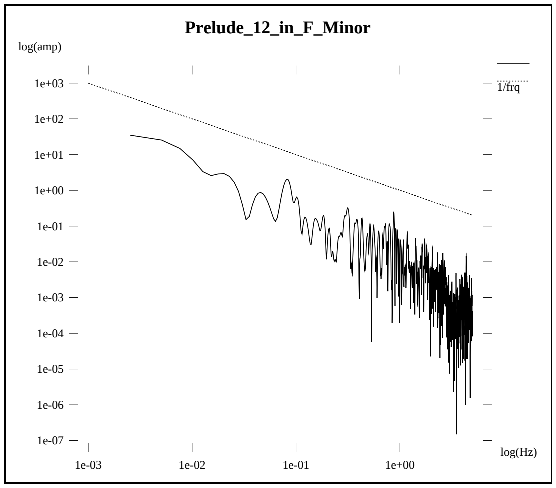
$(slope \approx -1.571)$



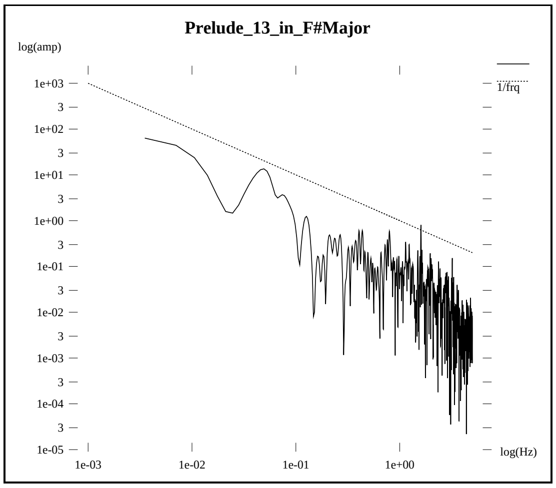
$(slope \approx -1.550)$



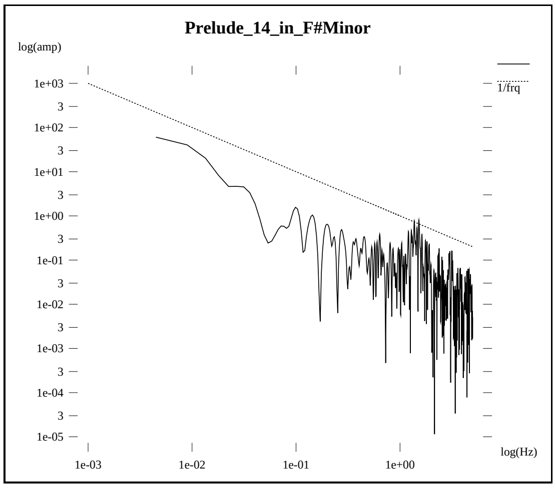
$(slope \approx -0.869)$



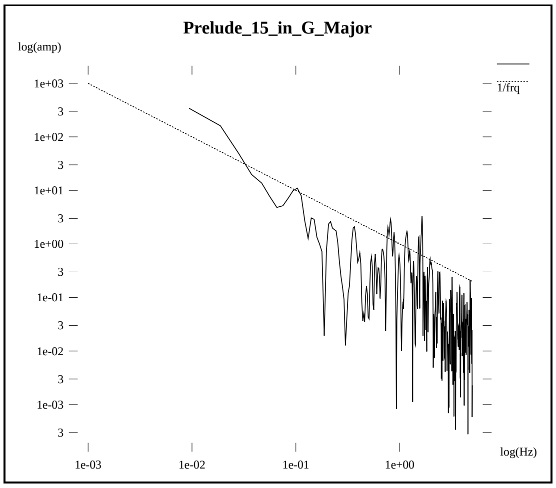
$(slope \approx -1.758)$



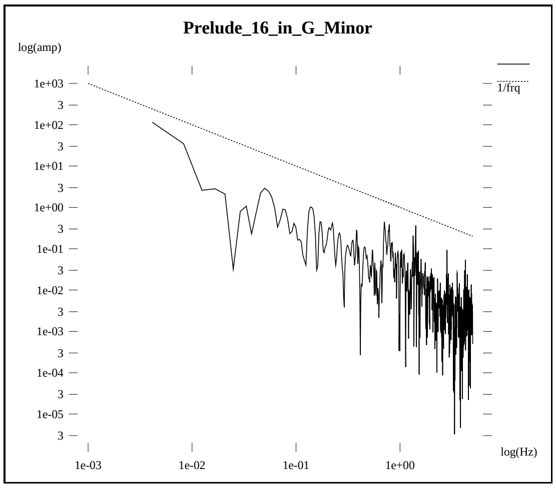
$(slope \approx -1.430)$



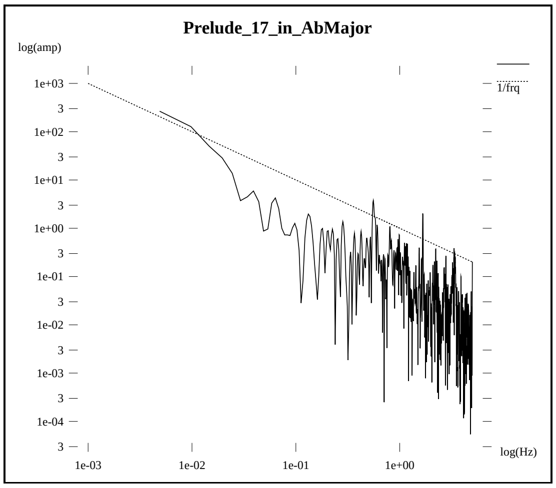
$(slope \approx -1.062)$



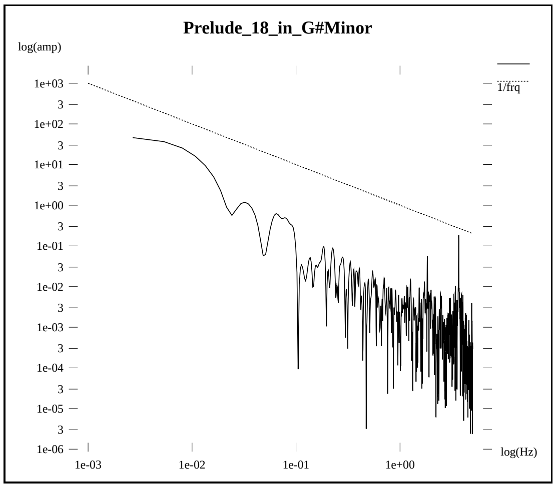
$(slope \approx -1.414)$



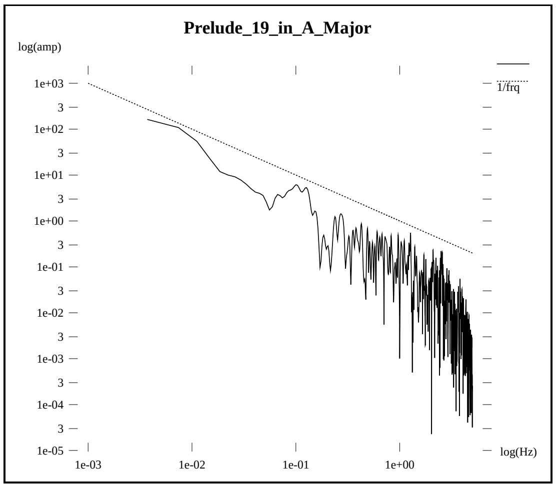
$(slope \approx -1.414)$



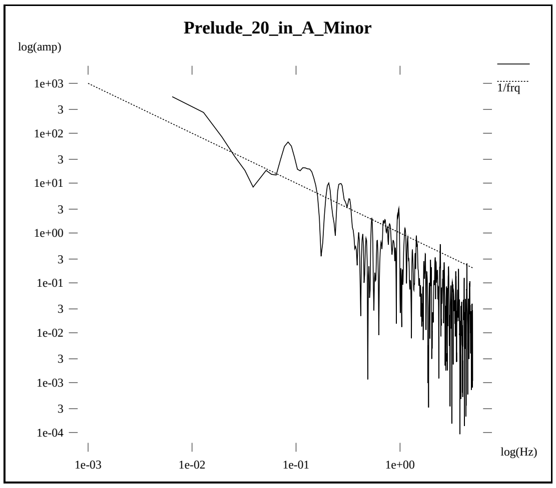
$(slope \approx -1.307)$



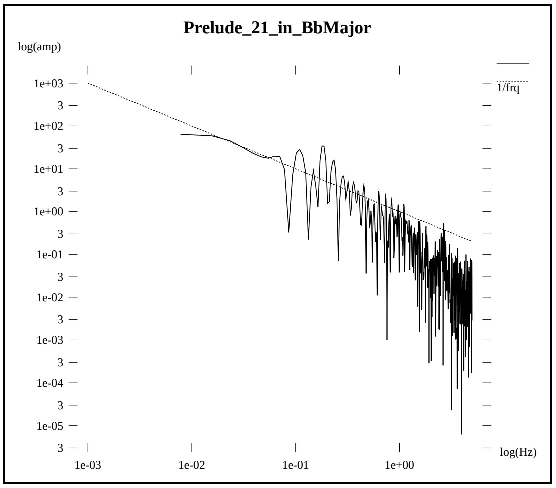
$(slope \approx -1.310)$



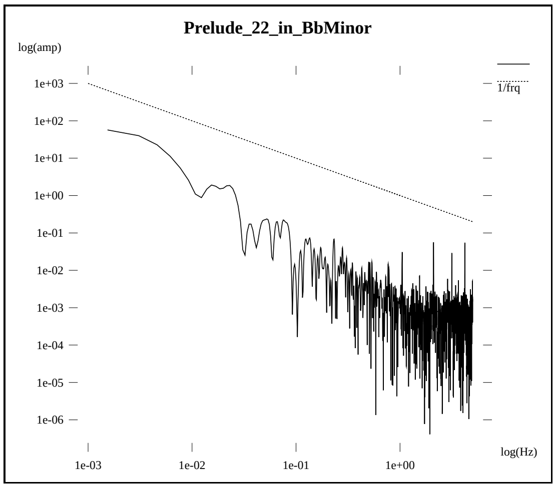
$(slope \approx -1.773)$



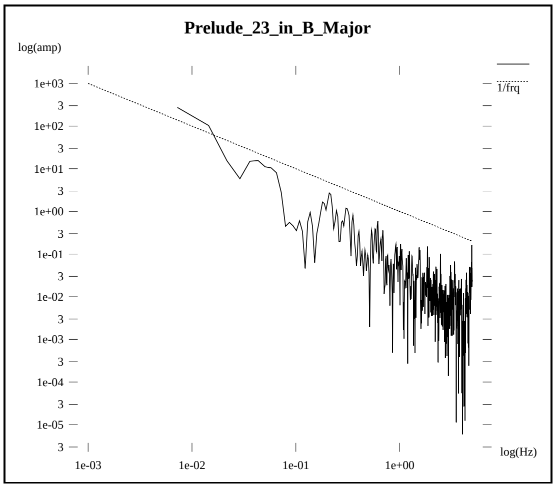
$(slope \approx -1.739)$



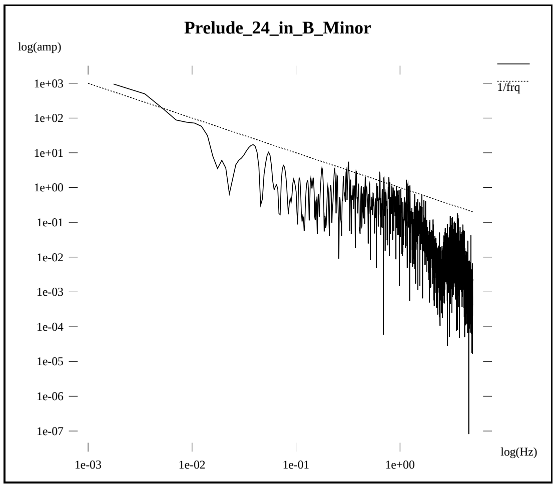
$(slope \approx -1.877)$



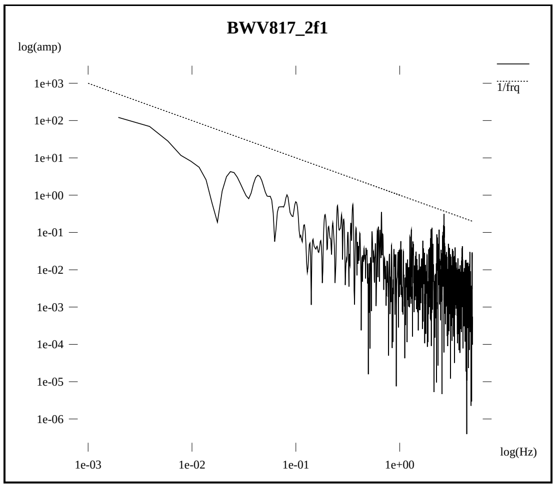
$(slope \approx -1.042)$



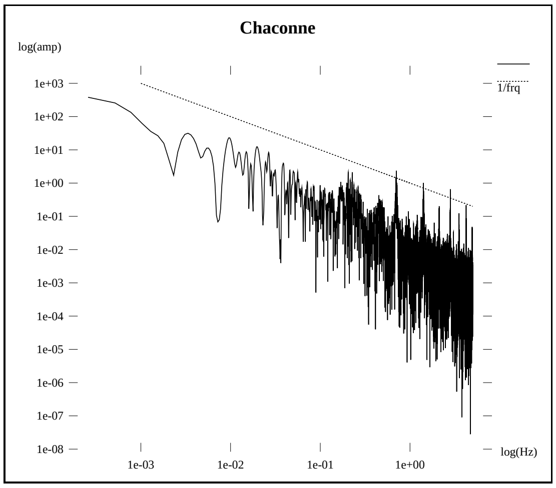
$(slope \approx -1.496)$



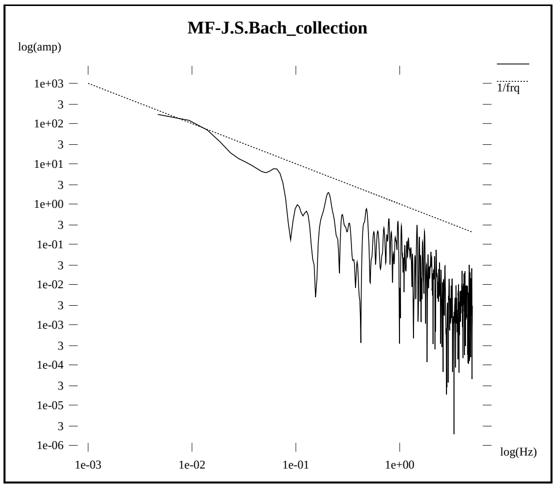
$(slope \approx -1.736)$



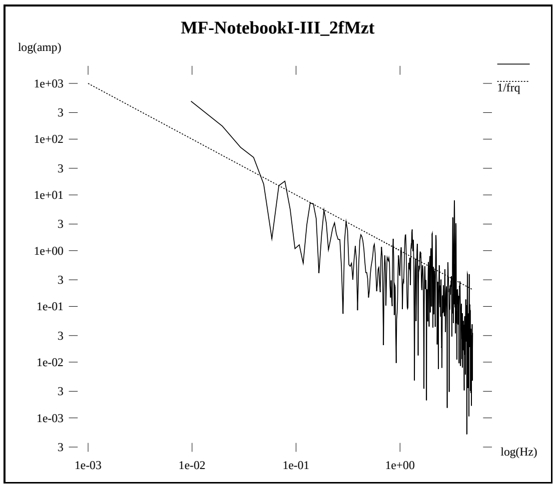
$(slope \approx -1.037)$



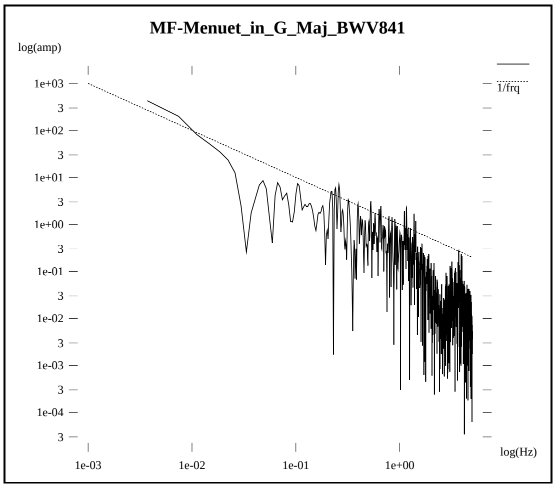
$(slope \approx -1.444)$



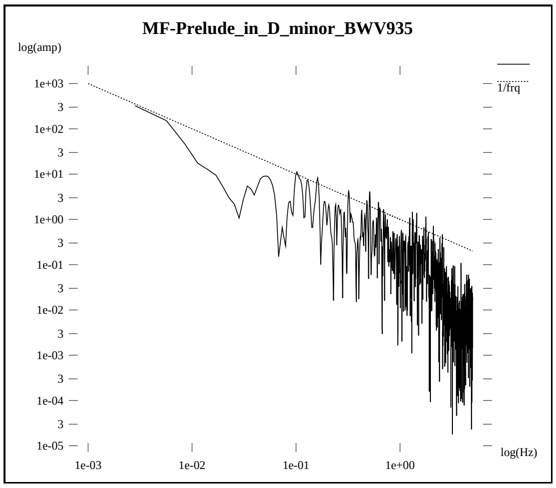
$(slope \approx -1.694)$



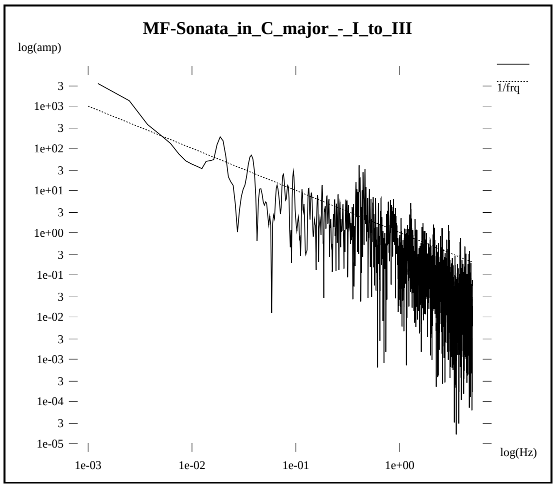
$(slope \approx -1.150)$



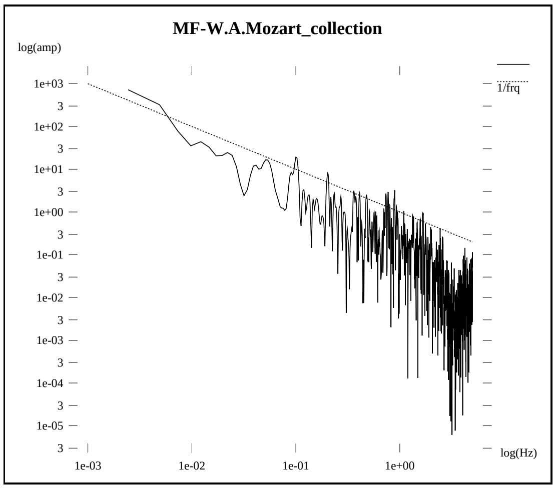
$(slope \approx -1.658)$



$(slope \approx -1.694)$



$(slope \approx -1.569)$



$(slope \approx -1.684)$

Appendix B

Descriptions of the Audio Examples

This appendix provides short desciptions for the accompanying audio examples. The text provided here is similar to the spoken words on tape preceeding each example.

Examples for Chapter 4

Ex. 1 A Shepard tone with 8 partials, starting at 32 Hz, is played first at normal speed and then twice as fast. The partials of this sound are geometricaly related to each other by a factor of 2. The claim is that the percieved pitch of the sound remains the same even though it is being played twice as fast. The example is played twice.

Ex. 2 A Shepard tone similar to the previous example, except with partials which are gemetrically related to each other by a factor of 2.12, is played first at normal speed and then twice as fast. The claim is that, paradoxically, the percieved pitch of the sound is decreased by a half step when the sound is played twice as fast. The example is played twice.

Ex. 3 This example is the resynthesis of the first 30 seconds of the extracted "top voice" from J. S. Bach's 3rd Brandenburg concerto.

Examples for Chapter 5

Ex. 4 This example illustrates the result of self-similar synthesis for a binary segmentation, and frequency factors of 1 and 2 is played. A graph of the frequency fluctuation of this example can be seen in figure 5-1.

- Ex. 5 This example illustrates the effect of a trinary segmentation with frequency factors of 1 and 1.5. A graph of the frequency fluctuation of this example can be seen in figure 5-2.
- Ex. 6 This example illustrates the effect of all-level synthesis for a trinary segmentation. The basic structure of this example is the same as the previous example except that all the levels are synthesized and added together, and lower amplitude factors are used.
- Ex. 7 This example illustrates the effect of all-level synthesis with binary segmentation and frequency factors of 1 and 2.
- Ex. 8 This example illustrates the effect of using a sinusoid window with time segmentation of 1 to 20. The self-similarity of this signal is illustrated in figure 5-3.
- Ex. 9 This example is a 5 second version of the previous example and it illustrates the scalability of the synthesis process.
- Ex. 10 This example illustrates the effect of an unequal trinary segmentaion. The different partials of this example are harmonically related.
- Ex. 11 This example illustrates the effect of magnifying the structures of a short sound by layering. The amplitude "window" for this example is a segment of a spoken word which is played before the example.
- Ex. 12 This example is similar to the previous example except that its amplitude "window" is a long cello melody. The original cello sound is played before the example.
- Ex. 13 This example illustrates the effect of layering many transposed copies of a looped piano note.
- Ex. 14 This example illustrates how the system can be used for creating rhythmical pieces. The amplitude "window" is a segment of a sampled powertom drum sound which is played before the example.
- Ex. 15 This example illustrates how a compound rhythm could be created and how multiple instruments could be used in a piece. This example uses three different segments of sampled sounds as instruments which are played before the example.

Appendix C

Morphosis

Morphosis (1992) is a piece composed by the author using the synthesis system described in this thesis. All the sounds were sculpted either from scratch, or from manipulation of short recorded acoustic sounds. The mixing for this piece was done using Csound[49]. *Morphosis* is a timbre melody (*Klangfarbenmelodie*), but it is not as subtle as Schoenberg had imagined (see page 39). The structure of the piece is based on simple geometrical shapes. The form of a triangle, which stands for the tonal form of "resolution — tension — resolution", is repeated in different scales in fairly symmetrical ways. The piece last about 4'20". The ending sound starts at 4'00". The shape of the piece is an isosceles triangle, with its highest point at 2'00". Four shapes starting from an isosceles interpolated to a right angle triangle form the first 2 minutes. The last two minutes is constructed by a large right angle triangle and a smaller one superimposed on top of it.

The piece starts with a quick build-up of a metallic sound which fades away to a sinusoid at 150 Hz through a noisy timbre. The sinusoid spawns other sinusoids which first drop in frequency and then rise to one of the first 5 harmonics of 700 Hz (700-3500 Hz) while fading away. The complex tones turn into a looped melody of a pitched female voice at 700 Hz. This voice spawns transposed copies of itself, where the transposition is done according to an exponential curve. The voices alternately rise or fall in pitch toward a full step higher or lower than the last spawned melody respectively. The effect is that not only the melody of every individual line is heard, but also a new melody is created by the interaction of transposed copies with each other. The same process is repeated 10 seconds later for a male voice three octaves lower. A percussive sound created from stretching the male voice enters 10 seconds later to create a sense of urgency and the sounds die down slowly.

The same male and female voices were used in a 20 to 1 time segmentation synthesis for the next opening sound, which sounds like a noisy "woosh" from which sounds of birds emerged. The spatialization effect at this moment was created by Csound. The bird songs drop in pitch while slowing down and turning into the male and female voices in the left and right channel. The voices are modulated with sinusoids at frequency of 0.1 Hz exponentially rising to 1000 Hz. The modulating signals are 90 degrees out-of-phase to create a movement in space. Meanwhile a metallic electronic sound, which goes through a few iterations of being timbral and becoming rhythmical and vice versa, is faded in. The sound creates a very clear and urgent need for resolution at 2' 00". The sound is resolved by an explosion from which a texture of falling piano notes emerges. The downward path is interrupted by many slashes of high frequency lines entering quickly one after the other, whose ensemble is pointing downward. A voice reading a sentence slowly emerges while the piano textures slowly turn into individual notes. The words of the voice become clear while creating a rhythmic texture. The muddy low frequency sound of a cello, which has been there since the explosion, can now be heard. The form of this sound is the same as that of the piano texture but with exactly half as many note attacks. The words fade away to single sentences while the background (piano and cello) is amplified. Another metallic sound fades in creating accelerating pulses of sounds, while geometrically-related partials are faded in one by one in the high spectrum. The pulses and the partials meet while fading away, from which the cello and piano timbre emerge and end the piece.

The main theme of the piece is the sense of tension and need for resolution built up over the first 2 minutes, and the connection to the dénouement in the second half created by voices falling in pitch. This piece was composed for and dedicated with much love to Isabella Khan.

Bibliography

[1] Jean-Baptiste Barrière. "Chréode I": the pathway to the new music with the computer. In Tod Machover, editor, *Contemporary Music Review*, volume 1, pages 181–201. Harwood Academic, New York, 1984.

[2] John Cage. *Silence*. Wesleyan University Press, Middletown, Connecticut, 1961.

[3] Herschell B. Chipp. *Theories of Modern Art*. University of California Press, Berkeley, California, 1968.

[4] Noam Chomsky. *Language and Problems of Knowledge*. MIT Press, Cambridge, Massachusetts, 1988.

[5] J. M. Chowning. The synthesis of comlplex audio spectra by means of frequency modulation. *Journal of Audio Engineering Society*, 21:526–534, 1973.

[6] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

[7] Joseph Dauben. *Georg Cantor*. Harvard University Press, Cambridge, Massachusetts, 1979.

[8] Richard Dawkin. *The Selfish Gene*. Oxford University Press, New York, 1976.

[9] Charles Dodge. A musical fractal. *Computer Music Journal*, 12(3), 1988.

[10] W. Jay Dowling and Dane L. Harwood. *Music Cognition*. Academic Press, Orlando, 1986.

[11] Edward FitzGerald. *The Rubáiyát of Omar Khayyám*. Dover Publications, New York, 1989.

[12] Patrick Flandrin. On the spectrum of fractional brownian motions. *IEEE Transactions on Information Theory*, 35(1):197–199, January 1989.

[13] Kahlil Gibran. *Spiritual Sayings of Kahlil Gibran.* Citadel Press, New York, 1962.

[14] James Gleick. *Chaos.* Penguin Books, New York, 1988.

[15] Kurt Gödel. *On Formally Undecideable Propositions Of Principa Mathematica and Related Systems.* Oliver and Boyd, Edinburgh and London, 1962.

[16] Paul Griffiths. Three works by Johathan Harvey. In Tod Machover, editor, *Contemporary Music Review*, volume 1, pages 87–109. Harwood Academic, New York, 1984.

[17] Jonathan Harvey, Denis Lorrain, Jean-Baptiste Barrière, and Stanley Haynes. Notes on the realization of bhakti. In Tod Machover, editor, *Contemporary Music Review*, volume 1, pages 111–129. Harwood Academic, New York, 1984.

[18] Douglas R. Hofstadter. *Gödel, Escher, Bach.* Basic Books, New York, 1979.

[19] Franz Kafka. *Parables and Paradoxes.* Schocken Books, New York, 1961.

[20] Marvin Keshner. $1/f$ noise. *Proceedings of the IEEE*, 70(3), March 1982.

[21] Soren Kierkegaard. *"Fear and Trembling" and "Repetition".* Princeton University Press, Princeton, New Jersey, 1983.

[22] L. Koblyakov. Notes préliminaires au sujet de la musique nouvelle. *Dissonanz*, 7, 1986.

[23] Lao-Tzu. *Te-Tao Ching.* Ballantine Books, New York, 1989. Translated by Robert Henricks.

[24] F. Lerdhal. Cognitive constraint of compositional systems. In John Sloboda, editor, *Generative Processes in Music: The psychology of Performance, Improvization, and composition*, pages 231–259. Clarendon Press, 1988.

[25] A. Lindenmayer. Mathematical models for cellular interaction in development, I-II. *Journal of Theoretical Biology*, 18:280–315, 1968.

[26] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, March 1963.

[27] Tod Machover. Computer music with and without instruments. In Tod Machover, editor, *Contemporary Music Review*, volume 1, pages 203–230. Harwood Academic, New York, 1984.

[28] Benoit B. Mandelbrot. Some noises with $1/f$ spectrum, a bridge between direct current and white noise. *IEEE Transaction, Information Theory*, 13(2):289–298, 1967.

[29] Benoit B. Mandelbrot. *The Fractal Geometry of Nature.* Freeman, San Fransisco, 1983.

[30] E. Nagel and J. R. Bewman. *Gödel's Proof.* New York University Press, New York, 1958.

[31] C. Vidal P. Bergé, Y. Pomeau. *Order Within Chaos.* Wiley-Interscience, New York, 1984.

[32] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988. sections 2.9 and 14.3.

[33] Miller Puckette. The patcher. *Proceedings of 1988 International Computer Music Conference*, 1988.

[34] Howard L. Resnikoff. *The Illusion of Reality.* Springer-Verlag, New York, 1989.

[35] Jean-Claude Risset. Mutations: a 33 rpm L.P. record including "trois moments newtoniens.". Collection INA-GRM (distributed by Harmonia Mundi), 1979. Cf. also compact disc INA C 1003.

[36] Jean-Claude Risset. Paradoxical sounds. In Max V. Mathews and John R. Pierce, editors, *Current Directions in Computer Music Research.* MIT Press, Cambridge, Massachusetts, 1989.

[37] Xavier Rodet, Yves Potard, and Jean-Baptiste Barrière. The CHANT project: From the synthesis of the singing voice to synthesis in general. *Computer Music Journal*, 8(3):15–31, 1984.

[38] Charles Rosen. *The Classical Style.* W. W. Norton, New York, 1972.

[39] Kaija Saariaho. Timbre and harmony: interpolations of timbral structures. In Stephan McAdams, editor, *Contemporary Music Review*, volume 2. Harwood Academic, New York, 1987.

[40] Arnold Schoenberg. *Style and Idea*. University of California Press, Berkeley, Los Angeles, 1975.

[41] Arnold Schoenberg. *Theory of Harmony*. University of California Press, Berkeley, Los Angeles, 1978.

[42] Erwin Schrödinger. *What is Life? & Mind and Matter*. Cambridge University Press, New York, 1967.

[43] Manfered R. Schroeder. *Fractals, Chaos, Power Laws*. W.H. Freeman, New York, 1991.

[44] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Illinois, 1949.

[45] R. N. Shepard. Circularity in judgments of relative pitch. *Journal of Acoustical Society of America*, 36(12):2346–2353, 1964.

[46] Karlheinz Stockhausen. *Die Reihe 3, ...how time passes.....*, volume 3. Theodore Presser Co., Bryn Mawr, Pennsylvania, english version edition, 1959.

[47] Karlheinz Stockhausen. *Stockhausen on Music*. Marion Boyars, London, New York, 1991.

[48] Ferrel G. Stremler. *Introduction to Communication Systems*. Addison-Wesley Pub. Co., Reading, Massachusetts, 1982.

[49] Barry L. Vercoe. *Csound: A manual for the audio processing system and supporting programs*. MIT Media Lab, Music and Cognition, Cambridge, Massachusetts, 1990.

[50] R. F. Voss and J. Clarke. "$1/f$ noise" in music: Music from $1/f$ noise. *Journal of Acoustical Society of America*, 63(1), 1978.

[51] G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. on Signal Processing*, March 1992.

[52] Shahrokh David Yadegari. Using self-similarity for sound/music synthesis. *Proceedings of 1991 International Computer Music Conference*, 1991.