INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Statistics and Music: Fitting a Local Harmonic Model to Musical Sound Signals

by

Rafael Angel Irizarry

B.S. (University of Puerto Rico at Río Piedras) 1993
M.A. (University of California at Berkeley) 1994

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor David R. Brillinger, Chair
Professor Kjell Doksum
Professor David Wessel

Spring 1998

UMI Number: 9902106

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Statistics and Music: Fitting a Local Harmonic Model to Musical Sound Signals

by

Rafael A. Irizarry

The dissertation of Rafael A. Irizarry is approved:

David R. Brillinger May 12, 1998

Chair Date

David Wessel May 13, 1998

Date

[illegible] May 19, 1998

Date

University of California, Berkeley

1998

Abstract

Statistics and Music: Fitting a Local Harmonic Model to Musical Sound Signals

by

Rafael A. Irizarry
Doctor of Philosophy in Statistics
University of California, Berkeley
Professor David R. Brillinger, Chair

Statistical modeling and analysis have been applied to different music related fields. One of them is sound synthesis and analysis. Sound can be represented as a real-valued function of time. This function can be sampled at a small enough rate so that the resulting discrete version is almost as good as the continuous one. This permits one to study musical sounds as a discrete time series, an entity for which many statistical techniques are available. Physical modeling suggests that many musical instruments' sounds are characterized by a harmonic and an additive noise signal. The noise is not something to get rid of rather it's an important part of the signal. In this research the interest is in separating these two elements of the sound. To do so a local harmonic model that tracks changes in pitch and of the amplitude of the harmonics is fit. Deterministic changes in the signal, such as pitch change, suggest that different temporal window sizes should be considered. Various ways to choose appropriate window sizes are studied. Amongst other things our analysis provides estimates of the harmonic signal and of the noise signal. Different musical composition applications may be based on the estimates.

David R. Brillinger

Professor David R. Brillinger
Dissertation Committee Chair

To Alex.

Contents

List of Figures

List of Tables

Acknowledgements

First I want to thank the people that were directly involved in helping with my thesis. I would especially like to thank David Brillinger who guided me through every single step of the way. Thanks to Adrian Freed for teaching me almost everything I know regarding sound and sound analysis. Also, thanks to David Wessel who first got me interested in musical applications. Finally, a special thanks to the people that helped me with all the computer and hardware problems: Ofer Licht, Matt Wright, Phil Spector and Hector Perez.

I also want to thank all the other people that helped me get to this stage. First, to my parents who helped with my education from an early age and to my family for all their support. Ivellise Rubio and Ani Adhikari, who helped me in choosing an undergraduate career in Math. Also to the Professional Development Program's Summer Math Institute for Minority Students and MIT's Minority Research Summer Program for helping me prepare for graduate school in the United States. Thanks to Jorge M. López, Dieter Rietz and Rafael Bras for helping in my transition to graduate school.

I would have never endured my first year of graduate school without the help of Rick Schoenberg and Joaquín Rodríguez. Also thanks to all the department professors, staff and my fellow grad students, who helped me in different aspects of my thesis work, especially to Kjell Doksum, Terry Speed, Sara Wong, Wesley Wang, Vlada Limic, Ben Hansen, and Michael Ostland. Finally, I want to thank Alexandra Nones for helping me with my "inglich" and for all the moral support.

Partial support for this work was provided by the National Science Foundation Graduate Research Fellowship, the University of California's Chancellor's Predoctoral Opportunity Fellowship and, and NSF grants DMS 96-25774, INT-9600251.

Chapter 1

Prelude: Introduction and Summary

Statistics has been applied in various ways to music. For example, various stochastic techniques have been applied in composition (Jones 1981). Stochastic techniques have also been used in forecasting unfinished works (Dirst and Weigend 1992). Voss and Clarke (1975) studied the spectral properties of different musical signals and speculated on the possibility of it being so called 1/f noise, see also Voss and Clarke (1978). In Brillinger and Irizarry (1998) this is studied in more detail, and in particular higher order statistics are examined. In this work the particular application that will be examined in detail is the analysis of sound signals produced by musical instruments. In this field, statistical techniques have been used, for example, to separate the signals into deterministic and stochastic parts and to deconstruct the deterministic part into *harmonic components*.

In these musical applications, as in many others, we need data. In Chapter 2 we will discuss how different musical entities can be represented as data, in particular how a sound wave $y(t)$ can be represented as discrete data.

Every sound we hear is the consequence of pressure fluctuations traveling through the air and hitting our ear drums. The function that describes the audible pressure fluctuations of air is called a "sound wave". In Chapter 2 we will also discuss some of the physical theory and psychoacoustic concepts that motivates the statistical modeling of the signals produced by "harmonic" instruments with what we will call a *local harmonic model*. This model asserts that the signal is a sum of sinusoids with frequencies equal to multiples of a *fundamental frequency.* The sinusoids are called *harmonics*, whereas the sounds related to these periodic components are called *overtones.* The amplitudes and phases of the frequency components together with the fundamental frequency are parameters, determined by the physical properties of the instrument and the person playing the instrument, which will change in time. Physical theory also suggests that the sound waves produced by instruments contain a *non-sinusoidal component* which is an integral part of the sound. We will assume that this non-sinusoidal component is somehow stochastic. The fact that the parameters of this model change with time and that the stochastic part is non-stationary makes the problem of estimating these parameters non-trivial and of interest.

In Chapter 3 we examine some of the previous work done on harmonic models and develop some results on weighted estimates which will be used when we perform local estimation of the parameters.

In Chapter 4 we define the signal plus noise statistical model that will be used in our analysis, (without loss of generality we assume the signal is one time unit long, $0 \le t \le 1$)

$$y(t) = s[t; \beta(t)] + \epsilon(t)$$

$$s[t; \beta(t)] = \sum_{k=1}^{K} \rho_k(t) \cos\left(k\lambda(t)t + \phi_k(t)\right)$$

where

$$\beta(t) = (A_1(t), \ldots, A_K(t), \phi_1(t), \ldots, \phi_K(t), \lambda(t))'$$

K is the number of partials, $\lambda(t)$ is the fundamental frequency (pitch), $\rho_k(t)$ is the amplitude of the k-th partial ($\rho_{k+1}(t)$ is the amplitude of the k-th harmonic) and $\phi_k(t)$ is the phase of the k-th partial. They are all assumed to be functions of time. The process $\epsilon(t)$ represents the non-sinusoidal component or the *noise.* Notice that we can rewrite the signal function

$$s[t; \beta(t)] = \sum_{k=1}^{K} \{A_k(t) \cos(k\lambda(t)t) + B_k \sin(k\lambda(t)t)\}$$

We assume that the signal $s[t; \beta(t)]$ is *locally approximately sinusoidal*, or equivalently that $\beta(t)$ is *locally approximately constant* and that the noise $\epsilon(t)$ is *locally stationary.* In the work to follow precise definitions are given.

For analytic purposes a discrete (sampled) version of the signal $y(t)$ is considered,

$$Y_{n,N} = y\left(\frac{n}{N}\right) \ , n = 0, \ldots, N.$$

Here n is time measured in units of $\Delta t = 1/N$, where N is the number of observations in the unit time interval. This is called the *sampling rate* in the technological literature. Notice that as N gets bigger the signal $y(t)$ is observed on a finer grid.

Fixing N, for any $n_0 \in \{1, \ldots, N\}$ consider a small enough segment, say h_N time units long, of the signal around $y(n_0/N)$ such that one is able to assume that the parameters are *approximately constant* within that segment.

To estimate $\beta(n_0/N)$ assume the parameters are actually constant in the time segment and use a method equivalent to weighted least squares. Namely seek

$$\hat{\beta}_N\left(\frac{n_0}{N}\right) = \min_{\beta} \sum_{n=1}^{N} w\left(\frac{|n - n_0|}{h_N \times N}\right)\left[Y_{n,N} - \mu\left(\frac{t}{T}; \beta\right)\right]^2$$

with w a window function having support in $[0, 1]$.

By repeating this procedure for each $n_0 \in \{0, \ldots, T\}$ we end up with an estimate $\hat{\beta}_N(t)$ of the function $\beta(t)$ for each $t \in \left\{\frac{1}{N}, \ldots, \frac{N}{N}\right\}$. By interpolation, $\beta(t)$ may be estimated for each $t \in [0, 1]$.

Under certain assumptions discussed in Chapter 4, including those already mentioned, it is shown that for any $t \in [0, 1]$ and for an appropriate window size sequence h_N the estimates are consistent and asymptotically normal as N goes to infinity. In current sound analysis research it is common to give estimates of harmonic parameters without an indication of their uncertainty. The asymptotic variance of the estimates provides a way to give standard errors and confidence intervals for our estimates. It is interesting to speculate on the meaning of these quantities in a music context.

Notice how in this estimation procedure, for each n_0, different values for window size h_N can be used. In practice the sample rate is finite and within any window the parameter function $\beta(t)$ is non-constant. If the window size is too small we might not have enough data points to perform meaningful estimation. On the other hand, if the window size is too big, the *approximately constant* assumption might not be appropriate. Many deterministic factors can make the assumption inappropriate, for example, a change in note creates a quick change in $\lambda(t)$. For finite sample rates we can improve our estimates by choice of h_N for every n_0.

In Chapter 5 methods similar to those used in the model selection literature are employed to decide on an optimal window size. One contribution of the thesis is that when taking the weights into consideration, criteria similar to Mallow's C_p, AIC and BIC are found for the case where the weight given to observations varies as well as the number of parameters.

Using the statistical computing package Splus we have created a program that realizes the analysis described above. It has been tested with various "harmonic" instrument sounds, including single notes played by an oboe, tenor saxophone, guitar, violin, pipe organ and shakuhachi flute, with encouraging results. Listening to the residuals of the fit ("residual analysis by ear") we hear sounds similar to what we expect; for example, in the case of the saxophone, we hear air and spit going through a tube. In many cases the residuals contain no audible pitch verifying the fact that we have removed the harmonic part successfully. The window size selection procedure appears to be working well in practice. Smaller window sizes are selected in parts of the sound signal where the pitch is changing. In Chapter 6 we present some of the examples studied. Future work is presented in Chapter 7.

There is an accompanying CD containing audio versions of some of these examples. The first track on the CD is an example of a stochastic composition created by the author. Tracks 2, 3, and 4 are examples of melodies created with an i.i.d. sequence, a random walk, and 1/f noise respectively. The CD is available through the Graduate Assistant of the Department of Statistics, 367 Evans Hall # 3860, Berkeley CA 94720-3860.

Chapter 2

Music and Statistics

2.1 Introduction

What is music? What is a musical sound? Nobody will probably ever find a definitive answer to these questions. It is beyond the scope of this thesis to seek to give precise definitions, but we can say various things about some of the sounds that are generally considered to be musical in nature, such as tones of orchestral instruments and the human voice when singing.

For centuries, understanding sound has been of interest. The Greeks and others must have noticed from the earliest times that plucked strings vibrate. Various Greek philosophers associated fast motion with high pitch and slow motions with lower pitches. In fact, the discovery of the relation between the lengths of strings and musical instruments is commonly attributed to Pythagoras.

Today the study of sound has become a popular research field and, with the advent of electronic music, a practical one too. Contemporary researchers are interested in, for example, the problem of determining what particular characteristics of the sound produced by musical instruments permit humans to distinguish one instrument from another (Grey 1975, Grey 1977, Risset and Wessel 1982, Deutsch 1982, Hartman 1997). Trying to answer this question has led to many new problems and interesting discoveries.

Every musical instrument has capabilities and limitations that help in distinguishing one instrument from the other. For example, a trumpet can play louder tones than a piano, but has a smaller range. But what really allows one to distinguish different instruments is a much more subtle characteristic that musicians call *tone quality*, *tone color* or *timbre*, (Pierce 1992, pages 196–199).

The human ear has been studied extensively and various theories have arisen to explain how it is able to "hear" timbre (Patterson et al. 1992). Sound signals have also been recorded and analyzed, using different approaches, with the goal of understanding what defines timbre. We will call this type of study *sound analysis.*

Recently scientists have also become interested in the creation or reproduction of musical sounds without the use of an acoustical instrument. This is called *sound synthesis.* The first attempt to synthesize musical sound was probably in 1906 with Thaddeus Cahill's Teleharmonium. Powered only by electricity the smoothly rotating tone generators of the Teleharmonium emitted synthetic tones purer than nature (Rhea 1984). More recently, the commercial music industry has become interested in reproducing sounds of acoustic instruments without the use of the actual instrument. Today, for less that US$1000 you can purchase a sound synthesizer that will reproduce sounds of a wide variety of instruments fairly well.

With today's technology we are finally able to process sounds in a data analytic fashion because the time is at hand when music can be treated directly as a quantity to be analyzed by contemporary statistical procedures and packages. Mathews (1963) was one of the first to successfully make use of sound analysis to produce effective sound synthesis. Mathews used the computer to analyze the sound produced by musical instruments with perceivable pitch, and then used the information obtained from the analysis to reproduce the sound. Nowadays we are also interested in using this information to facilitate the creation of new sounds based on the original sound. In our work we wish to analyze sound so as to be able to obtain some parametric representation of it that can later be manipulated to either reproduce the original sound or some version of it. We will call this procedure sound analysis/synthesis.

In this chapter we will discuss some of the procedures that have been used in sound analysis/synthesis. We will also discuss some of the physical and acoustical properties that motivate these methods.

2.2 Music as a time series

In order to speak about statistical analysis of music, we need somehow to represent the musical entities as data.

Every sound we hear is the consequence of pressure fluctuations traveling through the air and hitting our ear drums. The function that describes the audible pressure fluctuations of air is called a "sound wave". The energy transmitted by this "sound wave" can be transformed into a fluctuating voltage $V(t)$, which will be a continuous function in time. We will call the sound wave $V(t)$ produced by a musical sound its *signal representation.* Tape recorders work by storing the voltage function $V(t)$ on magnetic tape and then converting it back to air fluctuations through speakers.

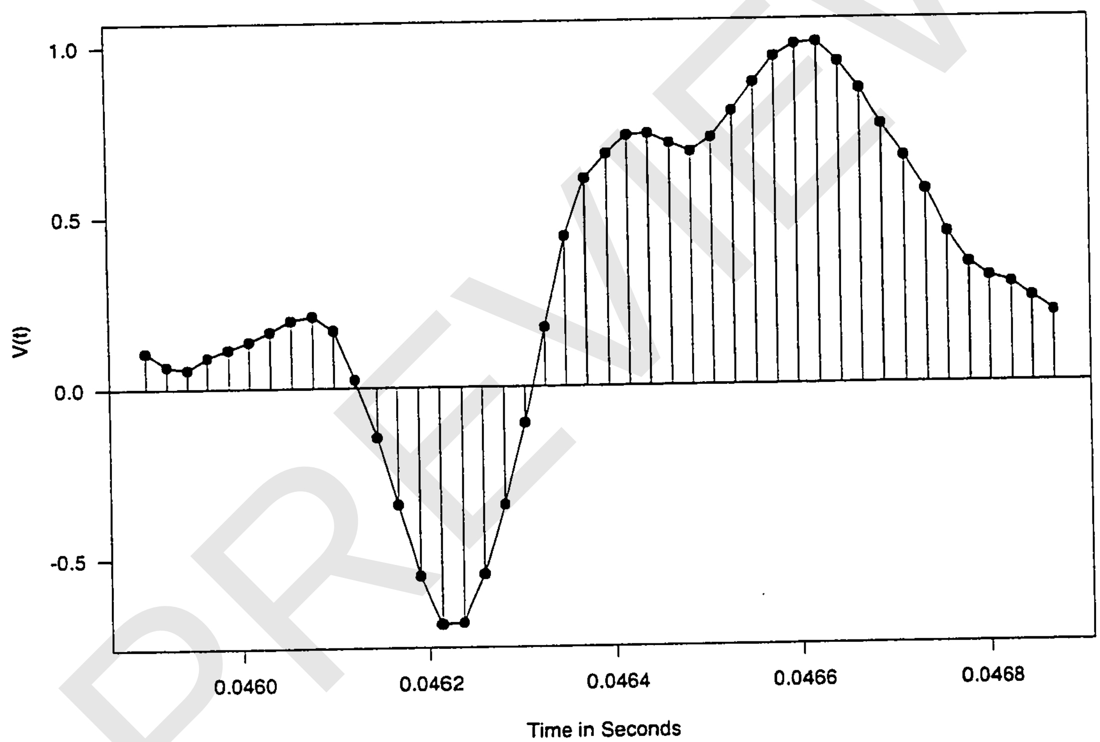

Figure 2.1: Function V(t) for a millisecond of a violin sound sampled at 44100 Hz.

One wants to have discrete data to facilitate statistical analysis. The obvious procedure is to take a discrete approximation of the continuous sound signal. Simply choose a sample rate Δt and consider the discrete time series $Y_n = V(n\Delta t)$ corresponding to the dots in Figure 2.1. This is called *digital sampling* or simply *sampling* by sound engineers. Compact Disc (CD) technology is proof of how well these discrete approximations of sound signals can work (track 5 on accompanying CD). In the case of commercial CDs, the sample rate is a standard at $\Delta t = 1/44100$ seconds i.e. 44100 Hz. The way CDs work is by *sampling* continuous sound signals using an *Analog to Digital Converter* (ADC). The discrete time series obtained from this procedure is then stored on the CD by making small and big indentations on it to represent the data in binary form. The CD player converts this series of numbers back into a continuous function (which is the approximation of the original) using a *Digital to Analog Converter* (DAC) which the speakers then fashion into air fluctuations.

2.3 The physics of musical sounds

Although not all existing sound synthesis and analysis techniques have found it necessary to use models that are in agreement with physical theory, most of them are essentially based on the physical properties of instruments.

The first important physical discovery related to music is that when fluctuations of air are approximately periodic, with period in the audible range, we perceive what musicians have defined as a *pitch* (Pierce 1992, Chapter 2). We will call the frequency related to this periodicity the *fundamental frequency*.

Instruments play different pitches by changing the fundamental frequency of the "sound wave" they are creating. Some cultures, e.g. Western cultures, have quantized these pitches and created *notes*. The pitch corresponding to 440 Hz has been called an *A note (A 440 Hz. concert pitch)*. Any frequency that holds a 2^n:1 relation with concert pitch A is also called an A note, but in another octave. In Figure 2.2 we see 10 milliseconds of the signal produced by a violin playing two C notes, one an octave above the other. Western music uses the 12 tone *equal-tempered scale* in which the frequencies between the notes an octave apart, say 440Hz (concert pitch A) and 880Hz (an octave above concert pitch A), are divided into 12 notes corresponding to frequencies with constant ratio between successive ones. These 12 notes are A, A♯ (A sharp), B, C, C♯, D, D♯, E, F, F♯, G, G♯ and that will bring us back to A (an octave above). If you look at a piano, where the black keys correspond to the sharps, you will see a twelve white-black key pattern repeating 7 times. We will refer to the fourth A from the left (concert pitch A) as A4, to the fourth C (middle C) as C4, the third D as D3, etc.. Adjacent notes are said to be a half-step apart or a semitone away (Pierce 1992, Chapter 4). This means that there is a logarithmic relation between note distance and frequency distance. Given a note with frequency f_1, we can find the frequency f_2 of a note that is k semitones away by solving a simple equation

$$12 \log_2(f_1/f_2) = k \tag{2.1}$$

Notice that k is not necessarily an integer. In fact, musicians call a hundredth of semitone a *cent.* Apparently the trained ear can distinguish two notes if they are 3 cents or more apart. (Pierce 1992, page 72).

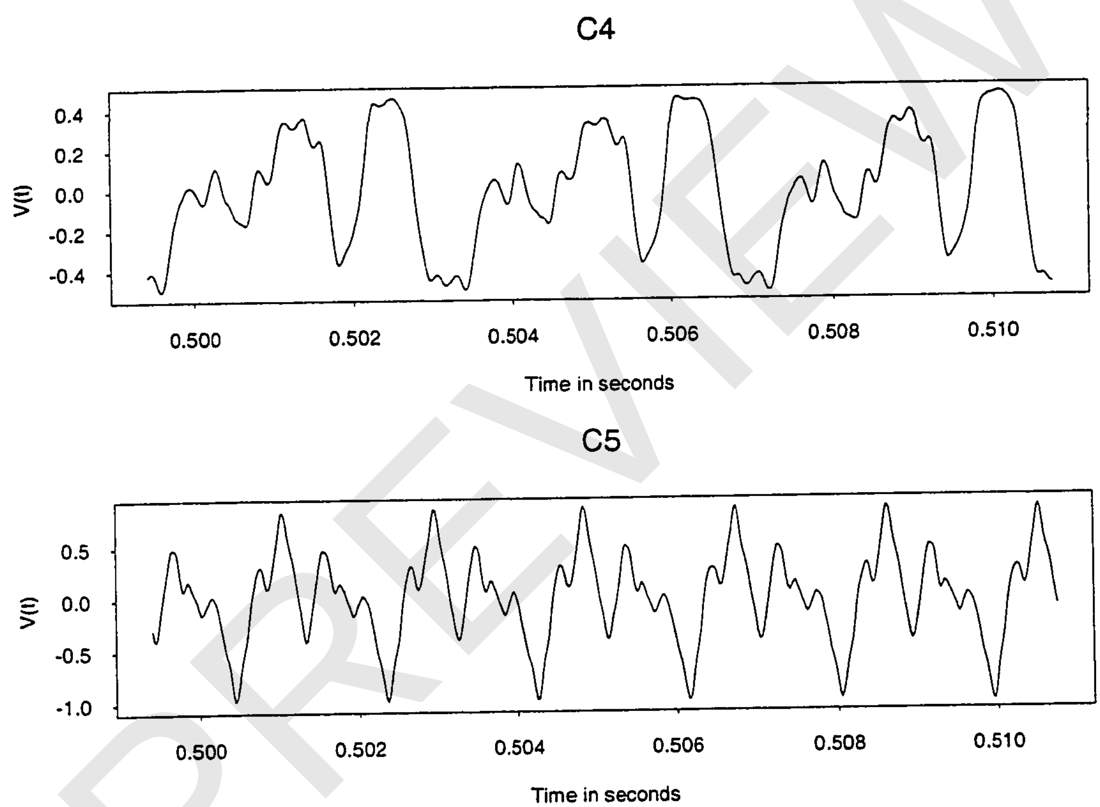

Figure 2.2: Function V(t) for 10 milliseconds of a violin sound playing C4 (middle C) and also playing an octave above, C5.

It should be noted that the frequency related to concert pitch A is not necessarily 440 Hz. For example the San Francisco Symphony tunes to *A 442 Hz. concert pitch.* In this work we will assume concert pitch A represents 440 Hz. and obtain the frequencies related to all other notes using equation (2.1).

The equal tempered scale convention has permitted composers to write with a notation that an instrumentalist can then turn into sounds. It provides another representation