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

The dissertation of Rafael Angel Irizarry is approved:

Chair Date

Date

Date

University of California, Berkeley

Spring 1998

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

by
Rafael Angel Irizarry

Abstract

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

by

Rafael Angel 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.

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, $0 \leq t \leq 1$)

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

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

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

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

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 $t \in \{\frac{1}{N}, \ldots, \frac{N}{N}\}$. 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.

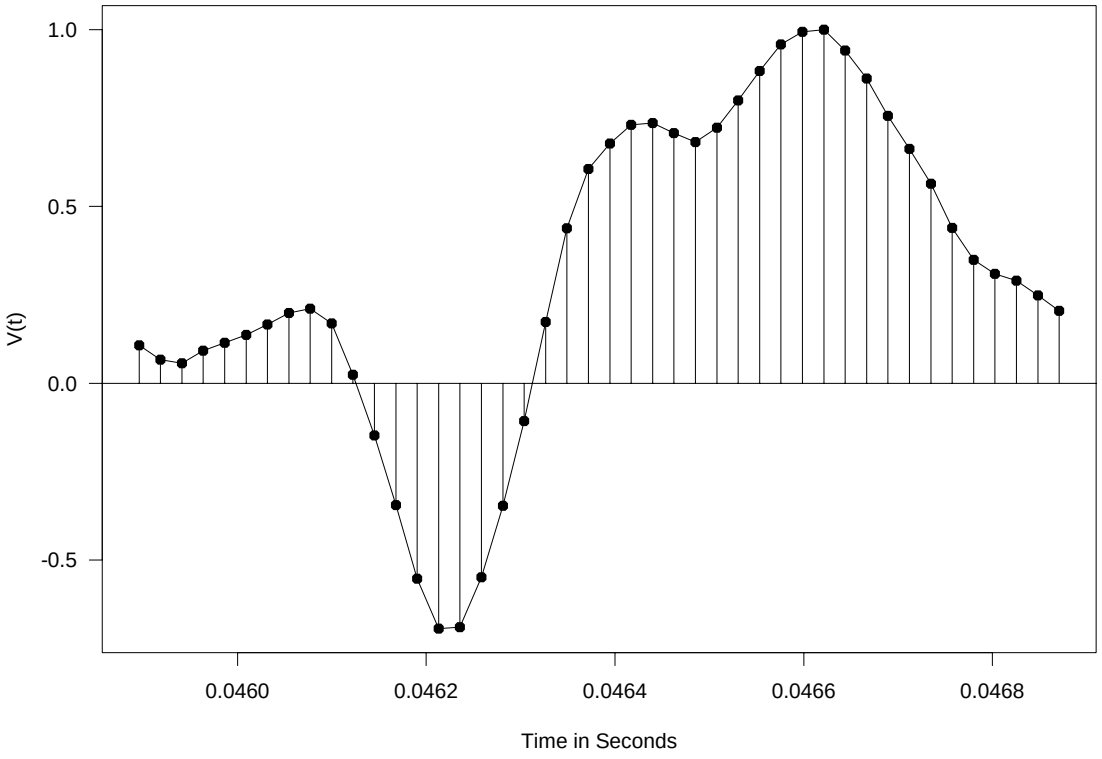

Figure 2.1: Function V(t) for a millisecond of a violin sound sampled at 44100 Hz.

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).

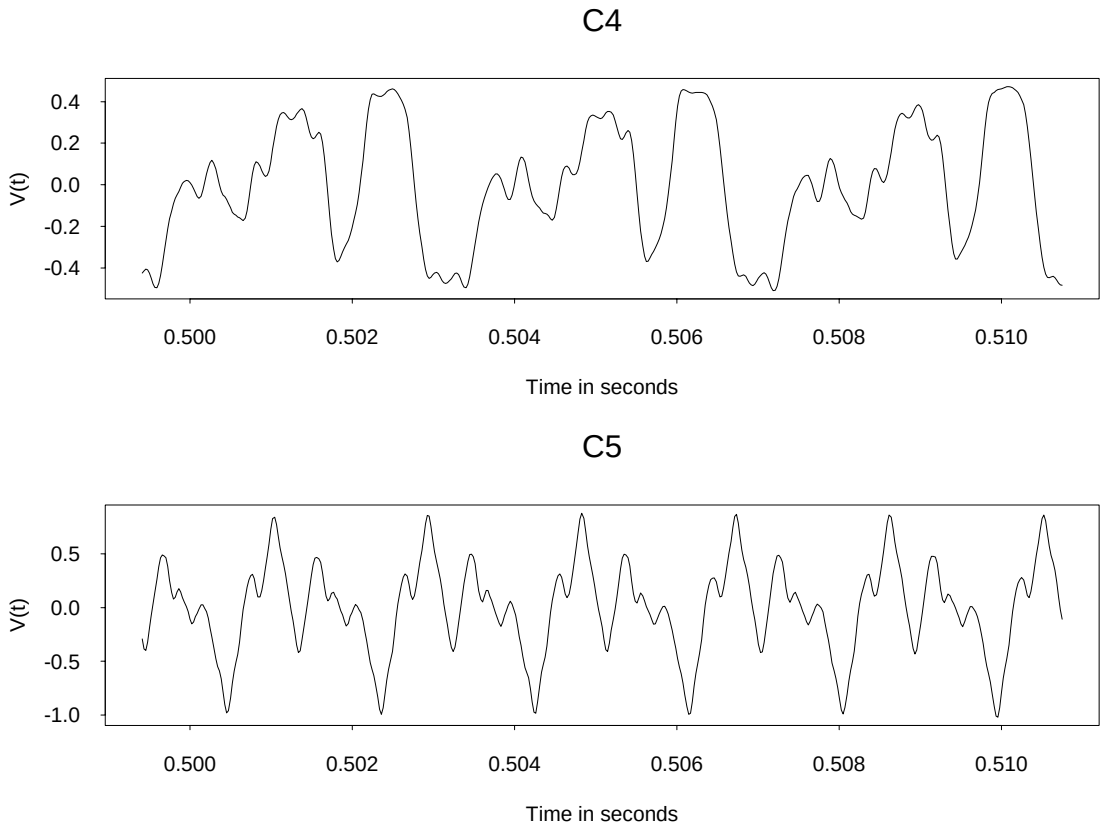

Figure 2.2: Function V(t) for 10 milliseconds of a violin sound playing C4 (middle C) and also playing an octave above, C5.

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

of music, the *score representation*. There are many ways in which scores can be converted into data, see for example Brillinger and Irizarry (1998), Bilmes (1993). This thesis focuses on the signal representation.

More recent discoveries have been related to the timbre of the instrument. For centuries, when listening to sounds produced by most instruments with perceivable pitch, musicians have been able to perceive tones at frequencies other than the fundamental frequencies. These tones are called *overtones*, and the periodic components related to them are called *partials*. For certain instruments, which we will call *harmonic instruments*, the keen ear can notice that the pitches of these overtones are related to frequencies that are multiples of the fundamental frequency being played (track 6 on accompanying CD). Sound analysis experiments confirm this fact (Brown 1996). When a harmonic structure is present, the first partial is the fundamental frequency and the partials at multiples of the first are called *harmonics*. Notice that in this case the first harmonic is the second partial. Instruments that don't have this harmonic structure, or *non-harmonic instruments*, will have partials, but not harmonics.

As early as the later half of the 19th century physicists were interested in the harmonic structure of musical sound signal (Rayleigh Reprinted 1945). Around this time a physicist named Hermann Ludwig Ferdinand von Helmholtz conducted an experiment that proved that sound signals produced by harmonic instruments actually had frequency components at multiples of the fundamental frequency, see von Helmholtz (1885). This discovery inspired physicists to seek an explanation for this phenomenon.

When a string is struck or plucked, it vibrates at different natural frequencies in accordance with their tension and diameter. The energy of vibration is then transfered to the air by way of a vibrating plate of wood and a resonating air chamber, with the sound eventually dying away. The musician can change the pitch by changing the length of the string using her/his fingers or hands.

The principles underlying the acoustics of bowed-string instruments, such as the violin, and wind instruments, such as the oboe, are different from the plucked strings. Here, a vibration is maintained by a feedback mechanism that converts the motion of the bow or the application of blowing pressure into an oscillatory motion that is converted into a sound signal. In the case of the wind and the bowed-string instruments, different tones are obtained by changing the length of the air column or the string respectively. The case of brass instruments is a bit more complicated, see Benade (1973) for a detailed exposition.

In Benade (1976), Fletcher and Rossing (1991) mathematical models of the physical acoustics of instrument sound productions are presented for practically all orchestral instruments and various non-orchestral ones. The equations of these physical models describe the mechanical and acoustic behavior of an instrument being played.

Recently researchers have become interested in the fact that not all the energy put into the physical system (that is the instrument) is converted into a sinusoidal signal, i.e. there is more in the signal than the periodic components related to the overtones (Cook et al. 1990). In some cases left over energy produces sound that is incorporated into the sound signal we hear as produced by the instruments. For example, the sound produced by a beginner on a flute may have a "windy" quality. This is the sound of air being blown into the flute and not converted into a harmonic signal. The component of the sound that is not produced by the partials or periodic components is referred to as the *residual*, *noisy* or *non-sinusoidal* component of the sound signal (Serra 1989). Some instruments are characterized by having a strong non-sinusoidal component, for example the shakuhachi flute discussed in Chapter 6.

The presence of non-sinusoidal components in a sound signal seem to be stronger during the beginning of a note, or what musicians refer to as the *attack*. For many instruments, the system that produces the harmonic signal takes an instant to fall into equilibrium. Researches have found that the attack is one of the most important factors in determining timbre (Grey and Moorer 1977, Charbonneau 1981, Masri and Bateman 1996). The non-sinusoidal component is therefore thought of as an important characteristic of the sound signal (Maganza and Caussé 1986, Cook et al. 1990, Chafe 1990).

Physical models also exist for non-harmonic instruments. Some non-harmonic instruments, such as the marimba, xylophone, timpani and piano have perceivable fundamental frequency. Some of these have partials and in the case of the piano, they are close to multiples of the fundamental. Others, such as percussion instruments like the gong, cymbals, bongos, and wood-blocks, have few or no frequency components. The physics that explains these facts can be be found in Benade (1976), Fletcher and Rossing (1991).

2.4 Psychoacoustics

In the second half of the 19-th century George Simon Ohm (of Ohm's law) conjectured that the human auditory system operates as a spectrum analyzer that displays the

power spectrum of a complex tone and is insensitive to the relative phases of the components (Hartman 1997). By spectrum, Ohm was referring to the Fourier transform of the signal

$$f(\lambda) = \int_{-\infty}^{\infty} V(t) \exp\{it\lambda\}\, dt$$

The power spectrum being the modulus $|f(\lambda)|^2$. At his time this conjecture was not accepted as plausible, but recent psychology experiments seem to confirm its validity, see Grey and Gordon (1978), Risset and Wessel (1982) for some examples and Pierce (1992, Chapter 7) for an overview. Furthermore, physiological studies of the ear have found evidence to support the validity of Ohm's conjecture, (Patterson et al. 1992). It is believed that a sort of *spectral analysis* is carried out in the cochlea to provide the brain with the necessary information to determine timbre see Pierce (1992, Chapter 7) for an overview.

We must notice that any musical signal $V(t)$ and its reverse $V(-t)$ will have the same power spectrum although in many cases the ear can definitely distinguish between a sound, say of a song, and its reverse. Ohm was perhaps referring to the way the ear operates within very small time windows. Studies have been conducted (Patterson and Green 1970) to determine the temporal resolution at which phase starts to make a difference. Green (1985) performed an experiment where subjects where played two different sounds separated in time by a number, Δt, of seconds. Then the same sounds where played but in reverse. Subjects were asked to say if they could hear a difference. After Δt was smaller than 1.5 milliseconds the difference was never noticed.

Psychoacoustic experiments also suggest that we hear "distances" in pitch on a logarithmic scale. For example, we perceive the distance between two notes that are an octave apart as the same no matter how high the frequency. The distance between 110 Hz. and 220 Hz., sounds the same as the distance between 4400 Hz. and 8800 Hz. For this reason it might be convenient to measure pitch distance in semitones instead of Hz.

2.5 Sound analysis and synthesis

Existing sound synthesis and analysis techniques are not necessarily meant to describe or model sound signals but rather to obtain useful parametric representations. A main goal is to obtain a parametric representation $\beta(t)$ of the sound signal $y(t)$ that provides a source for reconstruction or synthesis of the original signal, i.e. that $y(t) \approx s(t, \beta(t))$, and

that meaningful sound transformations can be obtained through the function $\beta(t)$. We will briefly describe the three basic models that are most commonly used for musical sound analysis and synthesis. For a detailed survey see Moorer (1977), Poli (1989), Smith (1991).

2.5.1 Abstract models

Abstract or black box models attempt to provide musically useful parameters in an general way. For example, Chowning's FM modulation (Chowning 1973) provides a way to generate synthetic sounds that have the harmonic characteristics of natural instrument sounds. In FM modulation we define a carrier frequency function $C(t)$, a modulator function $M(t)$, and peak amplitudes A and I for the carrier and modulator respectively. The parametric representation of a sound is then $\beta(t) = (C(t), M(t), A, I)$. To synthesize we use the formula

$$y(t) = A \sin(C(t) + [I \sin(M(t))])$$

The model is called abstract because the parameter function $\beta(t)$ here has no apparent physical interpretation. However, the synthesis obtained by altering the parameters until the sound produced by $y(t)$ is similar to the natural sound being synthesized works relatively well. Furthermore, we may obtain musically meaningful sound transformation via changes on the function $\beta(t)$. Other examples of abstract models can be found in Templaars (1977)

2.5.2 Physical models

Physical models attempt to parameterize sound in a fashion reflecting its source. Physical modeling synthesis starts from mathematical models of the physical acoustics of instrumental sound production. The parametric representation describes the mechanical and acoustic behavior of an instrument being played. Physical dimensions and constants of vibrating objects, such as their mass and elasticity, are specified. Boundary conditions to which the vibrating object is constrained are stipulated. Finally, the excitation is described algorithmically as a force disturbing the vibrating object in some way. In Figure 2.3 a clarinet is modeled using the *waveguide* technique (Hirschman et al. 1991, Hirschman 1991). Different notes and timbres can be obtained by changing parameters such as the size of the upper and lower bore (diameter of the instruments hole). For examples of physical modeling see Hiller and Ruiz (1971), Poli (1989), Jaffe and Smith (1983), Karplus and Strong (1983),

Rodet and Vergez (1996), Sullivan (1990), Borin et al. (1992), Verge (1996), Välimäki et al. (1996).

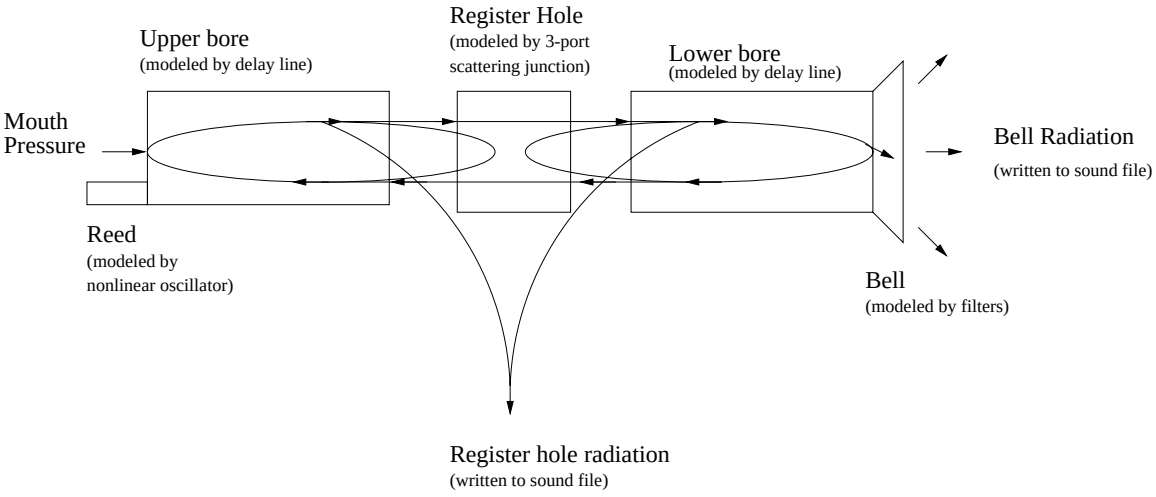

Figure 2.3: Clarinet modeled as a five-part structure.

2.5.3 Spectrum models

Spectrum models attempt to parameterize a sound by reflecting the way the human ear works. These models are based on the accepted belief that the human ear is able to decompose sound into different ranges of frequency. Example of spectrum models can be found in Flanagan and Golden (1966), Portnoff (1976).

Recent spectrum models combine ideas of the physical models. A common procedure assumes that the sound signal is the output of passing some simple waveform, for example a glottal excitation or blown air, through a linear time-varying filter that models the characteristics of the instrument in question. This model is of particular interest in the case of the human voice (Atal et al. 1978, Rodet et al. 1984). These characteristic are not modeled via physical theory describing the instrument, but rather by a spectrum type analysis of the sound. We summarize the characteristics describing the instrument by keeping only the information needed by the ear. The filter theory considered important to these techniques is described in Smith (1985). Some procedures assume that the original waveform is a stochastic processes and use statistical methods to motivate estimation techniques of the parameters of the filters, see for example Tabei et al. (1991), Yang and

Cabrera (1992), Dandawate and Giannakis (1992), Platonov et al. (1992).

2.6 Additive synthesis

Some of the first attempts at sound synthesis were based on *additive synthesis* as in Risset and Mathews (1969). This has proven to be one of the most effective methods available until now (Rodet 1997). Sound signals are modeled as summations of time-varying sinusoidal components. Additive synthesis is accepted perhaps as the most powerful and flexible method of sound/synthesis analysis. In fact, similar sinusoidal models have been proposed for speech signals (McAualay 1986). The details and reasoning leading to this technique are discussed below.

2.6.1 Periodogram analysis

We mentioned that when harmonic instruments produce musical sounds the keen ear can hear harmonic components at multiples of the fundamental frequency. Time series analysis provides a tool that allows us to check if the data are in agreement with this fact. For a stretch of signal Y_t the periodogram is defined by:

$$I^T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}\exp\{-i\lambda t\}Y_t\right|^2 , 0 \leq \lambda \leq \pi \tag{2.2}$$

When the signal Y_t has periodic components at certain frequencies, the periodogram will show peaks at these frequencies (Bloomfield 1976). Computed periodograms of sound signals produced by harmonic instruments verify the existence of overtones. Figure 2.4 presents the periodograms for the signal produced by a trumpet playing concert pitch A and for a clarinet playing that same note. Each signal is about 3 seconds long and the sample rate is 44.1 kHz, so $T \approx 132300$. Also note that the periodogram has peaks in the frequencies expected, mainly at $k \times 440$ Hz, with $k = 1, \ldots, K$. Notice that the trumpet seems to have more noticeable harmonics. This is in agreement with the fact that the timbre of the trumpet is *brighter* than that of a clarinet.

Helmholtz and other researchers of his time used methods based on the periodogram to analyze musical sounds that were "stable" (approximately fixed in amplitude and pitch). The intensity of each partial, as measured by the periodogram, namely

$$\text{intensity of partial } k = \sqrt{I^T(k\lambda)},\ \lambda \text{ the fundamental frequency}$$

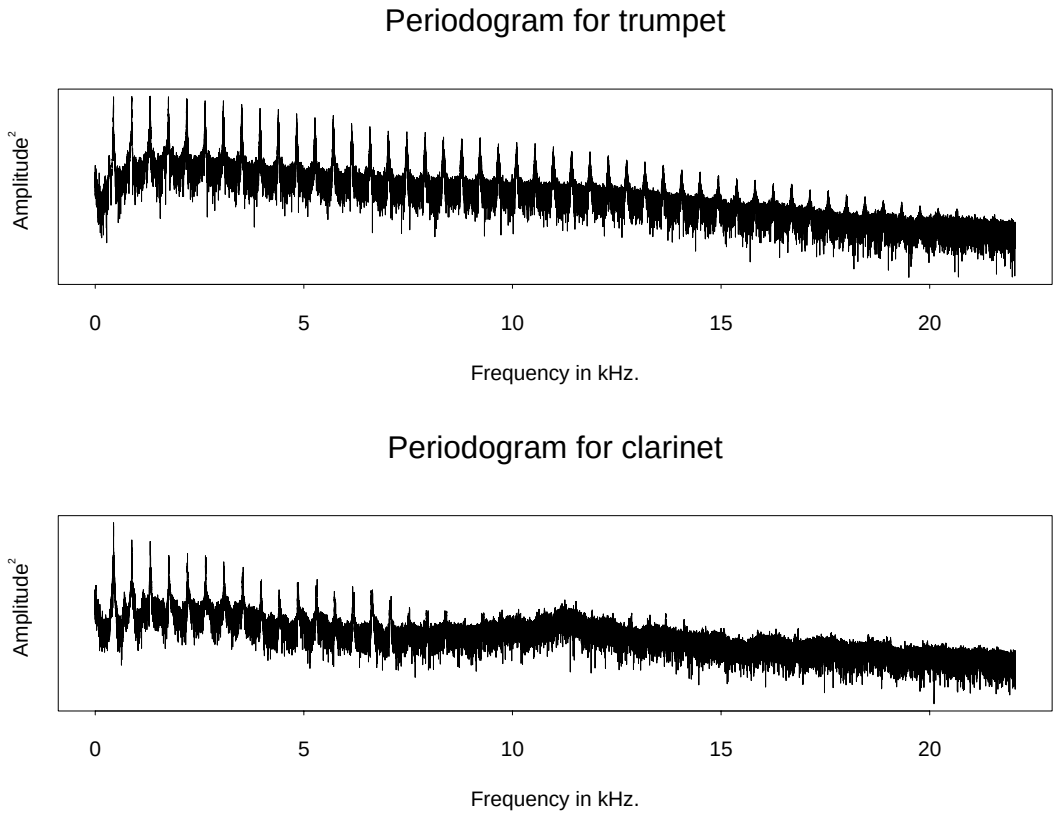

Figure 2.4: Periodogram for sound signals of a trumpet and a clarinet playing concert pitch A.

was used as a characterization of the sound. This characterization suggests the following parametric representation: define $\beta(t) = (\lambda, a_1, \ldots, a_k)$ where λ is the fundamental frequency and a_k is the value of the square root of the periodogram at the k-th partial. To obtain the approximation of the original sound from the parametric representation simply take $\hat{y}(t) = \sum_k a_k cos(k\lambda t)$. Instruments' sounds have been *synthesized* using this representations but the results are not satisfactory, in the sense that the synthetic sounds obtained sounded quite different from the original (track 7 on accompanying CD).

2.6.2 Dynamic periodogram analysis

In the early 1960s, Risset and Mathews (1969) made a discovery that greatly advanced the understanding of timbre. Risset and Mathews (1969) were the first to use the computer to analyze the sound produced by musical instruments in digital form. This

allowed him to study the local behavior of the harmonic components of signals. Mathews noticed that the intensity of the overtones varied relative to each other through time.

One way to verify this discovery using statistical tools is to compute dynamic periodograms, or *spectrograms*. We define the spectrogram of a signal at time t_0 by:

$$I^T(t_0, \lambda) = \frac{1}{2\pi T} \left| \sum_{t=t_0-M}^{t_0+M} \exp\{-i\lambda t\} Y_t \right|^2 \tag{2.3}$$

Here $T = 2M + 1$ is a suitable window size.

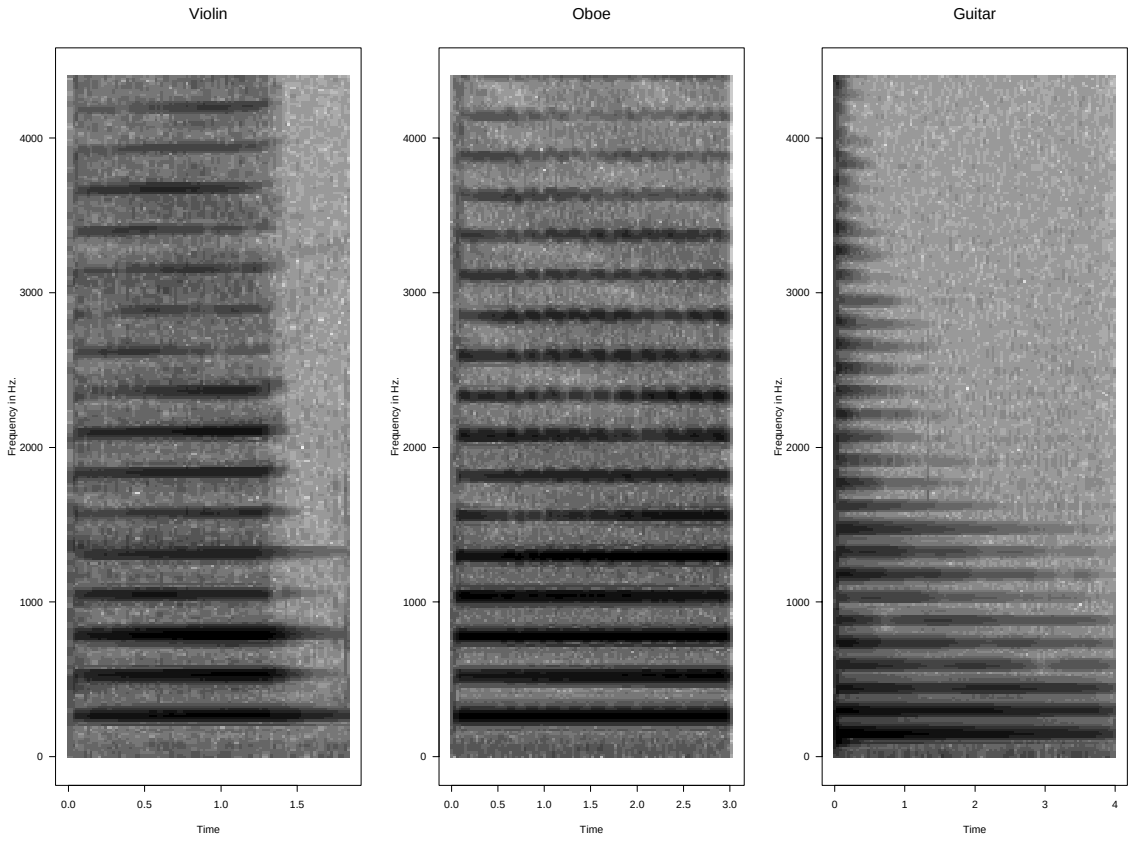

Figure 2.5: Spectrograms for harmonic instruments.

In Figure 2.5 we see the spectrogram for the signals produced by three harmonic instrument; a violin, an oboe, and a guitar (tracks 8, 9, and 10 on accompanying CD). The violin and oboe are both playing C4 (261.6256 Hz), while the guitar is playing D3 (146.8324 Hz). Dark shades of grey represents high power for the spectrogram. In these spectrograms the window size $2M + 1$ is taken to be 20 milliseconds. Notice that the spectrograms verify

Mathews' discovery. All the instruments show harmonic components at the frequencies we expect, yet the amplitudes of these harmonic components are definitely varying through time in different ways. This is particularly clear in the case of the guitar where the higher harmonics "die off" more rapidly in the sound.

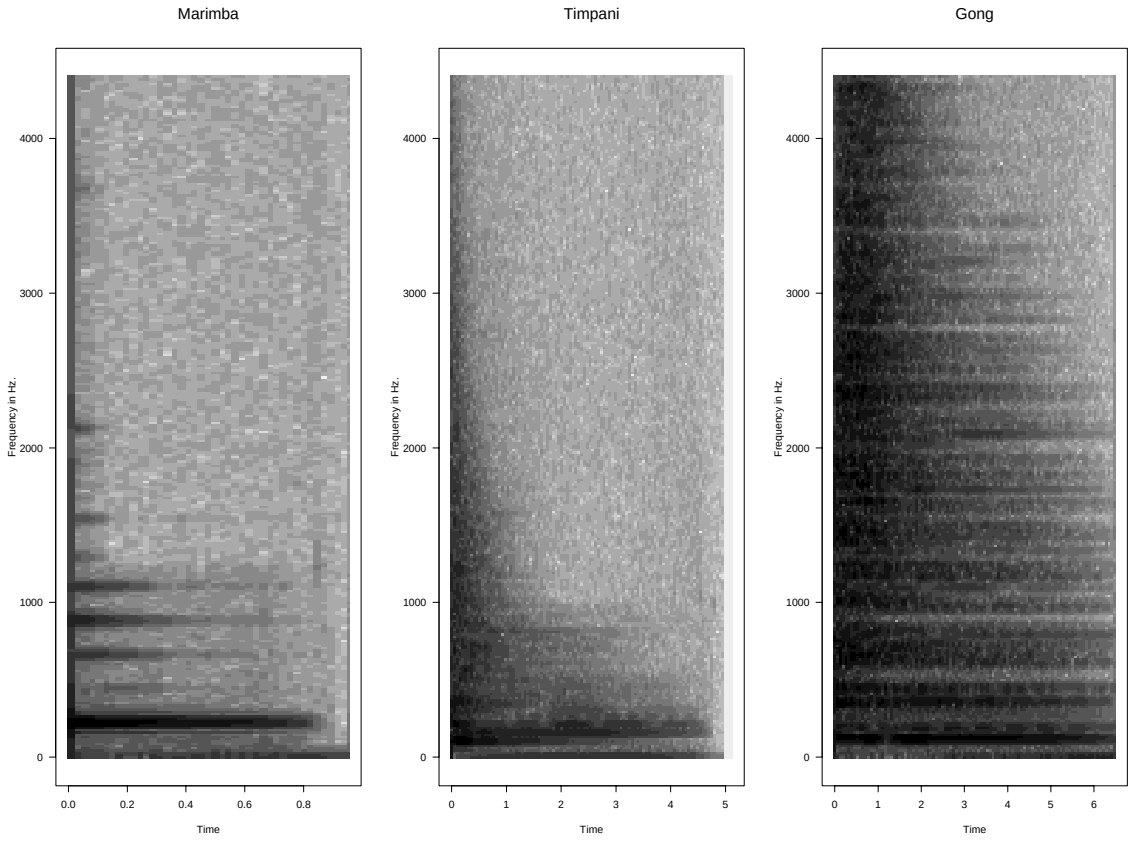

Figure 2.6: Spectrograms for non-harmonic instruments.

In Figure 2.6 we see the spectrogram for the signals produced by three non-harmonic instrument; a marimba, a timpani and a gong (tracks 11, 12, and 13 on accompanying CD). We notice that for these non-harmonic instruments the spectrograms show the lack of a harmonic structure. The graphs also give examples of signals with perceivable, ambiguous and non-perceivable pitch. Notice that the dark lines, representing high power in Figure 2.6, suggest that the marimba has a very definite fundamental frequency and some partials. In the case of the timpani there seems to be a fundamental frequency although it is not as clear as in the marimba and is more apparent in the later part of the signal. The gong seems to have no fundamental frequency or partials whatsoever. All these observations

are in agreement with what is perceived by our ears and described by the physical theory.

This analysis suggests the following parametric representation

$$\beta(t) = (\lambda, a_1(t), \ldots, a_K(t))$$

The signal reconstruction would then be $\hat{y}(t) = \sum_k a_1(t)cos(k\lambda)$. Mathews synthesized sounds using this parametric representation and found it greatly improved on the earlier mentioned technique.

The signals obtained from this reconstruction are still not exactly equal to the original, i.e. $y(t) - \hat{y}(t) \neq 0$. In fact, the difference between the original and the synthesized signals is perceived by a trained human ear (track 14 on accompanying CD). One possible explanation for this is the existence of a non-sinusoidal component. Some of the non-sinusoidal components that can be incorporated into the signals are the sound produced by fingers hitting keys, nails plucking strings, and surplus blown air. Some researchers have assumed this part of the signal to be stochastic and have proposed using an additive plus residual model.

2.7 Additive sinusoidal plus residual model

Serra (1989) incorporated a non-sinusoidal residual part to the additive synthesis and modeled it as an additive random signal. Since, many have proposed and used similar models (Serra and Smith 1991, DePalle and Poirot 1991, Rodet and Depalle 1992, Depalle and Tromp 1996, Solbach and Wöhrmann 1996, Rodet 1997). Notice that under this assumption one is dealing with a signal plus noise statistical model. In Serra (1989) the model presented is

$$y(t) = s\left[t; \beta(t)\right] + \epsilon(t) \tag{2.4}$$

with

$$s\left[t; \beta(t)\right] = \sum_{k=1}^{K} a_k \cos(\phi_k(t)) \tag{2.5}$$

with $\beta(t) = (a_1(t), \ldots, a_K(t), \phi_1(t), \ldots, \phi_K(t))'$. An implicit assumption is that the signal $s[t; \beta]$ resembles a sum of pure sinusoids.

Notice that in this case the parameter $\beta(t)$ can be thought of as the parametric representation of the deterministic part of the sound. Serra (1989) assumes the non-sinusoidal

part, $\epsilon(t)$ to be a stationary autoregressive process. The parameters of this stochastic process would then complete the parametric representation of the sound.

2.7.1 Estimation

One is interested in estimating $\phi_k(t)$ and $a_k(t)$. Estimation is done in two steps. In the first step the signal is divided into short, possibly overlapping segments, called *analysis frames* (Serra 1989). For each segment the peaks of the periodogram of the tapered data are considered as possible indication of a sinusoidal partial. The amplitude and frequency of each peak is recorded. In the second step, peaks of successive analysis frames are grouped into tracks. For a particular track, say the i-th track, the frequencies at which the peaks occur are considered to be estimates of the sinusoidal partial associated with $\phi_k(t)$. This is called *partial tracking* (Depalle et al. 1993a). This tracking is usually based on a heuristic approach (McAualay 1986, Serra 1989) that matches peaks of consecutive frames by the proximity of the frequencies associated with them. The algorithm allows *deaths* and *births* of partials for the case where one frame contains more peaks than the other. The problem with this technique is that it does not take into account the fact that some of the peaks might not be produced by sinusoidal components, but rather by the stochastic non-sinusoidal part of the signal. A procedure described in Depalle et al. (1993a), DePalle et al. (1993b) takes this into account and performs the tracking by globally optimizing over the set of all tracks via a Hidden Markov Model. The validity of this method under the model defined in (2.5) and the statistical properties of the estimates obtained are not discussed by DePalle et al. (1993b). Solbach and Wöhrmann (1996) test partial tracking techniques in simulated data, however theoretical exploration of this problem is not done and is left as future work.

2.7.2 Problems

Notice that we assume the existence of deterministic sinusoidal components (the partials). The strong peaks seen in the periodogram and spectrograms of signals produced by harmonic instruments, see Figures 2.4 and 2.5, agree with this assumption. The above mentioned partial tracking algorithms allow partials to exist at frequencies that are not multiples of the fundamental frequencies. Estimates obtained for the deterministic sinusoidal signal $s[t;\beta(t)]$ when many non-harmonic partials are "tracked" are hard to interpret.

Furthermore, the periodogram of a signal that is assumed to be stochastic will

also be stochastic. A peak in the periodogram can be due to chance and in particular when say only 256 observations are used when computing the periodogram the variation can be relatively large. Computing the statistical properties of periodogram peaks found by a tracking algorithm as the one presented in Rodet (1997) can be complicated, even when using a simple statistical model. For this reason finding an algorithm for partial tracking that provides useful estimates is not straightforward. In this work we do not intend to search for such an algorithm. Instead, for the case of signals produced by harmonic instruments, we assume a harmonic version of (2.5) and present an estimation procedure that, under regularity conditions, provides consistent asymptotically normal estimates. These do appear useful in practice. Notice that this will not only reduce the risk of incorporating unwanted partial tracks but also improve the accuracy of our estimates, as will be shown in section 3.4.3.

2.7.3 An example of partial tracking

For Figure 2.7, a stretch of length 0.4 seconds of a trumpet sound, playing concert pitch A, was analyzed. The stretch of sound was divided into 70 non-overlapping frames, each with 256 data points (0.006 seconds). For each frame, the figure shows various dots. These dots represent the frequencies (up to 4410 Hz) at which the periodogram has local maxima. The size of the points in the figure represents the amplitude of the local maxima. The lines seen in the figure represent the tracks believed to be the sinusoidal partials that a particular *partial tracking algorithm* detects. A straight forward method is employed. Within a given frame, say frame i, we do the following: for each frequency 440 x k, k an integer, the closest peak frequency $\phi_k(i)$ is considered to be part of the i-th track if $|\phi_k(i) - 440k| < 220$. If no such frequency exist, the track has a *death* at that frame. Notice that in some of the partials the tracks end at a particular frame and begin again in another. This is a feature of many partial tracking algorithms.

2.8 Applications

Obtaining parametric representations of sounds leads to many musical applications, (Mathews 1969, Mathews and Pierce 1989, Rowe 1994). In particular, the separation of the noise from the discrete part of the signal and the decomposition of the sinusoidal part into the separate partials can be used as in the following applications.

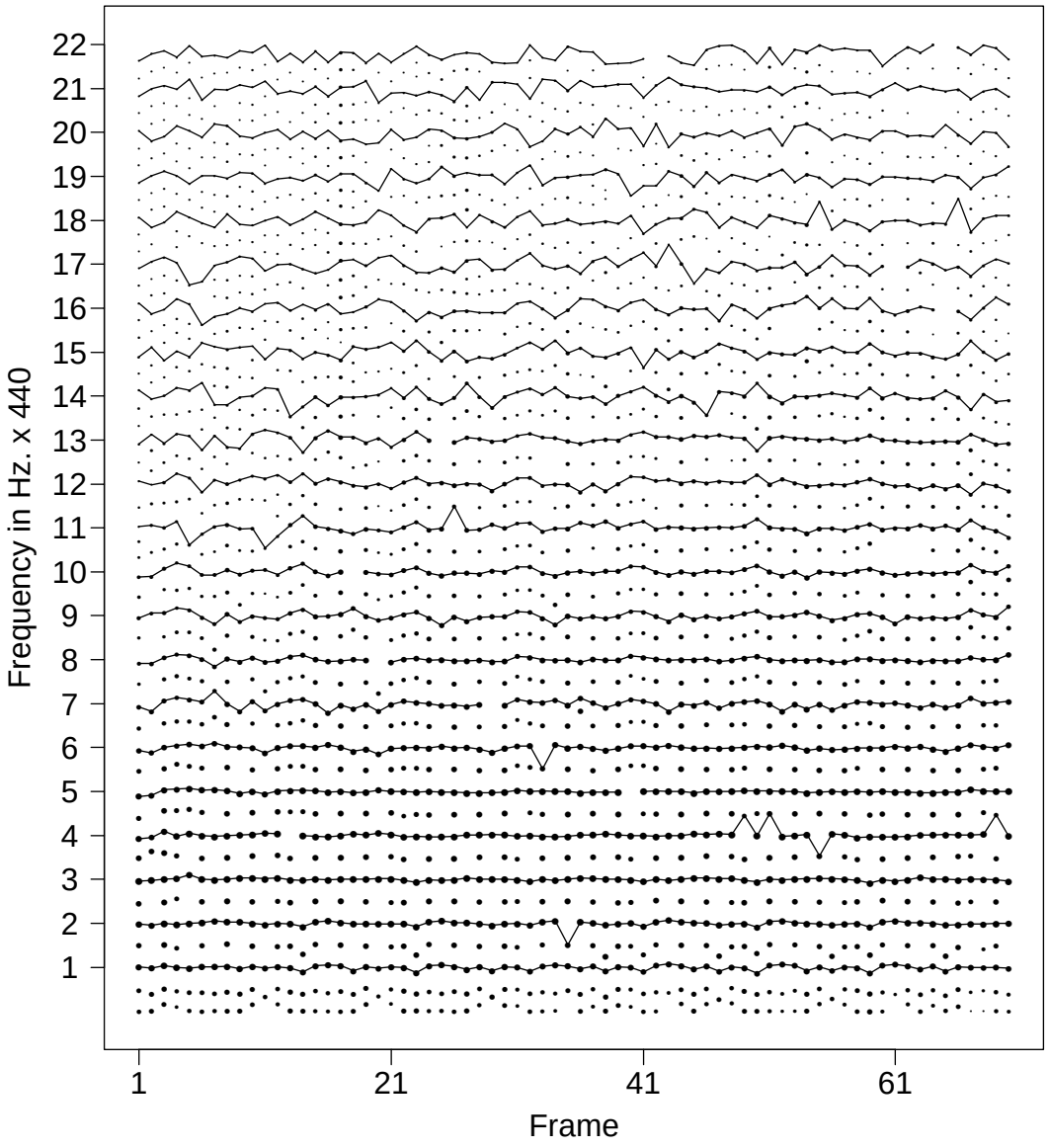

Figure 2.7: Periodogram maxima of analysis frames with partial tracks of a 0.4 second stretch of the sound signal of a trumpet playing A4.

2.8.1 Audio signal restoration

Music recordings can be damaged in different ways. For example, archived gramaphone recordings are usually of poor quality. Recovering the original signal from the available one is a problem that many in signal processing and music technology fields have tried to solve. In Vaseghi and Rayner (1988) some statistical tools are employed with multiple copies of the same recording. Sound synthesis techniques can be used as an alternative method, even when only one copy of the recording exists.

2.8.2 Sound recreation

As mentioned above, once we obtain a parametric representation of a sound signal, we can then use it to recreate the sound without the use of the musical instrument that created it. The parameters should be musically meaningful and should allow us to change the sound in musically meaningful ways.

Sound recreation techniques can provide more realistic synthetic sounds of natural instrument tones. The way most synthesizers work today is by recording a large number of an instrument's tones. When a key is pressed on the synthesizer, the appropriate recording is played back. Tones for which there are no recordings are synthesized by playing the recorded tone at a different volume and/or pitch. The result is for the most part easily distinguishable from the actual instrument. This is because there are spectral changes associated with dynamic and pitch changes. These changes are not captured by simply varying the amplitude and pitch of a recording. An actual loud piano note sounds different from a quiet piano note played at a high volume. This technique relates to timbre morphing which is discussed in the following section.

2.8.3 Timbre morphing

Timbre morphing is the process of combining two or more sounds to create a new sound with intermediate timbre and duration. This process differs from simply mixing sounds in that only a single sound, with characteristics from the two original ones, is audible as the morph sound (Tellman et al. 1995).

Morphing can be used to create interesting sounds that are not found in nature, but that have the characteristics of naturally occurring sounds. The resulting sounds can be used, for example, in electronic music composition.

An interesting example is the recreation of a castrato voice (Depalle et al. 1995). This was done to produce a sound-track for the film about Farinelli, the famous 18th century castrati. To create Farinelli's voice, the voice of a counter tenor and a soprano were analyzed and parametric representations were obtained. These two representations were combined in a way that produced a timbre similar to that of a castrato.

The representation of the signal obtained from fitting a harmonic model provides us with a direct method of morphing two sounds. In general if we have the parametric representations of two sounds, $\beta_1(t)$ and $\beta_2(t)$, we may then obtain a morph sound via

$$\tilde{y}(t) = s\left[t; \lambda(t)\beta_1(t) + [1 - \lambda(t)]\beta_2(t)\right]$$

with $\lambda(t) \in [0, 1]$ determining the "mix" of the morph at each time t.

2.8.4 Time-scale and pitch modification

By playing a 33 speed record at 45 speed, we can modify the pitch of our favorite artists so that they sing high enough so as to sound like the "Chipmunks". Yet in this case the sound signal is also played faster, i.e. the time-scale also changes. Fancy *samplers* used in the music industry use a technique similar to the record speed technique. The sample rate conversion technique used in digital samplers (Smith and Gossett 1984), achieves pitch alterations by changing the rate at which sounds are read from memory.

In many applications it is useful to vary the time scale of a signal without affecting pitch, or to conversely modify pitch without changing the time scale. An advantage of additive synthesis is that it permits the pitch and time-scale to be varied independently (Quatieri 1992). A simple example is to take the simple parametric representation presented in section 2.6.1, $\beta(t) = (\lambda, a_1, \ldots, a_k)$, and create a new signal with higher or lower pitch by simply changing the value of λ. To change the time-scale we could simply re-scale the time-unit through a time-scale function $z(t)$. In this case we reconstruct the new signal $y(t) = \sum_k a_k cos(k\lambda z(t))$

2.9 Conclusion

In this chapter we have shown how musical sound signals may be represented as discrete time series that can be analyzed statistically. We also discussed different applications that motivate sound analysis and synthesis. The physics of musical instruments and

psychology of music motivate different sound analysis and synthesis techniques. One such technique, additive sinusoidal plus residual model, is based on a statistical signal plus noise model which we will study in the work that follows.

Chapter 3

Frequency Estimation

3.1 Introduction

The physical modeling described in Chapter 2 suggests that sound signals produced by musical instruments may be characterized by a harmonic structure and noise. In this chapter we will present some results relating to previous work on frequency estimation that will subsequently be used in the following chapters.

Consider the model

$$y_t = s(t;\beta) + \epsilon_t \quad t = 1, \ldots, T$$

where the stationary ϵ_t has autocovariance function $c_{\epsilon\epsilon}(u) = \mathrm{Cov}\{\epsilon_{t+u}, \epsilon_t\}$, satisfies Assumption 1 below and has power spectrum

$$f_{\epsilon\epsilon}(\lambda) = \frac{1}{2\pi} \sum_u c_{\epsilon\epsilon} \exp\{-i\lambda u\}, \quad \infty < \lambda < \infty$$

Assumption 1 *$\{\epsilon_t\}$ is a strictly stationary real valued random process all of whose moments exist, with zero mean, and with $c_{\epsilon\ldots\epsilon}(u_1, \ldots, u_{L-1})$ the joint cumulant function of order L of the series ϵ_t for $L = 2, 3, \ldots$. Further the*

$$C_L = \sum_{u_1=-\infty}^{\infty} \cdots \sum_{u_{L-1}=-\infty}^{\infty} |c_{\epsilon\ldots\epsilon}(u_1, \ldots, u_{L-1})| \tag{3.1}$$

satisfy

$$\sum_k C_k z^k / k! < \infty$$

for z in a neighborhood of 0.

This assumption requires that the time series ϵ_t have a short span of dependence in a sense including that the measurements ϵ_t and ϵ_s are less statistically dependent on each other as they become more distant, i.e. as $t - s \to \infty$.

Many signals in nature have been statistically analyzed via sinusoidal regression models (Brillinger 1977), for example

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

where $\beta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\omega})' = (A_1, \ldots, A_K, B_1, \ldots, B_K, \omega_1, \ldots, \omega_K)'$. In a musical context, as indicated in previous chapters, we call the components of $\boldsymbol{\omega}$ the frequencies of the *partials*. The amplitudes of the partials are defined by $\rho_k = \sqrt{A_k^2 + B_k^2}$.

In Walker (1971), Hannan (1971), Hannan (1973), Hannan (1974), Brown (1990) estimates that are asymptotically equivalent to least squares estimates are presented. Consistency is shown for these estimates and asymptotic variance expressions are developed.

We intend to assume a model like the above holds locally in time for sound signals. Since we are going to be fitting this model in order to obtain estimates of parameter functions that depend on time, it is only natural to consider window based estimates. In this chapter we will be presenting the results obtained ibid, but for estimates that are asymptotically equivalent to weighted least squares. The results follow in almost the same way as in the unweighted case. We present the results needed for the more general weighted case.

3.2 One sinusoidal component

We will first consider the case of one sinusoidal component, that is

$$s(t;\beta_0) = A_0 \cos(\omega_0 t) + B_0 \sin(\omega_0 t) \tag{3.2}$$

with $\beta_0 = (A_0, B_0, \omega_0)'$ and then generalize to the case of several partials. Hannan (1973) and Walker (1971), amongst others, observe that the least squares method is asymptotically equivalent to the following, more direct, estimation procedure:

$$\hat{A}_T = \frac{2}{T} \sum_{t=1}^{T} y_t \cos(\hat{\omega}_T t) \tag{3.3}$$

$$\hat{B}_T = \frac{2}{T} \sum_{t=1}^{T} y_t \sin(\hat{\omega}_T t) \tag{3.4}$$

where $\hat{\omega}_T$ is such that

$$q_T(\hat{\omega}_T) = \max_{0 \leq \omega \leq \pi} q_T(\omega) \tag{3.5}$$

Here q_T is defined by

$$q_T(\omega) = \left| T^{-1} \sum_{t=1}^{T} y_t \exp\{it\omega\} \right|^2 \tag{3.6}$$

Notice that $q_T(\omega)$ is proportional to the periodogram, defined in (2.2). The authors above mentioned show that these estimates are consistent and asymptotically normal and the asymptotic variance matrix is obtained.

The weighted least squares method consists of choosing $\hat{\beta}$ to minimize the criterion

$$S_T(\beta) = \sum_{t=1}^{T} w(\frac{t}{T}) \left[y_t - s(t, \beta) \right]^2 \tag{3.7}$$

Here $w(s)$ is a weight function such that if we define the following constants

$$\begin{aligned} W_n &= \int_0^1 t^n w(t)\, dt \\ U_n &= \int_0^1 t^n w(t)^2\, dt \end{aligned} \tag{3.8}$$

then $w(s)$ satisfies the following assumption.

Assumption 2 *Assume that $w(s)$ is non-negative, bounded, of bounded variation, has support $[0,1]$, $W_0 > 0$ and, $W_1^2 - W_0 W_2 \neq 0$.*

Set

$$\Delta_n^T(\lambda) = \sum_{t=1}^{T} w(\frac{t}{T}) t^n \exp\{i\lambda t\} \tag{3.9}$$

Throughout we are going to need the following simple result

Lemma 1 *If $w(t)$ satisfies Assumption 2 then we have*

$$\lim_{T \to \infty} T^{-(n+1)} \Delta_n^T(\lambda) = W_n, \text{ for } \lambda = 0, 2\pi \tag{3.10}$$

$$\Delta_n^T(\lambda) = O(T^n), \text{ for } 0 < \lambda < 2\pi \tag{3.11}$$

for $n = 0, 1, 2$.

Proof: Fix n. To prove (3.10) notice that for $\lambda = 0, 2\pi$ we have that

$$T^{-(n+1)} \Delta_n^T(\lambda) = \sum_{t=1}^{T} (\frac{t}{T})^n w(\frac{t}{T}) (\frac{1}{T})$$

From the boundedness and bounded variation of $w(u)$ we have

$$\lim_{T\to\infty} T^{-(n+1)}\Delta_n^T(\lambda) = \int_0^1 u^n w(u)\,du = W_n$$

To prove (3.11) let $0 < \lambda < 2\pi$ and define

$$\Delta^t(\lambda) = \sum_{s=1}^{t} \exp\{i\lambda s\}$$

with the convention that $\Delta^0(\lambda) = 0$. Letting $h(u) = u^n w(u)$ and using summation by parts we have that

$$\begin{aligned}
\Delta_n^T(\lambda) &= T^n \sum_{t=1}^{T} h(\frac{t}{T}) \left[\Delta^t(\lambda) - \Delta^{t-1}(\lambda)\right] \\
&= T^n \left[\sum_{t=1}^{T} h(\frac{t}{T})\Delta^t(\lambda) - \sum_{t=0}^{T-1} h(\frac{t+1}{T})\Delta^t(\lambda)\right] \\
&= T^n \left[h(1)\Delta^T(\lambda) + \sum_{t=1}^{T-1}\left(h(\frac{t}{T}) - h(\frac{t+1}{T})\right)\Delta^t(\lambda)\right]
\end{aligned}$$

Notice that if $w(t)$ is bounded and has bounded variation on $[0,1]$, so does $h(t)$. Let M be $\sup_t |h(t)|$ and V be the total variation of $h(t)$. Then we have

$$\begin{aligned}
\left|\Delta_n^T(\lambda)\right| &= T^n \left|h(1)\Delta^T(\lambda) + \sum_{t=1}^{T-1}\left(h(\frac{t}{T}) - h(\frac{t+1}{T})\right)\Delta^t(\lambda)\right| \\
&\le T^n \left[|h(1)||\Delta^T(\lambda)| + \sum_{t=1}^{T}\left|h(\frac{t}{T}) - h(\frac{t+1}{T})\right|\,|\Delta^t(\lambda)|\right] \\
&\le T^n \left[M|\Delta^T(\lambda)| + V \max_{1\le t\le T}|\Delta^t(\lambda)|\right]
\end{aligned}$$

We know, see for example Bloomfield (1976), that $|\Delta^t(\lambda)| \le L = 1/|\sin(\frac{1}{2}\lambda)|$ for all t. Notice that L depends on λ, but given $0 < \lambda < 2\pi$ it is constant for all t thus

$$\left|\Delta_n^T(\lambda)\right| \le T^n L(M+V)$$

and this completes the proof of the lemma. □

As done in Walker (1971) for the unweighted case, we notice that if we define

$$\begin{aligned}
R_T(\beta) = &\sum_{t=1}^{T} w(\frac{t}{T})y_t^2 + \frac{1}{2}(A^2+B^2)\sum_{t=1}^{T} w(\frac{t}{T}) \\
&-2\sum_{t=1}^{T} w(\frac{t}{T})y_t\{A\cos(\omega t) + B\sin(\omega t)\} \qquad (3.12)
\end{aligned}$$

with $\beta = (A, B, \omega)$, then

$$S_T(\beta) - R_T(\beta) = \frac{1}{2}\sum_{t=1}^{T} w(\frac{t}{T})\{(A^2 - B^2)\cos(2\omega t) + 2AB\sin(2\omega t)\} \tag{3.13}$$

Here $S_T(\beta)$ is the weighted residual sum of squares of equation (3.7). The difference in equation (3.13) is deterministic and, using Lemma 1, we can show it is bounded as $T \to \infty$ if $0 < \omega < \pi$.

Now notice that the ω that maximizes the periodogram of the tapered data $w(t/T)Y_t$ also maximizes $2\sum_{t=1}^{T} w(t/T)y_t\{A\cos(\omega t) + B\sin(\omega t)\}$ in equation (3.12). Given ω we can directly find the A and B that minimize $R_T(\beta)$ by taking derivatives and solving when they are set to 0. This and (3.13) may be used to show that the following estimates are asymptotically equivalent to the least squares estimates,

$$\begin{aligned}
\hat{A}_T &= 2\sum_{t=1}^{T} w(\frac{t}{T})y_t\cos(\hat{\omega}_T t)/\sum_{t=1}^{T} w(\frac{t}{T}) \\
\hat{B}_T &= 2\sum_{t=1}^{T} w(\frac{t}{T})y_t\sin(\hat{\omega}_T t)/\sum_{t=1}^{T} w(\frac{t}{T})
\end{aligned}$$

Here $\hat{\omega}_T$ is such that

$$q_T(\hat{\omega}_T) = \max_{0\le\omega\le\pi} q_T(\omega)$$

with q_T now defined by

$$q_T(\omega) = \left| T^{-1}\sum_{t=1}^{T} w(\frac{t}{T})y_t\exp\{it\omega\}\right|^2 \tag{3.14}$$

Notice that these estimates are the exact same ones of equations (3.3), (3.4), (3.5), and (3.6) obtained in the unweighted case, but now using tapered data $w(t/T)Y_t$.

To prove consistency and asymptotic normality for these estimates we are going to need a result concerning the behavior of the periodogram of the noise and its derivatives with respect to ω.

Lemma 2 *Let the stationary noise process $\{\epsilon_t\}$ satisfy Assumption 1 and let the weight function $w(u)$ satisfy Assumption 2 then if*

$$p_T(\omega) = \left| T^{-(k+1)}\sum_{t=1}^{T} w(\frac{t}{T})t^k\epsilon_t\exp\{-it\omega\}\right|$$

one has

$$\lim_{T\to\infty}\sup_{0\le\omega\le\pi} p_T(\omega) = 0, \text{ in probability}$$

Lemma 2 has been shown to be true under different assumptions for the equally weighted case, $w(s) = 1$. Walker (1971) proves the Lemma for white noise with finite variance. Hannan (1973) proves it under ergodic and purely non-deterministic conditions. Wang (1991) simply assumes the periodogram of the noise is asymptotically bounded. Under this assumption we can readily obtain the result of Lemma 2. In this case letting

$$d^t(\lambda) = \sum_{s=1}^{t} \epsilon_s \exp\{i\lambda s\}$$

we know, by assumption, that $\max_{1\leq t\leq T} |d^t(\lambda)| = O_p(1)$ and by a summation by parts argument like that in the proof of Lemma 1, the result for the weighted case of Lemma 2 follows. Brillinger (1986) proves a version of this Lemma for spatial point processes. Under Assumptions 1 and 2, Lemma 2 follows directly from results of Brillinger (1981). Specifically defining $h(u) = u^k w(u)$ and

$$d_k^T(\omega) = \sum_{t=1}^{T} h(\frac{t}{T})\epsilon_t \exp\{-it\omega\}$$

we have by Theorem 4.5.1 in (Brillinger 1981, page 98)

$$\sup_{0\leq\omega\leq\pi} \left|d_k^T(\omega)\right| = O_p(\{T\log T\}^{-\frac{1}{2}})$$

Since

$$p_T(\omega) = T^{-1}\left|d_k^T(\omega)\right|$$

the lemma follows.

Now we are ready to prove the following theorem concerning the consistency of the estimates.

Theorem 1 *If ϵ_t satisfies Assumption 1 and weight function $w(t)$ satisfies Assumption 2, then for $0 < \omega_0 < \pi$*

$$\lim_{T\to\infty} \hat{A}_T = A_0, \ \lim_{T\to\infty} \hat{B}_T = B_0, \ \lim_{T\to\infty} T|\hat{\omega}_T - \omega_0| = 0, \ \textit{in probability}$$

Proof: We start by proving

$$\lim_{T\to\infty} T|\hat{\omega}_T - \omega_0| = 0, \text{ in probability} \tag{3.15}$$

which is stronger than ordinary consistency, but is needed to prove the consistency of the remaining two estimates and asymptotic normality.

Next write

$$A_0\cos(\omega_0 t)+B_0\sin(\omega_0 t)=D_0\exp\{i\omega_0 t\}+\overline{D}_0\exp\{-i\omega_0 t\}$$

where $D_0=\frac{1}{2}(A_0-iB_0)$. Now we have

$$\begin{aligned} q_T(\omega) &= \left|T^{-1}d_0^T(\omega)\right|^2+\left|T^{-1}[D_0\Delta_0^T(\omega_0+\omega)+\overline{D}_0\Delta_0^T(\omega_0-\omega)]\right|^2 \\ &+ 2\Re\left[\left(T^{-1}d_0^T(\omega)\right)\left(T^{-1}[D_0\Delta_0^T(\omega_0+\omega)+\overline{D}_0\Delta_0^T(\omega_0-\omega)]\right)\right] \end{aligned}$$

By Lemma 1 we have that for $0<\omega<\pi$

$$T^{-1}\Delta_1^T(\omega_0+\omega)=o(1)$$

and that

$$T^{-1}\Delta_1^T(\omega_0-\omega)=\begin{cases} W_0 & : \ \omega=\omega_0 \\ o(1) & : \ \text{otherwise} \end{cases}$$

Lemma 2 implies that for $0<\omega<\pi$

$$T^{-1}d_0^T(\omega)=o_p(1)$$

So we have that

$$q_T(\omega)=\frac{1}{4}\rho_0^2\left|T^{-1}\Delta_0^T(\omega-\omega_0)\right|^2+o_p(1)$$

and therefore

$$q_T(\omega_0)=\frac{1}{4}\rho_0^2W_0^2+o_p(1)$$

To prove (3.15), for any $b>0$, define

$$P_T(b)=\{\omega: T|\omega-\omega_0|\geq b\} \tag{3.16}$$

Notice that

$$\begin{aligned} \Pr\left(T|\hat{\omega}_T-\omega_0|\geq b\right) &\leq \Pr\left(\sup_{\omega\in P_T(b)}q_T(\omega)\geq q_T(\omega_0)\right) \\ &= \Pr\left(\sup_{\omega\in P_T(b)}\left|T^{-1}\Delta_0^T(\omega-\omega_0)\right|\geq W_0+o_p(1)\right) \end{aligned}$$

By proposition B.1.3 in Wang (1991, page 106) we have that

$$\begin{aligned} \sup_{\omega\in P_T(b)}\left|T^{-1}\Delta_0^T(\omega-\omega_0)\right| &= \sup_{\omega\in P_T(b)}\left||T^{-1}\sum_{t=1}^{T}w(\frac{t}{T})\exp\{iT(\omega-\omega_0)\frac{t}{T}\}\right| \\ &= \sup_{\omega\in P_T(b)}\left|\int_0^1 w(s)\exp\{iT(\omega-\omega_0)s\}\,ds\right|+o(1) \end{aligned}$$

Let ω^* be such that

$$\left|\int_0^1 w(s)\exp\{iT(\omega^*-\omega_0)s\}\,ds\right| = \sup_{\omega\in P_T(b)}\left|\int_0^1 w(s)\exp\{iT(\omega-\omega_0)s\}\,ds\right| \tag{3.17}$$

Let $b^* = T|\omega^*-\omega_0| \geq b > 0$. Then, by the definition of $P_T(b)$ given by equation (3.16), we have

$$\sup_{\omega\in P_T(b)}\left|\int_0^1 w(s)\exp\{iT(\omega-\omega_0)s\}\,ds\right| \leq \left|\int_0^1 w(s)\exp\{ib^*s\}\,ds\right| + o(1) \tag{3.18}$$

Thus

$$\lim_{T\to\infty}\Pr\left(T|\hat{\omega}_T-\omega_0|\geq b\right) =$$

$$\lim_{T\to\infty}\Pr\left(\sup_{\omega\in P_T(b)}\left|T^{-1}\Delta_0^T(\omega-\omega_0)\right| \geq W_0+o_p(1)\right) = \tag{3.19}$$

$$\lim_{T\to\infty}\Pr\left(\left|\int_0^1 w(s)\exp\{ib^*s\}\,ds\right| + o(1) \geq W_0+o_p(1)\right) = 0 \tag{3.20}$$

Since $W_0 > 0$ is a deterministic constant and $b^* > 0$

$$W_0 = \left|\int_0^1 w(s)\,ds\right| = \int_0^1 |w(s)\exp\{ib^*s\}|\,ds > \left|\int_0^1 w(s)\exp\{ib^*s\}\,ds\right| \tag{3.21}$$

we have (3.15).

Now we will prove consistency for $\hat{A}_T$ and $\hat{B}_T$. We have that

$$\hat{A}_T + i\hat{B}_T = 2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)\exp\{-i\hat{\omega}_T t\}$$

where

$$r(t,\beta) = \left[D_0\exp\{i\omega_0 t\} + \overline{D}_0\exp\{-i\omega_0 t\}\right]$$

and

$$W_0^T = \sum_{t=1}^{T} w(\frac{t}{T})$$

By the mean value theorem we have that for some $\tilde{\omega}_T$ satisfying $|\tilde{\omega}_T-\omega_0| \leq |\hat{\omega}_T-\omega_0|$

$$\begin{aligned}
&|\hat{A}_T - A_0 + i(\hat{B}_T - B_0)| = \\
&\left|2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)\left[\exp\{i\omega_0 t\} - it\exp\{i\tilde{\omega}_T t\}(\hat{\omega}_T-\omega_0)\right] - (A_0+iB_0)\right| \\
&\leq \left|2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)\exp\{i\omega_0 t\} - (A_0+iB_0)\right| \\
&+ \left|2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)it\exp\{i\tilde{\omega}_T t\}(\hat{\omega}_T-\omega_0)\right|
\end{aligned} \tag{3.22}$$

Looking at the first term of equation (3.22) and using Lemma 1 we have

$$\begin{aligned}
&2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)\exp\{i\omega_0 t\} \\
&\quad = 2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})\left[D_0\exp\{i\omega_0 t\}+\overline{D}_0\exp\{-i\omega_0 t\}\right]\exp\{i\omega_0 t\} \\
&\quad = 2(W_0^T)^{-1}\left[D_0\Delta_0^T(2\omega_0)+\overline{D}_0\Delta_0^T(0)\right] \\
&\quad = 2(W_0^T)^{-1}\frac{1}{2}(A_0+iB_0)\Delta_0^T(0)+o(1) \\
&\quad = (A_0+iB_0)+o(1)
\end{aligned}$$

For the second term of equation (3.22)

$$\begin{aligned}
&\left|2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})r(t;\beta)it\exp\{i\tilde{\omega}_T t\}(\hat{\omega}_T-\omega_0)\right| \\
&\quad \le 2(W_0^T)^{-1}\sum_{t=1}^{T}\left|w(\frac{t}{T})r(t;\beta)it\exp\{i\tilde{\omega}_T t\}(\hat{\omega}_T-\omega_0)\right| \\
&\quad \le 2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})|r(t;\beta)|t|\hat{\omega}_T-\omega_0| \\
&\quad \le 2(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T})[|D_0|+|\overline{D}_0|]T|\hat{\omega}_T-\omega_0| \\
&\quad \le \rho_0 T|\hat{\omega}_T-\omega_0|\,(W_0^T)^{-1}\sum_{t=1}^{T} w(\frac{t}{T}) \\
&\quad = \rho_0\, T|\hat{\omega}_T-\omega_0| = o_p(1)
\end{aligned}$$

And thus $|(\hat{A}_T-A_0)+i(\hat{B}_T-B_0)|=o_p(1)$, and because both the real and imaginary parts converge in probability to 0 the Theorem is proved. □

Theorem 2 *Under the same conditions as in Theorem 1 the vector of weighted estimates:*

$$\left(T^{1/2}(\hat{A}_T-A_0), T^{1/2}(\hat{B}_T-B_0), T^{\frac{3}{2}}(\hat{\omega}_T-\omega_0)\right) \tag{3.23}$$

converges in distribution to a normal vector with zero mean and variance matrix

$$\frac{4\pi f_{\epsilon\epsilon}(\omega_0)}{(A_0^2+B_0^2)}\mathbf{V} \tag{3.24}$$

where

$$\mathbf{V} = \begin{pmatrix} c_1 A_0^2 + c_2 B_0^2 & -c_3 A_0 B_0 & -c_4 B_0 \\ -c_3 A_0 B_0 & c_2 A_0^2 + c_1 B_0^2 & c_4 A_0 \\ -c_4 B_0 & c_4 A_0 & c_0 \end{pmatrix} \tag{3.25}$$

Here

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= U_0 W_0^{-2} \\ c_2 &= a_0 b_1 \\ c_3 &= a_0 W_1 W_0^{-2} (W_0^2 W_1 U_2 - W_1^3 U_0 - 2W_0^2 W_2 U_1 + 2W_0 W_1 W_2 U_0) \\ c_4 &= a_0 (W_0 W_1 U_2 - W_1^2 U_1 - W_0 W_2 U_1 + W_1 W_2 U_0) \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} a_0 &= (W_0 W_2 - W_1^2)^{-2} \\ a_1 &= (U_0 U_2 - U_1^2) \\ a_2 &= W_0^{-2} (W_0 U_1 - W_1 U_0)^2 \\ b_n &= W_n^2 U_2 + W_{n+1} (W_{n+1} U_0 - 2W_n U_1) \ , \ n = 0, 1 \end{aligned} \tag{3.27}$$

Here W_0, W_1, W_2, U_0, U_1 and U_2 are defined by (3.8).

Proof: If we let $\mathbf{X}_t = (\epsilon_t, \epsilon_t)$ we have that each component of $\mathbf{X}_t$ satisfies Assumption 1. Also, as mentioned above, the functions $h_1(s) = w(s)$ and $h_2(s) = s\, w(s)$ both satisfy Assumption 2. The assumptions for Theorem 4.4.2 in (Brillinger 1981, page 95) are then satisfied for the series $\mathbf{X}_t$ and the taper functions $h_1(t)$ and $h_2(t)$. Thus, we have that the vector $\mathbf{u}$, with components equal to the real and imaginary parts of the Fourier transform of the components of $\mathbf{X}_t$

$$\begin{aligned} u_1 &= T^{-\frac{1}{2}} \sum h_1(\frac{t}{T}) X_{1,t} \cos \omega_0 t = T^{-\frac{1}{2}} \sum w(\frac{t}{T}) \epsilon_t \cos \omega_0 t \\ u_2 &= T^{-\frac{1}{2}} \sum h_1(\frac{t}{T}) X_{1,t} \sin \omega_0 t = T^{-\frac{1}{2}} \sum w(\frac{t}{T}) \epsilon_t \sin \omega_0 t \\ u_3 &= T^{-\frac{1}{2}} \sum h_2(\frac{t}{T}) X_{2,t} \cos \omega_0 t = T^{-\frac{3}{2}} \sum w(\frac{t}{T}) \epsilon_t \, t \cos \omega_0 t \\ u_4 &= T^{-\frac{1}{2}} \sum h_2(\frac{t}{T}) X_{2,t} \sin \omega_0 t = T^{-\frac{3}{2}} \sum w(\frac{t}{T}) \epsilon_t \, t \sin \omega_0 t \end{aligned} \tag{3.28}$$

is asymptotically multivariate normal with zero mean and variance matrix

$$\mathbf{U} = \pi f_{\epsilon\epsilon}(\omega_0) \begin{pmatrix} U_0 & 0 & U_1 & 0 \\ 0 & U_0 & 0 & U_1 \\ U_1 & 0 & U_2 & 0 \\ 0 & U_1 & 0 & U_2 \end{pmatrix} \tag{3.29}$$

Expanding $q'_T(\omega)$ in the first two terms of its Taylor series, about ω_0 we can write:

$$T^{-\frac{1}{2}} q'_T(\omega_0) = -T^{\frac{3}{2}}(\hat{\omega}_T - \omega_0) T^{-2} q''_T(\tilde{\omega}_T), \quad |\tilde{\omega}_T - \omega_0| \leq |\hat{\omega}_T - \omega_0| \tag{3.30}$$

Notice that, calculating, the derivative

$$\begin{aligned} q'_T(\omega_0) = T^{-\frac{1}{2}} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \sin(\omega_0 t) \right) T^{-\frac{3}{2}} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \cos(\omega_0 t) \right) - \\ T^{-\frac{1}{2}} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \cos(\omega_0 t) \right) T^{-\frac{3}{2}} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \sin(\omega_0 t) \right) \end{aligned} \tag{3.31}$$

Using Lemmas 1 and 2 we can show that

$$\begin{aligned} T^{-\frac{1}{2}} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \cos(\omega_0 t) \right) &= T^{-\frac{1}{2}} \left(A_0 \sum_{t=1}^{T} w(\frac{t}{T}) \cos^2(\omega_0 t) \right. \\ &\quad \left. + B_0 \sum_{t=1}^{T} w(\frac{t}{T}) \sin(\omega_0 t) \cos(\omega_0 t) + \sum_{t=1}^{T} w(\frac{t}{T}) \epsilon_t \cos(\omega_0 t) \right) \\ &= T^{\frac{1}{2}} A_0 W_0 + u_1 + o\left(T^{-\frac{1}{2}}\right) \end{aligned}$$

as $y_t = A_0 \cos(\omega_0 t) + B_0 \sin(\omega_0 t) + \epsilon_t$. By calculating an expression like this for each term in (3.31) and multiplying out we find that

$$T^{-\frac{1}{2}} q'_T(\omega_0) = -W_1 B_0 u_1 + W_1 A_0 u_2 + W_0 B_0 u_3 - W_0 A_0 u_4 + o_p(1) \tag{3.32}$$

Taking the second derivative we have

$$\begin{aligned} T^{-2} q''_T(\tilde{\omega}) = \Bigg[& -T^{-1} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \sin(\tilde{\omega} t) \right) T^{-3} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t^2 \sin(\tilde{\omega} t) \right) \\ & + T^{-2} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \cos(\tilde{\omega} t) \right) T^{-2} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \cos(\tilde{\omega} t) \right) \\ & + T^{-2} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \sin(\tilde{\omega} t) \right) T^{-2} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t \sin(\tilde{\omega} t) \right) \\ & - T^{-1} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \cos(\tilde{\omega} t) \right) T^{-3} \left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \, t^2 \sin(\tilde{\omega} t) \right) \Bigg] \end{aligned} \tag{3.33}$$

Now

$$T^{-1}\left(\sum_{t=1}^{T} w(\frac{t}{T}) y_t \cos(\tilde{\omega}_T t)\right) = T^{-1}\left(A_0 \sum_{t=1}^{T} w(\frac{t}{T}) \cos(\omega_0 t)\cos(\tilde{\omega}_T t) + B_0 \sum_{t=1}^{T} w(\frac{t}{T}) \sin(\omega_0 t)\cos(\tilde{\omega}_T t) + \sum_{t=1}^{T} w(\frac{t}{T}) \epsilon_t \cos(\tilde{\omega}_T t)\right) \tag{3.34}$$

as $y_t = A_0 \cos(\omega_0 t) + B_0 \sin(\omega_0 t) + \epsilon_t$.

Since $T|\tilde{\omega}_T - \omega_0|$ converges to zero in probability, we have that the first term on the left of equation (3.34) converges to $A_0 W_0$ in probability and the second term to 0. The third term is dominated by

$$\sup_{0 \le \omega \le \pi} \left| T^{-1} \sum_{t=1}^{T} w(\frac{t}{T}) \epsilon_t \exp\{-it\omega\} \right|$$

which by Lemma 2 goes to 0 in probability. Similarly we find the limit in probability of each of the eight terms in (3.33). Multiplying out we have

$$T^{-2} q_T''(\tilde{\omega}_T) = \frac{1}{2}(A_0^2 + B_0^2)(W_1^2 - W_0 W_2) + o_p(1) \tag{3.35}$$

Using (3.30), (3.32) and (3.35) we have

$$\begin{aligned}
T^{\frac{3}{2}}(\hat{\omega}_T - \omega_0) &= \frac{2W_1 B_0 u_1 - 2W_1 A_0 u_2 - 2W_0 B_0 u_3 + 2W_0 A_0 u_4}{(A_0^2 + B_0^2)(W_1^2 - W_0 W_2)} + o_p(1) \\
T^{\frac{1}{2}}(\hat{A}_T - A_0) &= \frac{2}{W_0} u_1 - \frac{W_1}{W_0} B_0 T^{\frac{3}{2}}(\hat{\omega}_T - \omega_0) + o_p(1) \\
T^{\frac{1}{2}}(\hat{B}_T - B_0) &= \frac{2}{W_0} u_2 + \frac{W_1}{W_0} A_0 T^{\frac{3}{2}}(\hat{\omega}_T - \omega_0) + o_p(1)
\end{aligned}$$

By Assumption 2 we know that all the denominators are not 0, so now we can express the vector of standardized estimates as a linear combination of the vector $\mathbf{u}$, defined by equation (3.28), plus a quantity converging to 0 in probability.

$$\mathbf{z}_T = \mathbf{A}\mathbf{u} + \mathbf{o}_p(1)$$

with

$$\mathbf{z}_T = \begin{pmatrix} T^{\frac{1}{2}}(\hat{A}_T - A_0) \\ T^{\frac{1}{2}}(\hat{B}_T - B_0) \\ T^{\frac{3}{2}}(\hat{\omega}_T - \omega_0) \end{pmatrix}$$

and

$$\mathbf{A} = \begin{pmatrix} B_0^2 W_2 + A_0^2 (W_2 - \frac{W1^2}{W_0}) & -\frac{A_0 B_0 W_1^2}{W_0} & -B_0^2 W_1 & A_0 B_0 W_1 \\ -\frac{A_0 B_0 W_1^2}{W_0} & A_0^2 W_2 + B_0^2 (W_2 - \frac{W_1^2}{W_0}) & A_0 B_0 W_1 & -A_0^2 W_1 \\ -B_0 W_1 & A_0 W_1 & B_0 W_0 & -A_0 W_0 \end{pmatrix}$$

This implies that $\mathbf{Au}$ is asymptotically multivariate normal with variance matrix $\mathbf{AUA'}$. Specifically by Slutsky's theorem $\mathbf{z}$ has the same asymptotic distribution as $\mathbf{Au}$. By computing $\mathbf{AUA'}$ we obtain $\{4\pi f_{\epsilon\epsilon}(\omega_0)/(A_0^2 + B_0^2)\}\,\mathbf{V}$ and this completes the proof. □

Notice that the constants presented in equation (3.26) and (3.27) are quite complicated. However for certain window functions we have that

$$W_0 U_1 - W_1 U_0 = 0 \tag{3.36}$$

The uniform window function, $w(t) = 1$, and the Tukey triweight window function, $w(t) = (1 - |2t-1|^3)_+^3$, are examples where (3.36) holds. If this is the case we obtain the following simplifications for the constants

$$\begin{aligned} c_3 &= a_0 W_1^2 W_0^{-2} (W_1^2 U_0 - W_0^2 U_2) \\ c_4 &= a_0 W_1 (W_0 U_2 - W_1 U_1) \\ a_2 &= 0 \\ b_0 &= W_0^2 U_2 - W_0 W_1 U_1 \end{aligned}$$

Also observe that if $w(t) = 1$ for all t, the constants in (3.26) reduce to $c_1 = 1$, $c_2 = 4$, $c_3 = 3$, $c_4 = 6$ and $c_0 = 12$ and the variance matrix reduces to the variance matrix obtained in the equally weighted case by, for example, Walker (1971).

3.3 Several sinusoidal components

Now consider the model with more than one frequency

$$s(t;\beta_0) = \sum_{k=1}^{K} \{A_{k,0}\cos(\omega_{k,0}t) + B_{k,0}\sin(\omega_{k,0}t)\} \tag{3.37}$$

with $\beta_0 = (\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{\omega}_0)$ and $0 < w_{k,0} \neq w_{l,0} < 2\pi$ for all $1 \leq k \neq l \leq K$. The function corresponding to (3.12) whose minimization yields approximate weighted least squares

estimators becomes

$$R_T(\beta) = \sum_{t=1}^{T} w(\frac{t}{T}) y_t^2 + \frac{1}{2}\sum_{k=1}^{K}(A_k^2 + B_k^2)\sum_{t=1}^{T} w(\frac{t}{T}) \tag{3.38}$$
$$-2\sum_{k=1}^{K}\sum_{t=1}^{T} w(\frac{t}{T}) y_t\{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\}$$

Here $\beta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\omega})$. Similar to the case of one sinusoidal component we define the estimates $\hat{A}_{k,T}$, $\hat{B}_{k,T}$ and $\hat{\omega}_{k,T}$ for $k = 1, \ldots, K$ by

$$\hat{A}_{k,T} = 2\sum_{t=1}^{T} w(\frac{t}{T}) y_t \cos(\hat{\omega}_{k,T}\, t) / \sum_{t=1}^{T} w(\frac{t}{T}) \tag{3.39}$$

$$\hat{B}_{k,T} = 2\sum_{t=1}^{T} w(\frac{t}{T}) y_t \sin(\hat{\omega}_{k,T}\, t) / \sum_{t=1}^{T} w(\frac{t}{T}) \tag{3.40}$$

where if we write $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$ and $\hat{\boldsymbol{\omega}}_T = (\hat{\omega}_{1,T}, \ldots, \hat{\omega}_{K,T})$, $\hat{\boldsymbol{\omega}}_T$ is such that

$$q_T(\hat{\boldsymbol{\omega}}) = \max_{0 \le \boldsymbol{\omega} \le \pi} q_T(\boldsymbol{\omega}) \tag{3.41}$$

where q_T is defined by:

$$q_T(\boldsymbol{\omega}) = \sum_{k=1}^{K} \left| T^{-1} \sum_{t=1}^{T} w(\frac{t}{T}) y_t \exp\{it\omega_k\} \right|^2 \tag{3.42}$$

In this case to obtain (3.38) from the weighted least squares equation (3.7) we need to have that terms of the form $A_k A_l \sum_t \cos(\omega_k)\cos(\omega_l)$ and $B_k B_l \sum_t \sin(\omega_k)\sin(\omega_l)$ are bounded, since they are included in $S_T(\mathbf{A}, \mathbf{B}, \boldsymbol{\omega}) - R_T(\mathbf{A}, \mathbf{B}, \boldsymbol{\omega})$. Some conditions need to be imposed to avoid having the ω_k become too close together and thus prevent the estimators of two or more frequencies from converging in probability to the same value. Notice that if no constraint is imposed, the estimate for all the frequencies $w_{1,0}, \ldots, w_{K,0}$ will be the frequency that maximizes $q_T(\omega)$ of equation (3.14) in the previous section. An appropriate condition is

$$\lim_{T \to \infty} \min_{1 \le k \ne l \le K} (T|\omega_k - \omega_l|) = \infty \tag{3.43}$$

Walker (1971) proposes maximizing $q_T(\boldsymbol{\omega})$ subject to

$$\min_{k \ne l}(|\omega_k - \omega_l|) = T^{-\frac{1}{2}} \tag{3.44}$$

So we redefine the estimates of $\boldsymbol{\omega}_0$ as the value that maximizes

$$q_T(\hat{\boldsymbol{\omega}}) = \max_{0 \le \boldsymbol{\omega} \le \pi} q_T(\boldsymbol{\omega})$$

but under the constraint (3.44).

The following results for the weighted least squares estimates follow from Theorems 1 and 2.

Corollary 1 *Under the same assumptions as Theorem 1*

$$\lim_{T\to\infty} \hat{A}_{k,T} = A_{k,0},\ \lim_{T\to\infty} \hat{B}_{k,T} = B_{k,0},\ \lim_{T\to\infty} T|\hat{\omega}_{k,T} - \omega_{k,0}| = 0$$

in probability for each k

Proof: When (3.43) holds, then only K of the K^2 differences $\omega_k - \omega_{l,0}$ can be $O(T^{-1})$. If we label these differences $\omega_k - \omega_{k,0}$, then we have

$$\begin{aligned} q_T(\boldsymbol{\omega}) &= \sum_{k=1}^{K} \left| T^{-1} \sum_{l=1}^{K} w(\frac{t}{T}) \{ D_{l,0} \Delta_0^T(\omega_k + \omega_{l,0}) + \overline{D}_{l,0} \Delta_0^T(\omega_k - \omega_{l,0}) \} \right. \\ &+ \left. \sum_{t=1}^{T} w(\frac{t}{T}) \epsilon_t \exp\{i\omega_k t\} \right|^2 \end{aligned}$$

where $D_{k,0} = \frac{1}{2}(A_{k,0} - iB_{k,0})$ for $k = 1, \ldots, K$. Notice that $q_T(\boldsymbol{\omega})$ will be dominated by the sum of the terms $|\overline{D}_{k,0} \Delta_0^T(\omega_k - \omega_{k,0})\}|$ when the $\omega_k - \omega_{k,0}$, $k = 1, \ldots, K$ are small. In fact, we can show, just as in the proof of Theorem 1, that

$$q_T(\boldsymbol{\omega}) = \frac{1}{4} \sum_{k=1}^{K} \rho_k^2 \left| T^{-1} \Delta_0^T(\omega_{k,0} - \omega_k) \right|^2 + o_p(1)$$

and

$$q_T(\boldsymbol{\omega}_0) = \frac{1}{4} W_0^2 \sum_{k=1}^{K} \rho_k^2 + o_p(1)$$

For any $b > 0$, define

$$P_T(b) = \{ \boldsymbol{\omega} : T|\omega_k - \omega_{k,0}| \geq b, (1 \leq k \leq K) \}$$

Following in a similar way to that of the proof of Theorem 1, for some $b^* > 0$

$$\begin{aligned} &\lim_{T\to\infty} \Pr\left(T|\hat{\boldsymbol{\omega}}_T - \boldsymbol{\omega}_0| \geq b \right) = \\ &\lim_{T\to\infty} \Pr\left(\sup_{\boldsymbol{\omega} \in P_T(b)} \sum_{k=1}^{K} \rho_k^2 |T^{-1} \Delta_0^T(\omega_k - \omega_{k,0})| \geq W_0 \sum_{k=1}^{K} \rho_k^2 + o_p(1) \right) = \\ &\lim_{T\to\infty} \Pr\left(\sum_{k=1}^{K} \rho_k^2 \left| \int_0^1 w(s) \exp\{ib^* s\}\, ds \right| + o(1) \geq W_0 \sum_{k=1}^{K} \rho_k^2 + o_p(1) \right) = \\ &\lim_{T\to\infty} \Pr\left(\left| \int_0^1 w(s) \exp\{ib^* s\}\, ds \right| + o(1) \geq W_0 + o_p(1) \right) = 0 \end{aligned}$$

and the T-consistency of the frequency estimates $\hat{\boldsymbol{\omega}}$ is proved. Having established this, the consistency of each estimate $\hat{A}_{k,T}$ and $\hat{B}_{k,T}$, $k = 1, \ldots, K$ follows in the same way as the proof of Theorem 1. □

Corollary 2 *Under the same assumptions of Theorem 2 the vectors*

$$\left(T^{\frac{1}{2}}(\hat{A}_{k,T} - A_{k,0}), T^{\frac{1}{2}}(\hat{B}_{k,T} - B_{k,0}), T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_{k,0}) \right) \ , \ k = 1, \ldots, K \tag{3.45}$$

converge in distribution to mutually independent normal vectors with zero mean and variance matrices

$$\frac{4\pi f_{\epsilon\epsilon}(\omega_{k,0})}{A_{k,0}^2 + B_{k,0}^2} \mathbf{V}_k$$

where

$$\mathbf{V}_k = \begin{bmatrix} c_1 A_{k,0}^2 + c_2 B_{k,0}^2 & -c_3 A_{k,0} B_{k,0} & -c_4 B_{k,0} \\ -c_3 A_{k,0} B_{k,0} & c_2 A_{k,0}^2 + c_1 B_{k,0}^2 & c_4 A_{k,0} \\ -c_4 B_{k,0} & c_4 A_{k,0} & c_0 \end{bmatrix} \tag{3.46}$$

and the constants $c_0, \ldots, c_4$ are defined in equation (3.26)

Proof: By taking derivatives of $q_T(\boldsymbol{\omega})$ we notice that $\partial q_T(\boldsymbol{\omega})/\partial \omega_k$ doesn't depend on ω_l when $l \neq k$. We can then define, for each $k = 1, \ldots, K$

$$q'_{k,T}(\omega_k) = \left. \frac{\partial q_T(\boldsymbol{\omega}^*)}{\partial \omega_k} \right|_{\omega_k^* = \omega_k}$$

Then for each $k = 1, \ldots, K$ there exists a $\tilde{\omega}_{k,T}$ with $|\tilde{\omega}_{k,T} - \omega_{k,0}| \leq |\hat{\omega}_{k,T} - \omega_{k,0}|$ such that

$$T^{-\frac{1}{2}} q'_{k,T}(\omega_0) = -T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_{k,0}) T^{-2} q''_T(\tilde{\omega}_{k,T})$$

Now for each one of these we can directly compute the derivatives and, following the proof of Theorem 2, find that

$$T^{-\frac{1}{2}} q'_{k,T}(\omega_{k,0}) = -W_1 B_{k,0} u_{k,1} + W_1 A_{k,0} u_{k,2} + W_0 B_{k,0} u_{k,3} - W_0 A_{k,0} u_{k,4} + o_p(1)$$

and

$$T^{-2} q''_{k,T}(\tilde{\omega}_{k,T}) = \frac{1}{2}(A_{k,0}^2 + B_{k,0}^2)(W_1^2 - W_0 W_2) + o_p(1)$$

Here $u_{k,1}, \ldots, u_{k,4}$ are defined as in equation (3.28) but with using $\omega_{k,0}$ in place of ω_0. It is known, see for example Brillinger (1981), that the vectors $(u_{k,1}, \ldots, u_{k,4})'$ are asymptotically independent and that under condition (3.43) so are the $\hat{\omega}_k$'s. The result is then obtained in the same way as the proof of Theorem 2. □

3.4 Harmonic model

In the case of a "harmonic" musical instrument a more appropriate model for the sound it produces contains the constraint that all the component frequencies are a multiple of a fundamental frequency, a pitch. We will call this a *harmonic* model. Now the model comes down to

$$s(t;\beta) = \sum_{k=1}^{K} \{A_k \cos(k\lambda t) + B_k \sin(k\lambda t)\} \tag{3.47}$$

with $\beta = (\mathbf{A}, \mathbf{B}, \lambda)'$, λ the fundamental frequency or pitch.

For this model Brown (1990) finds the asymptotic distribution for the unweighted least squares estimates by redoing the whole problem under the new model. We will estimate using a technique similar to the one used by Brillinger (1980) to estimate a bifrequency and show that the same result is obtained in a simpler manner. We will also consider the weighted case.

The estimation procedure is the following. Define

$$\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K) = (\lambda, 2\lambda, \ldots, K\lambda)$$

Then, as for the estimation under model (3.37), we find $(\hat{\mathbf{A}}_T, \hat{\mathbf{B}}_T, \hat{\boldsymbol{\omega}}_T)$ from equations (3.39), (3.40), (3.41) and (3.42).

Notice that from Corollary 2 we know that for each $k = 1, \ldots, K$, $\hat{\omega}_{k,T}$ is asymptotically normal with mean $k\lambda$ and variance

$$4\pi c_0 \frac{f_{\epsilon\epsilon}(\omega_k)}{A_k^2 + B_k^2} \tag{3.48}$$

Furthermore, the $\hat{\omega}_{k,T}$'s are asymptotically mutually independent.

At the next step we can estimate λ via the the following regression model:

$$\hat{\omega}_k = k\lambda + \delta_k \quad k = 1, ..., K$$

where the δ_ks are independent errors with mean 0 and variance defined by (3.48). Obtaining the weighted regression equation estimates leads to:

$$\hat{\lambda}_T = \frac{\sum_{k=1}^{K} k\, \hat{\omega}_{k,T} (\hat{A}_{k,T}^2 + \hat{B}_{k,T}^2)/f_{\epsilon\epsilon}(\omega_k)}{\sum_{k=1}^{K} k^2 (\hat{A}_{k,T}^2 + \hat{B}_{k,T}^2)/f_{\epsilon\epsilon}(\omega_k)} \tag{3.49}$$

In practice we will be estimating assuming white noise, in which case this equation reduces to

$$\hat{\lambda}_T = \frac{\sum_{k=1}^{K} k\, \hat{\omega}_k (\hat{A}_{k,T}^2 + \hat{B}_{k,T}^2)}{\sum_{k=1}^{K} k^2 (\hat{A}_{k,T}^2 + \hat{B}_{k,T}^2)}$$

From equation (3.49), Corollary 1, and the fact that K is finite, it is apparent that

$$\lim_{T\to\infty} T|\hat{\lambda}_T - \lambda| = 0, \text{ in probability}$$

Theorem 3 *Under the same assumptions as for theorem 2, the vector*

$$\left(T^{\frac{1}{2}}(\hat{A}_{1,T} - A_1), T^{\frac{1}{2}}(\hat{B}_{1,T} - B_1), \ldots, T^{\frac{1}{2}}(\hat{A}_{K,T} - A_K), T^{\frac{1}{2}}(\hat{B}_{K,T} - B_K), T^{\frac{3}{2}}(\hat{\lambda}_T - \lambda)\right)'$$

converges in distribution to a multivariate normal with zero mean and variance matrix

$$\frac{4\pi}{\sum_k k^2(A_k^2 + B_k^2)/f_{\epsilon\epsilon}(k\lambda)} \begin{pmatrix} \mathbf{D} + c_0^{-1}\mathbf{E}\mathbf{E}' & \mathbf{E} \\ \mathbf{E}' & c_0 \end{pmatrix} \tag{3.50}$$

where the matrices appearing above are defined as follows:
$\mathbf{D}$ *is a* $2K \times 2K$ *matrix with entries*

$$\mathbf{D} = \left(\sum_k k^2(A_k^2 + B_k^2)/f_{\epsilon\epsilon}(k\lambda)\right) \begin{pmatrix} \mathbf{D_1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{D_K} \end{pmatrix} \tag{3.51}$$

with

$$\mathbf{D_k} = \frac{f_{\epsilon\epsilon}(k\lambda)}{b_0(A_k^2 + B_k^2)} \begin{pmatrix} c_1 b_0 A_k^2 + a_1 B_k^2 & a_2 A_k B_k \\ a_2 A_k B_k & a_1 A_k^2 + c_1 b_0 B_k^2 \end{pmatrix} \tag{3.52}$$

and

$$\mathbf{E} = c_4\,(-B_1, A_1, -2B_2, 2A_2, \ldots, -KB_K, KA_K)' \tag{3.53}$$

Proof: Let

$$\mathbf{Y} = \left(\hat{A}_{1,T}, \hat{B}_{1,T}, \ldots, \hat{A}_{K,T}, \hat{B}_{K,T}, \hat{\omega}_1, \ldots, \hat{\omega}_T\right)'$$

include the estimates found under the general model (3.37). From Corollary 2 we know the asymptotic variance matrix of these estimates. To find the variance matrix under the constraints we now use the fact that we find the estimates for model (3.47) by employing the regression model

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\delta}$$

where the matrices above are defined by:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \dots & 0 & 1\mathbf{X}_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \mathbf{X}_1 & K\mathbf{X}_2 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} A_1 \\ B_1 \\ \vdots \\ A_K \\ B_K \\ \lambda \end{pmatrix}$$

where δ has mean 0 and variance matrix $\mathbf{V}$ as in in Corollary 2. The matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are defined by

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } \mathbf{X}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Asymptotic normality follows from the fact that the estimates obtained from the regression are linear combinations of the estimates known to be jointly asymptotically normal from Corollary 2. To find the variance matrix we apply weighted regression and see that the new estimates have variance matrix equal to $(\mathbf{X'V^{-1}X})^{-1}$. Computing $\mathbf{C} = \mathbf{X'V^{-1}X}$ we obtain a matrix of the form

$$\begin{pmatrix} \mathbf{C_1} & \mathbf{C_2} \\ \mathbf{C_2'} & \mathbf{C_3} \end{pmatrix}$$

with $\mathbf{C_1}, \mathbf{C_3}$ symmetric matrices. Thus we can use the result, presented for example in Rao (1973, page 33), that states that when $\mathbf{C_1}, \mathbf{C_3}$ are symmetric matrices such that the inverse which occur in the expression below exist, we have that

$$\begin{pmatrix} \mathbf{C_1} & \mathbf{C_2} \\ \mathbf{C_2'} & \mathbf{C_3} \end{pmatrix} = \begin{pmatrix} \mathbf{C_1^{-1}} + \mathbf{FE^{-1}F'} & \mathbf{-FE^{-1}} \\ \mathbf{-E^{-1}F'} & \mathbf{E^{-1}} \end{pmatrix}$$

where $\mathbf{E} = \mathbf{C_3} - \mathbf{C_2'C_1^{-1}C_2}, \mathbf{F} = \mathbf{C_1^{-1}C_2}$. Using this we can directly compute $(\mathbf{X'V^{-1}X})^{-1}$ to obtain the desired result. □

3.4.1 Variance of the amplitudes

In some instances it might be useful to find estimates for the amplitudes of the harmonic components, the amplitudes being defined by

$$\rho_k = \sqrt{A_k^2 + B_k^2}$$

To estimate the ρ_k's we may use the estimates presented above and set

$$\hat{\rho}_k = \sqrt{\hat{A}_k^2 + \hat{B}_k^2}$$

That $\rho(A,B) = \sqrt{A^2+B^2}$ is a continuous function of (A,B) and 6.9.2.i of Rao (1973, page 124) imply that the $\hat{\rho}_k$'s are consistent. The linear approximation using Taylor expansions

$$\hat{\rho}_k \approx \rho_k + \frac{A_k(\hat{A}_k - A_k) + B_k(\hat{B}_k - B_k)}{\sqrt{A_k^2 + B_K^2}}$$

gives the variance of the asymptotic distribution as

$$\mathrm{Var}[\hat{\rho}_k] \approx \frac{A_k^2 \mathrm{Var}[\hat{A}_k] + B_k^2 \mathrm{Var}[\hat{B}_k] + 2A_k B_k \mathrm{Cov}[\hat{A}_k, \hat{B}_k]}{T(A_k^2 + B_k^2)} \tag{3.54}$$

Now, using the results of Theorem 3, we reduce this equation to

$$\mathrm{Var}[\hat{\rho}_k] \approx 4\pi c_1 f_{\epsilon\epsilon}(k\lambda)/T \tag{3.55}$$

where c_1 is as in the equations (3.26). In particular, notice that the variance of $\hat{\rho}_k$ is proportional to $f_{\epsilon\epsilon}(k\lambda)$.

3.4.2 Estimate of the spectrum

Notice that the asymptotic variance depends on the value of the spectrum, $f_{\epsilon\epsilon}(k\lambda)$, at the harmonic frequencies, $k\lambda$ for $k = 1, \ldots, K$. If we are to use the asymptotic approximation based on (3.50) to find standard errors and confidence intervals for our estimates, we need to estimate $f_{\epsilon\epsilon}(k\lambda)$ for $k = 1, \ldots, K$. We could assume the noise is white with variance σ^2, in which case the spectrum is $f_{\epsilon\epsilon}(\lambda) = \sigma^2/2\pi$ for $0 \leq \lambda \leq 2\pi$. However, in the case of music signals, there is evidence to suggest that the spectrum is higher for lower frequency, as we will see in Chapter 6. If we estimate the spectrum with a constant we can anticipate obtaining underestimates for the standard errors of the amplitudes of the lower harmonics and overestimates for the standard errors of the higher harmonics. We can instead use a *smoothed periodogram* estimate of the spectrum.

Let

$$I^T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} \exp\{-i\lambda t\}\hat{\epsilon}_t\right|^2 \quad 0 \leq \lambda \leq \pi$$

Choose a set of weights $w_j, j = 0, \pm 1, \ldots, \pm m$ where

$$\sum_{j=-m}^{m} w_j = 1$$

The smoothed periodogram estimate with weights $w_j, j = 0, \pm 1, \ldots, \pm m$ is

$$\hat{f}_{\epsilon\epsilon}(\lambda) = \sum_{j=-m}^{m} w_j I^T(\lambda + 2\pi j/T) \tag{3.56}$$

Notice that different values of m produce different estimates. The larger m the "smoother" the estimate.

3.4.3 Advantage of the harmonic model

The model defined by (2.5) is commonly used in the sound analysis procedures (Rodet 1997). Notice that in this model the harmonic constraint $\omega_k = k\lambda$ is not used. If we assume the harmonic model is true, then for each k, both estimators $\hat{\omega}_k$ and $k\hat{\lambda}$ are consistent estimates of the frequency related to the k-th partial $k\lambda$. The advantage of the latter is that it has smaller asymptotic variance. Observe that

$$\frac{\mathrm{Var}(\hat{\omega}_k)}{\mathrm{Var}(k\hat{\lambda})} \approx 1 + \frac{\sum_{l \neq j} l^2 (A_l^2 + B_l^2)/f_{\epsilon\epsilon}(l\lambda)}{(A_k^2 + B_k^2)/f_{\epsilon\epsilon}(k\lambda)}$$

which for practical purposes can be quite different from 1. In table 3.1 we see estimates of $\mathrm{EFR}_k = \sqrt{\mathrm{Var}(\hat{\omega}_k)/\mathrm{Var}(k\hat{\lambda})}$ for the 15 partials fitted to the violin segment analyzed in section 6.1. In this particular case we obtain estimates of the frequencies that are approximately between 5 and 250 times more "accurate".

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
EFR_k	6.4	12	4.8	30	40	33	18	18	45	290	54	120	230	82	220

Table 3.1: Estimated ratios between the standard errors of $k\hat{\lambda}$ and $\hat{\omega}_k$.

3.5 More than one fundamental frequency

In some applications we might want to consider a model where more than one fundamental frequency is present. In this case a pertinent model is

$$s(t;\beta) = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \{A_{jk}\cos(k\lambda_j t) + B_{jk}\sin(k\lambda_j t)\} + \epsilon(t) \tag{3.57}$$

where K_j represents the number of partials associated with the j-th fundamental frequency. If we estimate following the procedure used in Theorem 3 we have the following two results.

Corollary 3 *Under the same assumptions as Theorem 1*

$$\lim_{T\to\infty} \hat{A}_{jk,T} = A_{jk}, \ \lim_{T\to\infty} \hat{B}_{jk,T} = B_{jk}, \ \lim_{T\to\infty} T|\hat{\lambda}_{j,T} - \lambda_j| = 0$$

in probability for each j, k.

Corollary 4 *Under the same assumptions of Theorem 2 the vectors with k-th entry:*

$$\left(T^{\frac{1}{2}}(\hat{A}_{jk,T} - A_{jk}), T^{\frac{1}{2}}(\hat{B}_{jk,T} - B_{jk}), T^{\frac{3}{2}}(\hat{\lambda}_{j,T} - \lambda_j)\right) \tag{3.58}$$

for $k = 1, \ldots, K_j$ *with* $j = 1, \ldots, J$ *converge in distribution to mutually independent normal vectors with zero mean and variance matrices as in Theorem 3.*

Chapter 4

The Local Harmonic Model

4.1 Introduction

In the case of sound signals, the harmonic structure changes in time as the performer changes the sound being produced by the instrument. Examples of changes in the sound are changes of note or pitch, vibrato, and tremolo, to mention a few. For this reason the stationary model mentioned in Chapter 3 is not appropriate. Instead a non-stationary version with parametric functions changing in time needs to be considered.

$$y(t) = s\left[t, \beta(t)\right] + \epsilon(t) \quad t \in [0, D] \tag{4.1}$$

Here the process $\{\epsilon(t)\}$ is non-stationary noise, $\beta(t)$ is a real valued vector of functional parameters and D is the duration in seconds of the signal. Throughout the work, without loss of generality, we will assume every signal to have a one time unit duration, $D = 1$.

The work will always concern a discrete (sampled) version of the signal $y(t)$.

$$Y_{n,N} = s\left[\frac{n}{N}, \beta(\frac{n}{N})\right] + \epsilon_{n,N} \quad n = 1, \ldots, N$$

Here n is time measured in units of $\Delta t = 1/N$ seconds, where N is the number of observations per second, called the *sampling rate* in the audio technology literature. Notice that as N gets bigger the signal $Y(t)$, of fixed duration, is observed on a finer grid.

In practice this approach appears appropriate since longer signals will not give us better understanding of the local behavior of the parameter function $\beta(t)$. On the other hand a finer grid will give us more points closer to the estimation point. The goal of this setup is to develop useful approximations.

4.2 The locally harmonic model

For the case of sound signals produced by harmonic instruments we propose using a time-varying model, similar to the one proposed by Serra (1989), shown in equation (2.4). Two differences are that we will impose the constraint that the frequency of the partials are all multiples of a fundamental frequency and that the stochastic non-sinusoidal part will be allowed to be *locally stationary*.

We are interested in obtaining useful estimates of the bias and variance of our estimates. To do this, we will need a model with "smooth" enough functional parameters so that we can assume they are constant within "small" estimation window. We will also need the fundamental frequency to be "large" within the estimation windows so that we have enough oscillation to permit meaningful estimation of the sinusoidal parameters. The model we set down and the asymptotic theory we present has that goal. For a given sampling rate N, we propose the following *locally harmonic model*

$$Y_{n,N} = s\left[\frac{n}{N}, \beta_N(\frac{n}{N})\right] + \epsilon_{n,N} \text{ for } n = 1, \ldots, N \tag{4.2}$$

Here

$$s\left[t; \beta_N(t)\right] = \sum_{k=1}^{K} \left\{A_k(t)\cos(k\lambda_N(t)t) + B_k(t)\sin(k\lambda_N(t)t)\right\} \quad t \in [0,1]$$

with

$$\beta_N(t) = (A_1(t), B_1(t), \ldots, A_K(t), B_K(t), \lambda_N(t))'$$

where, for each k, $A_k(t)$ and $B_k(t)$ are continuous bounded functions for $t \in [0,1]$. We assure that the fundamental frequency is large using the following assumption.

Assumption 3 *There exist a continuous function $\lambda(t)$ with $0 < \lambda(t) < 2\pi$ for $t \in [0,1]$, such that the sequence of functions $\lambda_N(t) - N\lambda(t)$ converges uniformly to 0 for $t \in [0,1]$.*

In practice we have that the sample rate is large and may be chosen by the engineer. There are many observations per unit time, and the *instantaneous fundamental frequency* $\lambda_N(t)$ is many cycles per unit time. Notice that if $\lambda_N(t) = \lambda_N$ is constant in time for each N then the number of cycles per unit time $N\lambda_N$ tends to infinity with N. The signal $s[t, \beta_N(t)]$ is different for each N. Therefore, we must not interpret the asymptotics as having a fixed signal from which we can obtain better estimates as we increase the sample rate N.

A more reasonable interpretation of the asymptotics is that as N increases we consider smaller estimation window sizes, so that the functional parameter is closer to

constant, and larger instantaneous fundamental frequency so that within the estimation window we have a model that approximates the harmonic model of Chapter 3 with many observations. Consider a small enough estimation window size h_N then we can act as if the functional parameter is constant in time

$$\lambda_N(t) = \lambda_N(t_0) \quad t \in (t_0 - h_N/2, t_0 + h_N/2)$$

Letting $J_N = \lfloor h_N N \rfloor$, we have that within the estimation window the signal is approximately

$$s(n, \beta_N) = \sum_{k=1}^{K} \{A_k \cos(k\lambda_N(t_0)n/J_N) + B_k \sin(k\lambda_N(t_0)n/J_N)\} \quad n = 1, \ldots J_N \tag{4.3}$$

If $J_N \to \infty$ as $N \to \infty$, then Assumption 4 suggests that

$$\frac{\lambda_N(t_0)}{J_N} \approx \lambda(t_0)$$

with $0 < \lambda(t_0) < 2\pi$. Letting $\lambda_0 = \lambda(t_0)$ an approximation for (4.3) is

$$s(n, \beta) = \sum_{k=1}^{K} \{A_k \cos(k\lambda_0 n) + B_k \sin(k\lambda_0 n) \quad n = 1, \ldots J_N$$

Within the estimation window we have a model that approximates the harmonic model of Chapter 3 with a large number of observations J_N. Therefore, we should be able to obtain reasonable estimates of β for large values of N. Notice in particular that we are not interested in obtaining a consistent estimate of $\lambda(t)$ of equation (4.1), but rather find an estimate such that $J_N|\hat{\lambda}_N(t) - \lambda(t)|$ is small for all $t \in [0, 1]$ when N large. In this chapter we will make this asymptotic theory precise, but first we need some definitions and assumptions regarding the functional parameters and the noise processes

$$\{\epsilon_{n,N}, n = 1, \ldots, N; N \geq 1\}$$

The functional parameter $\beta_N(t)$ is assumed to be *locally approximately constant.* To make this definition precise we say that the functional parameters satisfy the following assumption.

Assumption 4 *There exist an M such that for each k,*

$$\sup_{t\in[0,1]} |A_k'(t)|, \quad \sup_{t\in[0,1]} |B_k'(t)|, \quad \sup_{t\in[0,1]} |\lambda_N'(t)| \ \leq \ M$$

for all N.

These assumption will permit us to assume the functional parameters are *approximately constant* within small enough segments of the signal and then to use the estimation procedure shown in Chapter 3 for this segment to obtain consistent asymptotically normal estimates. Intuitively this assumption prevents the local behavior of the function $s[t;\beta_N(t)]$ from being too different from a sum of sinusoids and thereby preserving some sort of local harmonic structure known to be present in sound signals. We say that $s[t;\beta_N(t)]$ is *locally approximately sinusoidal.*

The statistical models presented in the technological literature usually assume stationary noise, see for example Serra (1989), Rodet (1997). This assumption does not seem to hold in general for sound signals. For example, intuitively we would expect the amplitude of the non-sinusoidal part of the signal produced by a blown instrument to depend on how hard the instrumentalist is blowing. In this work we assume instead that the sequence of stochastic processes $\{\epsilon_{n,N}, n = 1,\ldots,N; N \geq 1\}$ is *locally stationary* as defined by Dahlhaus (1997).

Definition 1 *A sequence of stochastic processes $\{\epsilon_{n,N}, n = 1,\ldots,N; N \geq 1\}$ is called locally stationary with transfer function A' if there exists a representation*

$$\epsilon_{n,N} = \int_{-\pi}^{\pi} \exp\{i\lambda n\} A'_{n,N}(\lambda)\, d\xi(\lambda)$$

where $\xi(\lambda)$ is stochastic processes on $[-\pi,\pi]$ with $\overline{\xi(\lambda)} = -\xi(\lambda)$ and the cumulants satisfying

$$cum\{d\xi(\lambda_1),\ldots,d\xi(\lambda_k)\} = \eta\left(\sum_{j=1}^{k}\lambda_j\right) h_k(\lambda_1,\ldots,\lambda_{k-1})\, d\lambda_1 \ldots d\lambda_k$$

where

1. *$cum\{d\xi(\lambda_1),\ldots,d\xi(\lambda_k)\}$ is the cumulant measure of order k, $h_1 = 0$, $h_2 = 1$, $|h_k(\lambda_1,\ldots,\lambda_{k-1})|$ is bounded for each k and $\eta(\lambda) = \sum_{j=-\infty}^{\infty}\delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.*

2. *There exists a constant K and a 2π-periodic function $A : [0,1]\times\mathbb{R} \to \mathbb{C}$ with $A(u,-\lambda) = \overline{A(u,\lambda)}$ and*

$$\sum_{n,\lambda}\left|A'_{n,N}(\lambda) - A(\frac{n}{N},\lambda)\right| \leq KT^{-1}$$

for all N. Further $A(\lambda,u)$ is assumed to be continuous in u.

The smoothness of A in u guarantees that the process has locally a "stationary behavior".

As mentioned above, letting N tend to infinity no longer means extending the data to the future. It also does not mean that we sample more densely from $\epsilon(t)$, a continuous process. Notice that if this were the case we would need, for example, that the covariance $\mathrm{Cov}(\epsilon_{N/2,N}, \epsilon_{N,N})$, be the same for each N. This would make the concept of short span of dependence (Assumption 1) hard to interpret. In practice we have a fixed N, so we interpret $\{\epsilon_{n,N}\}$ as a non-stationary process but defined in such way that if N is large enough, approximately unbiased and normal distributed estimates can be obtained using local estimation. We will define such local estimation in the next section.

In the case of many music signals, we can think of the non-sinusoidal part of the signal as filtered stationary noise, but with the filter changing in time. For example, in a wind instrument the non-sinusoidal part of the signal can be attributed to the blown air that is not converted into a harmonic signal. This turbulent air passes through the register hole which can be thought of as a filter. Examples of ways that this filter changes are: different keys being pressed and spit getting stuck inside and moving around. The local stationarity assumption for the noise provides a way to model these changing filters, and will provide a way to synthesize the non-sinusoidal part of the signal via simulations.

4.3 Estimating parameters

Our purpose is to estimate the function

$$\beta(t) \equiv (A_1(t), B_1(t), \ldots, A_K(t), B_K(t), \lambda(t))'$$

for all $t \in [0,1]$, where $\lambda(t)$ is defined by Assumption 3. In this section we will describe how we will find an estimate $\hat{\beta}(t_0)$ for any $t_0 \in [0,1]$.

Notice that we don't necessarily observe the signal exactly at time t_0. Choose $n_0 = n_{0,N} \in \{1, \ldots, N\}$ such that $n_0 = \lfloor t_0 \times N \rfloor$, $\lfloor \cdot \rfloor$ being *integer part*. This implies that

$$\lim_{N \to \infty} \frac{n_0}{N} = t_0 \tag{4.4}$$

Now consider a small enough segment, say h_N time units long, of the signal around t_0 so that one is able to assume that the signal is *approximately sinusoidal* within that segment. We will call the interval $(t_0 - h_N/2, t_0 + h_N/2)$ the *estimation window*.

Next, define a window function $w_{h_N}(t)$ such that observations closest to the fitting point t_0 will receive full weight, observations near the ends $t_0 \pm h_N/2$ receive little weight, and observations outside the estimation window $(t_0 - h_N/2, t_0 + h_N/2)$ will receive no weight. Formally, assign weights through the function

$$w_{h_N}(t) = w\left(\frac{t}{h_N}\right)$$

where $w(t)$ is a symmetric function about t_0 satisfying Assumption 2 and decreasing on the interval $[t_0, t_0 + h_N/2]$. Notice that by changing the span of the function, h_N, we change the length of the segment where the *approximately sinusoidal* assumption is meant to hold, see Figure 4.1 for an example.

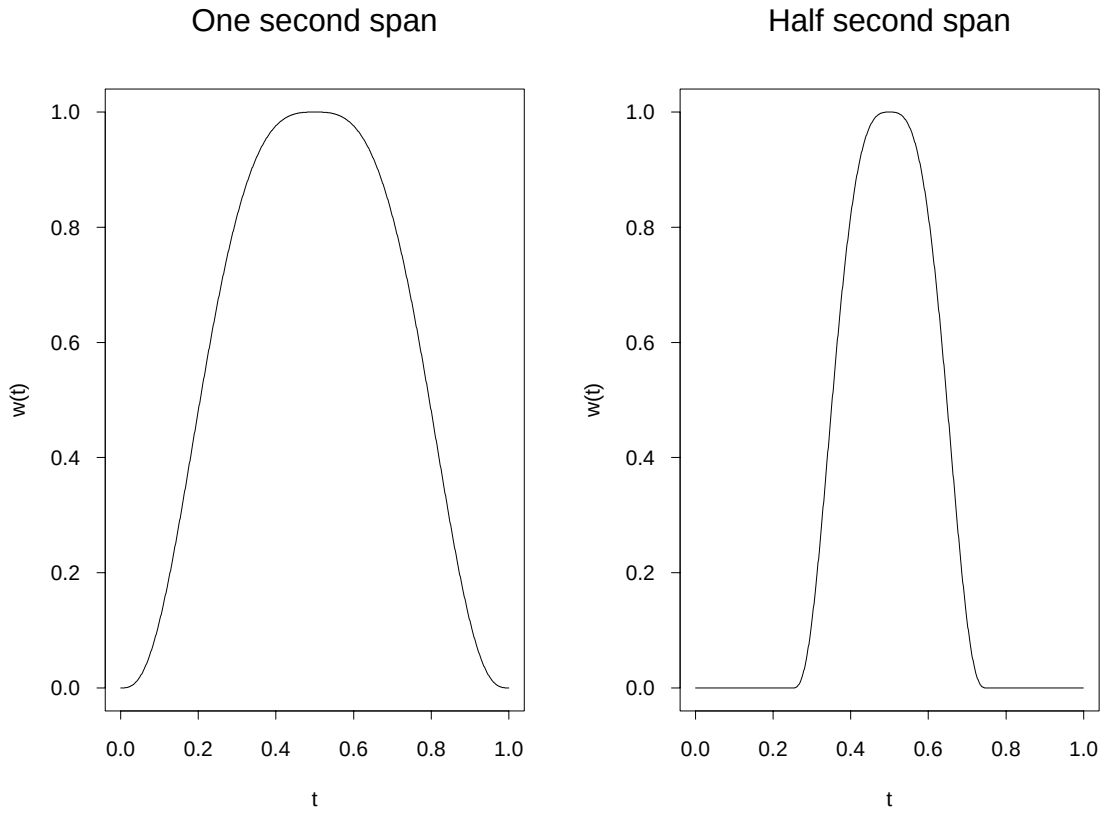

Figure 4.1: Tukey triweight window with two different spans

Now assume that the parameters are constant in time within the estimation window, i.e. $\beta(t) = \beta(t_0)$ for $t \in [t_0 - h_N/2, t_0 + h_N/2]$. Assume also that the noise $\{\epsilon_{n,N}\}$ is

stationary for $n/N \in [t_0 - h_N, t_0 + h_N]$. Now for the data set

$$\left\{ Y_{n,N}, \frac{n}{N} \in [t_0 - h_N, t_0 + h_N] \right\}$$

use the estimation methods described for the stationary case in Chapter 3.

Define the estimates $\hat{A}_{k,N}(t_0)$, $\hat{B}_{k,N}(t_0)$ and $\hat{\lambda}_N(t_0)$ for $k = 1, \ldots, K$. Letting $J_N = \lfloor h_N \times N \rfloor$ (without loss of generality assume J_N is even), $l = n_0 - J_N/2$, $u = n_0 + J_N/2$ and suppressing t_0, the estimates are defined by

$$\begin{aligned} \hat{A}_{k,N} &= 2 \sum_{n=l+1}^{u} w(\frac{n-l}{J_N})\, y_{n,N} \cos[\hat{\omega}_{k,N}\, n] / \sum_{n=0}^{J_N} w(\frac{n}{J_N}) \\ \hat{B}_{k,N} &= 2 \sum_{n=l+1}^{u} w(\frac{n-l}{J_N})\, y_{n,N} \sin[\hat{\omega}_{k,N}\, n] / \sum_{n=0}^{J_N} w(\frac{n}{J_N}) \end{aligned}$$

where if we write $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$ and $\hat{\boldsymbol{\omega}}_N = (\hat{\omega}_{1,N}, \ldots, \hat{\omega}_{K,N})$, $\hat{\boldsymbol{\omega}}_N$ is such that

$$q_N(\hat{\boldsymbol{\omega}}) = \max_{0 \le \boldsymbol{\omega} \le \pi} q_N(\boldsymbol{\omega})$$

under the constraint (3.44) where $q_N(\boldsymbol{\omega})$ is defined by:

$$q_N(\boldsymbol{\omega}) = \sum_{k=1}^{K} \left| (J_N)^{-1} \sum_{n=1}^{J_N} w(\frac{n}{J_N})\, y_{n+l,N} \exp\{in\omega_k\} \right|^2$$

Now $\hat{\lambda}_N$ is defined as in (3.49).

$$\hat{\lambda}_N = \frac{\sum_{k=1}^{K} k\, \hat{\omega}_{k,T} (\hat{A}_{k,N}^2 + \hat{B}_{k,N}^2)}{\sum_{k=1}^{K} k^2 (\hat{A}_{k,N}^2 + \hat{B}_{k,N}^2)} \tag{4.5}$$

By repeating this procedure for each $n_0 \in \{0, \ldots, N\}$ we end up with an estimate $\hat{\beta}(t)$ of the function $\beta(t)$ for $t \in \{\frac{1}{N}, \ldots, \frac{N}{N}\}$. Using linear interpolation, $\beta(t)$ is estimated for each $t \in [0, 1]$ by

$$\hat{\beta}\left(\lfloor t \times N \rfloor / N\right) + \left[\hat{\beta}\left(\lceil t \times N \rceil / N\right) - \hat{\beta}\left(\lfloor t \times N \rfloor / N\right)\right] \left(Nt - \lfloor Nt \rfloor\right)$$

where $\lceil \cdot \rceil$ is the closest integer greater than the number.

4.4 Asymptotics

For this section we will assume the sequence of stochastic processes

$$\{\epsilon_{n,N}, n = 1, \ldots, N; N \ge 1\}$$

is defined by

$$\epsilon_{n,N} = \int_{-\pi}^{\pi} \exp\{i\lambda n\} A(\frac{n}{N}, \lambda)\, d\xi(\lambda)$$

where $\xi(\lambda)$ is a stochastic process as in Definition 1 and $A(t,\lambda)$ is continuous in t. Notice that setting $A'_{n,N}(\lambda) = A(n/N, \lambda)$, the sequence is seen to be locally stationary.

We call

$$f_{\epsilon\epsilon}(t,\lambda) = |A(t,\lambda)|^2 \tag{4.6}$$

the time-varying spectral density of the process. Dahlhaus (1996) proves that under certain regularity conditions $f_{\epsilon\epsilon}(t,\lambda)$ is uniquely determined by the processes $\{\epsilon_{n,N}\}$.

To prove consistency and asymptotic normality we will need the following two conditions of the sequences $\{\epsilon_{n,N}, n = 1,\ldots,N; N \geq 1\}$. This work parallels Theorem 4.4.2 in Brillinger (1981, page 95) and Lemma 2 of Chapter 3. Finding general conditions where these assumptions hold is left as future work.

Condition 1 *The sequence of stochastic processes* $\{\epsilon_{n,N} : n = 1,\ldots,N; N \geq 1\}$ *is such that for any a sequence* $h_N \downarrow 0$, *with* $J_N = \lfloor h_N N \rfloor \to \infty$, *the following two conditions hold for every* $t_0 \in (0,1)$, *with* $n_0 = \lfloor t_0 \times N \rfloor$ *and* $l = n_0 - J_N/2$.

(a) The quantity defined by

$$p_N(\lambda) = \left| (J_N)^{-(k+1)} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) n^k \epsilon_{n+l,N} \exp\{i\lambda n\} \right|$$

is such that

$$\lim_{N\to\infty} \sup_{0\leq\lambda\leq\pi} p_N(\lambda) = 0, \text{ in probability}$$

(b) The vector **u** *defined by*

$$\begin{aligned}
u_1 &= (J_N)^{-\frac{1}{2}} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) \epsilon_{n+l,N} \cos(\lambda n) \\
u_2 &= (J_N)^{-\frac{1}{2}} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) \epsilon_{n+l,N} \sin(\lambda n) \\
u_3 &= (J_N)^{-\frac{3}{2}} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) \epsilon_{n+l,N}\, n \cos(\lambda n) \\
u_4 &= (J_N)^{-\frac{3}{2}} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) \epsilon_{n+l,N}\, n \sin(\lambda n)
\end{aligned}$$

is asymptotically multivariate normal with zero mean and covariance matrix

$$\mathbf{U} = \pi f_{\epsilon\epsilon}(t_0, \lambda) \begin{pmatrix} U_0 & 0 & U_1 & 0 \\ 0 & U_0 & 0 & U_1 \\ U_1 & 0 & U_2 & 0 \\ 0 & U_1 & 0 & U_2 \end{pmatrix}$$

with the constants U_0, U_1, *and* U_2 *defined as in equation (3.8).*

Notice that if we hold the time-varying spectrum constant over time, $A(t, \lambda) = A(\lambda)$, then each stochastic process $\{\epsilon_{n,N}\}$ can be thought of as being N observations of a stationary processes ϵ_n with cumulant functions

$$c_{\epsilon,\ldots,\epsilon}(u_1, \ldots, u_{k-1}) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp\left\{ i \sum_{j=1}^{k-1} \lambda_j u_j \right\} \left\{ \prod_{j=1}^{k-1} A(\lambda_j) \right\} h_k(\lambda_1, \ldots, \lambda_{k-1})\, d\lambda_1 \cdots d\lambda_{k-1} \quad (4.7)$$

If the cumulants defined by (4.7) satisfy equation (3.1) then Condition 1 holds by Lemma 2, Theorem 4.4.2 in Brillinger (1981), and the fact that $\{J_N\}_{N \geq 1}$ is a subsequence of $\{N\}_{N \geq 1}$.

Similar to what was done in Chapter 3 we let

$$\Delta_k^{J_N}(\lambda) = \sum_{n=1}^{J_N} w(\frac{n}{J_N}) n^k \exp\{i\lambda n\} \quad (4.8)$$

Next we develop the equivalent result found in Lemma 1 of Chapter 3 for the quantity in equation (4.8)

Lemma 3 *If* J_N *is a sequence of integers such that* $J_N \to \infty$ *then*

$$\lim_{J_N \to \infty} J_N^{-(k+1)} \Delta_k^{J_N}(\lambda) = W_k, \text{ for } \lambda = 0, 2\pi \quad (4.9)$$

$$\Delta_k^{J_N}(\lambda) = O(J_N^{k+1}), \text{ for } 0 < \lambda < 2\pi \quad (4.10)$$

with W_k *defined by 3.8 for* $k = 0, 1, 2$.

This follows by noticing that

$$\Delta_k^{J_N}(\lambda) = \sum_{n=1}^{J_N} w(\frac{n}{J_N}) n^k \exp\{i\lambda n\}$$

is a subsequence of

$$\sum_{n=1}^{N} w(\frac{n}{N}) n^k \exp\{i\lambda n\}$$

Then from the proof of Lemma 1 of Chapter 3, (4.9) and (4.10) hold.

We are now ready to prove the following Theorem.

Theorem 4 *Let the of sequence stochastic processes* $\{\epsilon_{n,N}, n = 1, \ldots, N; N \geq 1\}$ *be such that Condition 1 holds. Let the sequence of window sizes* $\{h_N,\ N > 1\}$ *be such that* $h_N \downarrow 0$ *and* $J_N = \lfloor h_N \times N \rfloor \to \infty$ *as* $N \to \infty$. *Then if Assumption 4 holds we have*

$$\lim_{N\to\infty} \hat{A}_{k,N}(t_0) = A_k(t_0),\ \lim_{N\to\infty} \hat{B}_{k,N}(t_0) = B_k(t_0),\ \lim_{N\to\infty} J_N|\hat{\lambda}_N(t_0) - \lambda(t_0)| = 0$$

in probability for each k *and each* $t_0 \in (0,1)$

Proof: We will prove this result for $K = 1$. The more general case follows in the same way that Corollary 1 followed from Theorem 1.

For this proof we will assume all the sums are over $1, \ldots, J_N$, unless otherwise specified. Without loss of generality assume that J_N is even. Notice that

$$\begin{aligned} q_N(\lambda) &= \left| J_N^{-1} \sum w(\frac{n}{J_N})\, \epsilon_{n+l,N} \exp\{in\lambda\} \right|^2 \\ &+ \left| J_N^{-1} \sum w(\frac{n}{J_N})\, s\left[\frac{n+l}{N}, \beta_N(\frac{n+l}{N})\right] \exp\{in\lambda\} \right|^2 \\ &+ 2\Re\left[\left(J_N^{-1} \sum w(\frac{n}{J_N})\, \epsilon_{n+l,N} \exp\{in\lambda\} \right) \cdot \right. \\ &\left. \left(J_N^{-1} \sum w(\frac{n}{J_N})\, s\left[\frac{n+l}{N}, \beta_N(\frac{n+l}{N})\right] \exp\{in\lambda\} \right) \right] \end{aligned} \tag{4.11}$$

Condition 1 requires that the first expression on the right goes to 0 in probability.

By the mean value theorem we have

$$\begin{aligned} s[t;\beta_N(t)] &= [A(t_0) + M_1(t-t_0)] \cos\left([\lambda_N(t_0) + M_3(t-t_0)]t\right) \\ &+ [B(t_0) + M_2(t-t_0)] \sin\left([\lambda_N(t_0) + M_3(t-t_0)]t\right) \end{aligned} \tag{4.12}$$

By assumption 4, the constants M_1, M_2 and M_3 are bounded. Since $\sin(t)$ and $\cos(t)$ are bounded functions we can write (4.12) as

$$s[t;\beta_N(t)] = A(t_0)\cos(\lambda_N(t_0)t + M_3 t(t-t_0)) + B(t_0)\sin(\lambda_N(t_0)t + M_3 t(t-t_0)) + M_4(t-t_0)$$

where M_4 is a bounded constant. Notice that by Assumption 3 and applying the mean value theorem for the first term on the right of equation (4.4) we have

$$\begin{aligned}\cos(\lambda_N(t_0)t + M_3 t(t-t_0)) &= \cos(N\lambda(t_0)t + o(1)\,t + M_3(t-t_0)) \\ &= \cos(N\lambda(t_0)t) + M_5[o(1)\,t + M_3(t-t_0)]\end{aligned}$$

where M_5 is a bounded constant. Similarly for the second term on the right of equation (4.4).

$$\sin(\lambda_N(t_0)t + M_3 t(t-t_0)) = \sin(N\lambda(t_0)t) + M_5[o(1)\,t + M_3(t-t_0)]$$

Let $A_0 = A(t_0)$, $B_0 = B(t_0)$, $\lambda_0 = \lambda(t_0)$ and suppressing the N, $\beta_0 = \beta_N(t_0)$. Then since $|t| < 1$ we have that by (4.4) and the continuity of $A(t)$, $B(t)$, and $\lambda(t)$.

$$s\left[\frac{n}{N}; \beta_N(\frac{n}{N})\right] = r(n, \beta_0) + M\,\frac{n-n_0}{N} + o(1)$$

with M bounded and

$$r(n, \beta_0) = A_0 \cos(\lambda_0 n) + B_0 \sin(\lambda_0 n)$$

Now let $M = \sup_t w(t)$, then

$$\begin{aligned}\left| J_N^{-1} \sum \frac{n - J_N/2}{N} \exp\{i\lambda n\} \right| &\le M(N\,J_N)^{-1} \sum |n - J_N/2| \\ &\le 2M(N\,J_N)^{-1} \sum_{n=1}^{J_N/2} n \\ &= M(N\,J_N)^{-1} \frac{J_N}{2}(\frac{J_N}{2} + 1) \\ &= M\frac{J_N + 2}{4N} = o(1)\end{aligned}$$

By a summation by parts argument like that in the proof of Lemma 1 in Chapter 3, we have

$$\left| J_N^{-1} \sum w(\frac{n}{J_N})\,\frac{n - J_N/2}{N} \exp\{i\lambda n\} \right| = o(1)$$

Now we can write the second term in equation (4.11) as

$$\left| J_N^{-1} \sum w(\frac{n}{J_N})\, s\left[\frac{n+l}{N}, \beta_N(\frac{n+l}{N})\right] \exp\{i\lambda n\} \right| =$$
$$\left| J_N^{-1} \sum w(\frac{n}{J_N})\, r(n+l, \beta_0) \exp\{i\lambda n\} + o(1) \right|$$

and

$$r(n+l, \beta_0) = \left(D_0 \exp\{i\lambda_0 n\} + \overline{D}_0 \exp\{-i\lambda_0 n\}\right) \exp\{i\lambda_0 l\}$$

where $D_0 = \frac{1}{2}(A_0 - iB_0)$ as before. Next notice that

$$J_N^{-1} \sum w_{h_N}(\frac{n}{N})\, r(\frac{n+l}{N}, \beta_0) \exp\{i\lambda n\} = J_N^{-1}[D_0 \Delta_0^{J_N}(\lambda_0 + \lambda) + \overline{D}_0 \Delta_0^{J_N}(\lambda_0 - \lambda)] \exp\{i\lambda_0 l\}$$

By Lemma 3 we have that for $0 < \lambda < \pi$

$$J_N^{-1} \Delta_0^{J_N}(\lambda_0 + \lambda) = o(1)$$

and that

$$J_N^{-1} \Delta_0^{J_N}(\lambda_0 - \lambda) = \begin{cases} W_0 & : \quad \lambda = \lambda_0 \\ o(1) & : \quad \text{otherwise} \end{cases}$$

Using this fact and Condition 1 we have that the third term in (4.11) converges to 0 in probability and that we can write

$$\begin{aligned} q_N(\lambda) &= \left| J_N^{-1}[D_0 \Delta_0^{J_N}(\lambda_0 + \lambda) + \overline{D}_0 \Delta_0^{J_N}(\lambda_0 - \lambda)] \exp\{i\lambda_0 l\} + o_p(1) \right| + o_p(1) \\ &= \frac{1}{4}(A_0^2 + B_0^2) |J_N^{-1} \Delta_0^{J_N}(\lambda - \lambda_0)|^2 + o_p(1) \end{aligned}$$

and therefore

$$q_N(\lambda_0) = \frac{1}{4}(A_0^2 + B_0^2) W_0^2 + o_p(1) \tag{4.13}$$

Finally, for any $b > 0$, define

$$P_N(b) = \{\lambda : J_N |\lambda - \lambda_0| \geq b\} \tag{4.14}$$

Notice that as in the proof of Theorem 2

$$\begin{aligned} \Pr\left(J_N |\hat{\lambda}_N(t_0) - \lambda(t_0)| \geq b \right) &\leq \Pr\left(\sup_{\lambda \in P_N(b)} q_N(\lambda) \geq q_N(\lambda_0) \right) \\ &= \Pr\left(\sup_{\lambda \in P_N(b)} |(J_N)^{-1} \Delta_0^{J_N}(\lambda_0 - \lambda)| \geq W_0 + o_p(1) \right) \end{aligned}$$

and that

$$\begin{aligned} J_N^{-1} \Delta_0^{J_N}(\lambda - \lambda_0) &= J_N^{-1} \sum_{n=1}^{J_N} w(\frac{n}{J_N}) n \exp\{i(\lambda - \lambda_0) n\} \\ &= \sum_{n=1}^{J_N} w(\frac{n}{J_N})(\frac{n}{J_N}) \exp\{i J_N (\lambda - \lambda_0) \frac{n}{J_N}\} \end{aligned}$$

The arguments of proposition B.1.3 in Wang (1991) now provide that

$$\sup_{\lambda \in P_N(b)} |J_N^{-1} \Delta_0^{J_N}(\lambda - \lambda_0)| = \sup_{\lambda \in P_N(b)} \left| \int_0^1 w(s) \exp\{i J_N (\lambda - \lambda_0) s\}\, ds \right| + o(1)$$

Following the steps of equations (3.17) through (3.21) in the proof of Theorem 1 we have

$$\lim_{N\to\infty} \Pr\left(J_N|\hat{\lambda}_N(t_0) - \lambda(t_0)| \geq b\right) = 0$$

Now we will prove consistency for $\hat{A}_N(t_0)$ and $\hat{B}_N(t_0)$. As above let $\beta_0 = \beta_N(t_0)$, $A_0 = A(t_0)$ and $B_0 = B(t_0)$. Then we have that

$$\hat{A}_N(t_0) = 2(W_0^{J_N})^{-1} \sum_{n=l+1}^{u} w(\frac{n-l}{J_N}) \left[s\left[\frac{n}{N}, \beta(\frac{n}{N})\right] + \epsilon_{n,N}\right] \cos[\hat{\lambda}_N\, n]$$

with $l = n_0 - J_N/2$, $u = n_0 + J_N/2$ and

$$W_0^{J_N} = \sum_{n=1}^{J_N} w(\frac{n}{J_N})$$

As before we use the mean value theorem to obtain

$$(W_0^{J_N})^{-1} \sum_{n=l+1}^{u} w(\frac{n-l}{J_N})\, s\left[\frac{n}{N}, \beta(\frac{n}{N})\right] \cos[\hat{\lambda}_N\, n] =$$

$$(W_0^{J_N})^{-1} \sum_{n=l+1}^{u} w(\frac{n-l}{J_N})\, r(n, \beta_0) \cos[\hat{\lambda}_N\, n] + o(1)$$

So we have that

$$\hat{A}_N(t_0) = (W_0^{J_N})^{-1} \sum_{n=l+1}^{u} w(\frac{n-l}{J_N}) \left[r(n, \beta_0) + \epsilon_{n,N}\right] \cos[\hat{\lambda}_N\, n] + o(1)$$

In the same way we obtain

$$\hat{B}_N(t_0) = (W_0^{J_N})^{-1} \sum_{n=l+1}^{u} w(\frac{n-l}{J_N}) \left[r(n, \beta_0) + \epsilon_{n,N}\right] \sin[\hat{\lambda}_N\, n] + o(1)$$

Since the parameter β_0 is constant over time the result now follows as the proof of Theorem 1. □

Theorem 5 *Under the same assumptions of Theorem 4 we have that for each* $t_0 \in (0,1)$ *the vector*

$$\begin{pmatrix} J_N^{\frac{1}{2}}(\hat{A}_{1,T}(t_0) - A_1(t_0)) \\ J_N^{\frac{1}{2}}(\hat{B}_{1,T}(t_0) - B_1(t_0)) \\ \vdots \\ J_N^{\frac{1}{2}}(\hat{A}_{K,T}(t_0) - A_K(t_0)) \\ J_N^{\frac{1}{2}}(\hat{B}_{K,T}(t_0) - B_K(t_0)) \\ J_N^{\frac{3}{2}}(\hat{\lambda}_T(t_0) - \lambda_0(t_0)) \end{pmatrix}$$

converges in distribution to a normal vector with zero mean and covariance matrix given by equations (3.50),(3.51),(3.52) and (3.53) in Theorem 3 as $N \to \infty$.

Proof: Expanding $q'_N(\lambda)$ in the first two terms of its Taylor series, about $\lambda(t_0)$ we can write.

$$J_N^{-\frac{1}{2}} q'_N(\lambda(t_0)) = -J_N^{\frac{3}{2}}(\hat{\lambda}_N(t_0) - \lambda(t_0)) J_N^{-2} q''_N(\tilde{\lambda}_N(t_0))$$

for some $|\tilde{\lambda}_N(t_0)|$ such that $|\tilde{\lambda}_T(t_0) - \lambda(t_0)| \leq |\hat{\lambda}_N(t_0) - \lambda(t_0)|$.

Using the result of equation (4.13), Lemma 3 and Condition 1 we can proceed as in the proof of Theorems 2 and Theorem 3 to arrive at the desired result. □

Chapter 5

Choosing the Window Size and the Number of Harmonics

5.1 Introduction

For a particular sound signal a variety of deterministic factors may affect the smoothness of the parameter function $\beta(t)$. A change in note creates a discontinuity in the fundamental frequency function $\lambda(t)$. A sudden change from playing softly to playing loudly produces a rapid change in all the amplitudes of the partials $\rho_k(t) = \sqrt{A_k(t)^2 + B_k(t)^2}$. Such phenomena suggest that h_N should not remain fixed for all $t \in [0,1]$. Say there are two values t_0 and t_1, the approximately constant assumption might be appropriate within the interval $(t_0 - h_N/2, t_0 + h_N/2)$, but not within $(t_1 - h_N/2, t_1 + h_N/2)$. For each t_0 one needs to make a decision on the size of the span h_N of the smoothing window. Also, it has been noted how for different sound signals the number of partials that seem meaningful varies.

We want a criteria that will permit us to choose amongst different possible estimates. Ideally, the estimates derived from using this criteria will be optimal/efficient in some sense. We will not deal with this matter now in any detail but leave it as future work. We will present intuitive arguments that lead to a useful criteria for practical estimates.

In this chapter we present three different criteria each based on existing methods of model selection. In general, we are given a set of N observations which we assume to be an outcome of some multivariate random variable whose probability distribution is

unknown to us but of which we have some knowledge. From the information provided by the data, we are to choose the number of partials K used in our model, the size h_N of the estimation window, and then estimate the parameters of the specified model. Since we are using weighted least squares, we might think of choosing K and h_N as the values that minimize the weighted residual mean-square error. The problem with this method is that larger values of K and smaller values of h_N will tend to have smaller weighted residual mean-square, regardless of the true model.

Mallows (1973) presents a criteria that is based on estimating the mean squared error. Akaike (1973) presents a criteria based on estimating the Kullback-Leibler Information Quantity. Akaike (1979), Schwarz (1978) and others present a Bayesian version of the latter. All these criteria were derived for the unweighted case and are used to choose among the competing models. In the case of linear regression they all come down to a criterion of mean squared error type but somehow penalizing for a large number of parameters. Some work has been done to find similar criteria when weights are used (Ronchetti 1985, Linhart and Zucchini 1986, Hampel et al. 1986, Machado 1993, Ronchetti and Staudte 1994, Hurvich 1997), but we need a criteria not only for the number of parameters but for window size also. In what follows we derive such criteria for the case of weighted linear regression. We will then see how this relates to the musical sound situation.

5.2 Weighted linear regression

Similar models to the one presented for musical sound signals in Chapter 4 can be found in the literature of general additive models and local likelihood estimation. See, for example, Hastie and Tibshirani (1990), Tibshirani and Hastie (1987). If we think of time as the predictor variables $\mathbf{x}_n$ and the sampled signal $\mathbf{y}$ as the dependent variable, the local harmonic model fits into the context of general additive models in which we have

$$y_n = \sum_{j=1}^{p} g_j(\mathbf{x}_n) + \epsilon_n$$

where the functions g_j are considered to be "smooth" functions and ϵ_n is noise.

In Chapter 4 we modeled the discrete part of the signal as a non-linear parametric function of time, where the parameters were also functions, but not parametric, of time. We can make a linear approximation of our model, as described in detail in section 5.6.

For this reason the concepts presented in this chapter are based on those of local weighted regression (Clevland 1979, Clevland and Devlin 1988). In this particular case we model the dependent variable as

$$y_n = \mathbf{x}_n' \boldsymbol{\beta}(\mathbf{x}_n) + \epsilon_n \tag{5.1}$$

where the $\mathbf{x}_n'$ are the $1 \times p$ rows of a $N \times p$ regression matrix $\mathbf{X}$ and $\boldsymbol{\beta}(\mathbf{x}_n)$ is a "smooth" $P \times 1$ functional vector.

Say we have N observations $y_1, \ldots, y_N$ and, at the moment, we are interested in estimating only $\boldsymbol{\beta} = \boldsymbol{\beta}(\mathbf{x}_{N/2})$. Because $\boldsymbol{\beta}(\mathbf{x})$ is assumed to be "smooth" we hope that all the data $\mathbf{Y}$ contain information about $\boldsymbol{\beta}$. To estimate $\boldsymbol{\beta}$, as in local weighted regression, we consider the weighted least squares estimate

$$\hat{\boldsymbol{\beta}} = \mathbf{HY} = \mathbf{X'WX^{-1}X'WY}$$

with $\mathbf{W}$ a diagonal matrix defined by $\mathbf{W} = \text{diag}\{w_n\}$ and $w_1, \ldots, w_N$ a set of weight coefficients defined by some window function

$$w_n = w\left(\frac{n}{N}\right)$$

where $w(u)$ satisfies Assumption 2 of Chapter 3. Because we have limited knowledge of the global behavior of the function $\boldsymbol{\beta}(\mathbf{x})$, we might want to consider different window matrices, each one using different weight coefficients depending on how much importance we want to give to certain parts of the data. In any case, it seems appropriate to weigh the central values more heavily.

By assuming that $\boldsymbol{\beta}$ is constant we are then able to obtain an estimate for $\boldsymbol{\beta}(\mathbf{x}_{N/2})$. To derive useful criteria we will make some further assumptions: as in the context of linear regression, assume $\mathbf{y} = (y_1, \ldots, y_N)'$ is a vector of mutually independent random variables. In the usual regression notation we can write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where the components of the $N \times 1$ vector $\boldsymbol{\epsilon}$ are i.i.d. random variables with mean 0 and variance σ^2. The distribution of $\mathbf{y}$ can be expressed as

$$f(y_n|\mathbf{x}_n, \boldsymbol{\beta}) = g(y_n - \mathbf{x}_n'\boldsymbol{\beta}) = g(\epsilon_n) \tag{5.2}$$

where $g(\cdot)$ is the error probability density function.

We are interested in estimating $\boldsymbol{\beta}$ and also in choosing from amongst P competing models that are generated by simply restricting the general parameter vector $\boldsymbol{\beta}$. In terms of the parameters, we represent the full model with P parameters as:

$$\text{Model(P): } f(\cdot|\mathbf{x}, \boldsymbol{\beta}_P), \boldsymbol{\beta}_P = (\beta_1, \ldots, \beta_p, \beta_{p+1}, \ldots, \beta_P)'$$

We denote the true value of the parameter vector $\boldsymbol{\beta}$ by $\boldsymbol{\beta}^*$ with $\boldsymbol{\beta}^* \in R^P$. Akaike (1973) formulates the problem of statistical model identification as one of selecting a model $f(\cdot|\mathbf{x}, \boldsymbol{\beta}_p)$ based on N observations from that distribution, where the particular restricted model is defined by the constraint $\beta_{p+1} = \beta_{p+2} = \ldots = \beta_P = 0$, or the model with p parameters,

$$\text{Model(p): } f(\cdot|\mathbf{x}, \boldsymbol{\beta}_p), \boldsymbol{\beta}_p = (\beta_1, \ldots, \beta_p, 0, \ldots, 0)' \tag{5.3}$$

We will refer to p as the actual number of parameters.

By assuming model(p) to estimate the non-zero components of the vector $\boldsymbol{\beta}^*$ we can now define an estimate $\hat{\boldsymbol{\beta}}_p$ using weighted least squares, namely

$$\hat{\boldsymbol{\beta}}_p = \mathbf{H}_p \mathbf{Y} = (\mathbf{X}_p' \mathbf{W} \mathbf{X}_p)^{-1} \mathbf{X}_p \mathbf{W}' \mathbf{Y}$$

where $\mathbf{X}_p = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ is the matrix formed by the first p columns of the regression matrix $\mathbf{X}$.

As mentioned earlier, we might want to consider different window matrices $\mathbf{W}_1, \ldots, \mathbf{W}_Q$. A convenient way to do so is by considering the J_q nearest points to $N/2$ when estimating β. We refer to $h_q = J_q/N$ as the span of the estimation window and define the diagonal matrix $\mathbf{W}_q = \text{diag}\{w_{q,n}\}$, $n = 1, \ldots, N$ with

$$w_{q,n} = w\left(\frac{n}{J_q}\right)$$

where $w(u)$ is a symmetric function about $1/2$ satisfying Assumption 2 and decreasing on the interval $[1/2, 1]$.

We can now define an estimate, assuming model(p) and using $\mathbf{W}_q$, by

$$\hat{\boldsymbol{\beta}}_{p,q} = \mathbf{H}_{p,q} \mathbf{Y} = (\mathbf{X}_p' \mathbf{W}_q \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{W}_q \mathbf{Y}$$

In this chapter we are concerned with the problem of how to choose between the $P \times Q$ competing estimates of $\boldsymbol{\beta}^*$. We present three criteria that will be used to choose p and q.

5.3 Weighted Mallow's C_p

Mallow's C_p is a technique for model selection in regression, (Mallows 1973). The C_p statistic is defined as a criteria to assess fits when models with different numbers of parameters are being compared. It is given by

$$C_p = \frac{\mathrm{RSS}(p)}{\sigma^2} - N + 2p \tag{5.4}$$

If model(p) is correct then C_p will tend to be close to or smaller than p. Therefore a simple plot of C_p versus p can be used to decide amongst models.

In the case of ordinary linear regression, Mallow's method is based on estimating the mean squared error (MSE) of the estimator $\hat{\boldsymbol{\beta}}_p = (\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{Y}$,

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}]^2$$

via a quantity based on the residual sum of squares (RSS)

$$\begin{aligned}
\mathrm{RSS}(p) &= \sum_{n=1}^{N}(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}_p)^2 \\
&= (\mathbf{Y} - \mathbf{X}_p\hat{\boldsymbol{\beta}}_p)'(\mathbf{Y} - \mathbf{X}_p\hat{\boldsymbol{\beta}}_p) \\
&= \mathbf{Y}'(\mathbf{I}_N - \mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p')\mathbf{Y}
\end{aligned}$$

Here $\mathbf{I}_N$ is an $N \times N$ identity matrix. By using a result for quadratic forms, presented for example as Theorem 1.17 in Seber (1977, page 13), namely

$$\mathrm{E}[\mathbf{Y}'\mathbf{A}\mathbf{Y}] = \mathrm{E}[\mathbf{Y}']\mathbf{A}\mathrm{E}[\mathbf{Y}] + \mathrm{tr}[\boldsymbol{\Sigma}\mathbf{A}]$$

$\boldsymbol{\Sigma}$ being the variance matrix of $\mathbf{Y}$, we find that

$$\begin{aligned}
\mathrm{E}[\mathrm{RSS}(p)] &= \mathrm{E}[\mathbf{Y}'(\mathbf{I}_N - \mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p')\mathbf{Y}] \\
&= \mathrm{E}[\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}]^2 + \mathrm{tr}\left[\mathbf{I}_N - \mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p'\right]\sigma^2 \\
&= \mathrm{E}[\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}]^2 + \sigma^2\left(N - \mathrm{tr}\left[(\mathbf{X}_p'\mathbf{X}_p)(\mathbf{X}_p'\mathbf{X}_p)^{-1}\right]\right) \\
&= \mathrm{E}[\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}]^2 + \sigma^2(N - p)
\end{aligned}$$

where N is the number of observations and p is the number of parameters. Notice that when the true model has p parameters $\mathrm{E}[C_p] = p$. This shows why, if model(p) is correct, C_p will tend to be close to p.

In the case of weighted regression it seems appropriate to base our estimate of the MSE of the estimate

$$\hat{\boldsymbol{\beta}}_{p,q} = \mathbf{H}_{p,q}\mathbf{Y} = (\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{W}_q\mathbf{Y}$$

on the weighted residual sum of squares (wRSS)

$$\begin{aligned} \text{wRSS}(p,q) &= \sum_{n=1}^{N} w_{q,n}(y_n - \mathbf{x}_n'\hat{\boldsymbol{\beta}}_{p,q})^2 \qquad (5.5) \\ &= \mathbf{Y}'(\mathbf{W}_q - \mathbf{W}_q\mathbf{H}_{p,q})'\mathbf{Y} \qquad (5.6) \end{aligned}$$

Since we assume that $\mathrm{E}[y_n] \approx \mathbf{x}_n\boldsymbol{\beta}$ and $\mathrm{Var}[y_n] \approx \sigma^2$, the weighted residual sum of squares has expected value:

$$\begin{aligned} \mathrm{E}[\text{wRSS}(p,q)] &\approx \mathrm{E}[\hat{\boldsymbol{\beta}}_{p,q} - \boldsymbol{\beta}]^2 + \mathrm{tr}(\mathbf{W}_q - \mathbf{W}_q\mathbf{H})\sigma^2 \\ &= \mathrm{E}[\hat{\boldsymbol{\beta}}_{p,q} - \boldsymbol{\beta}]^2 + \sigma^2\mathrm{tr}\left[\mathbf{W}_q\right] - \sigma^2\mathrm{tr}\left[\mathbf{W}_q(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{W}_q\right] \\ &= \mathrm{E}[\hat{\boldsymbol{\beta}}_{p,q} - \boldsymbol{\beta}]^2 + \sigma^2\mathrm{tr}\left[\mathbf{W}_q\right] - \sigma^2\mathrm{tr}\left[(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}(\mathbf{X}_p'\mathbf{W}_q\mathbf{W}_q\mathbf{X}_p)\right] \\ &= \mathrm{E}[\hat{\boldsymbol{\beta}}_{p,q} - \boldsymbol{\beta}]^2 + \sigma^2(W_q - V_{p,q}) \end{aligned}$$

where $W_q = \mathrm{tr}[\mathbf{W}_q]$ and $V_{p,q} = \mathrm{tr}[(\mathbf{X}_p'\mathbf{W}_h\mathbf{X}_p)^{-1}(\mathbf{X}_p'\mathbf{W}_h\mathbf{W}_h\mathbf{X}_p)]$. Defining the weighted version of C_p as

$$\mathrm{wC}_{p,q} = \frac{\text{wRSS}(p,q)}{\sigma^2} - W_q + 2V_{p,q} \qquad (5.7)$$

we have that when the true model has p parameters and the span of the weights being used h_q is appropriate we will have that $\mathrm{E}[\mathrm{wC}_{p,q}] \approx V_{p,q}$. As in the unweighted case a simple plot of $\mathrm{wC}_{p,q}$ versus $V_{p,q}$ can be used to decide amongst models and window sizes.

Notice that if we use equal weights, $\mathbf{W} = \mathbf{I}$, then $W_q = N$ (the number of observations), $V_{p,q} = p$ (the number of parameters) and $\mathrm{wC}_{p,q}$ is equivalent to the C_p defined in (5.4). For this reason we call W_q and $V_{p,q}$ the *equivalent number of observations* and *equivalent number of parameters*, respectively, for the weighted case.

The problem with the C_p and $\mathrm{wC}_{p,q}$ criteria is that we have to find an appropriate estimate of σ^2 to use for all values of p and q. An alternative method that does not present this problem is the next one presented.

5.4 Weighted AIC

Suppose $\mathbf{Y}$ is characterized by a probability function $f(\mathbf{y}|\mathbf{X},\boldsymbol{\beta})$ as in (5.2), which is assumed known apart from the P dimensional vector $\boldsymbol{\beta}$. Assume that there exists a true

parameter vector $\boldsymbol{\beta}^*$ defining a true probability density denoted by $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}^*)$. Within this setup we wish to select $\boldsymbol{\beta}$, from one of the models defined by (5.3), "nearest" to the true parameter $\boldsymbol{\beta}^*$ based on the observed data. The principle behind Akaike's Information Criterion (AIC) is to define "nearest" as the model that minimizes the Kullback-Leibler Information Quantity (Kullback 1959)

$$I(\boldsymbol{\beta}^*; \boldsymbol{\beta}) = \mathrm{E}\left[\log f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}^*) - \log f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})\right] \tag{5.8}$$

with the expectation taken over the true model. Since the first term on the right hand side of (5.8) is constant over all considered models we may consider only the expected log likelihood of the estimated model

$$H(\boldsymbol{\beta}^*; \boldsymbol{\beta}) = \mathrm{E}\left[\log f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})\right] \tag{5.9}$$

where again the expectation is taken over the true model. Akaike's procedure is based on estimating this quantity for each competing model and then choosing the one that minimizes it. Following Bickel and Doksum (1977), Lehmann (1983), if we assume $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})$ is regular to its first and second partial derivatives with respect to $\boldsymbol{\beta}$, then we have

$$H'(\boldsymbol{\beta}^*; \boldsymbol{\beta}) = 0$$
$$H''(\boldsymbol{\beta}^*; \boldsymbol{\beta}) = -J(\boldsymbol{\beta}^*)$$

where $J(\boldsymbol{\beta}^*)$ is the Fisher's information matrix. The analytical properties of the Kullback-Leibler Information Quantity are discussed in detail in Kullback (1959). Two important properties for Akaike's criterion are

1. $I(\boldsymbol{\beta}^*; \boldsymbol{\beta}) > 0$ if $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}^*) \neq f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})$
2. $I(\boldsymbol{\beta}^*; \boldsymbol{\beta}) = 0$ if and only if $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}^*) = f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})$

almost everywhere on the range of $\mathbf{y}$. The properties mentioned suggest that finding the model that minimizes the Kullback-Leibler Information Quantity is an appropriate way to choose the "closest" model. Equivalently, we can minimize $-2N\,H(\boldsymbol{\beta}^*; \boldsymbol{\beta})$.

Suppose that the observation $Y_1, \ldots, Y_N$ are described as coming from the model defined by the parameter $\boldsymbol{\beta}$. The log likelihood function $l(\boldsymbol{\beta})$ is defined by

$$l(\boldsymbol{\beta}) = \sum_{n=1}^{N} \log f(Y_n|\mathbf{x}_n, \boldsymbol{\beta})$$

The mean log likelihood, which is simply $\frac{1}{N}l(\boldsymbol{\beta})$, can also be interpreted as an estimator of the distance between the true probability density $f(y|\mathbf{X},\boldsymbol{\beta}^*)$ and $f(y|\mathbf{X},\boldsymbol{\beta})$. In fact, we have that the mean log likelihood is a natural estimator of $H(\boldsymbol{\beta}^*;\boldsymbol{\beta})$ since

$$\begin{aligned} \mathrm{E}\left[\frac{1}{N}l(\boldsymbol{\beta})\right] &= \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}[\log f(Y_n|\mathbf{x}_n,\boldsymbol{\beta})] \\ &= \mathrm{E}[\log f(\mathbf{y}|\mathbf{X},\boldsymbol{\beta})] \\ &= H(\boldsymbol{\beta}^*;\boldsymbol{\beta}) \end{aligned}$$

Notice that $\boldsymbol{\beta}$ is unobservable, and therefore so is $\frac{1}{N}l(\boldsymbol{\beta})$. A natural estimate of $H(\boldsymbol{\beta}^*;\boldsymbol{\beta})$ is $\frac{1}{N}l(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of $\boldsymbol{\beta}$. Akaike notices that in general $\frac{1}{N}l(\hat{\boldsymbol{\beta}})$ will overestimate $H(\boldsymbol{\beta}^*;\boldsymbol{\beta})$.

Let $\boldsymbol{\beta}_p$ be the parameter vector of the best fitting or approximating model under the constraint of Model(p), defined in equation (5.3). We are interested in finding the p that minimizes $-2N\,H(\boldsymbol{\beta}^*;\boldsymbol{\beta}_p)$. As above, $\boldsymbol{\beta}_p$ is unobservable. We find the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_p$ of $\boldsymbol{\beta}_p$ by describing the data with Model(p). Akaike's method is based on the fact that larger values of p will result in smaller values $l(\hat{\boldsymbol{\beta}}_p)$ or a "better" fit, regardless of the true model. We need to "penalize" for larger values of p. Akaike finds that

$$\mathrm{E}\left[-2l(\hat{\boldsymbol{\beta}}_p)\right] \approx -2N\,H(\boldsymbol{\beta}^*;\boldsymbol{\beta}_p) + 2p$$

This fact leads to the Akaike Information Criteria which is a bias corrected estimate of the loglikelihood given by

$$\mathrm{AIC}(p) = -2l(\hat{\boldsymbol{\beta}}_p) + 2p \tag{5.10}$$

See, for example, Akaike (1973), Akaike (1974), Bozdogan (1987) for the details.

We now generalize this criteria to the weighted case. Here we have that functional parameter defining the distribution of y_n depends on the regression variable $\mathbf{x}_n$. Assume that for each $\mathbf{x}_n$ there is a true parameter $\boldsymbol{\beta}^*(\mathbf{x}_n)$ and that the distribution of y_n is as defined by (5.1). At the moment we are interested only in estimating $\boldsymbol{\beta}^* = \boldsymbol{\beta}(\mathbf{x}_{N/2})$ and for the estimation we assume the functional parameter is fixed $\boldsymbol{\beta}(\mathbf{x}_n) = \boldsymbol{\beta}$. The Kullback-Leibler Information Quantity is then

$$I(\boldsymbol{\beta}^*;\boldsymbol{\beta}) = \sum_{n=1}^{N} I_n(\boldsymbol{\beta}^*(\mathbf{x}_n),\boldsymbol{\beta})$$

where

$$I_n(\boldsymbol{\beta}^*(x_n),\boldsymbol{\beta}) = \mathrm{E}\left[\log f(y_n|\mathbf{x}_n,\boldsymbol{\beta}^*(\mathbf{x}_n)) - \log f(y_n|\mathbf{x}_n,\boldsymbol{\beta})\right] \quad n = 1,\ldots,N$$

where for each n the expectation here is taken under the true model defined for y_n above.

In this case, to obtain a useful criteria, it seems appropriate to keep the true distribution constant by assuming $\boldsymbol{\beta}^*(\mathbf{x}_n) = \boldsymbol{\beta}^*$ and by defining a weighted information quantity

$$wI(\boldsymbol{\beta}^*;\boldsymbol{\beta}) = \frac{1}{W_q}\sum_{n=1}^{N} w_{q,n} I_n(\boldsymbol{\beta}^*;\boldsymbol{\beta})$$

with

$$I_n = \int \left[\log f(y_n|\mathbf{x}_n,\boldsymbol{\beta}^*) - \log f(y_n|\mathbf{x}_n,\boldsymbol{\beta})\right] f(y_n|\mathbf{x}_n,\beta)\, dy_n$$

We want to choose the estimate $\hat{\boldsymbol{\beta}}_{p,q}$ that minimizes

$$\mathrm{E}[wI(\boldsymbol{\beta}^*;\hat{\boldsymbol{\beta}}_{p,q})]$$

We justify this model selection method by intuitive reasoning, namely since we consider $\boldsymbol{\beta}(\mathbf{x})$ to be "smooth", it seems reasonable to consider a weighted version of the original method as is done in local likelihood estimation. A Bayesian justification could be possible but we leave that as future work.

As before the larger p, the smaller $wl(\hat{\boldsymbol{\beta}}_{p,q})$ will be regardless of the model. Furthermore, generally the smaller the span of the weight coefficients h_q, the smaller $wl(\hat{\boldsymbol{\beta}}_{p,q})$ will tend to be. To see this, consider the case where h_q is made small enough such that the number of observations receiving positive weight is less than or equal to p. The estimates in this case would probably be unreasonable, yet we would have a "perfect fit". We will derive a weighted version of AIC, for the case of weighted regression presented in section 5.2, that penalizes for both large values of p and small values of h_q.

To be able to obtain a specific criteria, consider the case where f is the normal density. Then we have that

$$\begin{aligned} wI(\boldsymbol{\beta}^*;\boldsymbol{\beta}) &= \frac{1}{2\sigma^2 W_q}\sum w_{q,n}(\mathbf{x}_n'\boldsymbol{\beta} - \mathbf{x}_n\boldsymbol{\beta}^*)^2 \\ &= \frac{1}{2\sigma^2 W_q}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(\mathbf{X}'\mathbf{W}\mathbf{X})(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \end{aligned}$$

If $\hat{\boldsymbol{\beta}}_{p,q}$ is the weighted least squares estimate $(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{W}_q\mathbf{Y}$ then it has variance matrix

$$\boldsymbol{\Sigma} = \sigma^2(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}\mathbf{X}_p'\mathbf{W}_q\mathbf{W}_q\mathbf{X}_p(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}$$

and we have that

$$\begin{aligned} 2\mathrm{E}[wI(\boldsymbol{\beta}^*;\hat{\boldsymbol{\beta}}_{p,q})] &= 2\mathrm{E}[(\hat{\boldsymbol{\beta}}_{p,q}-\boldsymbol{\beta}^*)'(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)(\hat{\boldsymbol{\beta}}_{p,q}-\boldsymbol{\beta}^*)] \\ &= \delta + \frac{1}{\sigma^2 W_q}\mathrm{tr}\left[(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)\boldsymbol{\Sigma}\right] \\ &= \delta + \frac{1}{W_q}\mathrm{tr}\left[(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}(\mathbf{X}_p'\mathbf{W}_q\mathbf{W}_q\mathbf{X}_p)\right] \\ &= \delta + V_{p,q}/W_q \end{aligned}$$

with

$$V_{p,q} = \mathrm{tr}[(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)^{-1}(\mathbf{X}_p'\mathbf{W}_q\mathbf{W}_q\mathbf{X}_p)]$$

the *equivalent number of parameters*, as defined above , $\boldsymbol{\beta}_{p,q} = \mathrm{E}[\hat{\boldsymbol{\beta}}_{p,q}]$ and,

$$\delta = \frac{1}{\sigma^2 W_q}(\boldsymbol{\beta}_{p,q}-\boldsymbol{\beta}^*)'(\mathbf{X}_p'\mathbf{W}_q\mathbf{X}_p)(\boldsymbol{\beta}_{p,q}-\boldsymbol{\beta}^*)$$

We are interested in estimating $2\mathrm{E}[wI(\boldsymbol{\beta}^*;\hat{\boldsymbol{\beta}}_{p,q})]$. Consider the weighted likelihood ratio statistic

$$\mathrm{wLR}(\mathbf{Y}) = -\frac{2}{W_q}\sum_{n=1}^{N} w_{q,n}\log\frac{f(y_n|\mathbf{x}_n,\hat{\boldsymbol{\beta}}_{p,q})}{f(y_n|\mathbf{x}_n,\hat{\boldsymbol{\beta}}_{P,q})}$$

For the normal case we have that

$$E[\mathrm{wLR}(\mathbf{Y})] = \delta + (V_{P,q}-V_{p,q})/W_q$$

This suggest that an estimate of $2\mathrm{E}[wI(\boldsymbol{\beta}^*;\hat{\boldsymbol{\beta}}_{p,q})]$ to consider is

$$\mathrm{wLR} + (2V_{p,q}-V_{P,q})/W_q \tag{5.11}$$

Now we can choose the values of p and q that minimize (5.11). By eliminating the constant terms for each model being compared, we find that a procedure to choose the number of parameters and the appropriate window matrix is to minimize the criterion

$$\mathrm{wAIC}(p,q) = -\frac{2}{W_q}\sum_{n=1}^{N} w_{q,n}\log f(Y_n|\mathbf{x}_n,\hat{\boldsymbol{\beta}}_{p,q}) + 2V_{p,q}/W_q$$

as a function of p and q.

In the case that f is the normal density we have

$$\mathrm{wAIC}(p,q) = \frac{2}{W_q}\sum_{n=1}^{N} w_{q,n}\left(\frac{1}{2}\log 2\pi + \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}(Y_n-\mathbf{x}_n\hat{\boldsymbol{\beta}}_{p,q})^2\right) + 2V_{p,q}/W_q$$

$$
\begin{aligned}
&= \log 2\pi + \log \sigma^2 + \frac{1}{\sigma^2 W_q} \sum_{n=1}^{N} w_{q,n} (Y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}_{p,q})^2 + 2V_{p,q}/W_q \\
&\approx \log 2\pi + \log \hat{\sigma}^2 + \frac{1}{\hat{\sigma}^2 W_q} \sum_{n=1}^{N} w_{q,n} (Y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}_{p,q})^2 + 2V_{p,q}/W_q
\end{aligned}
$$

Using the weighted residual mean-square estimate of the variance

$$\hat{\sigma}^2_{p,q} = \sum_{n=1}^{N} w_{q,n} (Y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}_{p,q})^2 / (W_q - V_{p,q}) \tag{5.12}$$

and removing the constants, the criteria reduces to

$$\text{wAIC}(p,q) = \log \hat{\sigma}^2_{p,q} + V_{p,q}/W_q$$

5.5 Weighted BIC

As we saw in the derivation of both the AIC and the wAIC, an important characteristic of the criteria is the penalty term they contain for too many parameters and in the case of the wAIC, for a small equivalent number of observations. This penalty permits us to choose between estimates without fitting too many parameters or using estimation windows that are too small. In the case of AIC a problem is that minimizing AIC does not produce an asymptotically consistent estimate of the model order, see for example Schwarz (1978).

Bhansali and Downsham (1977) also discuss this problem and consider using a constant other than 2 for the penalty factor in equation (5.10). The different constants are compared via simulation. Although this method seems arbitrary, it might prove to be practical.

Schwarz (1978), Akaike (1979), and Kashyap (1982) use a Bayesian approach to derive a criteria that is consistent. The criteria is asymptotically independent of the particular prior specification. This approach to model selection is based on the posterior probabilities of the alternative models, given the observations (Sclove 1994). The criterion is

$$\text{BIC}(p) = -2l(\hat{\boldsymbol{\beta}}_p) + 2p(\log N + 1) \tag{5.13}$$

Bozdogan (1987) points out that when the mean log likelihood is used to estimate the Kullback-Leibler information quantity, the bias introduced by the maximum likelihood estimates of the parameters needs to be corrected for. In the derivation of AIC, this bias

comes out as a noncentrality parameter δ, which is an unknown but deterministic constant. It depends not only on the number of observations, but also on the specific estimation method. Moreover, the change in δ over different sample sizes is also important in order to justify the correction of the bias further. That is, we do not want to have a very large δ, since it varies with the basic model. As is well known, noncentrality parameters determine the power of test procedures, and the estimation of δ on the basis of preliminary data may be necessary to choose among competing models.

We note from equation (5.11) that one such correction in δ is already given in deriving the wAIC, that is, $\delta \approx \text{wLR} - (V_{P,q} - V_{p,q})/W_q$. Also, we note that the correction factor $(V_{P,q} - V_{p,q})/W_q$ is independent of the sample size N. However, in testing a null hypothesis (or a model) distinguishing from the alternative hypothesis by the value of a parameter, if the test statistic has a non-central chi-square distribution, which is the case in the non-weighted case, then the degrees of freedom is an increasing function of the sample size N, see Kendal and Stuart (1967). This suggest that to make wAIC consistent, the multiplier of the number of free parameters in the penalty term must be made to depend on the sample size, e.g. by setting the penalty to

$$a(N)(V_{P,q} - V_{p,q})$$

where $a(N)$ is an increasing function of N. In wAIC, we note that $a(N) = 1$. As discussed in Davis and Vinter (1985), the selection of the function $a(N)$ is important, and it should be chosen so that it has various desirable properties for the corresponding estimates. One of the choices suggested is $a(N) = \log N$. Notice that in equation (5.13) $a(N) = \log N + 1$. We therefore suggest the following criteria

$$\text{wBIC} = -\frac{2}{W_q}\sum_{n=1}^{N} w_{q,n} \log f(Y_n|\mathbf{x}_n, \hat{\boldsymbol{\beta}}_{p,q}) + V_{p,q}(\log N + 1) \tag{5.14}$$

In the normal case it reduces to

$$\text{wBIC} = \log \hat{\sigma}^2 + (V_{p,q}/W_q) \log N$$

5.6 Linear approximation

We have presented three criteria for the case of linear regression. Now we will show how these criteria can be used in the case of the local harmonic model.

In the context of Chapter 4, we will fix the sample rate N and the estimation point t_0. Say that we chose a maximum span H such that we are not willing to consider any interval bigger than $(t_0 - H/2, t_0 + H/2)$ as appropriate for the approximately sinusoidal assumption for the signal. Assume that the parameter function is constant within that interval, say

$$\boldsymbol{\beta}(t) = \boldsymbol{\beta} \text{ , for } t \in (t_0 - H/2, t_0 + H/2)$$

This implies that we are willing to assume that there is a fixed subset of the data $\mathbf{Y}_H$

$$\{Y_{j,N}, j = n_0 - J_N/2, \ldots, n_0 + J_N/2\}$$

with expected value

$$E[Y_{j,N}] = s_{j,N} = s(\frac{j}{N}, \boldsymbol{\beta})$$

Here $J_N = \lfloor H \times N \rfloor$ and $n_0 = \lfloor t_0 N \rfloor$.

Define the vector $\mathbf{s} = \{s_j = s_{j,N}\}$ for $j = 1, \ldots, J_N$. By considering a linear approximation of the signal, we can now express the model as linear regression. The linear approximation will be

$$\mathbf{Y}_H \approx \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Where $\mathbf{X}$ is the gradient matrix defined by

$$\mathbf{X} = \frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}$$

Say we are considering different spans $h_1, \ldots, h_Q < H$ and different numbers of partials $k = 1, \ldots, K$. We can decide on which h_q and k to use in the estimation of $\boldsymbol{\beta}$ using the criteria defined above.

5.6.1 Simplification of $V_{p,q}$

Notice that to compute the value $V_{p,q}$ we need to perform matrix multiplications and inversions. When N and p are large this can be computationally expensive. It is useful to find an approximation that saves some computational work.

For the case where the function $s(t; \boldsymbol{\beta})$ is as defined by (3.47) we can compute the derivatives and use Lemma 1 from Chapter 3 to obtain a useful approximation of $V_{p,q}$ when N is sufficiently large, namely

$$V_{p,q} \approx p\frac{U_0}{W_0} + \frac{2W_1[W_1U_0/W_0 - U_1] + W_0U_2 - W_2U_2}{W_2W_0 - W_1^2} \tag{5.15}$$

where the constants, W_n and U_n for $n = 0, 1, 2,$ are defined in equation (3.8) for the function $w(u/h_q)$ and the number of parameters by $p = 2K + 1$, where K is the number of partials. Notice that when we use the square window $w(t) = 1$ this approximation reduces to p, the number of parameters, as expected.

5.7 Simulations

The lack of a precise theoretical justification and the fact that generally we are dealing with relatively small values of N motivates the use of simulations to check the effectiveness of the model selection methods presented above. In this section, we illustrate some simulation results for the determination of the number of partials and appropriate window sizes. It must be remembered that the simulations results are approximations for particular parameter values, yet they often prove helpful in studying complex procedures.

For the simulation we want to use a signal that is representative of some musical sound. We try to imitate a 50 millisecond stretch of the sound signal of a clarinet playing A4. In Figure 5.1 we see that it appears to be approximately sinusoidal, leading to a model of the form

$$y_n = \sum_{k=1}^{K} \rho_k \cos(k\lambda n + \phi_n)$$

We "synthesize" the simulation signal in the following way: Based on the number of clear peaks appearing in the periodogram of the clarinet signal, seen in Figure 2.4, we choose $K = 15$ as the number of partials in the simulation signal y_n. Next, we find the weighted estimates of the harmonic model for the stretch of the original signal following the procedure of Chapter 3. The amplitude estimates found are then chosen as the true amplitudes ρ_k of the simulated signal. The phases ϕ_k of the simulated signal are chosen at random from a uniform distribution on $[0, 2\pi]$. The resulting model for the simulation signal is

$$\begin{aligned} y_n \;=\; & 122\cos(\lambda n + 2.65) + 20.3\cos(2\lambda n + 3.11) + 18.8\cos(3\lambda n + 4.56) \\ & +5.97\cos(4\lambda n + 0.452) + 4.09\cos(5\lambda n + 1.76) + 5.37\cos(6\lambda n + 6.13) \\ & +2.96\cos(7\lambda n + 3.2) + 1.52\cos(8\lambda n + 4.85) + 0.66\cos(9\lambda n + 6.16) \\ & +0.25\cos(10\lambda n + 4.49) + 0.564\cos(11\lambda n + 2.96) + 0.63\cos(12\lambda n + 0.753) \\ & +0.27\cos(13\lambda n + 4.5) + 0.36\cos(14\lambda n + 1.91) + 0.28\cos(15\lambda n + 4.87) \\ & n = 1, \ldots, N \end{aligned} \tag{5.16}$$

Here the ϵ_t are i.i.d normal with mean 0 and variance 1. The amplitudes have been rescaled so that the overall signal to noise ratio, defined by

$$\text{SNR} = \left(\frac{1}{2} \sum_{k=1}^{K} \rho_k^2 \right) / \text{Var}(y_t)$$

is the same as the estimated signal to noise ratio for the original signal, namely $\hat{\text{SNR}} = 7896$. We are now ready to create simulation signals by generating ϵ_n's. In Figure 5.1 we see a comparison of the original stretch of signal and a simulation signal.

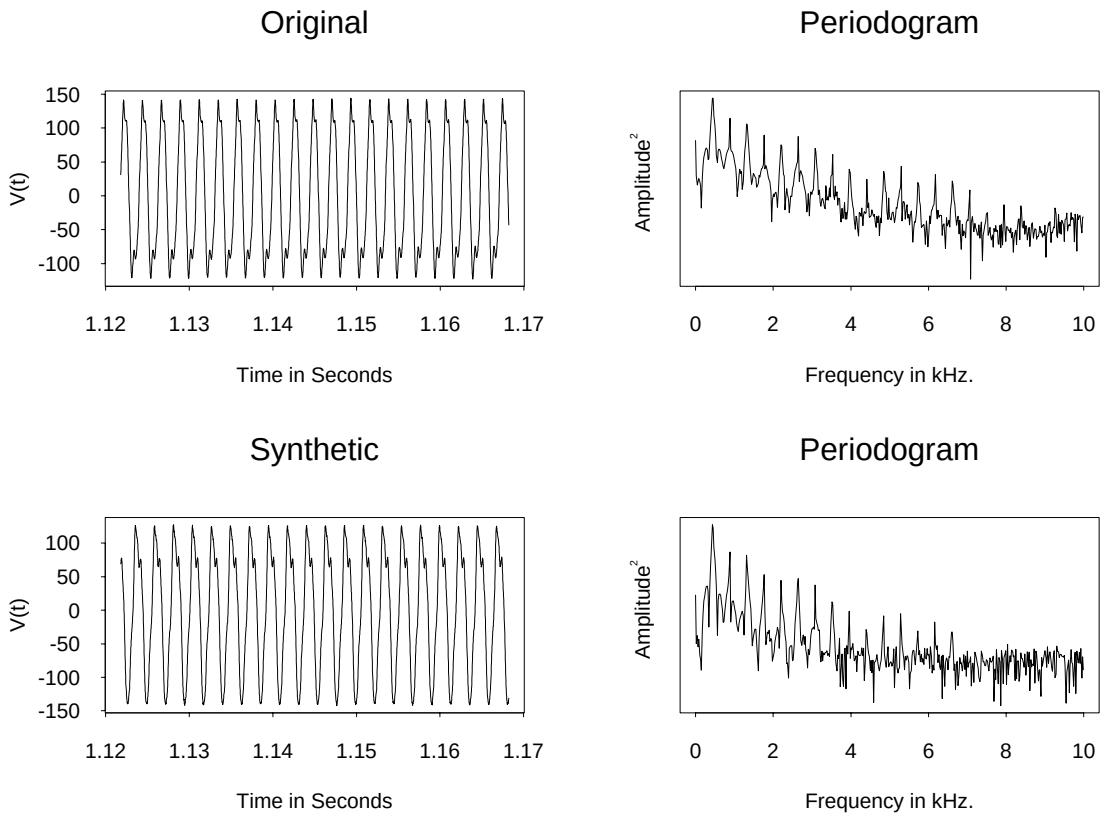

Figure 5.1: Comparison of original and synthetic clarinet signals.

5.7.1 Choosing the number of partials

To study the effectiveness of our criteria at estimating the number of partials K we did the following: For a given value of N (number of observations) we simulated 1000 signals y_n using model (5.16). Since the original signal was of a clarinet playing concert

pitch A we let $\lambda = 440$ Hz. For each y_n we fit a harmonic model with $K = 1, \ldots, 48$ partials using Tukey's triweight window $w(t) = (1 - |2t - 1|^3)^3_+$. For each $p = 2K + 1$ we used the weighted residual mean-squares estimate $\hat{\sigma}^2_p$ to compute the criteria

$$\log \hat{\sigma}^2_p + a(N)(V_{p,q}/W_q)$$

with $a(N)$ the *penalty factor*. Notice that for $a(N) = 1$ the criteria is equivalent to the wAIC and when $a(N) = \log N$ it is equivalent to the wBIC. We compared the *hit rate* (percentage of time the criteria was minimized at $K = 15$) for $a(N) = 0$, 1, 2, 4, $\log N$, $2 \log N$ and $4 \log N$. We repeated this experiment for $N = 440$, 880, 2200, and 4400. Notice that if we sample a sound signal at 44.1 kHz, the number of data points in 10, 20, 50 and 100 milliseconds are approximately $N = 440$, 880, 2200, and 4400 respectively. The results are seen in Table 5.1.

Penalty	Number of observations			
Factor	N=440	N=880	N=2200	N=4400
0	46.6%	34.8%	35.3%	25.0%
1	79.2%	81.0%	80.4%	64.9%
2	80.1%	93.8%	93.5%	88.4%
4	53.4%	97.4%	99.4%	98.9%
$\log N$	30.4%	92.4%	100.0%	100.0%
$2 \log N$	0.5%	45.1%	99.0%	100.0%
$4 \log N$	0.0%	0.1%	59.2%	100.0%

Table 5.1: Hit rate for the estimated number of partials.

Notice that for larger values of N the wBIC performs better than the other criteria. In fact, for $N = 2200$ and $N = 4400$ the wBIC achieved a perfect hit rate. The wAIC seems to be performing better for the smaller N. Further simulation for other cases, e.g. other fundamental frequencies, are left as future work.

5.7.2 Choosing the window size

In this section we will study the effectiveness of our criteria in estimating appropriate window sizes when the parameters of the simulation model are not constant in time.

$$y_n = s(n, \beta_n) + \epsilon_n, \ n = 1, \ldots, N \tag{5.17}$$

We use a similar model for $s(n, \beta_n)$ to that of equation (5.16) specifically let

$$\begin{aligned} s(n,\beta_n) &= a_n \{122\cos(\lambda_n n + 2.65) + 20.3\cos(2\lambda_n n + 3.11) + 18.8\cos(3\lambda_n n + 4.56) \\ &+ 5.97\cos(4\lambda_n n + 0.452) + 4.09\cos(5\lambda_n n + 1.76) + 5.37\cos(6\lambda_n n + 6.13) \\ &+ 2.96\cos(7\lambda_n n + 3.2) + 1.52\cos(8\lambda_n n + 4.85) + 0.66\cos(9\lambda_n n + 6.16) \\ &+ 0.25\cos(10\lambda_n n + 4.49) + 0.564\cos(11\lambda_n n + 2.96) + 0.63\cos(12\lambda_n n + 0.75) \\ &+ 0.27\cos(13\lambda_n n + 4.5) + 0.36\cos(14\lambda_n n + 1.9) + 0.28\cos(15\lambda_n n + 4.9)\} + \epsilon_n \end{aligned}$$

Again the ϵ_t are i.i.d normal with mean 0 and variance 1.

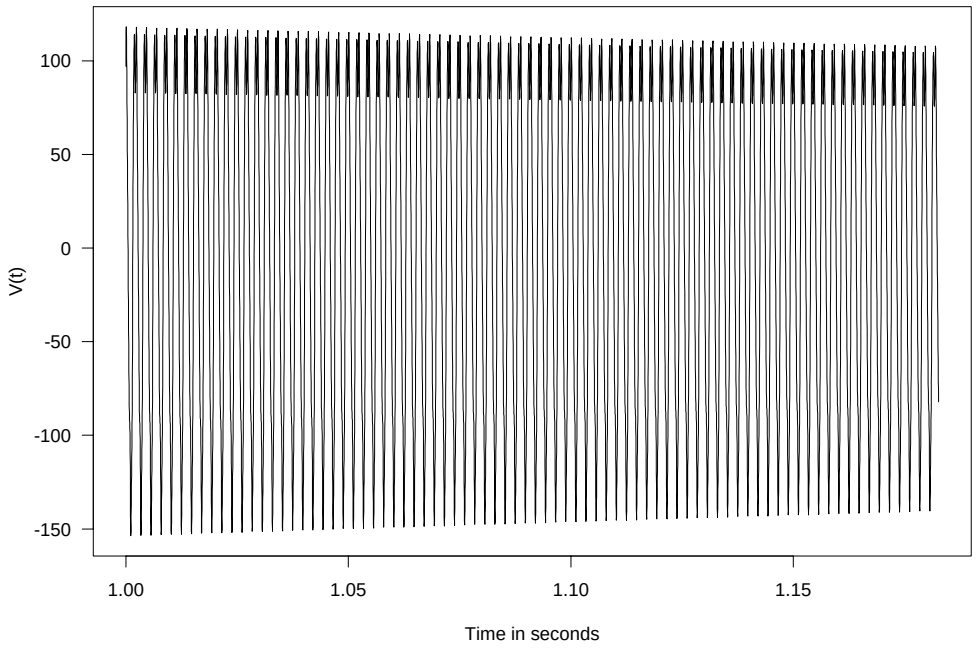

Figure 5.2: A simulated decay signal.

Given a deterministic signal $s_n = s(n, \beta_n)$, for example Figure 5.2, we hope that our criteria will choose the window size for which the mean squared error (MSE) of the estimate $\hat{y}_{N/2}$ is minimized (since we are simulating ϵ_n as Gaussian, this is equivalent to minimizing the Kullback-Leibler Information Criteria). To test this we do the following: We simulate 1000 signals y_n using model (5.17). For each y_n we fit a harmonic model with

15 partials using Tukey's window $w(t) = (1 - |2t/h_q - 1|^3)^3_+$ with different spans h_q. For each q we use the weighted residual mean-square (wRMS) $\hat{\sigma}^2_q$ to compute the wAIC(q) and the wBIC(q). We then compute the average over all simulation of the quantities wRMS(q), wAIC(q), and wBIC(q). We consider a criteria to be working well if the q that minimizes it is close to the q that minimizes the MSE.

In the three following examples we see that the wBIC performs as well or better than the other two criteria.

Case of a stable note

The first example is simply a stable note with $a_n = 1$ and $\lambda_n = 440$ Hz. for all n. In this case the signal being consider is the same as that of equation (5.16). In Figure 5.3 we plot the MSE and the average values of the wRMS, wAIC and wBIC against the span h_q. We see that the wAIC and wBIC, on average, choose the "correct" window size.

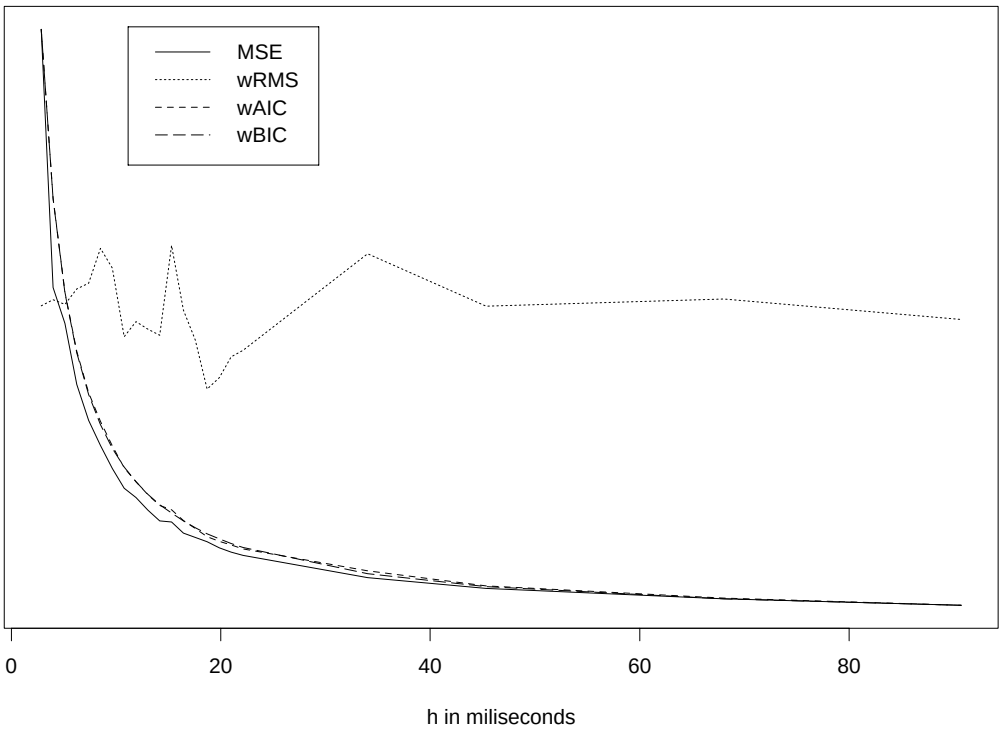

Figure 5.3: Comparison of wRMS, wAIC and wBIC in the case of a stable note.

Change of pitch

The second example tries to imitate a change of pitch. The amplitude is held constant, $a_n = 1$, but the fundamental frequency function λ_n is defined in the following way

$$\lambda_n = \begin{cases} \lambda_1 & : \quad 0 < n <= N' \\ \lambda_2 & : \quad N' < n < N \end{cases}$$

We ran the simulation with $\lambda_1 = 440$ Hz., $\lambda_2 = 466.1638$, $N = 2161$ and $N' = 2053$. The musical interpretation of this is that a clarinet starts playing concert pitch A for about 47 milliseconds and then changes to A♯ (a semitone above the previous note). In Figure 5.4 we plot the MSE and the average values of the wRMS, wAIC and wBIC as before. The symbols $\times$, $\triangle$, $\Diamond$ and $\square$ denote the location of the value that minimizes the MSE, wRMS,wAIC and wBIC respectively. In average, the wAIC and wBIC perform better at choosing a window size than the wRMS.

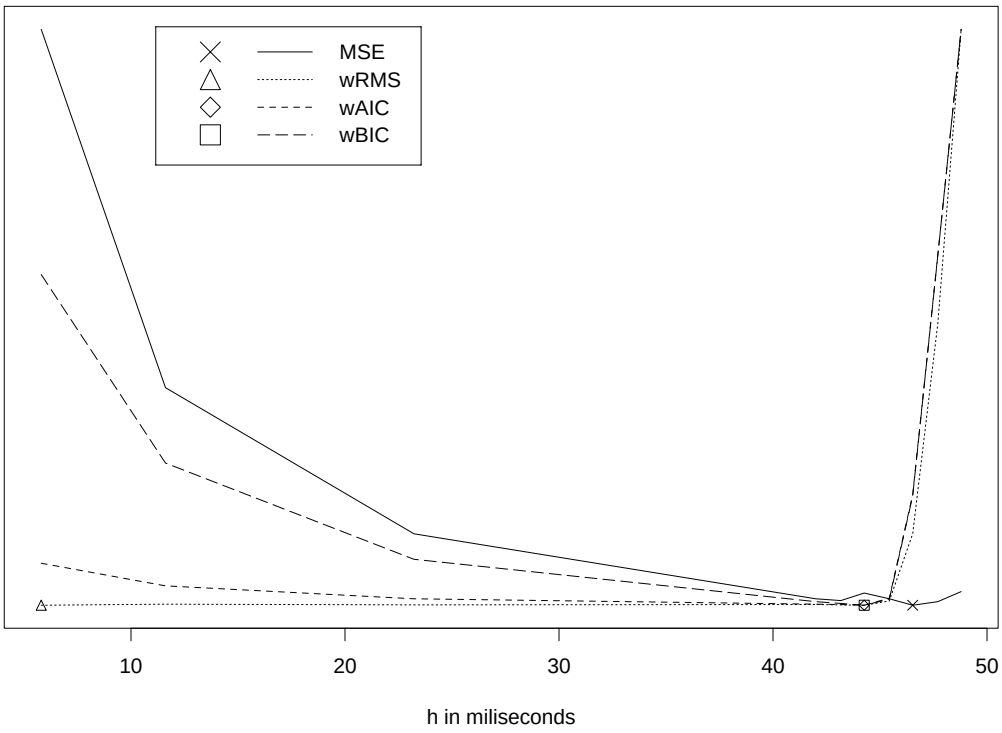

Figure 5.4: Comparison of weighted RMS, AIC and BIC for a change of pitch.

Decaying amplitude

Finally, we construct signals keeping the frequency constant, $\lambda_n = 440$ Hz, but letting a_n be a decaying exponential.

$$a_n = \exp\{-\alpha n\}$$

To construct the signals choose $N = 8055$ and $\alpha = 0.5/44100$. The musical interpretation is that we have a signal with similar characteristics to the part of the signal corresponding to the decay. In figure 5.2 we see the simulated signal y_n used for the simulation. In figure 5.5 we see the result of the simulation. The symbols $\times$, $\triangle$, $\Diamond$ and $\square$ denote the location of the value that minimizes the MSE, wRMS,wAIC and wBIC respectively. Notice that the wBIC performs the best.

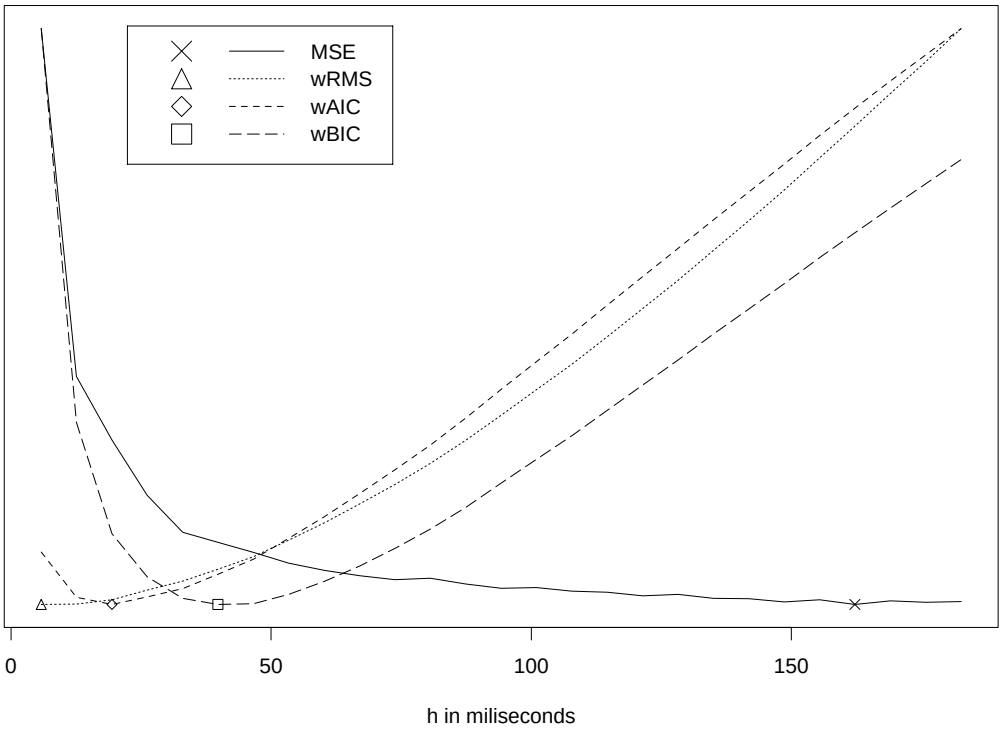

Figure 5.5: Comparison of weighted RMS, AIC, and BIC for a decay.

Chapter 6

Finale: Examples and Possible Compositional Uses

6.1 Introduction

A variety of sound signals were recorded in a recording studio at the Center for New Music and Audio Technology (CNMAT). These recordings were converted into digital signals using an Analog to Digital Converter (ADC). Software synthesis programs generated a sound file as their output (a sound file is simply a data file stored on a disk). After all the samples for a composition are calculated the sound file can be played and heard through a Digital to Analog Converter (DAC). For the details on how sound can be recorded into computer sound files see Roads (1996), Wagner (1978).

Many different sound file formats exist. The differences between such formats are mainly the sample rate at which the sound signal was sampled and the number of bits per stored sample. This information is usually contained in the header of such files. The sound files used in this work were sampled at 44.1 kHz. and used 16-bit words. The particular format used by the SGI computers, at CNMAT, is the Audio Interface File Format (aiff). The CNMAT aiff files are in stereo (two channels are recorded). For this chapter's work only the left channel was used and for these files the left and right channels were practically equal. The aiff sound files used in this work, are available via ftp from the CNMAT server `cnmat.berkeley.edu`.

The sound files can be converted into text files readable to Splus using the sound

exchange program `sox` (free software available for UNIX and MS-DOS on the Internet). Similarly, any data set contained in a text file can be converted into a sound signal.

An Splus program performing the estimation procedure described in Chapters 3 and 4 was used to analyze sound signals. In practice the procedure worked very well at separating the harmonic signal from the non-sinusoidal signal. In this chapter we will review some of the examples that we found most interesting.

The procedure was tested with many sound signals including examples produced by a clarinet, guitar, oboe, pipe organ, tenor saxophone, shakuhachi flute, trumpet, and violin. In each one of these cases the sound produced by the estimated harmonic part of the signal sounds very much like the original signal. The sound produced by the residuals is what we might expect the non-sinusoidal part of the signal to sound like. For example, in the case of a saxophone, the sound produced by the amplified residuals is similar to the sound of air and spit going through a tube (track 15 on accompanying CD). The estimates of the functional parameters and non-sinusoidal part of the signal provide ways to create new sounds based on the original.

We also incorporated into the Splus program a dynamic window selection procedure that uses the wBIC criteria, described in Chapter 5, to choose amongst different window sizes for each estimation time. In practice the procedure worked very well at selecting smaller window sizes in parts of the signal where the functional parameter appeared not to be near constant.

The two fundamental model, described in Chapter 6, proved to be useful in the case of a sound signal with reverberation. By fitting a local harmonic model to each tone included in the sound signal we were able to obtain estimates that appear to provide a separation of the two tones from each other and from the non-sinusoidal part of the signal.

6.2 Appropriateness of local fitting

A basic goal of our analysis is to estimate the functional parameter $\beta(t)$ of Chapter 4. First we need to check if the locally approximately constant assumption is sensible for sound signals. We illustrate appropriateness of local fitting with an example.

For the sound signal of a violin, sampled at 44.1 kHz. ($N = 44100$), playing the note C4 we consider a 50 millisecond stretch (2200 observations) around $t_0 = 1.145$ seconds. Notice in Figure 6.1 how the signal, within the considered segment, does seem

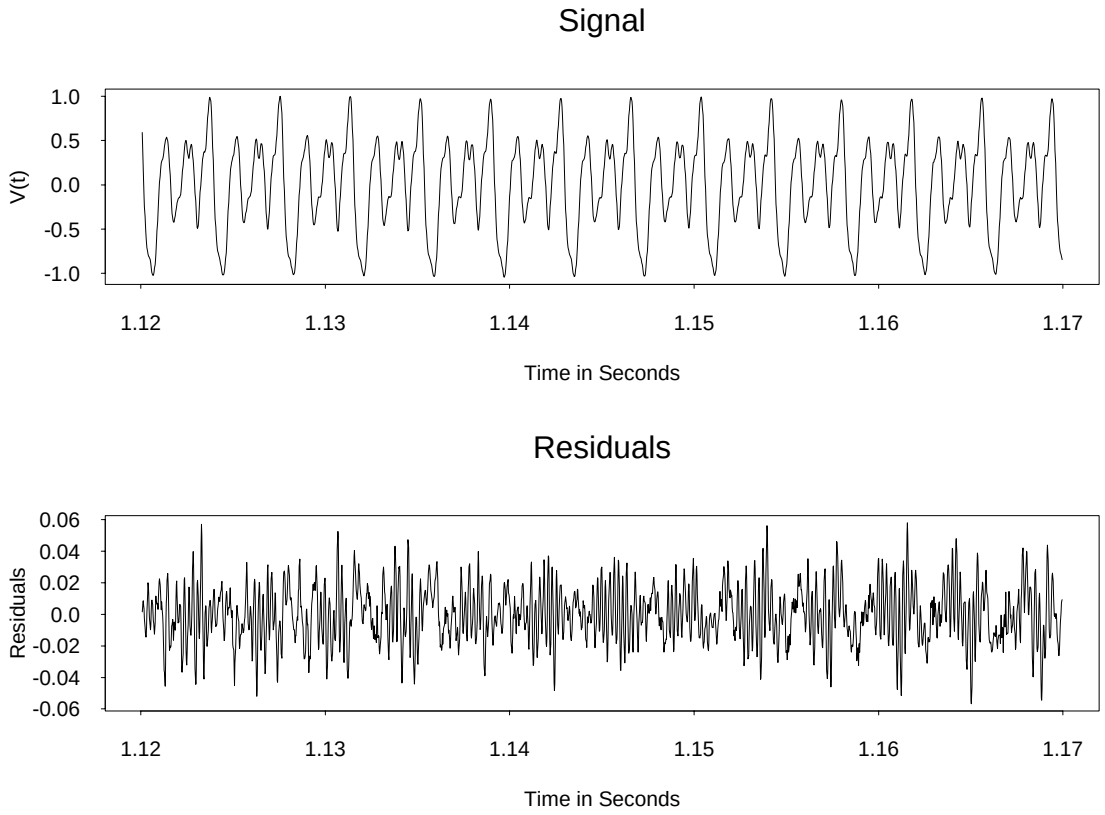

Figure 6.1: Local fit for the sound signal of a violin playing C4 and corresponding residuals.

to be approximately a sum of sinusoids, i.e. that the parameter function $\beta(t)$ seems to be usefully constant.

We fit a harmonic model, like that of equation (3.47), with 15 partials (K=15) to the stretch of data presented in Figure 2.7 and obtain a reasonable fit. The weighted residual mean-square is $\hat{\sigma}^2 = 0.0003$. Comparing this to the weighted estimate of the variance of the original signal $\sum_{n=1}^{N} w(n/N)y_n^2 / \sum_{n=1}^{N} w(n/N) = 0.26$ shows that the fitted model explains a large amount of the variation of the original signal.

The residual plot, also seen in Figure 6.1, suggests that the noise could be considered stationary in the given stretch. The periodogram of the residuals and a *smoothed periodogram* estimate of the spectrum $f_{\epsilon\epsilon}$ are shown in Figure 6.2. The value of m in equation (3.56) for the smooth periodogram estimate is 12.

Using the asymptotic approximation for the variance of the estimated parameters,

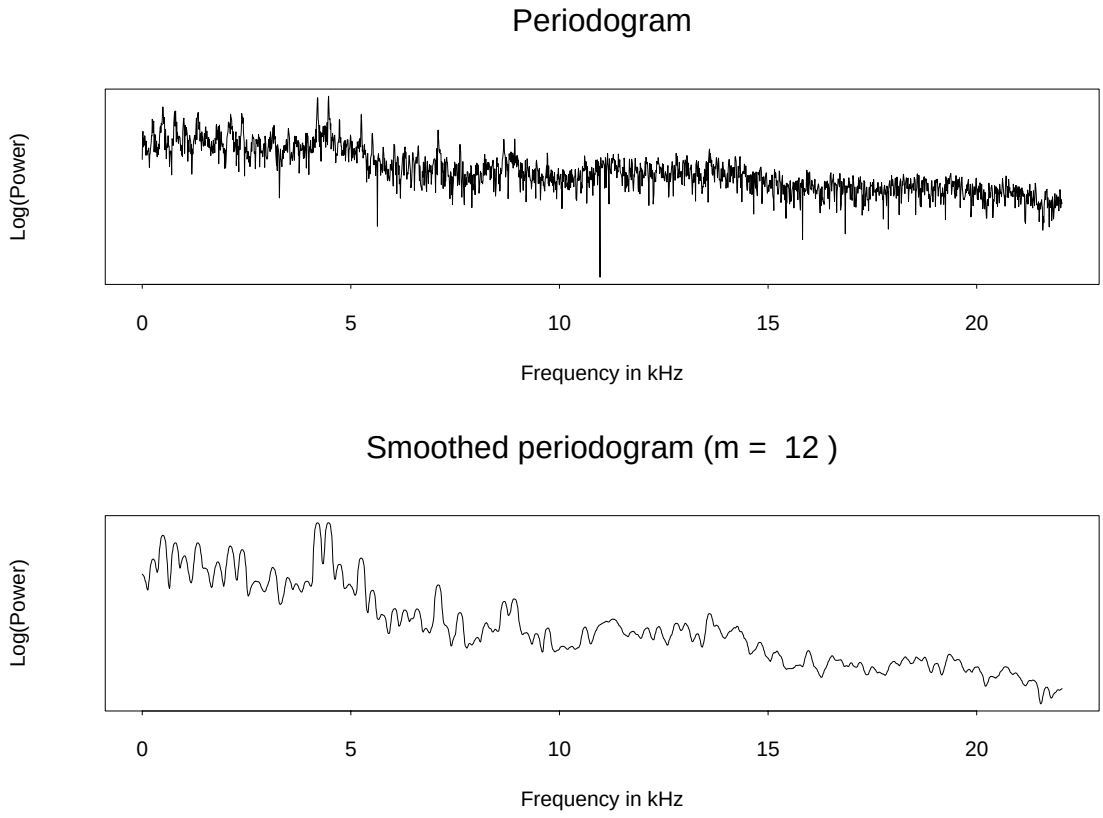

Figure 6.2: Smoothed periodogram estimates for the spectrum of the noise for the sound signal of a violin playing C4.

given by equation (3.50), and the smoothed periodogram estimate of the spectrum, shown in Figure 6.2, we can give estimated standard errors for the estimates. In Table 6.1 we see the estimates for all the components of β and their respective estimated standard errors.

The estimate of the fundamental frequency is 262.795 Hz. The note being played is C4. If the instrument is tuned to A 440 Hz. concert pitch then C4 is equivalent to 261.6256 Hz. Notice that our estimate is not included in the approximate 99% confidence interval around the frequency corresponding to C4, $[261.6590, 261.5921]$. Is the violin out of tune? In section 6.4 we discuss possible musical interpretations of standard errors.

Estimate	Value	$\widehat{\text{se}}$	Estimate	Value	$\widehat{\text{se}}$
$\hat{A}_1$	0.246	0.00121	$\hat{B}_1$	0.0999	0.00126
$\hat{A}_2$	-0.294	0.00249	$\hat{B}_2$	0.0344	0.00267
$\hat{A}_3$	0.546	0.00212	$\hat{B}_3$	-0.137	0.00338
$\hat{A}_4$	0.0218	0.00128	$\hat{B}_4$	0.0554	0.00124
$\hat{A}_5$	-0.0642	0.00196	$\hat{B}_5$	0.0246	0.00202
$\hat{A}_6$	0.0194	0.000846	$\hat{B}_6$	0.0284	0.00082
$\hat{A}_7$	-0.0198	0.00134	$\hat{B}_7$	0.0718	0.00108
$\hat{A}_8$	-0.0637	0.00239	$\hat{B}_8$	-0.119	0.00198
$\hat{A}_9$	-0.0189	0.00164	$\hat{B}_9$	-0.0422	0.00154
$\hat{A}_{10}$	-0.00115	0.000503	$\hat{B}_{10}$	-0.00235	0.000502
$\hat{A}_{11}$	-0.00487	0.000444	$\hat{B}_{11}$	-0.00768	0.00043
$\hat{A}_{12}$	-0.00964	0.000889	$\hat{B}_{12}$	-1.56e-05	0.000909
$\hat{A}_{13}$	-0.00138	0.000421	$\hat{B}_{13}$	-0.00152	0.00042
$\hat{A}_{14}$	-0.00508	0.000546	$\hat{B}_{14}$	-0.00557	0.000543
$\hat{A}_{15}$	-0.00363	0.000579	$\hat{B}_{15}$	-0.000779	0.000586
$\hat{\lambda}$	262.796	0.010578			

Table 6.1: Parameter and standard error estimates for the local harmonic model.

6.2.1 Heuristic window size and number of partials selection

For our estimation procedure to make sense we need to consider appropriate stretches of sound signals. For each local fit we need to find a stretch of signal that is appropriate for the approximately sinusoidal assumption. Rodet and Depalle (1992) use stretches of 256 data points. When the sample rate is 44.1 kHz. this is equivalent to about 5.8 milliseconds of sound. Since we are fitting a model with up to hundreds of parameters ($2 \times$ number partials $+ 1$) it seems appropriate to use longer stretches of data. In fact, if we are interested in estimating the parameter function at t_0, it is convenient to find long stretches of signal around that point where the approximately sinusoidal assumption seems appropriate. To do this we may examine plots of such stretches to determine which is more appropriate.

In Figure 6.3 we see two stretches of the signal of a violin playing C4 around time $t_0 = 0.17$ seconds, one with a duration of 20 milliseconds and the other with a duration of 50 milliseconds. Notice that in the second plot the functional parameter doesn't appear to be approximately constant, but rather that the amplitude is growing with time. Looking at the residuals, also seen in Figure 6.3, produced from fitting a harmonic model with 15

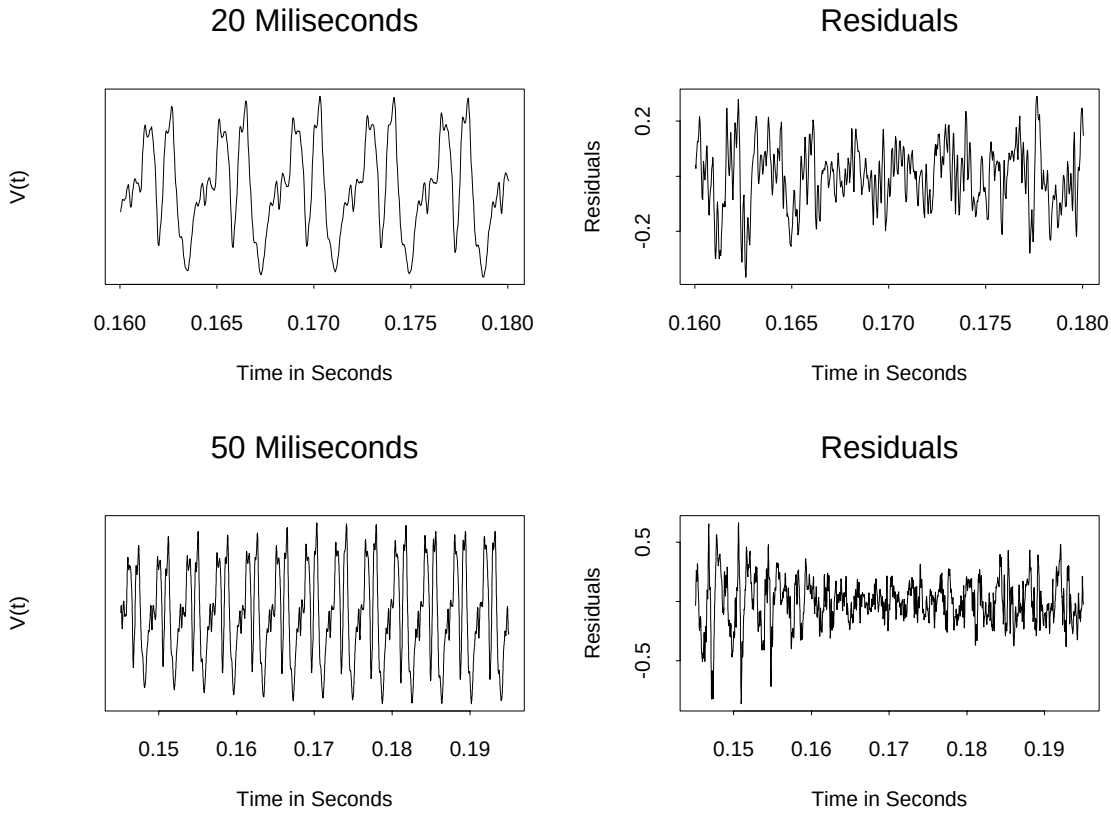

Figure 6.3: Comparison of two stretches of different duration of the sound signal of a violin playing C4.

partials, we see that they do not appear to be stationary for the 50 millisecond stretch. In this case we would pick the 20 millisecond stretch over the 50 millisecond one.

Another decision that we need to make is how many partials K to consider in our model. Previous estimation procedures, (Serra 1989, Depalle et al. 1993a), usually fit many partials. If the sample rate is N Hz. and the fundamental frequency is λ Hz. at least $N/2\lambda$ partials (frequencies above $N/2$ are aliased) are considered. Fitting too many parameters may result in estimates that are hard to interpret. We need to decide how many partials to include in our model.

The number of "peaks" in the periodogram plot may be used to obtain a general idea of how many partials to consider. As we saw in Figure 2.4, the clarinet seems to have less "significant" harmonics than the trumpet. We may also use the estimation procedure described in Chapter 3 in the following way.

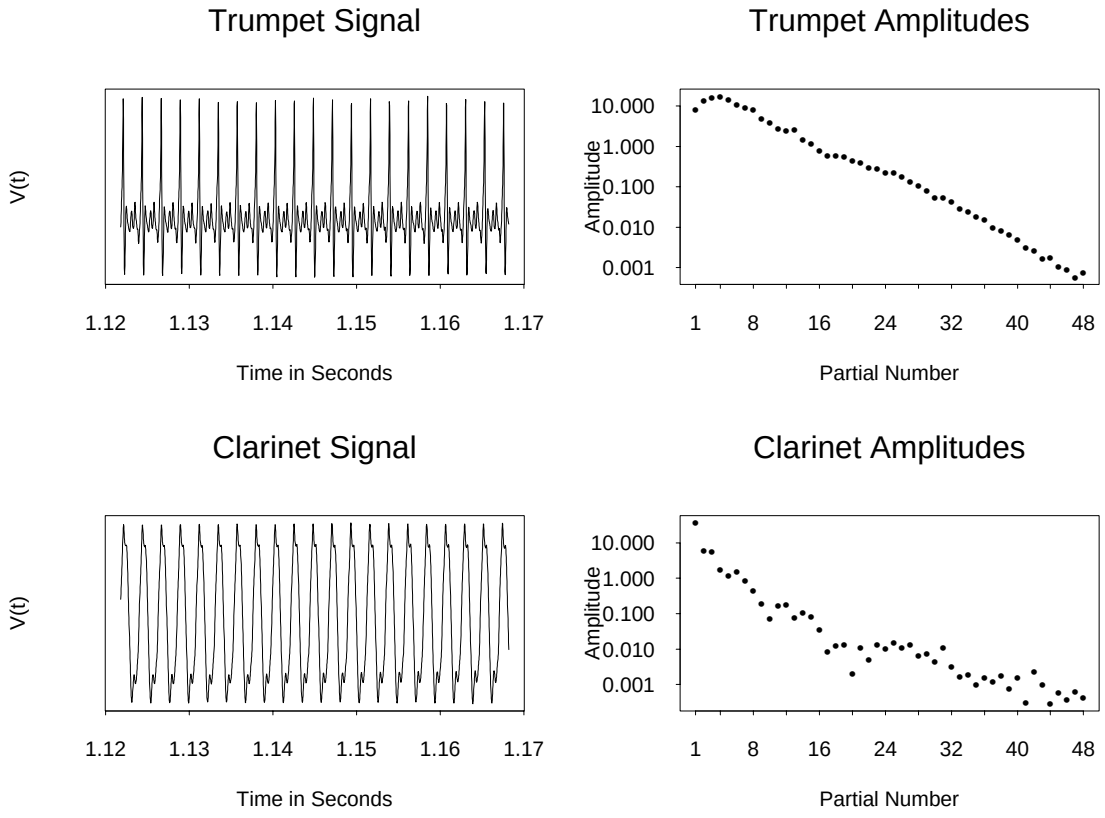

Figure 6.4: Stretches of 50 millisecond duration of the sound signals of a trumpet and a clarinet playing concert pitch A and estimated amplitudes

We examine stretches of about 50 millisecond durations of a clarinet and a trumpet playing concert pitch A (440 Hz.). The plots in Figure 6.4 suggest that the approximately sinusoidal assumption is appropriate in these stretches. We fit a harmonic model with 48 partials to each one. In Figure 6.4, we also see the estimated amplitudes for each partial of the harmonic model. Notice that the amplitudes estimated for the higher partials are close to 0, relative to the amplitude estimated for the lower partials.

In Figure 6.5 we see the estimates of the higher amplitudes surrounded by 99% confidence intervals. The confidence intervals are constructed using the asymptotic approximation for the variance of the estimates given by equation (3.54). Notice that in some cases 0 is included in such intervals. As we expected the clarinet has fewer amplitudes that appear *statistically significant*.

We may test the hypothesis that the amplitude estimates are 0. Using the asymp-

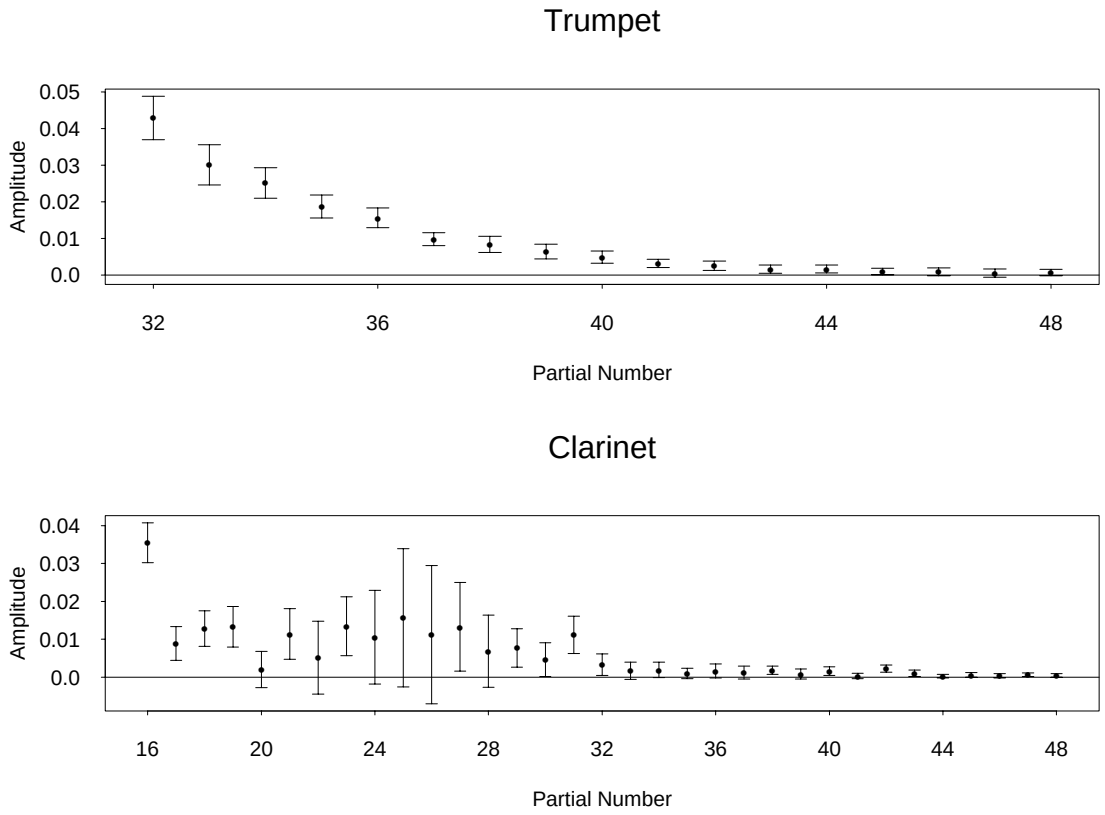

Figure 6.5: Estimated amplitudes of higher partials with estimated 99% confidence intervals.

totic normality result of equation (3.55) we obtain an estimate for the standard error of $\hat{\rho}_k$

$$\widehat{\mathrm{se}}(\hat{\rho}_k) = \sqrt{4\pi c_1 \hat{f}_{\epsilon\epsilon}(k\hat{\lambda})/T}$$

Using this, we construct a z-test by defining the z-statistic for each k by

$$z_k = \frac{(\hat{\rho}_k - 0)}{\widehat{\mathrm{se}}(\hat{\rho}_k)}$$

The results of Chapter 4 suggest that under the assumption that p_k is 0, the z_k's are approximately normal. In Table 6.2 we present the p-values obtained from the z-statistics for $k = 21, \ldots, 48$ when estimating using different window sizes. The p-values that are higher than 0.01 are in bold-faced. Notice that for the smaller window size $N = 256$, we obtain that for 17 out of the 27 partials the hypothesis is not rejected at the 1% level. In the case of the larger window sizes $N = 1024$ and $N = 2048$, we obtain 8 and 10 respectively.

A problem with this z-test is that the amplitudes ρ_k are non-negative and that

Partial	p-value			Partial	p-value		
number	N=256	N=1024	N=2048	number	N=256	N=1024	N=2048
21	**0.091**	0.000	0.000	35	**0.011**	0.010	**0.014**
22	**0.202**	0.007	**0.055**	36	**0.248**	0.001	0.005
23	0.000	0.000	0.000	37	**0.065**	**0.191**	**0.020**
24	**0.206**	0.010	0.005	38	0.000	0.000	0.000
25	**0.017**	0.009	0.005	39	**0.019**	**0.029**	**0.035**
26	0.007	**0.034**	**0.033**	40	0.000	0.000	0.000
27	0.006	0.000	0.000	41	**0.015**	**0.234**	**0.104**
28	0.004	**0.135**	**0.015**	42	0.000	0.000	0.000
29	**0.102**	0.000	0.000	43	**0.020**	0.002	0.000
30	0.001	0.000	0.001	44	**0.013**	0.004	**0.036**
31	0.002	0.000	0.000	45	**0.151**	**0.029**	0.002
32	**0.058**	0.000	0.000	46	**0.015**	**0.034**	**0.025**
33	**0.236**	**0.082**	**0.014**	47	0.006	0.000	0.000
34	0.001	0.000	0.002	48	**0.097**	0.000	0.003

Table 6.2: P-values when testing if an amplitude estimate is 0.

they could be correlated. We do not intend to use hypothesis testing as a tool to choose how many partials to include in our model, but rather as a descriptive illustration of why we need to consider different values for different sound signals. In the next section we present the wBIC as a criteria to decide among different choices.

6.2.2 Using the wBIC

We have presented heuristic ways of determining appropriate window sizes and number of partials to consider when performing estimation. In Chapter 5, we introduced a criteria that can help. In the examples presented in this section, the wBIC of equation (5.5) is used to automatically decide how many partials to use in our model and how big a window size to consider for the estimation.

In the previous section we saw how for a clarinet signal the assumption that the parameter is locally constant appeared reasonable within a 50 millisecond temporal window around $t_0 = 1.15$ seconds. Thus, for the sound signal stretch of a clarinet playing A4 (presented in Figure 6.4) we fit 48 different models, one for each of the vales $K = 1, \ldots, 48$. We pick 48 as the maximum number of partials because the fundamental frequency is around 440 Hz. and 48×440 Hz. $= 21120$ Hz. which is close to the Nyquist frequency

22050 Hz. Notice that 50 is the highest possible number of partials that we may consider without the partials exceeding 22050 when the fundamental frequency is 440 Hz., but since the fundamental frequency estimate may be a bit higher than 440 Hz. we are safe by considering 48 as the maximum. The resulting wBIC criteria for each one of the competing models are shown in Figure 6.6. Notice that the model with 15 partials minimizes the criteria.

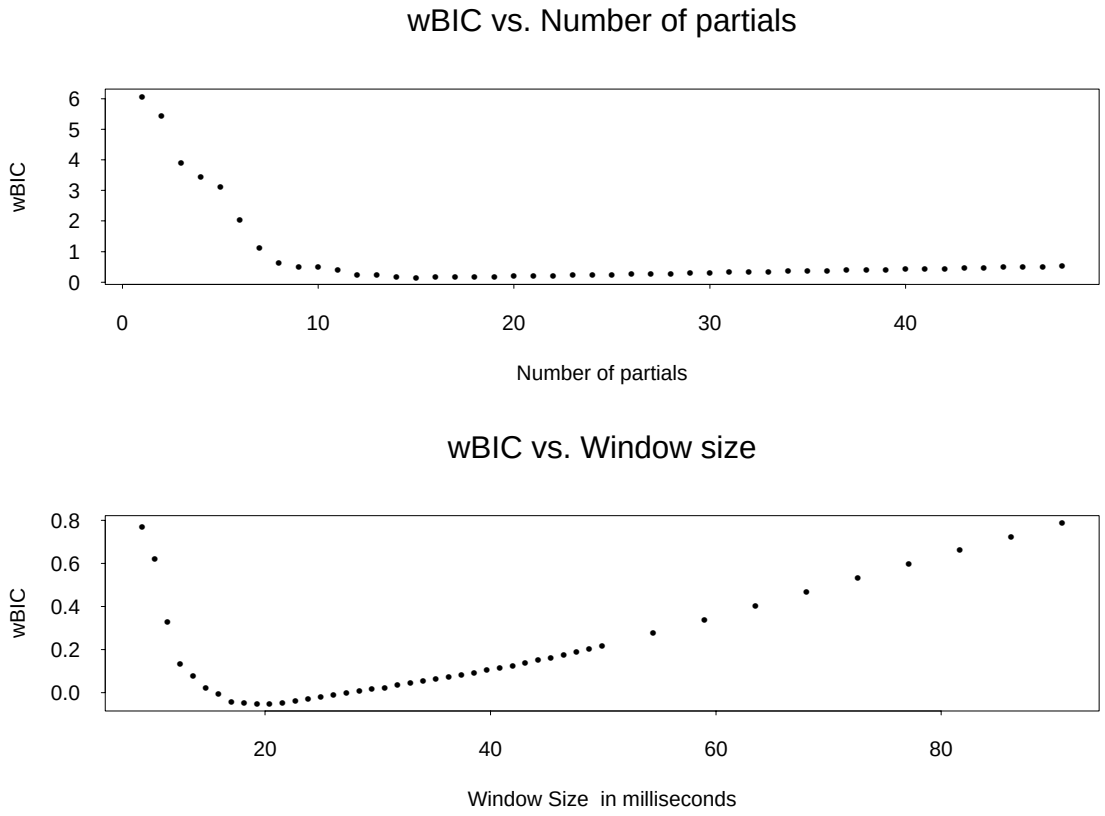

Figure 6.6: wBIC values attained by fitting models with different number of partials and using different window sizes to a sound signal stretch of a clarinet playing A4.

This information may be useful in practice. Existing additive synthesis methods track many partials. Fitting only 15 will make estimation procedures faster and might possibly even provide more accurate estimates.

Now, assuming that our model has 15 partials, we can use the wBIC criteria to automatically choose a window size. We fit the harmonic model with 15 partials using different window sizes around $t_0 = 1.15$ and for each fit we calculate the wBIC. In Figure 6.6,

we see the resulting values of the wBIC for window sizes ranging from 1 to 100 milliseconds. Notice that the wBIC is minimized for spans around 20 milliseconds.

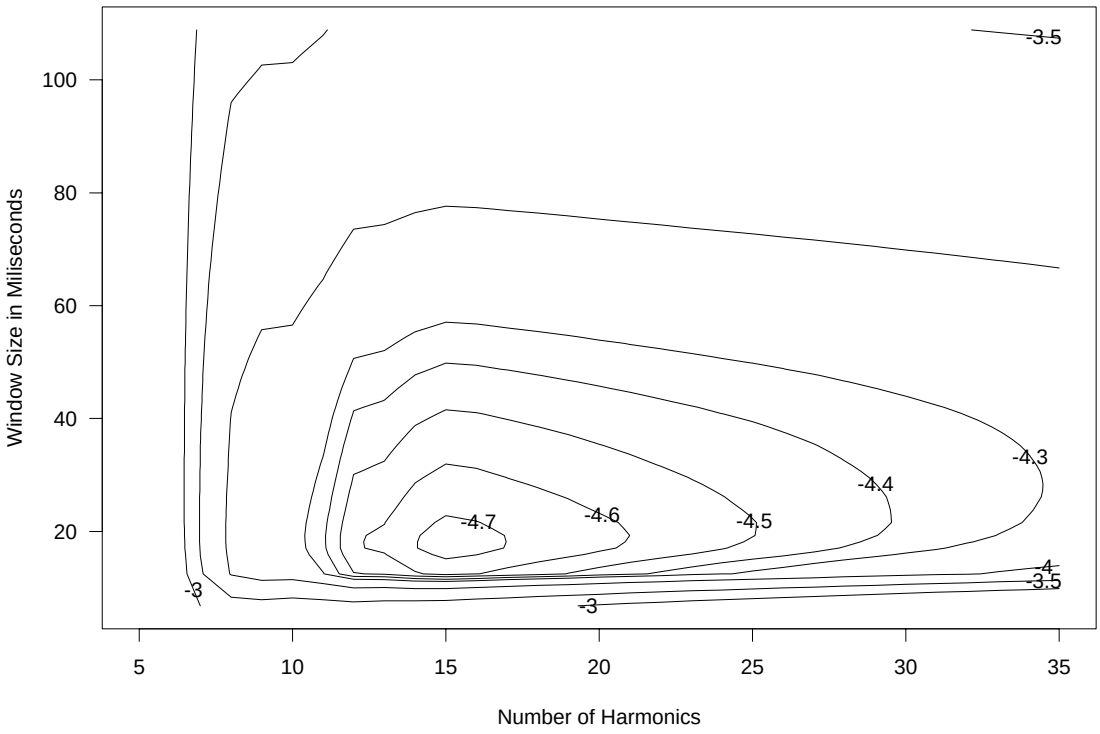

Figure 6.7: Contour plots of the wBIC when fitting with different numbers of partials and different window sizes for sound signal stretches of a clarinet playing A4 around $t_0 = 1.15$ seconds.

In Figure 6.6 we minimize the wBIC over p while leaving h_q constant at what we believe to be reasonable values, and vice versa. In Figure 6.7, we present a contour plot of the value of the wBIC for different pairs of values (p, q) for the number of partials and window sizes. Notice that the pairs that seem to minimizes the wBIC are around 15 partials and window sizes of about 20 milliseconds.

When the same instrument plays different notes the number of "significant" partials in the signal might change. A harmonic instrument playing C7 (2093.005 Hz.) will have as many as 10 harmonics that are below 22050 Hz. If the instrument plays C5 (523.2511 Hz.) it will may have as many as 41 harmonics that are below 22050 Hz. In general, we expect

lower pitch notes to have more partials. In Figure 6.8 we see the contour plot of the value of the wBIC for a violin playing C5 and C7. Notice that for C5 the wBIC chooses 17 partials and for C7 it chooses 6. In Chapter 2 we saw how for the case of a guitar sound higher harmonics "die off" more rapidly as time progresses, see Figure 2.6. This might suggest that we consider a harmonic model with less partials during the later part of the signal. In Figure 6.8 we see the contour plot of the value of the wBIC for two sets of stretches of the same guitar signal. The first set is taken from the beginning of the note (stretches around $t_0 = 0.40$ seconds), the second is taken from the end (stretches around $t_0 = 3.4$ seconds). Notice how in the first case 12 partials are chosen and in the second only 6.

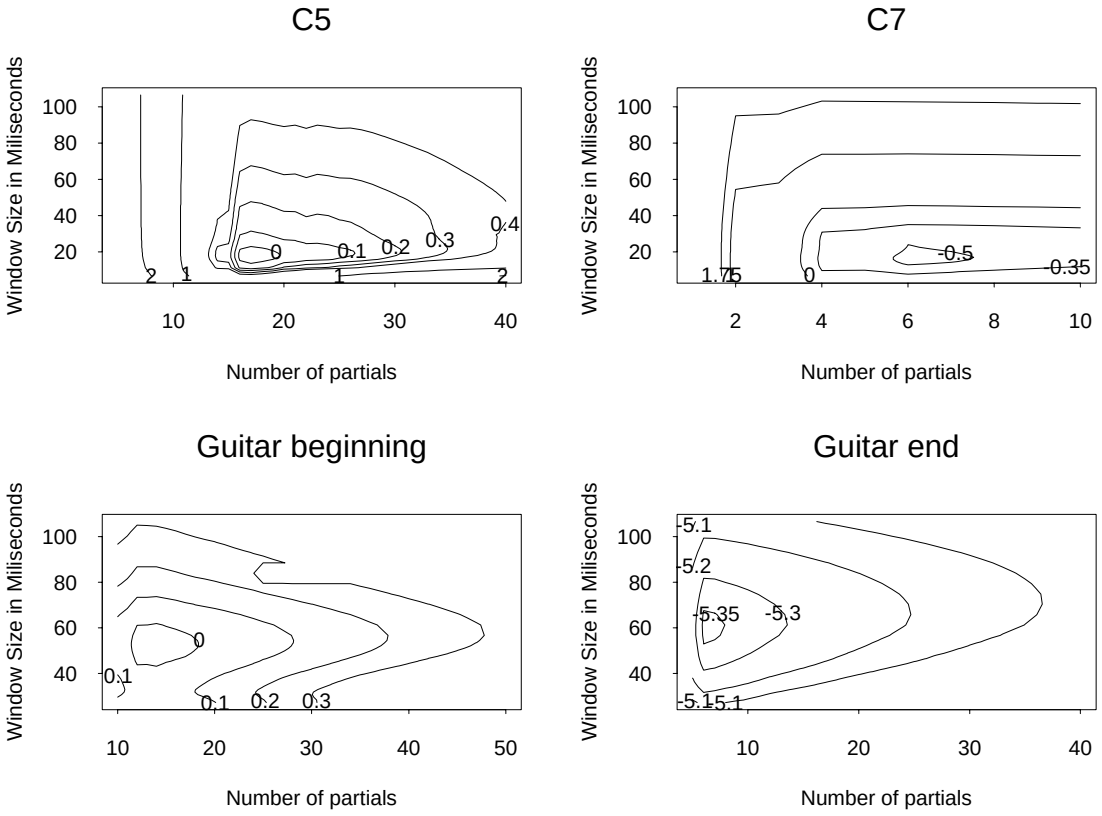

Figure 6.8: Contour plots of the wBIC for sound signal stretches of violin playing C5 and of a violin playing C7. Also for stretches at the beginning and end of the sound signal of a guitar playing D3.

6.3 Estimating the functional parameter

Given that local fitting appears to be a sensible way of estimating parameters of a harmonic model we can use the procedure of Chapter 4 to estimate the functional parameter $\beta(t)$ for all $t \in [0, 1]$. As an example we run the analysis on the sound signal of an oboe playing C4 (261.6256 Hz.) for a duration of 3 seconds. We obtain an estimate $\hat{\beta}(t)$ by keeping the number of partials in the model and estimation window sizes constant in time, at 15 partials and 20 milliseconds respectively. In Figure 6.9 we see the estimate $\hat{\lambda}(t)$ and the estimates for the amplitudes $\hat{\rho}_k(t) = \sqrt{\hat{A}_k^2 + \hat{B}_k^2}$ for $k = 1, \ldots, 5$. The estimated fundamental frequency is between 258 Hz. and 260.5 Hz, close to the frequency related to C4, (261.6256 Hz.). The oboist might be playing a bit out of tune. The dotted lines in Figure 6.9 are 3 cents away from the average fundamental frequency (259.25 Hz.).

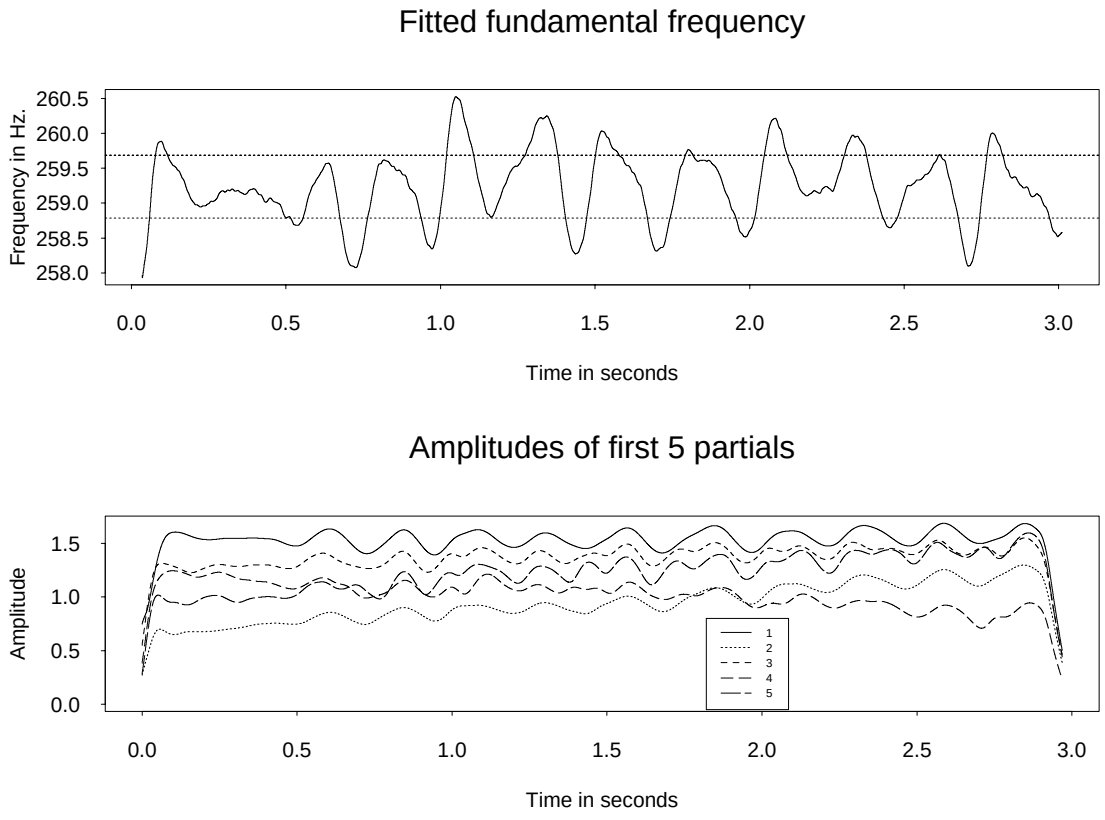

Figure 6.9: Estimated fundamental frequency and amplitude of first five partials for the sound signal of a violin playing C4.

The figure seems to suggest that there are variations in pitch perceivable to the

ear. This is in agreement with what we hear; the oboist is playing a *vibrato*, slight and rapid variations in pitch. Figure 6.9 also suggests that there are variation in the amplitudes of the first five partials. This is in agreement with what we hear; the oboist is playing a *tremolo*, slight and rapid variations in amplitude.

Once we find an estimate $\hat{\beta}(t)$ we can construct estimates for the harmonic part of the signal,

$$\hat{y}(t) = \sum_{k=1}^{\hat{K}} \{\hat{A}_k(t)\cos(k\hat{\lambda}(t)t) + \hat{B}_k(t)\sin(k\hat{\lambda}(t)t)\} \tag{6.1}$$

The estimate of the non-sinusoidal signal, represented in our model by the noise $\epsilon(t)$, is provided by the residuals

$$\hat{\epsilon}(t) = \hat{y}(t) - y(t) \tag{6.2}$$

In many of the cases studied, the sounds of the signals $y(t)$ and $\hat{y}(t)$ were almost indistinguishable. When amplified, the sound of residuals sounded much as we expected: specifically a sound like that of air and spit going through a tube for the saxophone, clarinet and trumpet, a screechy metallic sound for a violin, a pluck with no tone for the guitar, etc.. (tracks 16–30 on accompanying CD) We can assess the fit further by studying the residuals.

6.3.1 Residual analysis

We need a way to assess our estimation procedure. We may use spectrum and time-varying spectrum estimates based on the residuals to do so. Two types of residuals are available for the estimation of such quantities.

First we define the *global residuals* as the residuals obtained from subtracting the fitted signal from the original.

$$\hat{\epsilon}_n = Y_n - \hat{Y}_n = Y_n - s[\frac{n}{N}; \hat{\beta}(\frac{n}{N})] \text{ for } n = 1, \ldots, N$$

In Figure 6.10 we see the global residuals for the fit of a 0.20 second sound signal of a trumpet playing D♯4 (329.6276 Hz.). Notice that these residuals don't appear to be stationary.

Acting as if the series $\{\epsilon_n\}$ is stationary, we can estimate the spectrum $f_{\epsilon\epsilon}(\lambda)$ via a smoothed periodogram based on the series $\{\hat{\epsilon}_n\}$.

If instead we assume that $\{\epsilon_n\}$ is locally stationary, we may use the spectrogram defined in (4.6) as a basic estimate for the time-varying spectrum $f_{\epsilon\epsilon}(t, \lambda)$.

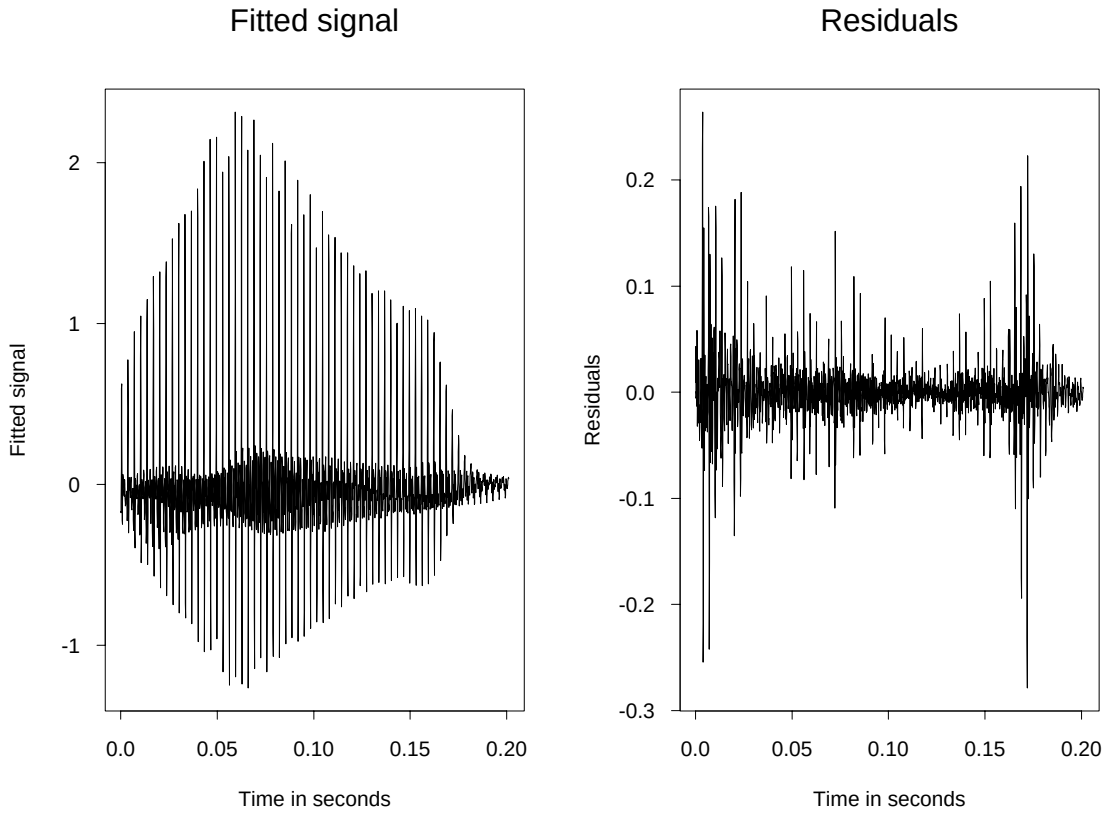

Figure 6.10: Global residuals and fitted signal for the sound signal of a trumpet playing D4.

The other type of residuals we can consider are the residuals obtained in the local fit for each $n_0 = 1, \ldots, N$. Define the *local residuals* as

$$\hat{\epsilon}_{j,n_0} = Y_j - s[\frac{j}{N}; \hat{\beta}(\frac{n_0}{N})] \text{ for } j = n_0 - J_N, \ldots, n_0 + J_N$$

Notice that we obtain a set of local residuals for each estimation time n_0/N. Note also that, in practice, we might be using different J_N's for each n_0 as described in Chapter 5.

Acting as if the series $\{\epsilon_n\}$ is stationary, we can estimate the spectrum $f_{\epsilon\epsilon}(\lambda)$ using the local residuals via the averaged periodogram. By letting

$$I^{J_N}(\lambda; n) = \frac{1}{2\pi J_N} \left| \sum_{|j| \leq J_N} \exp\{-i\lambda j\} \hat{\epsilon}_{j,n} \right|^2, \; 0 \leq \lambda \leq \pi$$

the averaged periodogram estimate of the spectrum of the noise is

$$\hat{f}_{\epsilon\epsilon}(\lambda) = \frac{1}{N}\sum_{n=1}^{N} I^{J_N}(\lambda; n)$$

If instead we assume that $\{\epsilon_n\}$ is locally stationary, we may construct an estimate of the time-varying spectral density by estimating $f(t, \lambda)$ with the smoothed periodogram estimate of local residuals of the fit at estimation time $n/N \approx t$.

In Figure 6.11 we present the two spectrum estimates and in Figure 6.12 we present the two time-varying spectral density estimates for the trumpet residuals of Figure 6.10. Both time-varying spectral density estimates suggest that the non-sinusoidal part of the signal is not stationary. Notice that the residuals seem to have sporadic spikes. This shows up in the local residual time-varying spectral density estimate as dark vertical lines. This seems to be a characteristic of the non-sinusoidal part of trumpet signals. The spectrogram of the global residuals seems to have "smoothed" out this characteristic.

In Figures 6.13 through 6.17 we present the plots of the global residuals and their spectrograms for a clarinet, guitar, oboe, and a violin. Figures 6.10 through 6.17 show that the power of the noise is stronger during the beginning of the note or the attack. This is in agreement with the physical theory discussed in Chapter 2. The horizontal dark lines seen in the spectrograms and the peaks in the spectrum estimates seem to suggest that the residuals have some sort of frequency components at the harmonic frequencies. This might suggest that the non-sinusoidal stochastic component is signal related in some way. The stochastic part of the signal might not be additive. In the case of many instruments it might not make sense to assume that the noise is additive (Maganza and Caussé 1986, Chafe 1990, Cook et al. 1990). The frequency components seen in the figures might also be the result of lack of fit, and the underlying periodic functions might not be sinusoids. Further assessment of the residuals is left as future work.

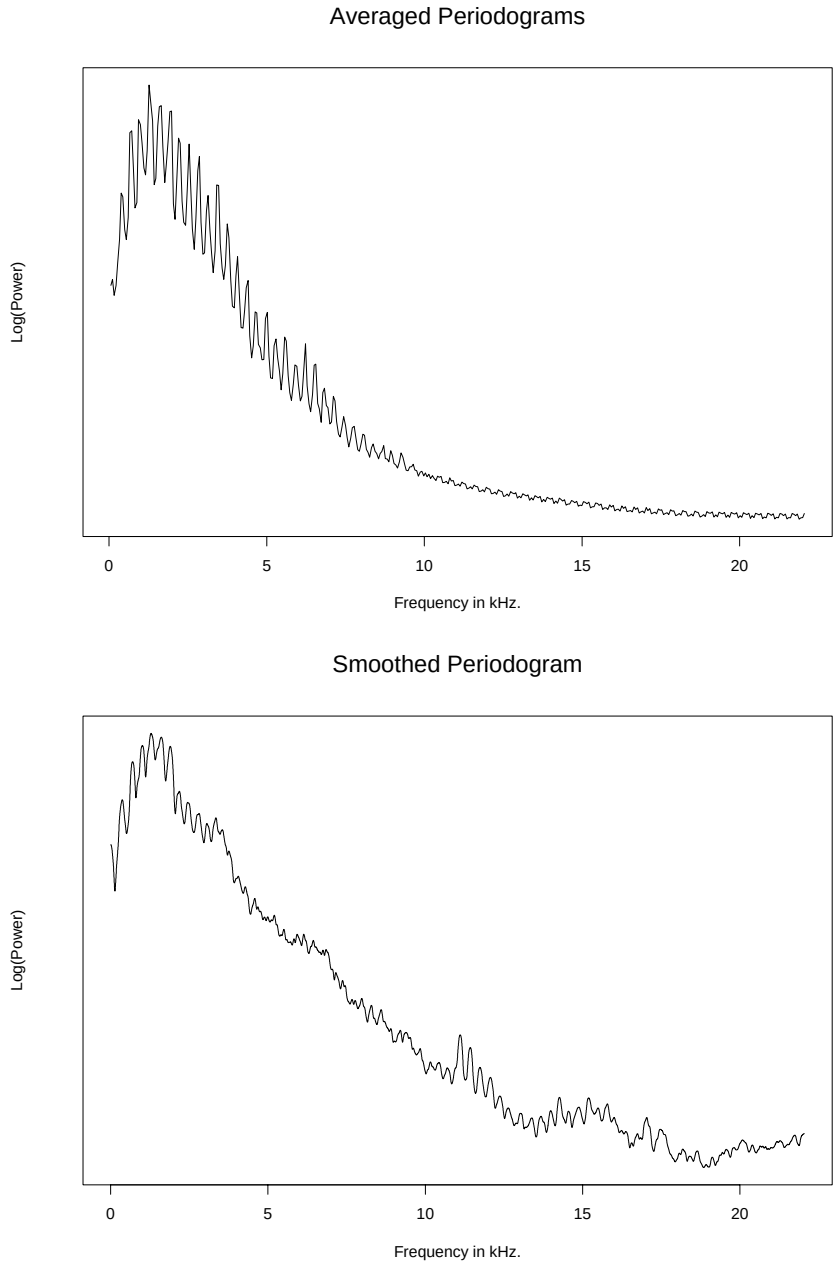

Figure 6.11: Averaged periodogram of the local residuals and smoothed periodogram of the global residuals for the sound signal of a trumpet playing D♯3.

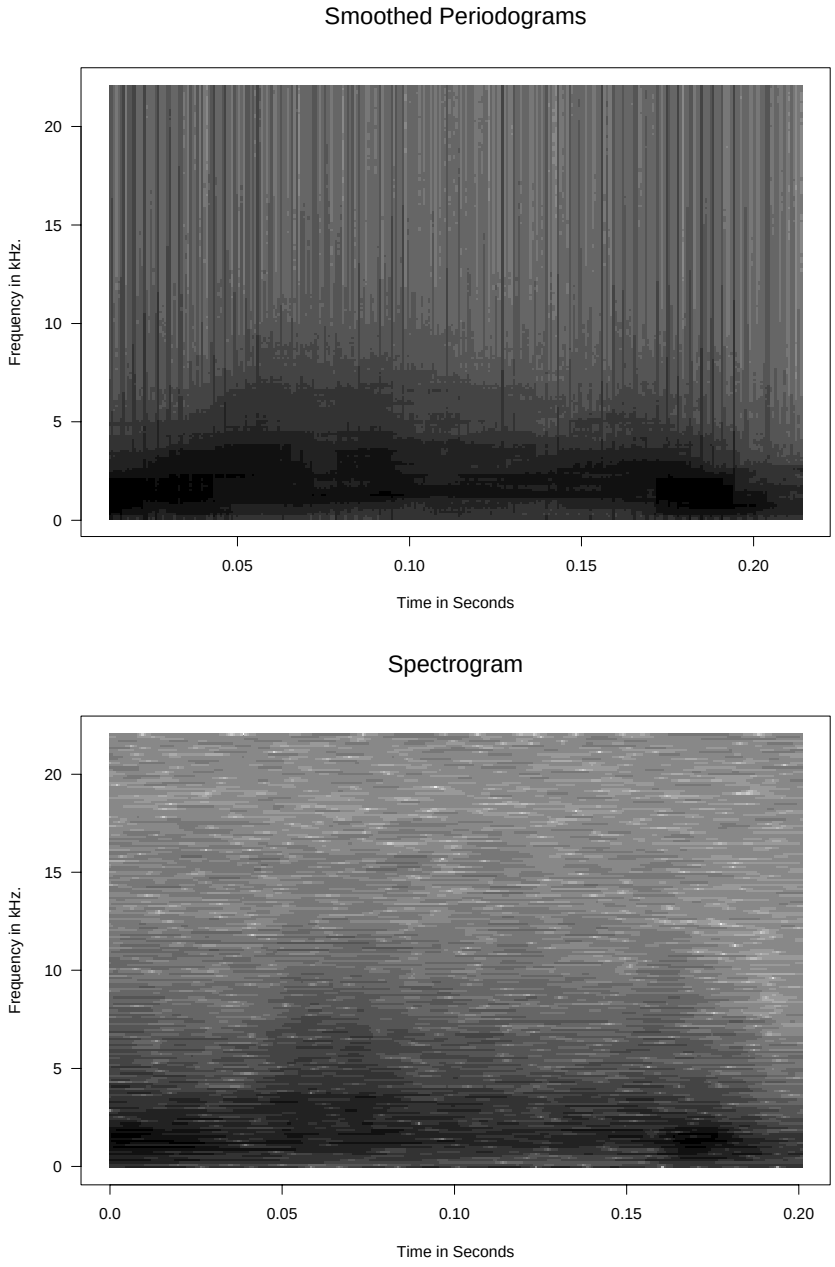

Figure 6.12: Smoothed periodograms of local residual for each sampled time and spectrogram of the global residuals for the sound signal of a trumpet playing D3.

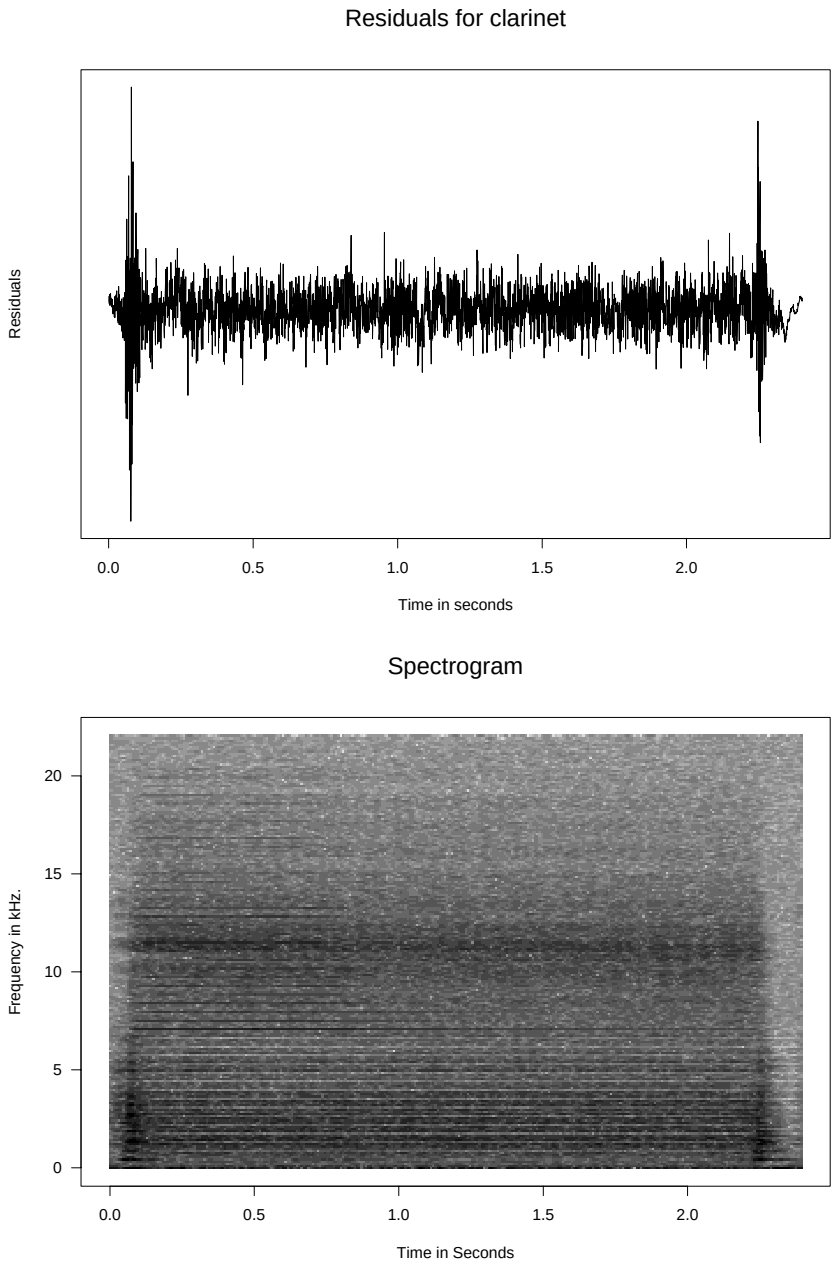

Figure 6.13: Global residuals and spectrogram for the sound signal of a clarinet playing A4.

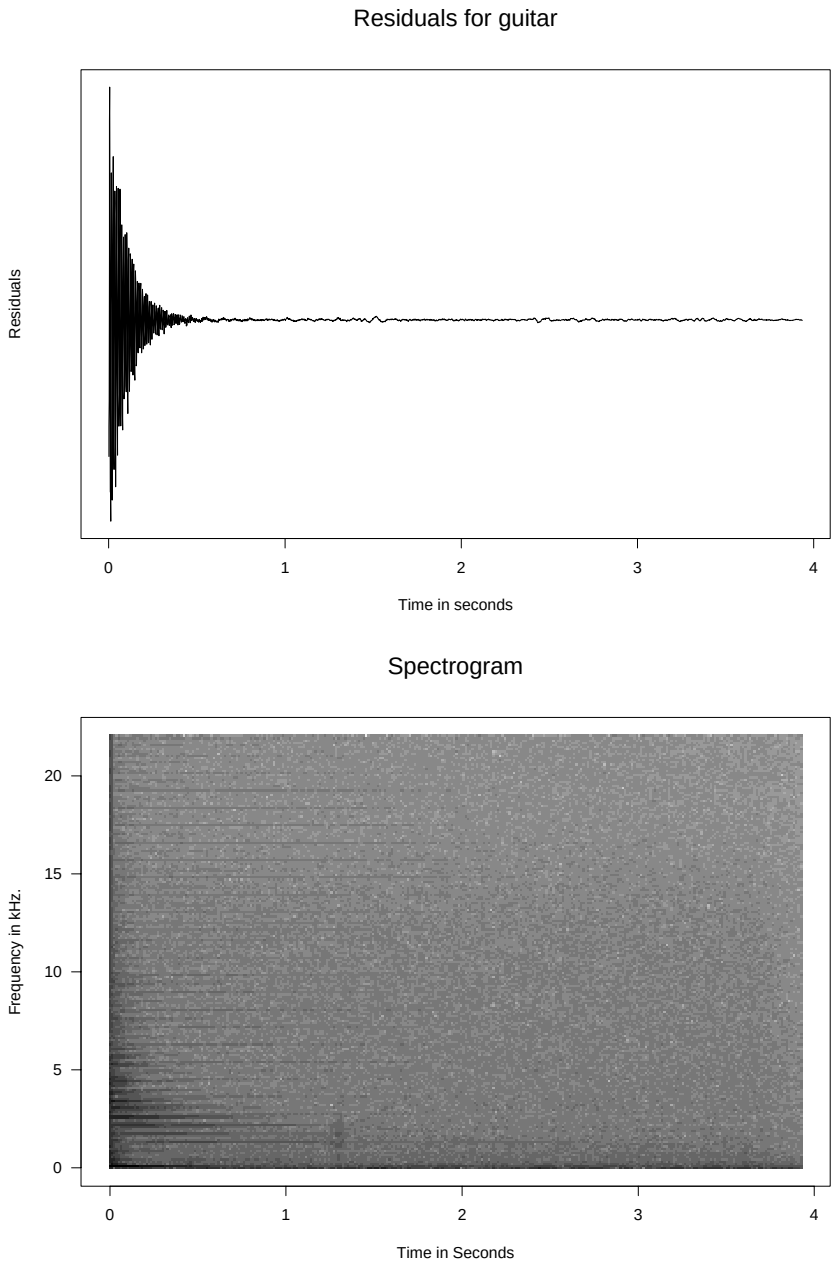

Figure 6.14: Global residuals and spectrogram for the sound signal of a guitar playing D3.

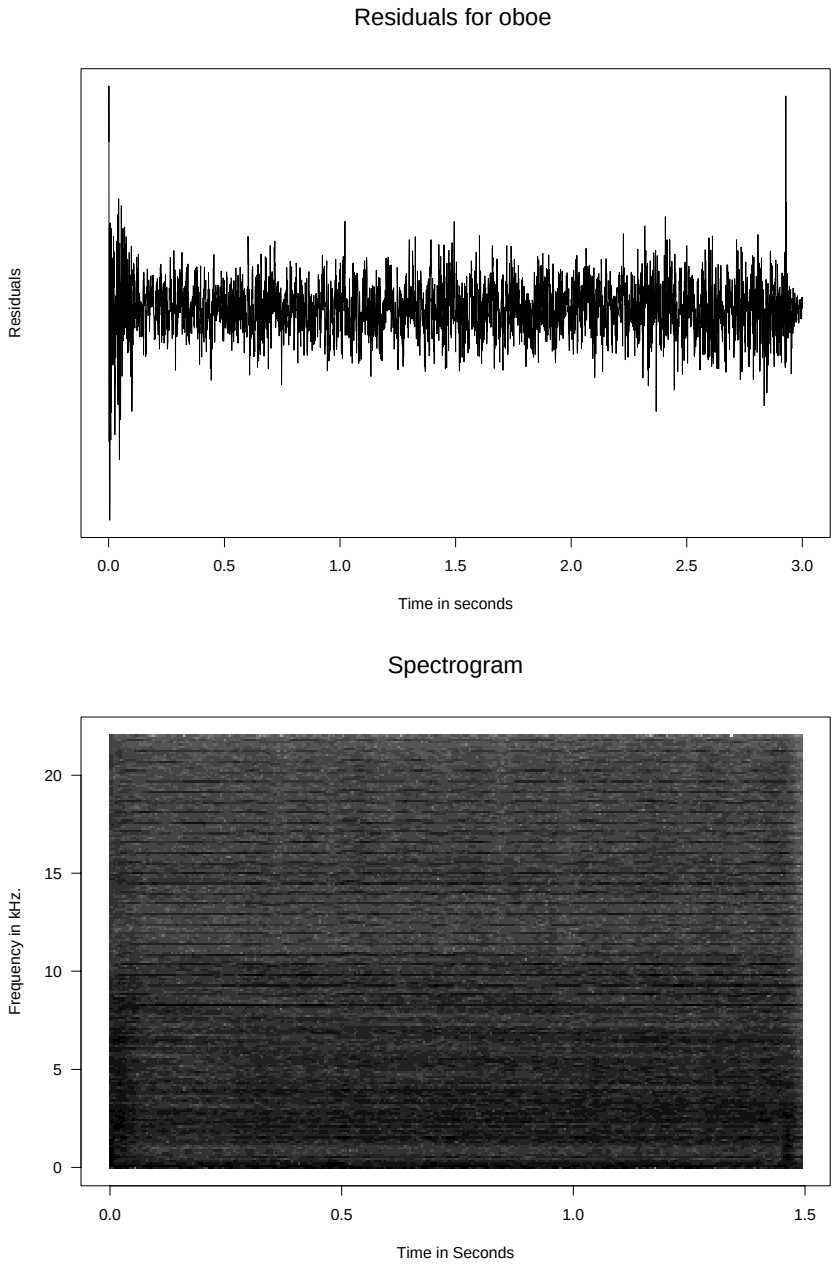

Figure 6.15: Global residuals and spectrogram for the sound signal of an oboe playing C4.

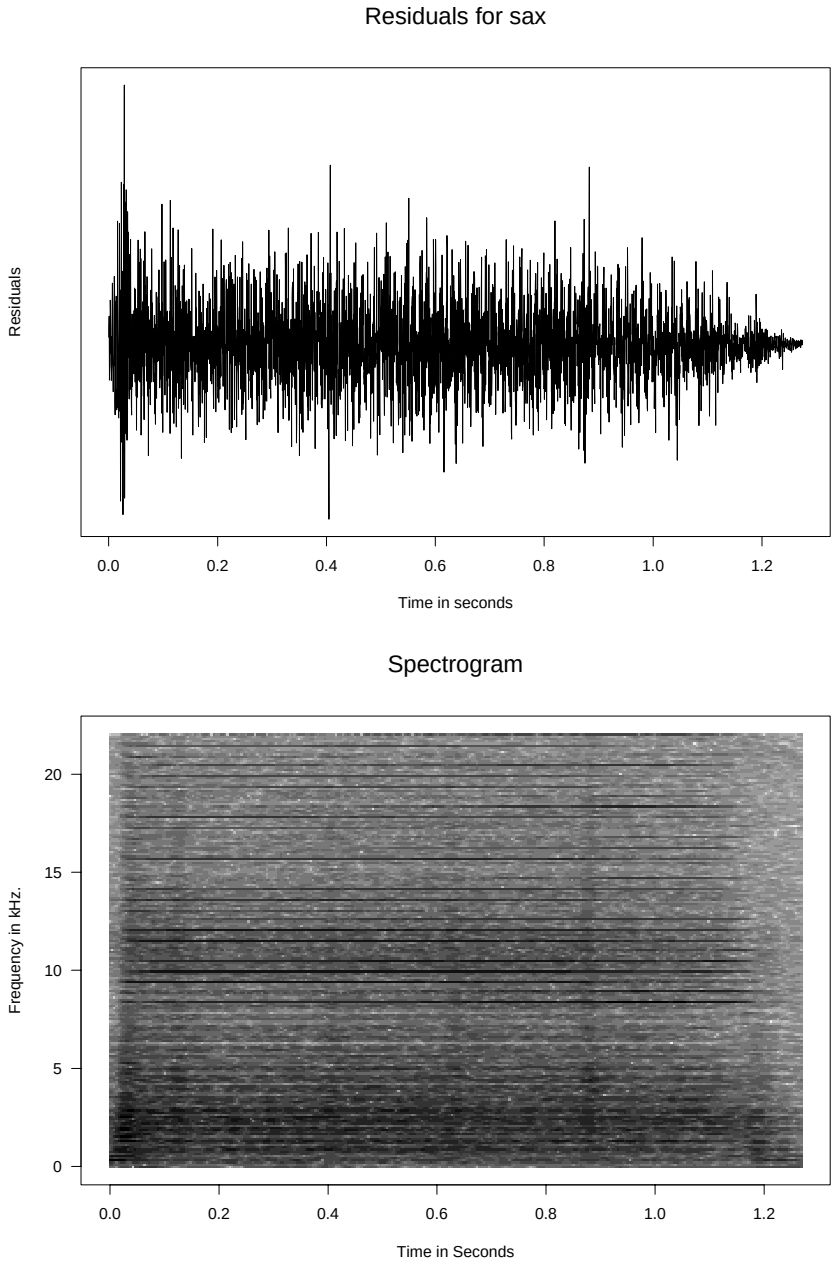

Figure 6.16: Global residuals and spectrogram for the sound signal of a tenor saxophone playing C5.

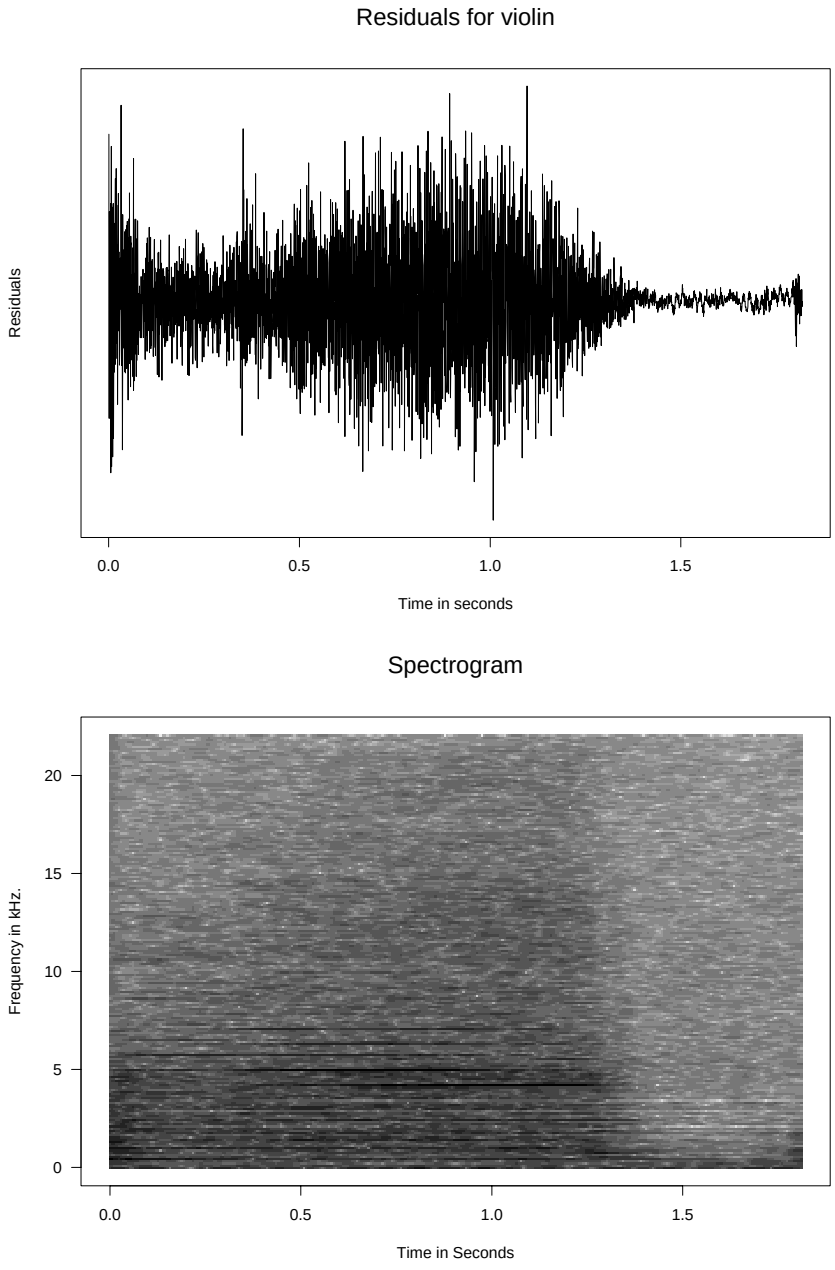

Figure 6.17: Global residuals and spectrogram for the sound signal of a violin playing C4.

6.3.2 Dynamic window selection

As described in Chapter 5, when performing the estimation for each $t \in (0,1)$ we might need to consider different window sizes. In our algorithm, for each t_0, we calculated $\hat{\beta}_{p,q}(t_0)$ for various windows sizes h_q via the procedure described in Chapter 4. We then choose amongst the $\hat{\beta}_{p,q}(t_0)$'s using the wBIC criteria, given by equation (5.5). We next illustrate that dynamic window selection can improve our procedure via an example.

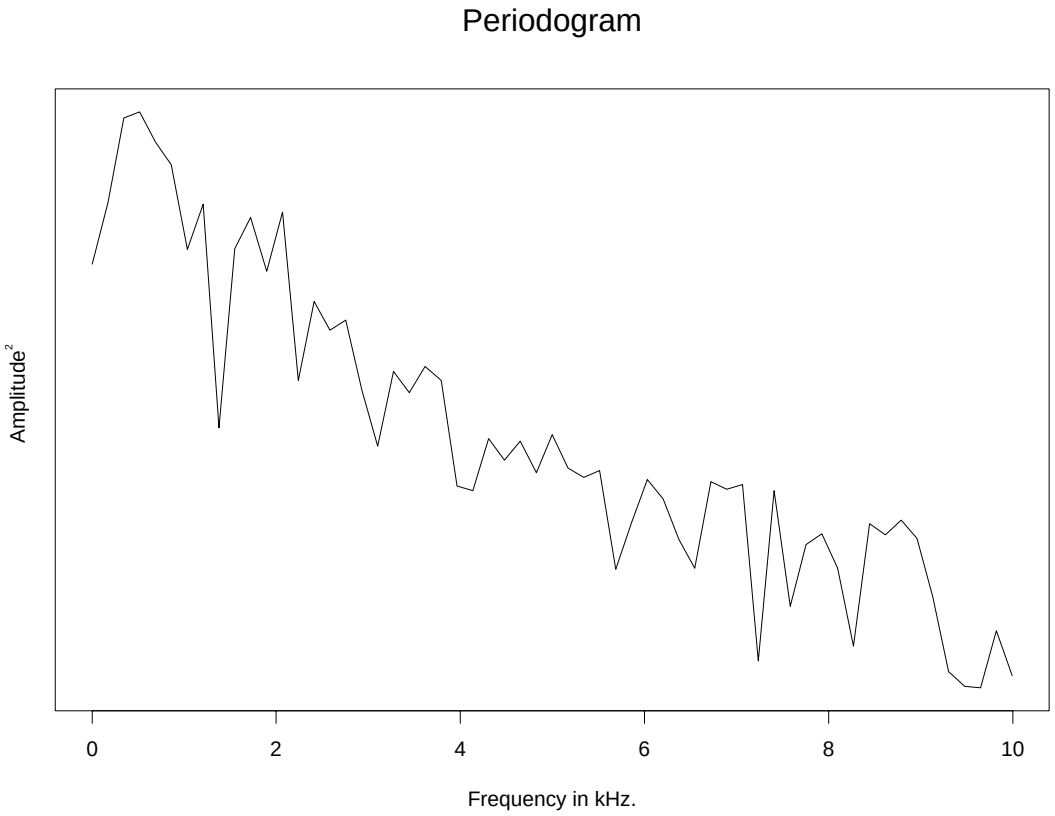

Figure 6.18: Periodogram for a 256 data points segment of the sound signal of a shakuhachi flute.

The shakuhachi flute is a Japanese instrument characterized as being "noisy". The sound of the performer blowing is one of its distinguishing characteristics. By listening to the sound signal studied in this example (track 31 on accompanying CD), we notice that it is characterized by a rapid change of pitch for the first half second, then the pitch is held steady for about 3.5 seconds, then a vibrato is played for about half a second after which the pitch is held fixed again.

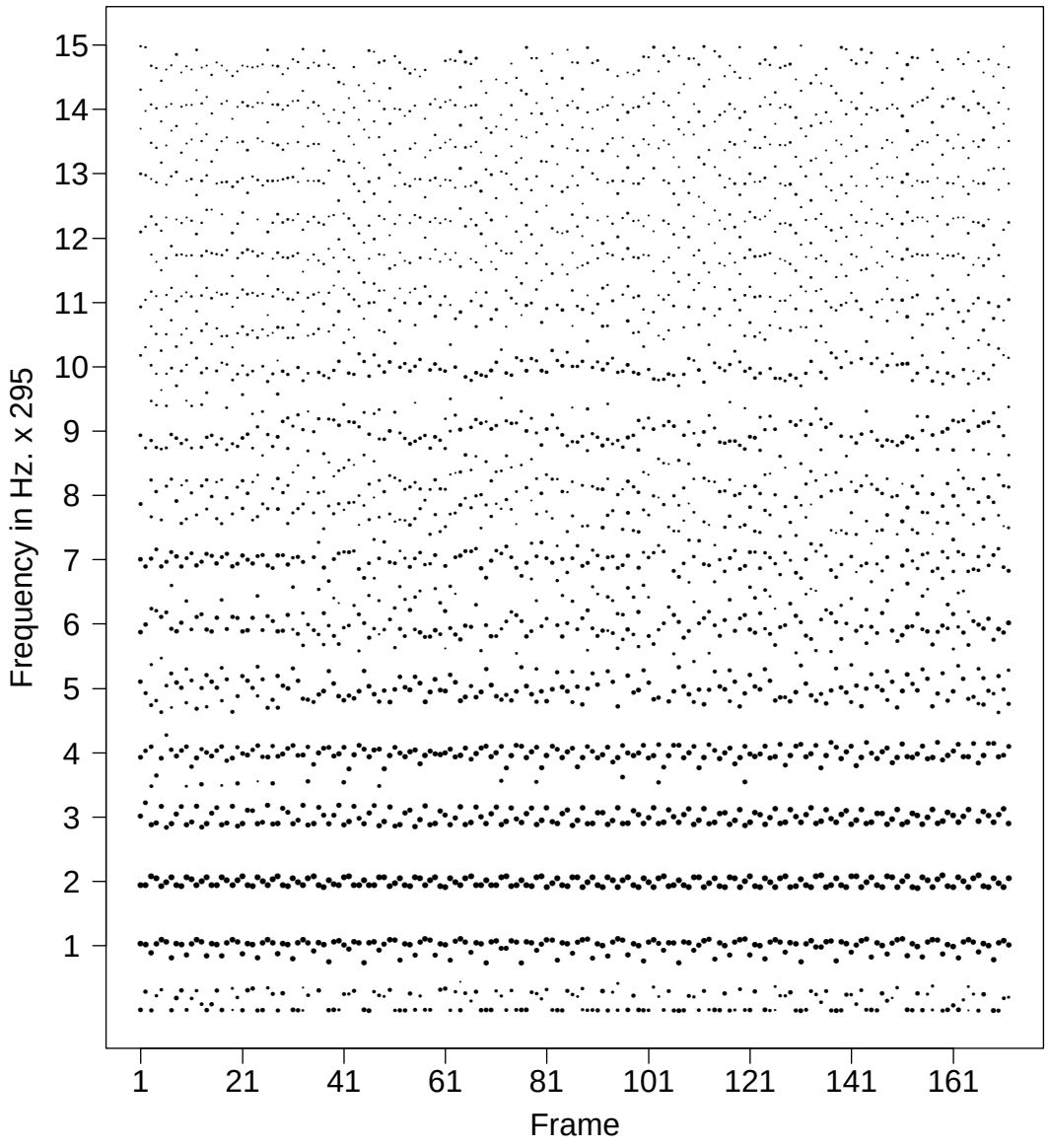

Figure 6.19: Location of peaks of periodograms for non-overlapping segments of a 5.8 milliseconds of the sound signal of a shakuhachi flute.

This particular sound is interesting to study for two reasons. First, the noisy character of the sound makes the partial tracking techniques, described in Chapter 2, difficult to implement because peaks in the periodogram are hard to interpret with small amounts of data. The power of the non-sinusoidal part of the signal is high and this makes the peaks in the periodogram unclear, seen in Figures 6.18 and 6.19. The size of the points in the figure represents the amplitude of the local maxima.

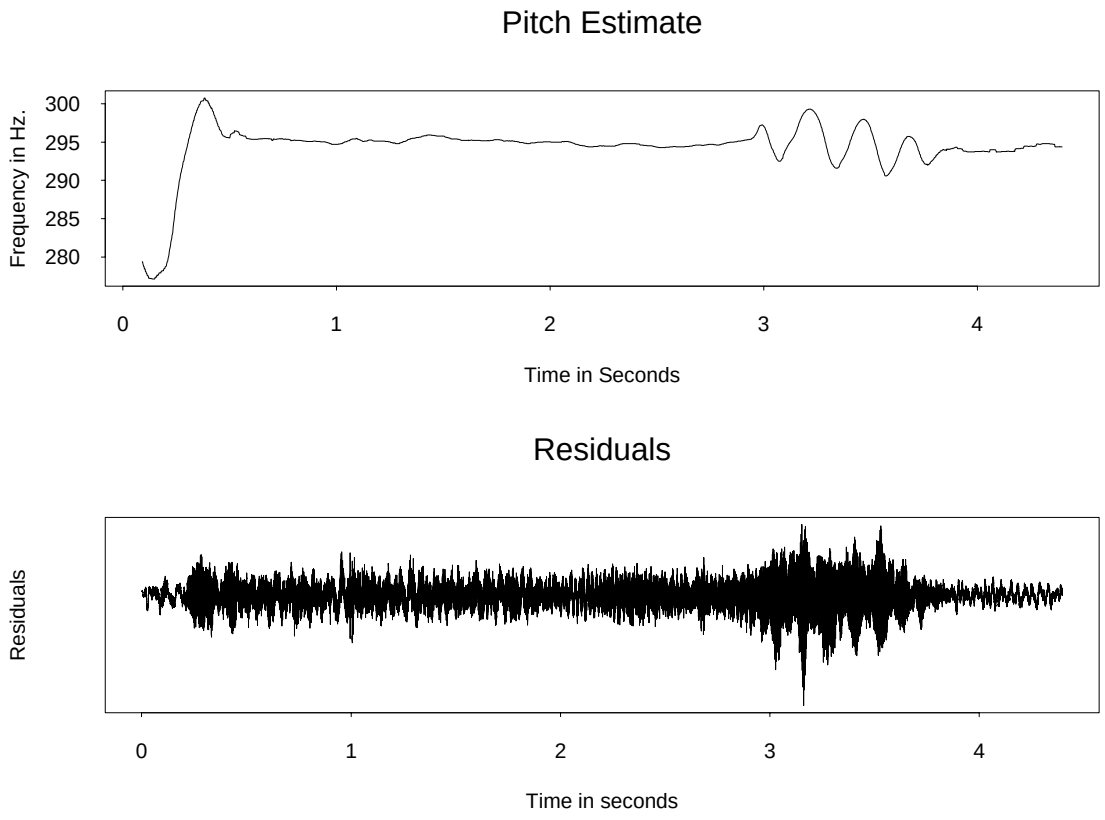

Figure 6.20: Estimated pitch when using a fixed window size and the residuals of the fit.

Second, the different behavior of the pitch function in different parts of the signal suggests that a fixed, large window size is inappropriate and thus that different window sizes should be used in different parts of the signal. The shakuhachi flute example provides a test for our window size selection criteria.

We fitted a local harmonic model with 15 partials to the shakuhachi flute using a fixed window size of about 20 milliseconds. In Figure 6.20, we see the estimated fundamental frequency and a residual plot (tracks 32 and 33 on accompanying CD). We notice that during

the vibrato part, i.e. the part where the signal is not near constant, the fit is not as good (the residuals are bigger).

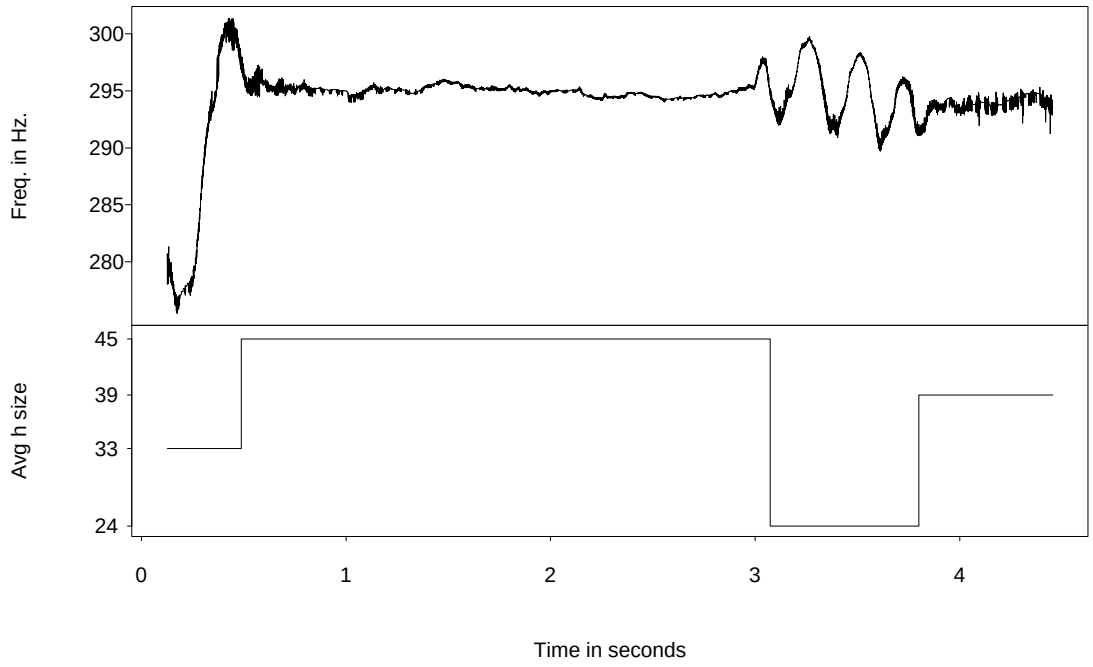

Figure 6.21: Estimated pitch when using a dynamic window size and the average window size (in milliseconds) used in four different sections.

The dynamic window procedure, using the wBIC to choose between windows, provides a solution to this problem. We fitted a local harmonic model with 15 partials to the shakuhachi flute and choosing between various possible window sizes. In Figure 6.21 we see how the procedure, on average, chooses smaller window sizes during the parts of the signal where the parameter function is not near constant, as we would expect.

Finally, we notice the improvement of the dynamic window method by comparing the residual plots, seen in Figure 6.22, of the fixed window and dynamic window procedures (tracks 33 and 34 on accompanying CD).

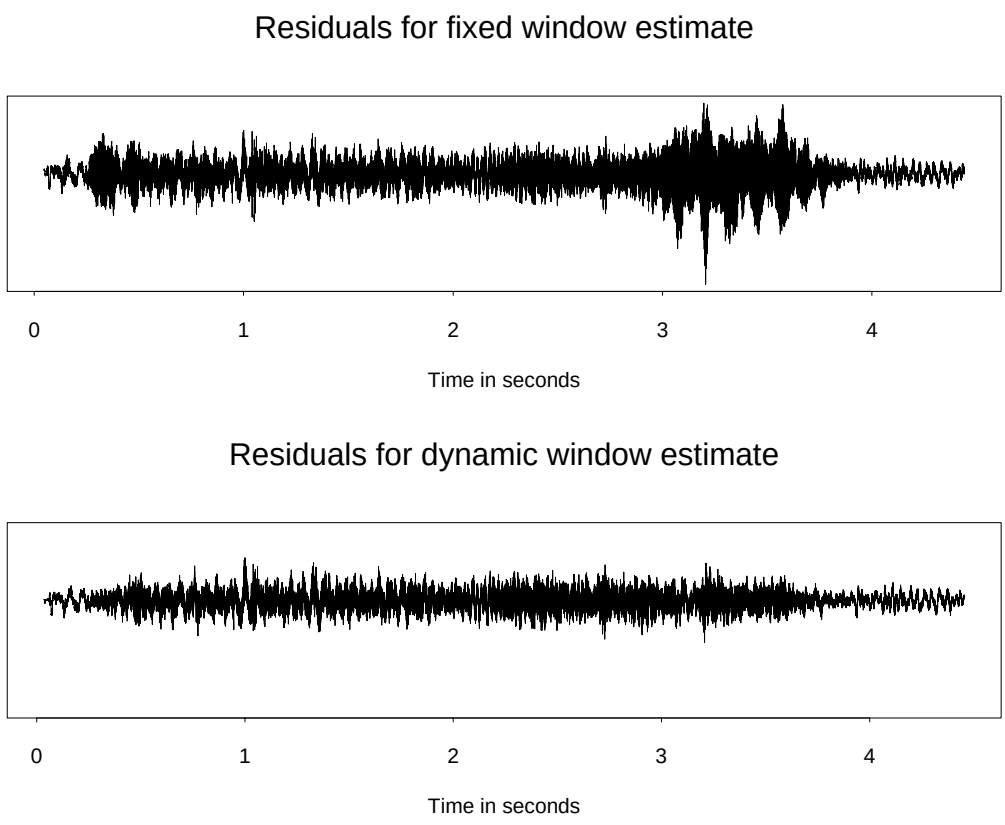

Figure 6.22: Comparison of the two residual plots on the same scales.

6.4 Standard errors

In the music technology literature a variety of pitch estimators are presented (Harris and Weiss 1963, Martin 1982, Junhar 1997). When such procedures are used to estimate the fundamental frequency of sound signals produced by one instrument playing one note, the movement of these estimates are sometimes explained with deterministic factors. Conclusions like: "an instrumentalist can never play at exactly the concert pitch A frequency of 440 Hz. and can never hold the exact frequency for extended periods of time" are drawn from such estimates. Some recent sound analysis procedures (Rodet 1997) assume that signals contain a stochastic element, mainly the non-sinusoidal part of the signal. The stochastic assumption is not made for the functional parameter, which is assumed to be deterministic. If we assume the presence of such a stochastic element, the possibility exists that the variation over time of the pitch estimates can be explained by chance as

opposed to deterministic reasons.

In current sound analysis research it is common to give estimates of sinusoidal parameters without indications of their uncertainties. The asymptotic variances of the estimates resulting from the models provides a way to obtain standard errors and to construct confidence intervals for our estimates. It is interesting to speculate on the meaning of these quantities in a musical context. In the following section we discuss a possible interpretation of these statistical variation measures.

6.4.1 Statistically significant out of "tuneness"

As mentioned in Chapter 2, it is convenient to measure pitch in a logarithmic scale. Given a base frequency f_1 we can transform any frequency f_2 into semitones using equation (2.1). For example, the Musical Instrument Digital Interface (MIDI) standard (Loy 1989) assigns the number 69 to concert pitch A (440 Hz.) and then using formula (2.1) assigns a *MIDI note number* to any frequency λ via

$$\text{MIDI number of } \lambda = 69 + 12\log_2(\lambda/440)$$

Apparently the trained ear can distinguish two notes if they are 3 cents (a hundredth of a tone) or more apart.

We fit a local harmonic model to a signal produced by a trumpet playing (or trying to play) concert pitch A (440 Hz). The recording was made by a professional trumpet player and the trumpet was tuned to A 440 Hz. concert pitch using a commercial *tuner*. In Figure 6.23 we see that the estimated pitch found by our procedure is closer than 3 cents from concert pitch A for most of the signal. Figure 6.23 also shows approximate 99% confidence intervals around 440 Hz. The figure suggest that for most of the signal the trumpet player is statistically significantly out of tune. Is it reasonable that the statistical variation of our estimates is "small"? One possible interpretation is the following.

In general, the human ear/brain is quite accurate at determining pitch. Suppose that the stochastic part of the signal made the variation in this "pitch estimate" large. Changes in pitch might then be detected even when hearing a sound with deterministic constant pitch.

If we consider the estimated pitch found by our procedure ignoring its statistical variability one could conclude that at the beginning and end of the trumpet signal, the trumpet player is more out of tune than during the middle. However, we must also notice

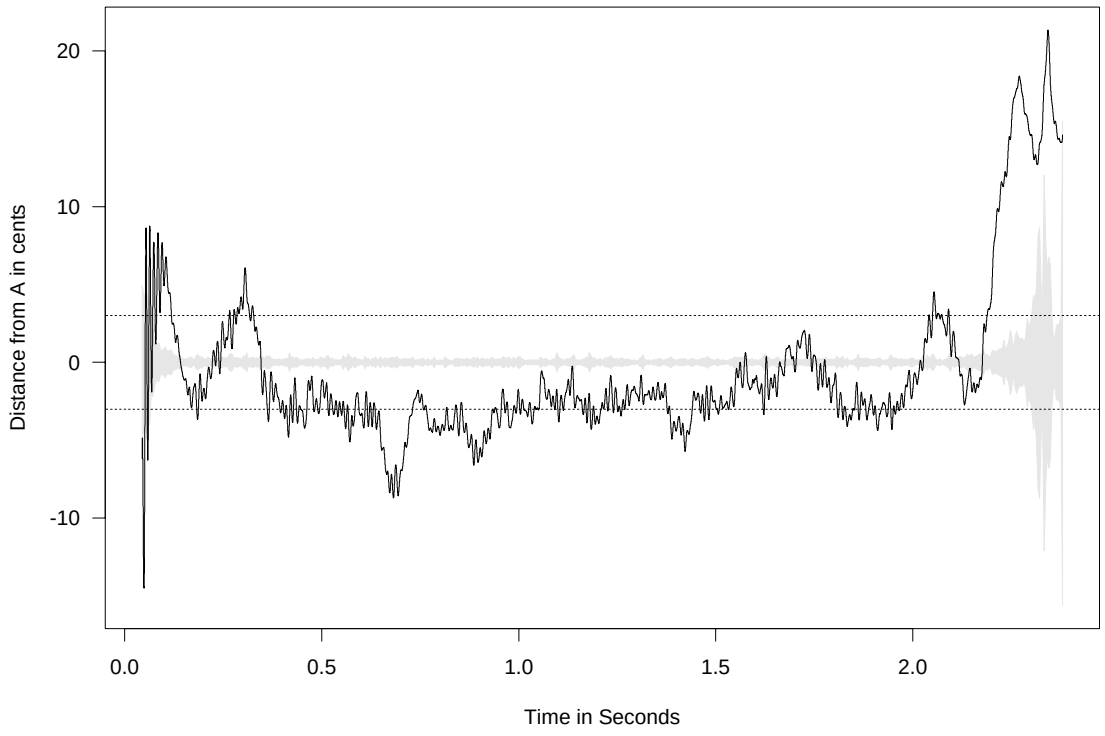

Figure 6.23: Pitch estimate for trumpet sound and confidence interval around 440 Hz.

that during these sections the standard error is also bigger reflecting the possibility that the larger deviation is due to chance. In Figure 6.24 we see the distance between the estimated fundamental frequency from 440 Hz. in standard units

$$\frac{\hat{\lambda} - 440}{\widehat{\text{se}}[\hat{\lambda}]}$$

Notice that the variability in the estimated pitch function at the beginning and end of the trumpet sound signal doesn't appear to be high anymore. In figure 6.24 we also see the distance between the estimated fundamental frequency of the oboe sound signal, presented in Figure 6.9, and 259.25 Hz in standard units. The figure suggests that the oboist is playing a *statistically significant* vibrato.

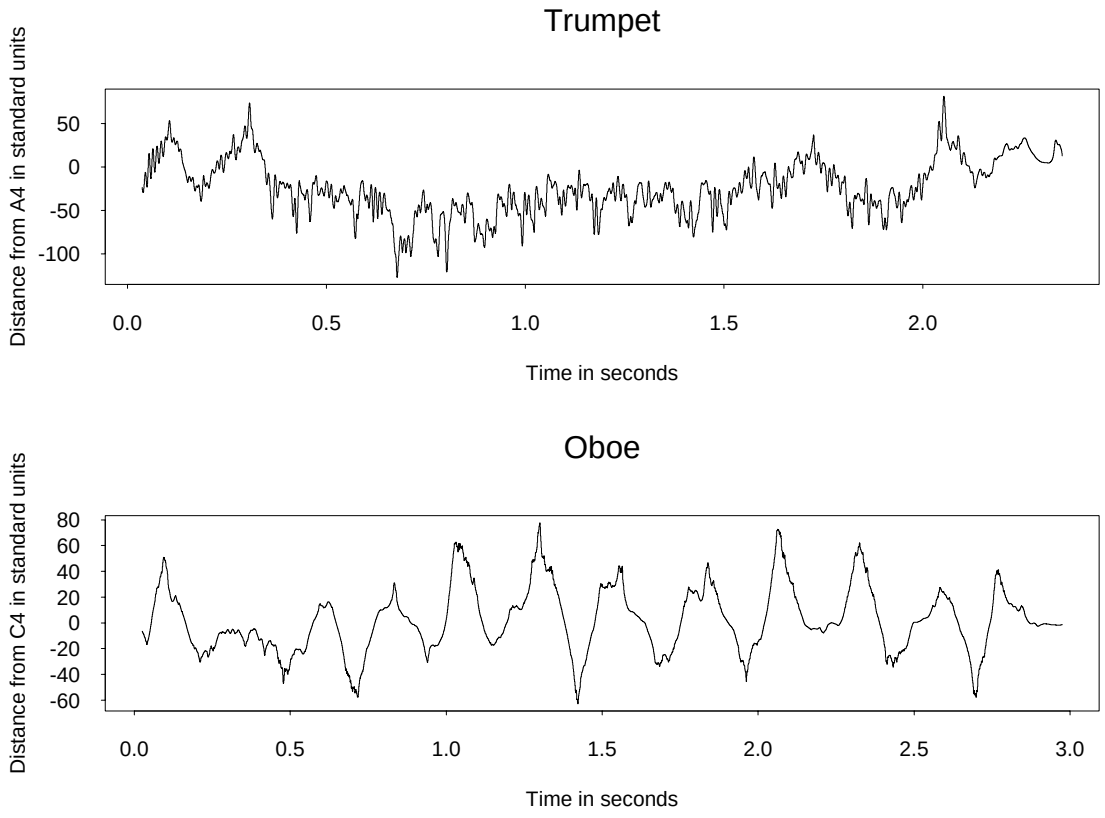

Figure 6.24: Estimated fundamental frequency in standard units for trumpet and oboe signals.

6.4.2 Problems

Confidence intervals estimates are constructed using the asymptotic approximation. In our case N is usually between 800 and 2000 observations. Thus there is a possibility that the variance of our estimates might be quite different from the approximation used. Simulation and bootstrap methods can be used to check this. This is left as future work. Furthermore, we obtain variance estimates under the assumption of additive noise. For many instruments this assumption appears inappropriate. The noise seems to be signal related and possible not additive (Maganza and Caussé 1986). Finding variance structures under assumptions like these is also left as future work.

6.5 Creating new sounds

Our analysis method provides estimates of the fundamental frequency $\hat{\lambda}(t)$, of the amplitude of each partial

$$\hat{\rho}(t) = \sqrt{\hat{A}_k^2(t) + \hat{B}_k^2(t)}$$

and of the phase of each partial

$$\hat{\phi}_k(t) = \arctan\left\{-\frac{\hat{B}_k(t)}{\hat{A}_k(t)}\right\}$$

The estimate of the harmonic signal given in equation (6.1) can now be written as

$$\hat{y}(t) = \sum_{k=1}^{\hat{K}} \hat{\rho}_k(t)\cos(k\,\hat{\lambda}(t)t + \hat{\phi}_k(t))$$

As mentioned in Chapter 2, it is believed that the human ear is unable to notice the difference when the phase function $\phi_k(t)$ changes. To study this we construct an estimate of the signal ignoring the phase

$$\tilde{y}(t) = \sum_{k=1}^{\hat{K}} \hat{\rho}_k(t)\cos(k\hat{\lambda}(t)t) \tag{6.3}$$

In many cases, the sounds of the signal $y(t)$ and $\tilde{y}(t)$ were almost indistinguishable from each other (tracks 25 and 26 on accompanying CD). Although to the human ear these two signals sound very similar, analytically (because of the difference in phase) they are quite different. Therefore $\tilde{\epsilon}(t) = \tilde{y}(t) - y(t)$ can't be a expected to be a useful estimate of the non-sinusoidal part of the signal.

We can create new sounds by altering the parameter function in different ways. In general, we can create a new signal based on the estimates of the original

$$z(t) = \sum_{k=1}^{\hat{K}} r_k(t)\hat{\rho}_k(t)\cos(k\,l(t)\hat{\lambda}(t)t) \tag{6.4}$$

where $r_k(t)$ and $l(k)$ are functions that will permit us to control the change we wish to perform on the original sound. We can now

- Change pitch through the function $l(t)$ (Pitch Modification)
- Change duration of certain parts of the signal by using $z(d(t))$where d is a time substitution function (Wessel 1987). For example, if d(t)=2t, the signal lasts half as long as the original. (Time scale modification)
- Change the energy of a specific harmonic through the functions $r_k(t)$.

6.5.1 The hidden soprano

It is said that if one listens to an oboe very closely one can hear the voice of a soprano singing at an octave above the fundamental frequency of the oboe signal. Our analysis provides a way of bringing the hidden soprano out.

The signal produced by $\tilde{y}(t)$ as defined in (6.3) sounds practically the same as the original. As mentioned above, the soprano sound is heard at an octave above the original oboe sound. This implies that the soprano sound will have fundamental frequency twice that of the original. Assuming that the sound of the soprano is harmonic, the harmonics of the hidden soprano sound will be at frequencies corresponding to the even partials of the original oboe sound. This leads us to believe that if we recreate a sound containing only the even partials of an oboe, it should sound like a soprano.

It is known to musicians that an instrumentalist or singer can use vibrato to make her/his part stand out. We hope that by creating a sound with the even partial components sounding as a vibrato we can make the hidden soprano "come out". By letting

$$r_k(t) = \begin{cases} 1 & : \quad k \text{ odd} \\ 1 + a_k \cos(\mu t) & : \quad k \text{ even} \end{cases}$$

with μ the vibrato frequency and a_k the maximum deviation from $k\lambda$, we can create the desired signal $z(t)$ with equation (6.4). We produce the signal $z(t)$ by letting $\mu = 10$ Hz. and $a = 20$ Hz. Listening to the signal produced by converting $z(t)$ into sound the hidden soprano is clearly heard by the author and other trained musicians (track 35 on accompanying CD).

6.5.2 Beginner's violin sound

As mentioned in Chapter 2, the non-sinusoidal component is thought of as an important characteristic of the sound signal. Equations (6.1) and (6.2) provide estimates of the harmonic and non-sinusoidal parts of the signal. The separation of the the noise from the discrete part of the signal has many applications. In this section we provide a simple example.

The sound produced by a beginner on a violin may have a "screechy" quality. This is the sound of the bowing that is not converted into a harmonic signal. In this case the noise characterizes the sound as that of a beginner playing the violin as opposed to a

professional. This leads us to believe that if we "amplify" the non-sinusoidal part of the sound signal played by a professional violin player (track 28 on accompanying CD), it should sound like a beginner's violin sound.

We hope that we can create the beginner's violin sound by letting

$$z(t) = \hat{y}(t) + M\,\hat{\epsilon}(t)$$

Here $M > 1$ controls how much we amplify the non-sinusoidal part. We produce the signal $z(t)$ by letting $M = 10$. Listening to the signal produced by converting $z(t)$ into sound the beginner's violin sound is heard by the author (track 46 on accompanying CD).

6.6 Two fundamental frequencies

Sometimes sounds produced by a previous note can be heard because of an echo. The auditory term for this type of echo is *reverberation*. To a certain extent this effect is pleasing to the human ear (this is the reason people like to sing in the shower). When there is reverberation, the existence of harmonic components related to a second fundamental frequency unrelated to the present note, the local harmonic model with one fundamental frequency is inappropriate. Other analysis methods will also have problems analyzing the sound. For example, analysis methods using sinusoidal tracking will be unable to distinguish between the partials of the two signals. The solution is to fit a local harmonic model with more than one fundamental.

6.6.1 Removing reverberation in pipe organ sounds

The sound studied in this example is a pipe organ playing two consecutive notes. The room where the recording was made, Hertz Hall in UC Berkeley, is a concert hall characterized as having quite a bit of echo. When the second note is played, the first note can still be heard. This is reverberation. If we look at the spectrogram of the signal, see Figure 6.25, we can see that after 1.2 seconds or so the second note begins (track 36 on accompanying CD). The vertical line is at the note change. In this figure we can see the frequency component related to the main fundamental frequency change to a smaller value after 1.2 seconds, from about 368 Hz. to about 325 Hz. The vertical line is at the note change. This is due to the note change from F♯4 to E4. We also notice that frequency components of the first note remain during the playing of the second note. This is due to

the reverberation effect. Although this can be seen clearly in the spectrogram it requires close attention to be heard. This is an example of how statistical plots can be be useful in the analysis of sound.

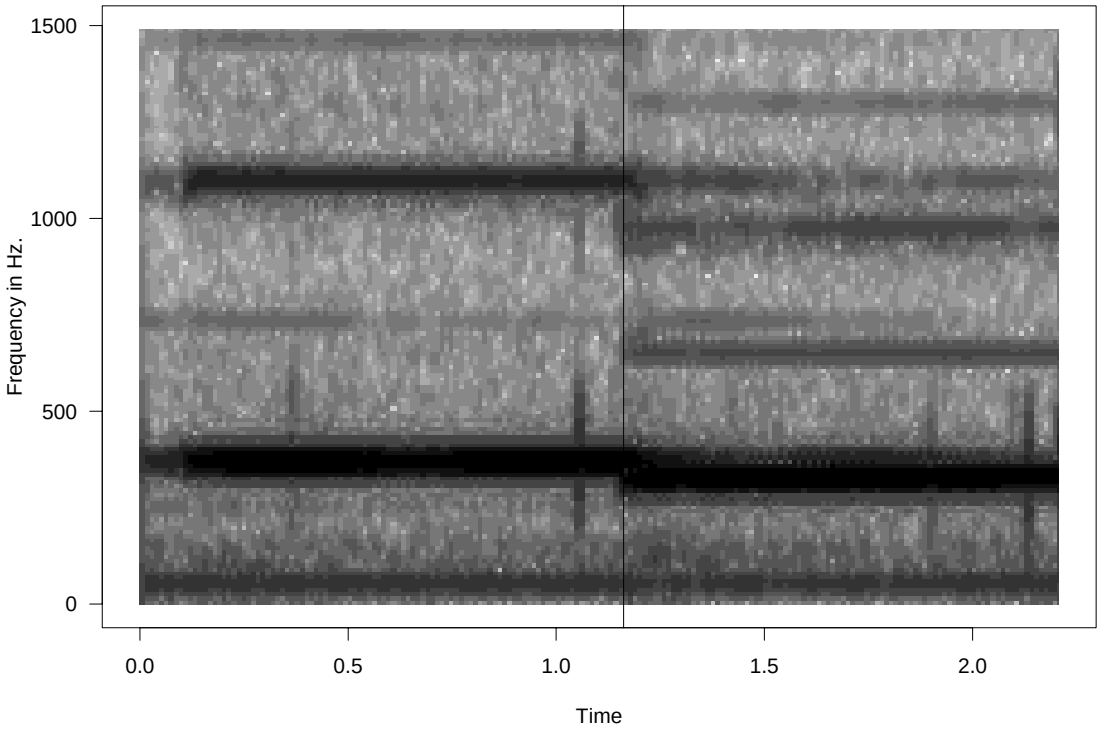

Figure 6.25: Spectrogram for the sound signal with reverberation of a pipe organ playing F♯4 and E4.

In Figure 6.26 we see the location of the periodogram maxima for a stretch of the organ signal around the time where the new note starts. The vertical line is at the note change. Notice that partial tracking procedure, as those described in Chapter 2, will be able to distinguish partials with periodicities around the two fundamental frequencies, however they will not be able to identify them as fundamental frequencies.

A solution is to fit a local harmonic model with two fundamental frequencies. Figure 6.27 shows the estimate of the fundamental frequency function when a harmonic model with one fundamental and 15 partials is fitted. Notice than during the second note the estimate seems to be varying more that during the first. This could be a bad fit due to

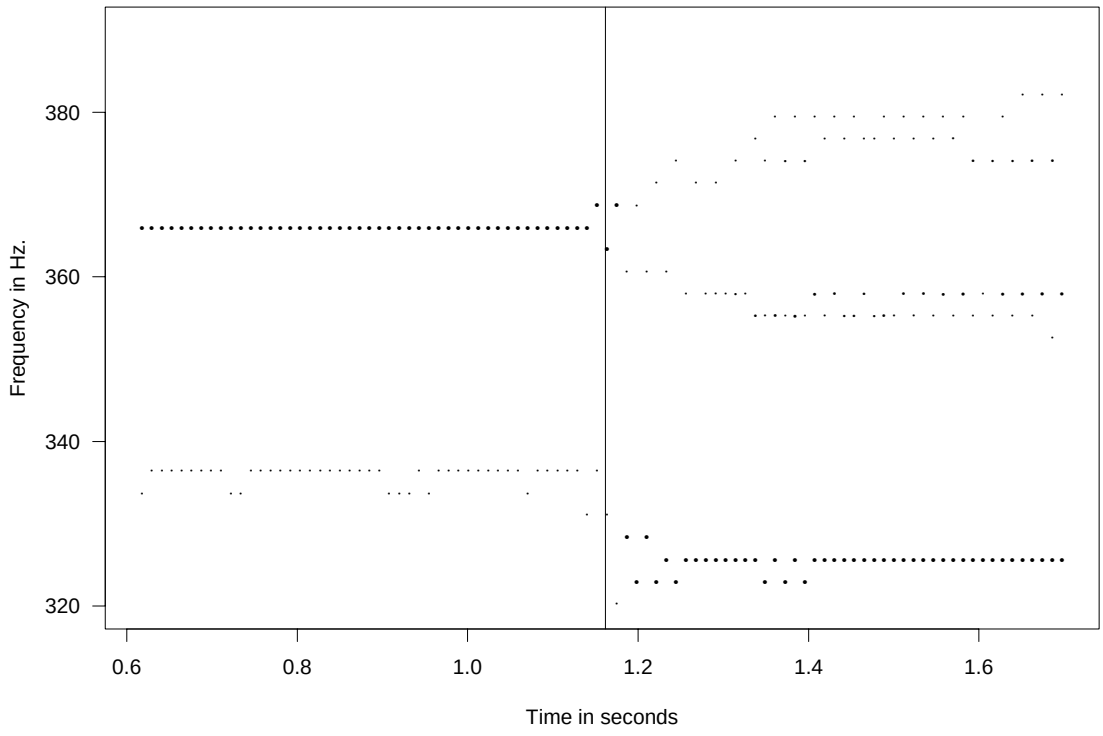

Figure 6.26: Location of periodogram maxima near the fundamental frequencies of the sound signal with reverberation of a pipe organ playing F♯4 and E4.

the inappropriateness of the model (track 37 on accompanying CD).

In Figure 6.27 we show a spectrogram for the residuals obtained by fitting the one fundamental model. Notice that during the part of the spectrogram corresponding to the part of the signal where the reverberation was occurring, the harmonic structure produced by the echo of the previous note can be seen (track 38 on accompanying CD).

The estimation procedure is greatly improved by fitting a two-fundamental model to the second part of the signal (track 39 on accompanying CD). We fitted a two fundamental local harmonic model with 10 and 5 partials related to the E4 and F♯4 fundamental frequencies respectively. In Figure 6.28 we see the estimates for the fundamental frequencies obtained (tracks 40 and 41 on accompanying CD). The size of the lines reflect the strength of the overall amplitude of the signal related to that fundamental. Notice that this estimate seems more "stable" than the one seen in Figure 6.27. By looking at the residual spec-

trogram in Figure 6.28 we see that the harmonic structure due to reverberation has been substantially removed (track 42 on accompanying CD).

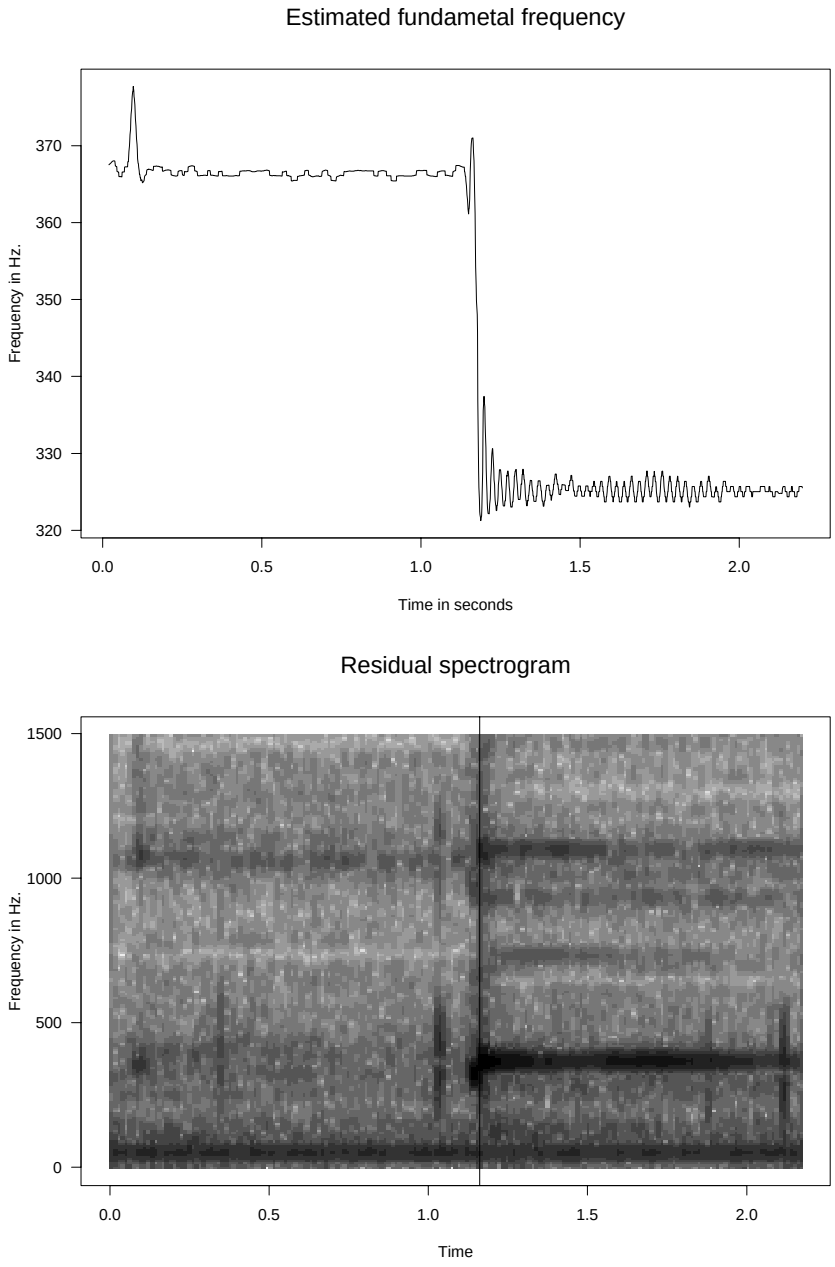

Figure 6.27: Frequency estimate for the sound signal of the pipe organ sound using one fundamental model and spectrum for the residuals.

Another interesting observation is that the residuals seem to have a strong fre-

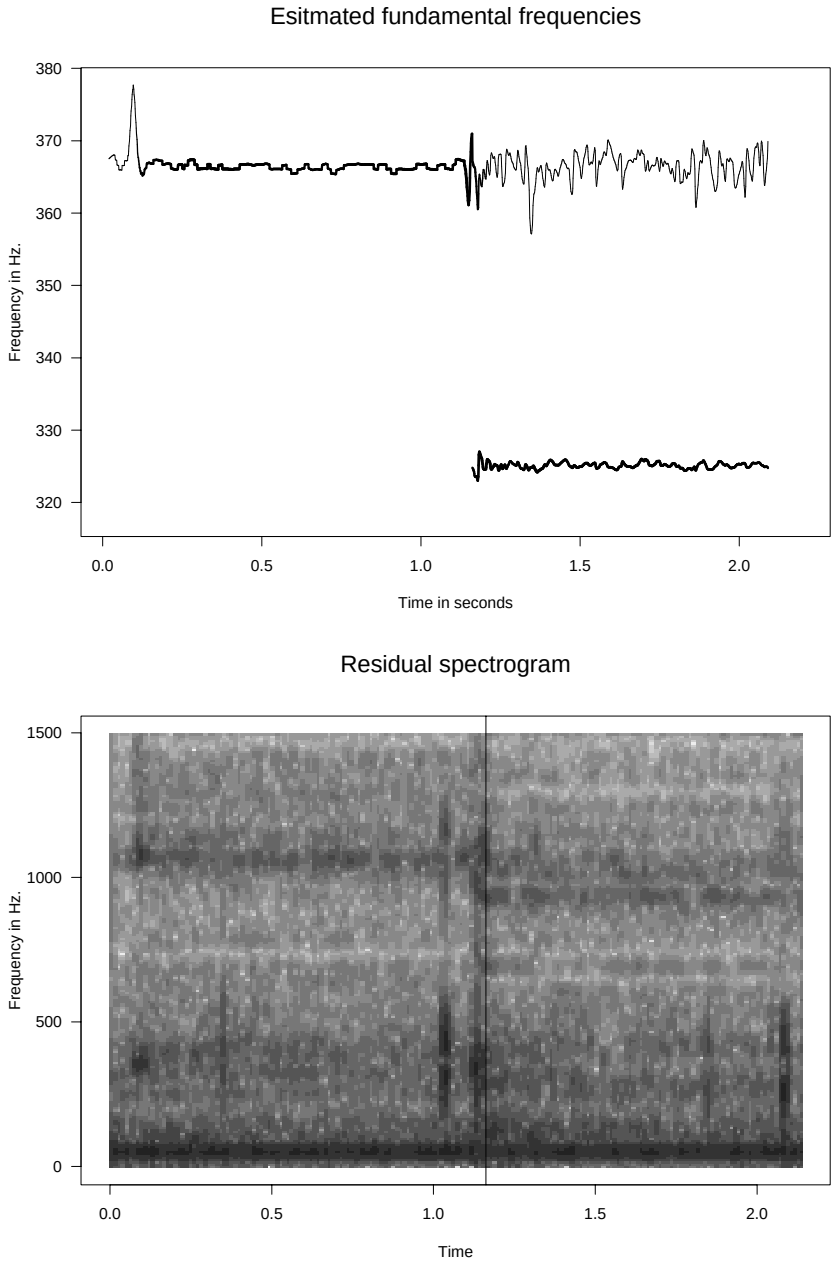

Figure 6.28: Estimates of the two fundamental frequencies for the sound signal of the pipe organ sound using a two fundamental model and residual spectrogram.

quency component at a low frequency, approximately 50 Hz. This can be seen not only in the spectrograms of the residuals, but also in an overall spectrum estimate, see Figure 6.29. This might be a characteristic of the sound of the wind going through the pipes.

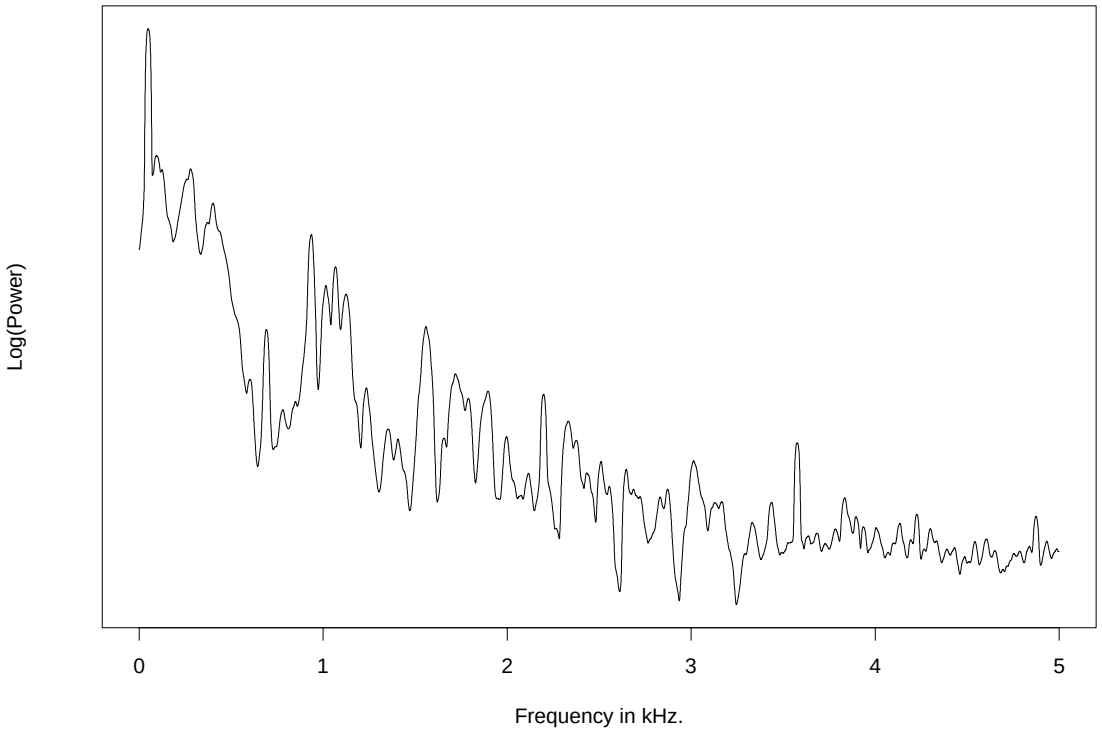

Figure 6.29: Smooth periodogram estimate (m=22) of the spectrum using the residuals of the two fundamental model.

Chapter 7

Coda: Future Work

Some general projects that are suggested as future research include: a) examining efficiency of the weighted estimates in the stationary case, b) deriving model selection criteria under more general assumptions as done in Konishi and Kitagawa (1996), c) finding a Bayesian justification for the wBIC, d) further simulations to test the criteria of Chapter 5, e) performing bootstrap simulations to obtain alternate variance estimates to compare with the variance obtained via asymptotic approximation, f) incorporating techniques similar to those used in nonparamteric estimation, for example those used in Fan and Gijbles (1995). The following sections discuss five specific examples of possible future work.

7.1 Other assumptions on the local behavior

In the current analysis we have assumed that the parameter function $\beta(t)$ is approximately constant. The approach of likelihood-based local regression methods, see for example Staniswalis (1989) and Tibshirani and Hastie (1987), is to assume approximate linearity. Although this adds more parameters, in certain cases it allows for the examination of longer stretches of data.

For the signal of a note played on a blown "harmonic" instrument the pitch is usually stable, but the amplitude is constantly changing because of fluctuations in air pressure. We could assume that the amplitude of the partials are locally linear by assuming that locally the model is

$$s(t;\beta) = \sum_{k=1}^{K} (a_k + b_k t)\cos(k\lambda t + \phi)$$

for t near t_0. Notice that this is the same model as before except for the fact that locally the amplitude of each harmonic is changing linearly in time

$$\rho_k(t) = a_k + b_k t$$

Because the new parameters, β_k, appear linearly, preparing a program for their estimation should not prove difficult.

In the case of instruments, like the guitar, where after the note starts the sound just "decays", we could model the amplitude of the harmonics with a decaying exponential as done in Bolt and Brillinger (1979), Dahlen (1981), Hasan (1979), Hassan (1982), Hasan (1983):

$$\rho_k(t) = \alpha_k + \beta_k \exp\{-\gamma_k t\}$$

In this case we could probably consider much larger time windows given that the pitch is stable enough. The asymptotic properties of these new estimates would be studied.

7.2 Other assumptions on the harmonic structure

Another possibility for future work is to consider the case where the signal is not exactly harmonic. Some instruments, for example the piano, have partials at frequencies that are not quite multiples of the fundamental frequency but are relatively close. We could then consider a local *quasi-harmonic* model where we impose the following type of constraint on the frequencies of the partials

$$\omega_k(t) = k\lambda(t) + \delta_k$$

Here the δ_k's are assumed to be constant in time, throughout the signal. We could use the estimates $\hat{\boldsymbol{\omega}}_N(n/N)$ of equation (3.41), $n = 1, \ldots, N$ to estimate the fundamental frequency function $\lambda(t)$ and the constants δ_k. Furthermore, we could examine the possibility of δ_k being random.

7.3 Further study of the noise and residuals

Further theoretical development for the non-stationary case is needed. Finding alternative, more general assumptions, under which Condition 1 hold is of particular interest.

For the model presented in Chapter 4 we assumed the noise processes $\{\epsilon_{n,N}\}$ to be *locally stationary*. Using techniques similar to those presented by Dahlhaus (1997) could provide a way to represent the noise parametrically. In turn, this would provide a way to synthesize the non-sinusoidal part of the signal by simulating noise with the structure estimated using the *local stationarity* techniques.

Assessment of the procedure, in particular via the study of the residuals is possible. The non-stationary structure of the noise is something that interests computer music researchers. Residual analysis would not only serve as an assessment procedure, but would also provide insight into the nature of the non-sinusoidal signals produced by instruments. Ideally we would want to find possible ways of synthesizing the non-sinusoidal signal. In particular, a comparison of the global and local residuals is of interest.

In Chapter 2 we mentioned how some procedures assume that sound signals are the output of passing some simple waveform through a linear time-varying filter. There is really no physically based explanation for this assumption. In fact, recently researchers have become interested in studying the third- and higher- order moments in sound signals (Wilson et al. 1992, Dubnov and Tishby 1996, Dubnov and Rodet 1997). The techniques used in Brillinger (1965), Brillinger and Rosenblatt (1967) could be used to analyze the residuals obtained from our analysis. Furthermore we can study the possibility of the stochastic part not being additive.

7.4 Optimal estimates

In Wang (1991) an AIC type estimator of K (the number of harmonics) was found to be consistent under the assumption that the error process $\{\epsilon_t\}$ is ergodic. An interesting problem is to find conditions under which a criteria like the wBIC produces consistent estimates of the number of harmonics. Furthermore, conditions could be found for the local stationary process $\{\epsilon_{n,N}\}$ so that an optimal window function and window size exists for each t_0, i.e. that the estimates found using such window function are efficient. It would be interesting to investigate if the estimates found using the wBIC are equivalent to such efficient estimates.

7.5 Other loss functions

We are considering least squares estimates. By doing this, in a sense we are acting as if the best estimates are the estimates that minimize the residual sum of squares. But what our ears consider the best estimate might not be the same. A possibility for future work is to consider minimizing other loss functions. For example, we could consider estimates that minimize the "harmonic" structure of the periodogram of the residuals. We would search the general literature for work on psychological aspects.

7.6 Residual Analysis

Throughout the work done for this thesis residual analysis "by ear" was used as a way to assess fits, i.e.. residuals were converted into sound signals. It turned out to be very useful to find, for example, some signal in the noise amongst other characteristics.

Can trained musicians distinguish between the sounds produced by different statistical processes? Can they distinguish between white noise and an AR processes? If we construct signals with the estimated deterministic signal plus its SE multiplied by some constant M, what is the smallest value of M that would make the difference audible? These and other similar questions are interesting psychoacoustics problems to consider as future work.

7.7 More efficient computational tools

It is important to carry out the estimation procedures on many different sound signals so as to assess the model and also corroborate its effectiveness. Some of the ways the algorithm created in Splus and the existing partially linear algorithm, (Bates and Lindstrom 1986), can be improved include:

- The current estimation tool could be written using the C language in order to make it faster and more efficient.
- An effective algorithm can be created to find starting values for the fundamental frequency. This is known in the computer music literature as a pitch estimator. The fact that the partials for the fundamental frequency of $k\lambda$ are included in those with fundamental frequency λ makes this a non-trivial problem.

- Using the current program, we do not have the option of specifying the covariance structure of the errors. Therefore, all the estimation is performed under the assumption of independent Gaussian errors. In a new program this could be made an option. We might include prewhitening as part of this.

- An interesting statistical computing problem arises from this estimation procedure. In the current procedure we run the estimation procedure with the data points $Y_j, \ldots, Y_k$ and then run the same procedure on the data points $Y_{j+1}, \ldots, Y_{k+1}$. We observe that these two data sets differ only in two points. Currently the only information obtained from the first data set to estimate the second are the starting values. An interesting task would be to find a way to make the algorithm faster through the use of this information.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov and B. Csaki (eds), *Second International Symposium on Information Theory*, Academiai Kiado, Budapest, pp. 267–281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Auto Control* **AC-19**: 716–723.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika* **66**(2): 237–242.

Atal, B., Chang, J., Mathews, M. and Tukey, J. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique, *Journal of the Acoustical Society of America* **65**(5): 1535–1555.

Bates, D. and Lindstrom, M. (1986). Nonlinear least-squares with conditionally linear parameters, *Proceedings, Statistical Computing Section* pp. 152–157.

Benade, A. H. (1973). The physics of brasses, *Scientific American* **229**(1): 24–35.

Benade, A. H. (1976). *Fundamentals of Musical Acoustics*, Oxford University Press, New York, New York.

Bhansali, R. and Downsham, D. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criteria, *Biometrika* **64**: 547–551.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics*, Holden-Day, San Francisco.

Bilmes, J. (1993). *Timing is of the essence*, Master's thesis, MIT.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*, John Wiley and Son, New York.

Bolt, B. A. and Brillinger, D. R. (1979). Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series, *Geophysics Journal of the Royal Astronomy Society* **59**: 593–603.

Borin, G., Poli, G. D. and Sarti, A. (1992). Algorithm and structures for synthesis using physical models, *Computer Music Journal* **16**(4): 30–41.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrica* **52**(3): 345–370.

Brillinger, D. R. (1965). An introduction to polyspectra., *Annals of Mathematical Statistics* **36**: 1351–1374.

Brillinger, D. R. (1977). Fitting cosines: Some procedures and some physical examples, *in* I. MacNeil and G. Umphrey (eds), *Applied Probability, Stochastic Processes, and Sampling Theory*, Reidel, Boston, MA, pp. 75–100.

Brillinger, D. R. (1980). The comparison of least squares and third–order periodogram procedures in the estimation of bifrequency, *Journal of Time Series Analysis* **1**: 95–102.

Brillinger, D. R. (1981). *Time Series Data Analysis and Theory*, Expanded Edition Holden–Day, San Francisco.

Brillinger, D. R. (1986). Regression for randomly sampled spatial series: The trigonometric case, *Applied Probability Trust* pp. 275–289.

Brillinger, D. R. and Irizarry, R. A. (1998). An investigation of the second- and higher-order spectra of music, *Signal Processing* **39**.

Brillinger, D. R. and Rosenblatt, M. (1967). Computation and interpretation of the k-th order spectra, *in* B. Harris (ed.), *Spectral Analysis of Time Series*, John Wiley, pp. 189–232.

Brown, E. N. (1990). A note on the asymptotic distribution of the parameter estimates for the harmonic regression model, *Biometrika* **77**(3): 653–656.

Brown, J. C. (1996). Frequency ratios of spectral components of musical sounds, *Journal of the Acoustical Society of America* **99**(2): 1210–1218.

Chafe, C. (1990). Pulsed noise in self-sustained oscillations of musical instruments, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2, Albuquerque, NM, pp. 1157–1160.

Charbonneau, G. R. (1981). Timbre and the perceptual effects of three types of data reduction, *Computer Music Journal* **5**(2): 10–19.

Chowning, J. M. (1973). The synthesis of complex audio spectra by means of frequency modulation, *Journal of the Audio Engineering Society* **21**(7): 46–54.

Clevland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**(368): 829–836.

Clevland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association* **83**(403): 596–610.

Cook, P., Chafe, C. and Smith, J. (1990). Pulsed noise in musical systems, techniques for extraction, analysis and visualization, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Glasgow, UK, pp. 63–65.

Dahlen, F. (1981). The effect of windows on the estimation of free oscillation parameters, *Geophysical Journal of the Royal Astronomy Society* **69**: 537–549.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes, *Stochastic Processes and their Applications* **62**: 139–168.

Dahlhaus, R. (1997). *Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes*, Institute of Statistical Science Academia Sinica, Heidelberg, Germany.

Dandawate, A. V. and Giannakis, G. B. (1992). Cyclic-cumulant based identification of almost periodically time-varying systems: Parametric methods, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, IEEE, San Francisco, CA, pp. 229–232.

Davis, M. and Vinter, R. (1985). *Stochastic Modelling and Control*, Chapman and Hall, New York.

Depalle, P., García, G. and Rodet, X. (1993a). Analysis of sound for additive synthesis: Tracking of partials using hidden markov models, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Waseda University Center for Scholarly Information, pp. 94–97.

DePalle, P., García, G. and Rodet, X. (1993b). Analysis tracking partials, *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 225–228.

Depalle, P., García, G. and Rodet, X. (1995). The recreation of a castrato voice, Farinelli's voice, *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, New Paltz, NY, pp. 242–245.

DePalle, P. and Poirot, G. (1991). SVP: A modular system for analysis, processing and synthesis of sound signals, *Proceeding of the International Computer Music Conference*, International Computer Music Association, Montreal, Canada, pp. 161–164.

Depalle, P. and Tromp, L. (1996). An improved additive analysis method using parametric modelling of the short-time Fourier transform, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 297–300.

Deutsch, D. (1982). *Psychology of Music*, Academic Press, New York.

Dirst, M. and Weigend, A. (1992). Baroque forecasting: On completing J.S. Bach's last fugue, *in* A. S. Weigend and N. A. Gershenfeld (eds), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Santa Fe Institution Studies in the Sciences of Complexity, Addison-Weseley, Reading, MA, pp. 151–172.

Dubnov, S. and Rodet, X. (1997). Statistical modeling of sound aperiodicities, *Proceedings of International Computer Music Conference*, International Computer Music Association, Greece, pp. 43–50.

Dubnov, S. and Tishby, N. (1996). Influence of frequency modulation jitter on high order moments of sound residual with applications to synthesis and classification, *Proceed-*

ings of the International Computer Music Conference, International Computer Music Association, Hong Kong, pp. 378–385.

Fan, J. and Gijbles, I. (1995). Data-driven selection in local polynomial fitting: Variable bandwidth and spatial adaptation, *Journal of the Royal Statistical Society* **57**(2): 371–394.

Flanagan, J. and Golden, R. (1966). Phase vocoder, *The Bell System Technical Journal* **45**: 1493–1509.

Fletcher, N. H. and Rossing, T. D. (1991). *The Physics of Musical Instruments*, Springer-Verlag, New York.

Green, D. (1985). Temporal factors in psychoacoustics, *in* A. Miohelsen (ed.), *Time Resolution in Auditory Systems*, Springer–Verlay, New York.

Grey, J. (1975). *An exploration of musical timbre*, PhD thesis, Department of Music Stanford University.

Grey, J. and Gordon, J. W. (1978). Perceptual effects of spectral modifications of musical timbres, *Journal of the Acoustical Society of America* **63**(5): 1493–1500.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbre, *Journal of the Acoustical Society of America* **62**: 1270–1277.

Grey, J. and Moorer, J. (1977). Perceptual evaluations of synthesized musical instrument tones, *Journal of the Acoustical Society of America* **63**: 454–462.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions*, John Wiley & Sons, New York.

Hannan, E. J. (1971). Non-linear time series regression, *Journal of Applied Probability* **8**: 767–780.

Hannan, E. J. (1973). The estimation of frequency, *Journal of Applied Probability* **10**: 510–519.

Hannan, E. J. (1974). Time series analysis, *IEEE Transactions on Auto Control* **AC-19**: 706–715.

Harris, C. and Weiss, M. (1963). Pitch extraction by computer processing of high resolution Fourier analysis, *Journal of the American Statistical Association* **35**: 339–343.

Hartman, W. M. (1997). *Signals, Sound, and Sensation*, AIP Press, Woodburg, New York.

Hasan, T. (1979). *Complex Demodulation: Some Theory and Some Applications*, PhD thesis, University of California, Berkeley.

Hasan, T. (1983). Complex demodulation: Some theory and applications, *in* D. R. Brillinger and P. Krishnaiah (eds), *Handbook of Statistics*, Vol. 3, Elsevier Science Publishers, pp. 125–156.

Hassan, T. (1982). Nonlinear time series regression for a class of amplitude modulated cosinusoids, *Journal of Time Series Analysis* **3**: 109–122.

Hastie, T. J. and Tibshirani, R. J. (1990). *General Additive Models*, Chapman and Hall, London.

Hiller, L. and Ruiz, P. (1971). Synthesizing sounds by solving the wave equation for vibrating objects, *Journal of the Audio Engineering Society* **19**: 463–470.

Hirschman, S. (1991). *Digital Waveguide Modeling and Simulation of Reed Woodwind Instruments*, PhD thesis, Stanford University Department of Electrical Engineering.

Hirschman, S., Cook, P. and Smith, J. (1991). Digital waveguide modeling of reed woodwinds: An interactive development, *in* B. Alphonce and B.Pennycook (eds), *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, pp. 300–303.

Hurvich, C. M. (1997). Mean square over degrees of freedom: New perspectives on a model selection treasure, *in* D. R. Brillinger, L. Fernholz and S. Morgenthaler (eds), *The Practice of Data Analysis: Essays in Honor of John W. Tukey*, Princeton University Press, Princeton, NJ, pp. 203–215.

Jaffe, D. A. and Smith, J. O. (1983). Extensions of the Karplus-Strong plucked-string algorithm, *Computer Music Journal* **7**(2): 56–69.

Jones, K. (1981). Compositional applications of stochastic processes, *Computer Music Journal* **5**(2): 381–396.

Junhar, J. (1997). Advanced pitch detection algorithms, *in* D. Kocur, D. Levicky and M. S. (eds), *3rd International Conference on Digital Signal Processing*, DSP, Kosice, Slovakia.

Karplus, K. and Strong, A. (1983). Digital synthesis of plucked-string and drum timbres, *Computer Music Journal* **7**(2): 467–479.

Kashyap, R. L. (1982). Optimal choice of AR and MA parts in autoregressive moving average models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-4**(2): 99–104.

Kendal, M. and Stuart, M. (1967). *The Advanced Theory of Statistics*, Vol. 2, second edn, Hafner Publishing, New York.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika* **83**(4): 875–890.

Kullback, S. (1959). *Information Theory and Statistics*, John Wiley & Sons, New York.

Lehmann, E. L. (1983). *Theory of Point Estimation*, John Wiley and Sons, New York.

Linhart, H. and Zucchini, W. (1986). *Model Selection*, John Wiley & Sons, New York.

Loy, G. (1989). Musicians make a standard: The MIDI phenomenon, *in* C. Roads (ed.), *The Music Machine*, Vol. 9, The MIT Press, Cambridge, MA.

Machado, J. A. (1993). Robust model selection and M-estimation, *Econometrics* **9**: 478–493.

Maganza, C. and Caussé, R. (1986). Bifurcations, period doubling and chaos in clarinet-like systems, *Europhysics Letters* **1**(6): 295–302.

Mallows, C. (1973). Some comments on C_p, *Technometrics* **15**(4): 661–675.

Martin, P. (1982). Comparison of pitch detection by cepstrum and spectral comb analysis, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, IEEE, New York, pp. 180–183.

Masri, P. and Bateman, A. (1996). Improved modelling of attack transient in music analysis-resynthesis, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 100–104.

Mathews, M. V. (1963). The digital computer as a musical instrument, *Science* **142**(11): 553–557.

Mathews, M. V. (1969). *The Technology of Computer Music*, MIT Press, Boston, MA.

Mathews, M. V. and Pierce, J. R. (1989). *Current Directions in Computer Music*, The MIT Press, Cambidge, MA.

McAualay, R. J. (1986). Speech analysis/synthesis based on a sinusoidal representation, *IEEE Transactions of Acoustics, Speech and Signal Processing* **ASSP-34**(4): 744–754.

Moorer, J. A. (1977). Signal processing aspects of computer music: A survey, **65**(8): 1108–1137.

Patterson, R. and Green, D. (1970). Discrimination of transient signals having identical energy spectra, *Journal of the Acoustical Society of America* **48**: 894–905.

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. and Allerhand, M. (1992). Complex sounds and auditory images, *in* Y. Cazals, L. Demany and K. Horner (eds), *Auditory Psychology and Perception*, Pergamon, Oxford, pp. 426–446.

Pierce, J. (1992). *The Science of Musical Sound*, Freeman, New York.

Platonov, A., Gajo, Z. and Szabatin, J. (1992). General method for sinusoidal frequencies estimation using ARMA algorithms with nonlinear prediction error transformation, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, IEEE, San Francisco, pp. 441–444.

Poli, G. D. (1989). A tutorial on digital sound synthesis techniques, *in* C. Roads (ed.), *The Music Machine : Selected Readings from Computer Music Journal*, The MIT Press, pp. 429–446.

Portnoff, M. L. (1976). Implementation of the digital phase vocoder using the fast Fourier transform, *IEEE Transactions Acoustics, Speech, and Signal Processing* **ASSP–24**(3): 243–248.

Quatieri, T. F. (1992). Shape invariant time-scale and pitch modification of speech, *IEEE Transactions of Signal Processing* **40**(3): 497–510.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, second edn, John Wiley & Sons, New York, p. 33.

Rayleigh, J. (Reprinted 1945). *The Theory of Sound*, Dover, New York.

Rhea, T. (1984). The history of electronic musical instruments, *in* T. Darter (ed.), *The Art of Electronic Music*, Quill, New York, pp. 1–63.

Risset, J.-C. and Mathews, M. V. (1969). Analysis of musical-instrument tones, *Physics Today* **22**(2): 23–30.

Risset, J.-C. and Wessel, D. L. (1982). Exploration of timbre by analysis and synthesis, *in* D. Deutsch (ed.), *The Psychology of Music*, Academic Press, New York.

Roads, C. (1996). *The Computer Music Tutorial*, The MIT Press, Cambridge, MA.

Rodet, X. (1997). Musical sound signals analysis/synthesis: Sinusoidal+residual and elementary waveform models, *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop (TFTS'97)*, IEEE, Coventry, UK.

Rodet, X. and Depalle, P. (1992). A new additive synthesis method using inverse Fourier transform and spectral envelops, *Proceedings of International Computer Music Conference*, International Computer Music Association, San Jose, California, pp. 410–411.

Rodet, X., Potard, Y. and Barrière, J.-B. (1984). The CHANT project: From the synthesis of the singing voice to synthesis in general, *Computer Music Journal* **8**(3): 15–31.

Rodet, X. and Vergez, C. (1996). Physical models of trumpet-like instruments detailed behavior and model improvement, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 448–453.

Ronchetti, E. (1985). Robust model selection in regression, *Statistics & Probability Letters* **3**: 21–23.

Ronchetti, E. and Staudte, R. G. (1994). A robust version of Mallow's C_p, *Journal of the American Statistical Association* **89**(426): 550–559.

Rowe, R. (1994). *Interactive Music Systems*, The MIT Press, Cambridge, Massachusetts.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**(2): 461–464.

Sclove, S. L. (1994). Some aspects of model-selection criteria, *in* H. Bozdogan (ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling*, Kluwer Academic Publisher, Netherlands, pp. 37–67.

Seber, G. A. (1977). *Linear Regression Analysis*, Wiley, New York.

Serra, X. (1989). *A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition*, PhD thesis, Stanford University.

Serra, X. J. and Smith, J. O. (1991). Spectral modeling synthesis: A sound analysis/synthesis system based on deterministic plus stochastic decomposition, *Computer Music Journal* **14**(4): 12–24.

Smith, J. and Gossett, P. (1984). A flexible sampling-rate conversion method, *Proc. IEEE ICASSP*, Vol. 2, IEEE, San Diego, pp. 19.4.1–19.4.2.

Smith, J. O. (1985). Fundamentals of digital filter theory, *Computer Music Journal* **9**(3): 13–23.

Smith, J. O. (1991). Viewpoints on the history of digital synthesis, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Montreal, Canada, pp. 1–10.

Solbach, L. and Wöhrmann, R. (1996). Sound onset localization and partial tracking in gaussian white noise, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 324–327.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models, *Journal of the American Statistical Association* **84**(405): 276–283.

Sullivan, C. R. (1990). Extending the Karplus-Strong filter to synthesize electric guitar timbres with distortion and feedback, *Computer Musicl Journal* **14**(3): 26–37.

Tabei, M., Musicus, B. and Ueda, M. (1991). A maximum likelihood estimator for frequency and decay rate, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, IEEE, Ontario, Canada, pp. 3125–3128.

Tellman, E., Haken, L. and Holloway, B. (1995). Timbre morphing of sounds with unequal numbers of features, *Journal of the Audio Engineering Society* **42**(9): 678–689.

Templaars, S. (1977). The VOSIM signal spectrum, *Interface* **6**: 81–96.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation, *Journal of the American Statistical Association* **82**: 559–567.

Välimäki, V., Hänninen, R. and Karjalainen, M. (1996). An improved digital waveguide model of a flute - implementaion issues, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 1–4.

Vaseghi, S. V. and Rayner, P. J. W. (1988). A new application of adaptive filters for restoration of archived gramophone recordings, *International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, IEEE, New York, pp. 2548–51.

Verge, M.-P. (1996). Physical modeling of aeroacoustics sources in flute-like instruments, *Proceedings of the International Computer Music Conference*, International Computer Music Association, Hong Kong, pp. 100–104.

von Helmholtz, H. (1885). *On the Sensation of Tone*, (trans. A.J. Ellis), London.

Voss, R. F. and Clarke, J. (1975). '1/f noise' in music and speech, *Nature* **258**: 317–318.

Voss, R. F. and Clarke, J. (1978). '1/f noise' in music: Music from 1/f noise, *Journal of the Acoustical Society of America* **63**(1): 258–263.

Wagner, M. J. (1978). *Introductory Musical Acoustics*, Contemporary Publishing Company, Raleigh, N.C.

Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals, *Biometrika* **58**: 21–36.

Wang, X. (1991). *On the Estimation of Trigonometric and Related Signals*, PhD thesis, University of California at Berkeley.

Wessel, D. (1987). Control of phrasing and articulation in synthesis, *Proceedings of the International Computer Music Conference*, International Computer Music Association, pp. 108–116.

Wilson, G. R., Hardwicke, K. R. and Trochta, R. T. (1992). Coherent harmonic detection using non-stationary higher order spectra, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. ICASSP-92, San Francisco, CA, pp. 201–204.

Yang, J.-T. and Cabrera, S. (1992). Adaptive weighted norm linear prediction for sinusoids incorporating a priori frequency information, *International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, IEEE, San Francisco, CA, pp. 201–204.