UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Institute of Sound and Vibration Research

Acoustic Features of Piano Sounds

by

Christos Karatsovis

Thesis for the degree of Doctor of Philosophy

November 2011

UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING AND THE ENVIRONMENT
INSTITUTE OF SOUND AND VIBRATION RESEARCH

ABSTRACT

ACOUSTIC FEATURES OF PIANO SOUNDS

by Christos Karatsovis

To date efforts of music transcription indicate the need for modelling the data signal in a more comprehensive manner in order to improve the transcription process of music performances. This research work is concerned with the investigation of two features associated with the reproduced sound of a piano; the inharmonicity factor of the piano strings and the double decay rate of the resulting sound. Firstly, a simple model of the inharmonicity is proposed and the factors that affect the modelled signal are identified, such as the magnitude of the inharmonicity, the number of harmonics, the time parameter, the phase characteristics and the harmonic amplitudes. A formation of a so-called "one-sided" effect appears in simulated signals, although this effect is obscured in real recordings potentially due to the non-uniformly varying amplitudes of the harmonic terms. This effect is also discussed through the use of the cepstrum by analysing real piano note recordings and synthesized signals. The cepstrum is further used to describe the effect of the coupled behaviour of two strings through digital waveguides. Secondly, the double decay rate effect is modelled through coupled oscillators and digital waveguides. A physical model of multiple strings is also presented as an extension to the simple model of coupled oscillators and various measurements on a real grand piano are carried out in order to investigate the coupling mechanism between the strings, the soundboard and the bridge. Finally, a model, with reduced dimensionality, is proposed to represent the signal model for single and multiple notes formulated around a Bayesian framework. The potential of such a model is illustrated with the transcription of simple examples of real monophonic and polyphonic piano recordings by implementing the Metropolis-Hastings algorithm and Gibbs sampler for multivariate parameter estimation.

Contents

Declaration of authorship

I, Christos Karatsovis, declare that the thesis entitled "Acoustic Features of Piano Sounds" and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly while in candidature for a research degree at this University;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help.

Signed:……

Date:……

Acknowledgements

I have been part of the ISVR since 1996, when I originally came to the UK to study for a Bachelor's Degree in Engineering Acoustics and Vibration and then after 3 years I moved on to do a Master's Degree in Sound and Vibration Studies. Over the past 10 most recent years, I have been working full-time in ISVR Consulting whilst studying for my Doctor of Philosophy candidature. ISVR has been very much of a home for me.

Throughout these 15 years, I have met, without a doubt, the most helpful and friendliest people that I could ever have imagined. The person, who has been the most supportive throughout my academic and career path, is Emeritus Professor Joe Hammond. Without his constant help and support I would not have been able to complete this work. He has always been my mentor and will always be.

I would like to express my deepest gratitude to my supervisor Professor Paul White, who has really helped me to raise the level of this research and Dr Neil Ferguson, who has helped me numerous times in understanding complicated structural dynamics concepts associated with this work.

I cannot thank enough Mrs Maureen Mew; she has been extremely helpful as always. She would always find time to help you out no matter what. I would also like to thank Mrs Norma Martin, who was the Personal Assistant of Emeritus Professor Joe Hammond as Dean of the Engineering Faculty at the time. She was always willing to arrange dates for meetings and venues.

I would also like to equally thank ISVR and ISVR Consulting for sponsoring this work over the years and a big thank you goes to Mrs Diane Farrenden for sorting out printed and bound copies over the years.

I would also like to thank all my friends who supported me throughout these years.

Last but not least, I would like to thank my mother, Katerina, and my sister, Maria, for their immeasurable love. Also, I would like to say that my father would have been very proud of me.

List of main variables

α_n , β_n	Amplitudes of the modelled sines and cosines of the signal
ζ_n	Decay rates (or damping factors) of the modelled sines and cosines
Θ	Parameter space
ξ_i	Random perturbation of estimate at the i^{th} iteration of the M-H algorithm
ρ	Linear density associated with the material of a piano string
ϕ	Phase term
b	Harmonic amplitudes matrix
d	Signal matrix
e	Additive Gaussian noise matrix
G	Basis function matrix containing the signal parameters
B	Inharmonicity associated with a piano note
E	Young's modulus of elasticity of a piano string
F	Force applied to an oscillator
f	Frequency
K	Radius of gyration of a piano string
L	Number of modelled signal amplitudes
l	Effective length of a vibrating piano string
M	Signal length
m	Mass of an oscillator
N	Number of harmonics present in the signal (inc. the fundamental frequency)
P	Power associated with an oscillator
Q	Acceptance function of the parameter space Θ
R	Number of notes present in the signal
T	Tension of a piano string
t	Time parameter
$Y_{x_1 x_2}$	Mobility between points x_1 and x_2

Chapter 1

Project introduction

1.1 Introduction

In all civilisations, music represents an important form of expression. In western cultures, the written form of music, by means of a written score or a manuscript, plays a central role in the way of allowing composers to share their intentions and communicate them with performing musicians.

Music transcription is the process of converting a live or recorded performance into a written score. Different information is represented in a score: the notes associated with their pitch, the duration, the tempo and dynamics. This information, in conjunction with the individual skills of the performing musician, can give rise to different interpretations. As a matter of fact, the personal interpretation of a musical piece can differentiate a masterful from a poor performance.

Historically, the problem of automatic music transcription relating to pitch identification has been in existence for many decades now. It is considered to be a multidisciplinary task blending different areas of science, art and engineering, such as signal processing, psychoacoustics, and musical acoustics.

Music instruments that can only play a single note at a time are called monophonic, whereas instruments that can play a multitude of notes are called polyphonic. These two different types of instrument require different techniques when transcribing music performances. Monophonic performances are relatively straightforward to analyse, whilst polyphonic performances are more involved since a multitude of notes can share a number of same harmonics and therefore make the transcription process difficult to implement in practice. Standard frequency analysis techniques in the latter case are not always adequate.

The original motivation for this research work was to develop a novel technique for automatically transcribing polyphonic performances specifically written for the piano.

In the literature, many methods have been proposed for transcribing different music performances, some more successful than others as will be discussed later in detail. Some methods utilise sophisticated signal processing methods, where statistical prior knowledge of the music is incorporated in a model, whilst others propose methods of "training" the data through a set of parameters. Other methods implement psychoacoustic models for emulating the ability of the human ear to distinguish between different sounds. Finally, there are methods that may combine a number of the above methods.

The majority of these methods and techniques use relatively simple forms of describing the signal, whereas only a few incorporate some of the unique physical characteristics relating to the sound reproduction mechanism of a music instrument, such as the inharmonicity factor relating to the bending stiffness found in piano strings. These simple signal models may be regarded as an oversimplification of the true representation of the actual reproduced sound rendering the transcription process difficult, if not impossible.

Therefore, it has become more evident over the years that more effort should be placed on understanding the complex mechanism of sound generation in polyphonic instruments through the investigation of some of their unique physical characteristics. The focus of this research work is to assess in particular the sound generation mechanism of the piano, as an important example of polyphonic music instrument, rather than proposing yet another music transcription method for polyphonic music.

1.2 Approaches and difficulties of automatic music transcription

In this section, an attempt is made to briefly reveal the difficulties experienced by the different approaches in the automatic transcription of music and will make it evident as to why the need of modelling the physics of the signal in a more comprehensive matter is important to the transcription process. A more detailed description of the different methods and limitations of transcribing monophonic and polyphonic music performances is presented in chapter 2.

The methods and techniques in automatic music transcription can be categorised in many different ways. We have chosen to present them in three broad groups, i.e. neural network, probabilistic and blind separation methods. Hybrid methods may combine any of the three groups.

1.2.1 Auditory perception

The trained ear of a musician is capable of distinguishing between different notes and different instruments in a recording. However, the combined processing mechanism of the human ear coupled with that of the brain is not yet fully understood.

In simple terms, the functionality of the human ear is divided into two main parts (Klapuri and Virtanen, 2008). First, the signal is passed through a number of band pass filters or "channels" (approximately 100 in number) that are used to represent the frequency response (or selectivity) of the human ear. Second, the signal in each channel produces the neural impulses to the auditory nerve associated with the different hair cells and sound is ultimately perceived by the brain.

It is believed that the auditory information associated with pitch identification takes place both in each auditory channel and then combined across all channels (Cheveigné, 1999). In the past, a model was developed as a means of emulating the functionality of the human ear (Meddis, 1986). In particular, the harmonic deduction, i.e. the identification of the harmonics, can be based on a filter bank by splitting the signal into several frequency channels, and then the output of each channel can be coupled to Meddis's model of hair cell induction. Meddis and Hewitt (1990) have in

fact proposed a process by which autocorrelation functions are computed in each channel of the inner ear and then they are summed across all the channels in order to predict the maximum of the autocorrelation function representing the perceived pitch. However, despite significant efforts in developing even more accurate pitch perception models over the years, there are still functions in the auditory perception mechanism that are not yet fully understood.

1.2.2 Neural network methods

The automatic music transcription method utilising a neural network formalism is based on an iterative training process for the parameter estimation. These methods normally use a segmentation routine in the form of an average Short Time Fourier Transform (STFT) feeding a blackboard system (Bello *et al*., 2000). The blackboard system would consist of the database with all the hypotheses of the model, the scheduler and the knowledge sources executing the required actions. Polyphonic music transcription is not successful with this method when there is strong harmonic overlap. Most of these methods are focused on revealing the frequency information of the signal, despite the main physical phenomena taking place in the time domain. However, some more advanced neural network models take into account some of the physical characteristics of the music instrument, such as the piano. For instance, in one of the methods (Ortiz-Berenguer *et al.*, 2005), 88 patterns are used, one for each piano note, in a neural network training process coupled with a simple acoustical model of the piano. In this model, the bending stiffness of the strings, or inharmonicity, of the piano is also modelled as part of the process. Training is carried out on a few notes to compute their associated inharmonicity factors and then a model is used to obtain the interpolated values of the inharmonicity factor for the remaining notes of the piano.

An automatic music transcription technique could also be used as a combination of a comprehensive auditory model with a neural network framework (Marlot, 1999 and 2001). However, such hybrid models, although very comprehensive and useful, are limited to the interpretation of the output of the signal and do not assess the sound at its origin point before being perceived as an auditory signal.

1.2.3 Probabilistic methods

Another method of music transcription is the one in which the emphasis is placed on probabilistic methods, such as Bayesian and Markov Chain Monte Carlo (MCMC) techniques. In particular, the parameters of an unknown number of notes with an unknown number of harmonics, based on Bayesian formalism, is calculated by allowing prior knowledge about the nature of the data to be incorporated into the model. The data is modelled in frames and joint parameter estimation can be performed across multiple adjacent frames for obtaining estimates of notes, music intervals and chords (Walmsley *et al.*, 1999). The majority of these methods utilise a fairly basic model for the description of the signal. They are primarily based on the principles of the generalised linear model. For instance, the signal is modelled as a sum of sine and cosine waves with random Gaussian noise.

More advanced attempts have been implemented in a time-varying amplitude process for each harmonic present in the signal and their inharmonic relationship (Godsill and Davy, 2002; Davy and Godsill, 2002). Such representations may considerably increase the dimensionality of the model rendering it expensive to compute numerically. These methods are most accurate in transcribing monophonic performances, as well as having some success in transcribing polyphonic performances. Limited success in the case of certain music intervals, such as octaves and fifths, is achieved due to the considerable overlap of the shared harmonics of the notes (Godsill and Davy, 2002; Davy and Godsill, 2002).

1.2.4 Blind separation methods

Over the years, “blind” separation techniques, such as Independent Component Analysis (ICA) and “sparse coding”, have been gaining ground as a means of obtaining the parameters from mainly mixed observations, e.g. polyphonic performances, where there is no statistical (or prior) knowledge of any of the signals (Klapuri and Virtanen, 2008).

In the basic form of ICA, the mixed signal is analysed in the time-frequency domain using typical time-frequency analysis techniques, such as the STFT. The weighted

sum of the basis spectra that represent the mixed signal (including a “noise” term) can be analysed to obtain musically important information, such as the fundamental frequency of the signal (Klapuri and Virtanen, 2008). A special case of ICA, is sparse coding, where a cost function can be minimised to obtain the frequencies in the signal provided the data can be modelled in terms of a small number of active elements chosen out of a large set.

1.3 Main themes of the project

Significant effort has been placed in the development and understanding of music transcription models based on signal processing approaches, such as Bayesian, neural network and blind separation methods, as mentioned briefly in the previous sections. Most methods employ simplistic models to describe the sound reproduction mechanism of a piano and very little emphasis has been placed on the development of a comprehensive physical model of a piano based on its rather unique sound reproduction features that could in turn be used in a transcription method.

In the past, the characteristic features associated with the mechanism of sound reproduction of the piano have been addressed and investigated by many authors. These features concern the interaction between the hammers and strings of a piano and the different sound produced due to different playing dynamics (Helmholtz, 1877; Askenfelt and Jansson, 1988; Hall, 1987), the inharmonicity in piano strings (Fletcher, 1964; Taylor, 1965; Rossing, 1990), and the double decay rate characteristics of piano notes (Weinreich, 1977; Naganuma *et al.*, 2004; Tanaka *et al.*, 1999; Nishiguchi *et al.*, 2003 and 2004).

This research serves as a basis for a further investigation of some of these acoustic features through a new perspective. In particular, novel observations are made regarding the inharmonicity factor of piano strings and the importance of the double decay rate in piano notes, as well as appropriate analytical models are developed for these features combining dynamics theory, signal processing techniques and real experimental data. These modelled features may be used as part of a newly proposed signal model, based on a Bayesian formalism framework, which could ultimately be incorporated in a more comprehensive transcription method in the future. The proposed signal model is brought together in chapter 5 of this research.

1.4 Original contributions

The following original contributions for the candidature for a research degree at the University of Southampton are summarised below:

- This work is mainly focused in describing "imperfections" in piano sounds. These imperfections may constitute part of a more comprehensive signal model as a means of potentially improving the transcription accuracy in monophonic and polyphonic music performances. This conceptual approach may be extended to other instruments provided other characteristic imperfections are identified.

- The effect of inharmonicity, found in piano strings, is therefore modelled as a means of providing a more comprehensive signal model for representing the reproduced sound of a piano. Novel observations and analytical formulations associated with the inharmonicity are also presented in the time and frequency domain in order to further unravel its effect on the reproduced sound.

- The effect of the double decay rate is modelled and presented as a means of understanding the resulting piano sound. This effect is discussed in the frequency and time domain through the modelling of coupled oscillators. Also, a physical modelling extension is presented for the total number of strings present in a piano instrument.

- The coupling mechanism between the string, the bridge and the soundboard of the piano is investigated through measurements carried out on a real grand piano. Observations associated with this complicated coupling mechanism are presented and compared with theoretical formulations.

- A signal model is proposed for single and multiple piano notes in a probabilistic Bayesian framework, the dimensionality of which is considerably smaller than existing attempts in the literature (Godsill and Davy, 2002; Davy and Godsill, 2002). Indeed, in the existing literature, a "de-tuning" parameter is used for each individual harmonic present in the signal due to their non-integer frequency spacing relative to the fundamental frequency. Hence, for N number of harmonics, there will be N de-tuning parameters that need to be

computed. However, in this research, we propose a single inharmonicity factor for all *N* harmonics.

- The potential of such a model is illustrated with the transcription of simple real monophonic and polyphonic piano recordings. Multivariate estimation of the parameter space is achieved through the implementation of novel criteria embedded in known algorithms, such as the Metropolis-Hastings (M-H) algorithm and Gibbs sampler.

1.5 Conclusions

This chapter has provided the introduction and rationale for the project. The original objective was to develop a method for automatically transcribing polyphonic music performances for the piano. However, it has become clear from the literature that there is little to suggest that certain important features of the piano have been considered in detail in the signal models of the transcription methods. This may explain, to a certain extent, the poor success rate of the different methods of polyphonic music transcription as will be discussed in chapter 2.

The focus of this research work in the next chapters would be to further investigate some of the important features of the piano instrument, such as the inharmonicity and the double decay rate effect associated with the sound generation mechanism of a piano through the use of modelling and experimental validation. A newly proposed signal model, based on Bayesian formalism, will be discussed in chapter 5 following these investigations.

Chapter 2

Existing knowledge review

2.1 Introduction

This chapter provides an overview of the methods of automatic music transcription found in the literature and presents the various important acoustic features associated with the piano's sound generation mechanism as discussed by other authors in the past.

This chapter is divided into two main sections. The first section describes the methods of automatic music transcription with special emphasis on the problem of octave detection and the lack of physical modelling in those methods, whilst the other section describes the acoustic features of the sound generation mechanism of the piano through the hammer-string interaction, the radiation from the soundboard, the inharmonicity found in pianos strings, the double decay rate effect and the modelling methods of dynamically coupled strings.

2.2 The problem of octave detection in musical signals

Pitch recognition of monophonic music is fairly straightforward, since there is no need to identify notes with shared harmonics (or also referred to as "partials" in the literature). On the other hand, polyphonic music is a very complex subject since a multitude of notes can share a number of coincident harmonics. The most problematic musical interval for transcription is the octave. This is a musical interval between two notes played simultaneously in which the fundamental frequency and the harmonics of the higher note coincide with all the harmonics of the lower note. In the latter case, a simple analysis of the Fourier spectrum of the signal is not sufficient to separate the notes and transcribe them into a written form. Also, the task of polyphonic music transcription can become even more difficult when trying to differentiate instruments that are being played simultaneously in a music piece.

Many methods have been proposed and developed in order to solve the problem of polyphonic music transcription. However, the separation of notes in an octave has not been resolved yet despite considerable effort. As a result, the octave detection problem has become one of the ultimate challenges in polyphonic music transcription.

2.3 Methods of automatic music transcription

2.3.1 Early approach

Moore (1977) researched the transcription of music played by two instruments. The method he developed was able to detect the notes played by the two instruments, although there were restrictions since the instruments were not allowed to "cross". This meant that the fundamental frequency of a note played on one instrument was not allowed to be greater, at any given time, than the fundamental frequency of the played note on the other instrument. The detection of an octave interval using Moore's method was not possible.

2.3.2 Neural network approach

Bello *et al.* (2000) suggested a method based on neural network formalism in order to analyse simple polyphonic tracks. The method uses a segmentation routine in the form of an average STFT feeding a blackboard system. The blackboard system consists of the database with the hypotheses of the model, the scheduler, and the knowledge sources executing the intending actions. One of the knowledge sources is a network chord recogniser. The authors stated that octave detection was not possible with this method due to the high number of coincident harmonics associated with this musical interval.

Chien *et al.* (2002) suggested another method, based on a neural network formalism, for octave detection in the case of the piano instrument. In this method a constant Q time-frequency analysis method is implemented via a nonorthonormal discrete wavelet transform. A Support Vector Machine (SVM) technique is used as an octave detector in the system. By implementing this method for the transcription of polyphonic music, 3 out of 4 octaves are recognised showing the potential success of this method in octave detection.

Marlot (1999) also developed a chord recognition method for the piano instrument based on neural network formalism. The training set is a large database, which was developed by gathering recordings of single piano notes covering the whole playing

range of the piano at different dynamic levels. These recordings were generated through commercially available CD audio recordings and several synthesizer sounds. The chords were then generated from the individual note recordings. As far as the neural network architecture is concerned, four different feed forward neural networks were tested in this method; multi-layer perceptrons, radial basis function networks, SVMs and time-delay networks. Again, the octave detection problem was not successfully resolved due to the high number of coincident harmonics in the octave music intervals. Marlot (2001) also suggested another transcription method called SONIC by using an auditory model based on Meddis's theory (1986). The latter theory describes the simulation of the basilar membrane in the inner ear using a bank of filters. A filter bank effectively splits the auditory signal into frequency channels. The output of the auditory model is an autocorrelogram representing the signal with respect to time, channel centre frequency and autocorrelation lag. Autocorrelograms can be summed up to estimate the periodicity in a signal and hence explain the pitch perception of the human ear. In Marlot's method the auditory model is combined with adaptive oscillators instead, formed into networks, in order to determine harmonics in a music piece. The conclusion from this method is that the vast majority of errors are associated with misjudged repeated notes and octave intervals.

Pertusa *et al.* (2005) suggested an interesting transcription method based on the identification of a pattern of a given instrument in the frequency domain. In particular, band-grouped spectrograms of polyphonic music performances are combined with time-delay neural networks for obtaining estimates. In this way, a complex auditory model and a signal processing method is avoided. A learning algorithm with these grouped spectral bands is used to detect polyphonic performances, where a dynamic neural network is employed for the note detection and characterisation process. Four categories of timbre have been investigated in this method; sinusoidal, sawtooth, clarinet and Hammond organ waveshapes. A near perfect accuracy with this transcription method is achieved for recognising these specific instrument categories. Note that all four categories are characterised by sounds that are nearly stable in time along the duration of a note. On the other hand, the sound of a piano has time-varying amplitude characteristics due to its transient nature, so the transcription with this method may be limited.

Poliner *et al.* (2006) have proposed a transcription method of combining probabilistic methods with neural network formalism. In this method, SVM classifiers with Hidden Markov Models (HMM) are combined in order to transcribe both synthesized and real piano recordings. In particular, the classifier outputs of the SVM are temporally smoothed by an HMM as part of a post-processing stage. Over two thirds of the transcription results are accurate with this method. The advantage of this method is that it does not require prior knowledge of how the harmonic features of the signal are represented in the model. This feature of this method minimises the analysis time of the music transcription process, but perhaps at the expense of accuracy. Physical modelling could potentially improve the accuracy of this particular music transcription method through the use of a more comprehensive signal model.

2.3.3 Probabilistic approach

In this section, the probabilistic approach is investigated in more detail as opposed to the neural network methods in which mere 'training' of the data is normally performed. One of the main aims of this research project is to aid in the development of a more comprehensive signal model, hence the understanding of the probabilistic methods offers perhaps a more effective approach to achieve this.

Many authors have suggested the parameter estimation of monophonic and polyphonic music performances through the use of probabilistic models. This section is primarily focused on the use of a Bayesian formalism allowing prior knowledge about the nature of the data to be incorporated into the transcription model. The majority of these models simplify the representation of the signal as a sum of sinusoids.

An explanation of Bayesian formalism is presented in Appendix A.

In particular, Walmsley *et al.* (1999) suggested a method of estimating the parameters of an unknown number of notes with an unknown number of harmonics based on Bayesian formalism allowing prior knowledge about the nature of the data. This approach is based on the use of a harmonic model in order to estimate parameters,

such as the fundamental frequency and detect the model order, such as the number of concurrently sounding notes and the number of harmonics in each one.

The data is segmented into frames $\mathbf{d_i}$ of length M_i during which the data is considered stationary. Each frame consists of a maximum number of R notes and each note r consists of the fundamental frequency ω_i^r, number of harmonics H_i^r and the harmonic amplitudes $\mathbf{b_i^r}$. Each note can also be switched in and out of the model using a binary indicator variable λ_i^r.

The parameter estimation is based on MCMC methods using the Metropolis-Hastings (M-H) algorithm to produce the maximum posterior parameter estimation. A number of transition kernels are proposed to explore the parameter space. The recognition of octaves and fifths is not successful with this method.

The model is constructed in terms of the generalised linear model with the addition of matrix $\mathbf{e_i}$ with random entries chosen from a normal distribution with zero mean, standard deviation one, and having variance $\sigma_{e_i}^2$. The formulation representing the generalised linear model is given by

$$\mathbf{d_i} = \sum_{r=1}^{R} \lambda_i^r \mathbf{G_i^r} \mathbf{b_i^r} + \mathbf{e_i}$$

where $\mathbf{G_i^r}$ is the harmonic basis matrix of the model.

The likelihood function $p(\mathbf{d_i}|\Theta_i^r, \sigma_{e_i}^2)$ of the above model can be maximised in order to estimate the note parameters $\Theta_i^r = \{\lambda_i^r, \omega_i^r, H_i^r, \mathbf{b_i^r}\}$. However, this method does not account for any prior knowledge of the parameters. A prior knowledge is incorporated into the model using a set of "hyperparameters" $\{\Delta_\Theta^r\}$. In this model, these so-called block hyperparameters, v^r and $\sigma_{\omega^r}^2$, represent the spread of the pitch over a data block and Λ^r denotes whether the note is active over each data block.

Various types of prior distribution are chosen for the model

$$p(\mathbf{b_i^r}),\ p(H_i^r),\ p(\lambda_1^r|\Lambda^r),\ p(\omega_i^r|\mathrm{v}^r, \sigma_{\omega^r}^2),\ p(\sigma_{e_i}^2),\ p(\Lambda^r) \text{ and } p(v^r).$$

The joint posterior distribution of the model would be

$$p\left(\Theta_i^r, \Delta_\theta^r, \sigma_{e_i}^2 \middle| \mathbf{d_i}\right) \propto p\left(\Theta_i^r, \Delta_\theta^r, \sigma_{e_i}^2\right) \prod_{i=1}^{M_f} p\left(\mathbf{d_i} \middle| \Theta_i^r, \sigma_{e_i}^2\right)$$

The joint posterior density is difficult to optimise analytically and therefore a more appropriate numerical approach is considered. The M-H algorithm is used to model the successive states of the Markov chain, where the final state of the chain is used to estimate the parameters in question.

The M-H acceptance function for a parameter space Θ can be written as follows

$$Q(\Theta^k, \Theta^*) = \frac{p(\Theta^* | \mathbf{d}) T(\Theta^*, \Theta^k)}{p(\Theta^k | \mathbf{d}) T(\Theta^k, \Theta^*)}$$

where $T(\Theta^k, \Theta^*)$ produces a proposal state Θ^* from the current state Θ^k that is accepted with probability $\min\left(1, Q\left(\Theta^k, \Theta^*\right)\right)$. The rationale for the acceptance function is discussed further in Appendix A.

The state space move is proposed with the use of local and global transition kernels. Global kernels will move the Markov chain into high probabilistic regions, whilst local ones will obtain more accurate parameter estimation.

Local kernels are simply random perturbations about the current value of the M-H acceptance function.

In terms of global kernels, the independence sampler is used to define a proposal distribution $r\left(\omega^{r*} \middle| H^*\right)$. The proposal distribution of the latter sampler mimics the target posterior distribution having a high acceptance function, rather depending on the current state. So, the distribution would have its modes in similar locations to the posterior density. The modes of this distribution are the fundamental frequencies of the signal and in turn the harmonic amplitudes are calculated from a least-squares projection of the model. A multiple step is also used to overcome problems with octaves that rise due to harmonic overlap by carrying out a joint move for the following parameters of note *r*: $\left\{v^r, \left\{\omega_i^r, \mathbf{b_i^r}, H_i^r\right\} i\right\}$. In that way, the kernel traverses harmonically related modes of the posterior distribution.

The above model is used on a recording, where major chords and octaves may be detected, although the recognition process becomes very difficult when octaves and chords are played together due to the large number of common harmonics.

Davy *et al.* (2002) re-developed the latter model by implementing a time-varying amplitude process for each of the harmonics present in the signal. Further flexibility was also incorporated into the process by modelling non-stationary error and inharmonicity in the signal, through the use of a parameter named "de-tuning". The estimation of the parameters was obtained by using a reversible jump MCMC algorithm.

Results with this method have been reported for monophonic and polyphonic music transcriptions. In the latter case, the successful transcription of polyphonic music consisting of a 2-note mixture of a saxophone and a trumpet, with different fundamental frequencies, has been reported.

Leistikow *et al.* (2004) developed a Bayesian framework model for identifying music intervals and chords operating on single-frame STFT peaks. In principal, the pitch component information is evaluated by an MCMC approach accounting for overlapping harmonics and spurious peaks. To obtain the posterior probability estimates, the input signals are derived by mixing several single-note piano recordings with additive Gaussian noise. None of the unique acoustic features of the piano instrument are modelled in this method. The estimated posterior densities of the parameter space are accurate, although the effect of overlapping harmonics, especially in octave intervals, has an effect in the estimation process. Also, the results are sensitive to user specific settings (noise variance scaling, spectral decay parameter and sampling distribution) for the chosen examples. The authors suggest the enlargement of the MCMC parameter space and the learning of the user defined parameter settings may be carried out using the Expectation-Maximisation (EM) technique.

More recent efforts by Peeling *et al.* (2007) proposed another probabilistic approach, where peaks detected in the frequency domain spectrum of a chord are modelled as realisations of a non-homogeneous Poisson point process. In particular, the number of peaks that are detected by a STFT is modelled as a Poisson random variable and

hence the likelihood function may easily be formulated without associating peaks to particular note fundamentals or harmonics. As an effect, the computational complexity of a full probabilistic model is avoided and the transcription accuracy can be high according to the authors. In general, successful results have been reported with this method for up to a 4-note mixture recording, although there were notes associated with octave music intervals that were not accurately transcribed.

2.3.4 Blind separation approach

In the case of blind source separation techniques, such as ICA, there is the assumption of statistical independence of sources. In our case, these sources may represent the notes of an instrument as part of monophonic or polyphonic music performances.

In simple terms, the signal is modelled as a collection of STFT spectra. The signal is represented as the weighted sum of the basis spectra including a "residual", i.e. a noise-related, parameter (Klapuri and Virtanen, 2008). Each pitch value would correspond to a different spectrum. "Components" are therefore formed which represent instances of the basis spectrum and the time-varying pitch. A collection of such components ultimately represent the mixed signal. The components are further analysed in order to obtain musically important information, such as their onset and offset times, and the fundamental frequency of the signal (Klapuri and Virtanen, 2008).

Another technique linked to ICA is called "sparse coding", where the signal is modelled in terms of a small number of active elements chosen out of a large set and a cost function can be minimised to estimate the desired parameters. Learning can also be employed, where a linear generative model is used and the observed data is represented as a weighted sum of elements chosen from a "dictionary" of available "features" or "atoms" (Abdallah and Plumbley, 2006). In such techniques, a cost function can be minimised as a means of estimating the frequencies in the data. However note that the data might not be sparse enough, where several components might not be active in order to yield meaningful frequency estimates (Abdallah and Plumbley, 2006).

Like other methods reviewed in this chapter, where the STFT spectra is used, blind source separation methods cannot necessarily model features of the reproduced sound. As a result, this might limit the success rates of the transcription of music performances with these methods.

2.4 Features of the sound generation mechanism of the piano

The sound generation mechanism of a piano is complicated and there have been many attempts in describing its intricate elements through various methods and techniques. First, the pianist by depressing a piano key will trigger a hammer and that in turn will interact with the string, then the vibration of the string will propagate along its length and will interact with the piano bridge. Finally, the sound will radiate from the soundboard (Aramaki *et al.*, 2001).

The piano was invented in 1709 by the Italian Bartolomeo Cristofori and its modern form was finalised around the middle of the nineteenth century (Taylor, 1965). The piano is also perhaps the most versatile musical instrument of all. It is capable of delivering both monophonic and polyphonic music performances with a vast range of dynamics. As mentioned earlier, the piano is also an instrument that covers a large frequency range of about eight octaves. This corresponds to a range of fundamental frequencies from 27.5 Hz to 4,186 Hz. The sound quality of a piano depends on the type of piano, that is being either an upright or a concert grand, the level of craftsmanship, sophistication and design that goes into each one and the proper tuning of it.

The piano normally consists of 88 keys, although some concert grands may have more. The piano falls into the category of percussive instruments since the note is produced by percussive means. The sound of piano note is of a transient nature and initiates with a violent percussive attack, when a hammer strikes a string, followed by a decay of vibration (Taylor, 1965). Figure 2.1 shows a simplified version of the action mechanism associated with the sound generation mechanism of a vibrating piano string:

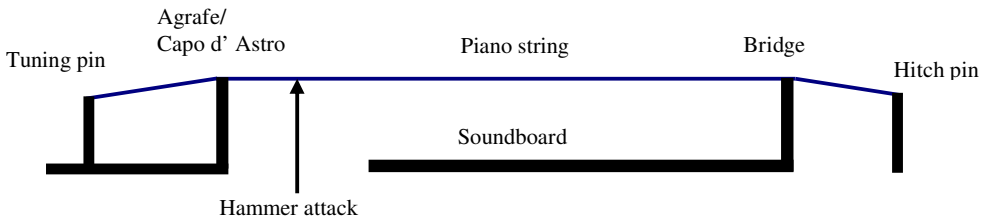


Figure 2.1 – The mechanism of piano sound reproduction

From Figure 2.1, it can be seen that the string is anchored to the large cast iron frame of the piano. On one side, it is anchored to the frame by means of using an "agrafe" or a "capo d' astro". A "tuning pin" is used to adjust the tension for tuning purposes. On the other side, it is fixed to the frame by means of a "hitch pin". Note that the bridge strongly couples the vibrating string with the soundboard. The soundboard is the main source of sound radiation from the musical instrument, as is the top wooden plate of string instruments, such as the guitar, the violin, the viola and so forth. For that reason also, the choice of wood and the shape of the soundboard is of great importance when determining its final tonal characteristics (Taylor, 1965). Normally, the soundboard is made of 1 cm thick spruce with ribs in the cross-grain direction, in order to stiffen it (Rossing, 1990).

When a piano key is depressed, the damper resting on the string will be raised and the hammer will strike the corresponding string, setting it into a free vibrating motion. As a result, vibrations of the string will be transmitted from the bridge to the soundboard of the piano and sound will be produced.

The 88 piano keys correspond to an excess of 200 strings. This is because most of the notes have more than a single string attached to them. The reason for having multiple strings for a single note is mainly because the sound associated with that note would be "weak" otherwise. This happens in the case of the treble notes corresponding to high frequency notes. These notes do not have very long strings and therefore the restorative energy associated with these shorter vibrating strings is not sufficient to produce a well-defined sound. On the other hand, bass notes can be as long as 2 metres and can have enough energy to produce sound before ultimately being amplified by the soundboard.

2.4.1 Hammer-string interaction

One of the most controversial topics of discussion has always been how far can the influence of a performer be stretched when a tone is produced by means of depressing a key. It is believed that it is not very likely that the performer can have any further control over the hammer other than the velocity with which it strikes the key of a

piano (Taylor, 1965). It is also accepted that the greater the dynamics, i.e. the greater force when striking a key, the greater the number of harmonics produced in that way.

This effect can be seen by examining the Fourier transform of piano note C_1, played at two different dynamic levels, specifically "piano" and "fortissimo". The results of such an analysis are shown in Figure 2.2. Note that the Fourier analysis is carried out on two different 10 sec recordings sampled at 44.1 kHz. The recordings were derived from a currently discontinued sample-based library named "Gigastudio, version 3" (originally trademarked by Tascam), which contains real piano note recordings of a "Kawai" grand piano.

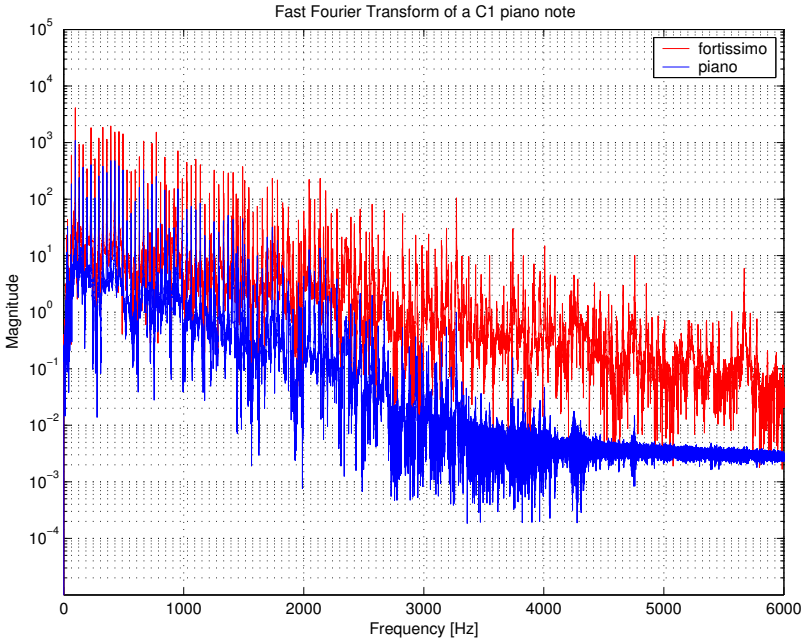


Figure 2.2 – Magnitude of a single FFT of the sound of a depressed piano note at two different dynamic levels (fortissimo and piano)

From the Figure 2.2, it can be seen that in the case of the *fortissimo*, the piano key is depressed with a large force so that higher harmonics manifest themselves from about 3,200 Hz and above in contrast to the *piano* case, where the key is depressed in a more gentle manner.

This is because when a key is struck with a greater force, or in other words with greater dynamics, then the velocity of the triggered hammer is greater and the contact

time of the hammer with the strings will decrease. Shorter contact time will result in a greater number of harmonics being produced. Hence, there is a direct relationship between the contact time of the hammer and string with the hammer velocity at different dynamic levels (Askenfelt and Jansson, 1988).

Consequently, the independent control of the loudness and the harmonic content cannot be achieved at the same time (Taylor, 1965).

The interaction between the hammer and string of a piano has been a topic of discussion for over a century now (Helmholtz, 1877). In general, hard and narrow hammers produce a -6 dB/octave spectrum envelope at high frequencies, whereas in the case where the mass of the hammer is less than the mass of the string, the high frequencies will attenuate with a -12 dB/octave roll-off (Hall, 1987).

Chaigne and Askenfelt (1994) have modelled the vibration of a string in a piano, when there is a frequency dependent loss and nonlinear hammer excitation. Boutillon (1987) explained this interaction between a piano hammer and a string by modelling the hammer as a point mass and the felt as a hysteretic spring. Through a number of experiments and investigations, the following relationship was obtained for the hammer force F on the felt as a function of the compression of the felt

$$F = a(\Delta y)^a$$

where a is an empirically derived coefficient and Δy is the change in the compression of the felt.

In simple terms, this model is an effective way of modelling the interaction between the hammer and the strings through the nonlinear and hysteretic compliance of the felt when in contact with the strings.

Finally, the dynamic behaviour of the soundboard can be described through the equations of motion for a thin orthotropic plate, where the two principle axes of the elastic constant tensor lie in the plane of the board (Giordano *et al.*, 2004).

2.4.2 Inharmonicity

In the case of an "idealised" vibrating string, where the only force controlling its motion is the tension, the frequencies of the fundamental and harmonics for clamped ends are given from the following formula (Blevins, 1979)

$$f_n = n\frac{1}{2l}\sqrt{\frac{T}{\rho A}} = nf_0 \qquad \textbf{(2.4.2 – 1)}$$

where T, l, ρ and A are the tension force, length, linear density and cross-sectional area of the vibrating string. f_0 is the fundamental frequency and n is an integer number representing the harmonic number.

From equation **(2.4.2 – 1)**, it can be seen that the resulting harmonics will be exact multiples of the fundamental frequency. However, piano strings are made of high strength steel wires in order to sustain high dynamic levels when keys are depressed by a performer. In practice, strings also have bending stiffness as well as tension and that affects the way harmonics manifest themselves in the frequency domain specifically. The inharmonicity factor, which is a result of the bending stiffness of a string, is given by (Fletcher, 1964)

$$B = \frac{\pi^2 EAK^2}{Tl^2} \qquad \textbf{(2.4.2 – 2)}$$

where E is the Young's modulus of elasticity and K is its radius of gyration.

Equation **(2.4.2– 2)** for the inharmonicity parameter B can alternatively be written as follows (Nishiguchi, 2004)

$$B = \frac{\pi^3 d^4 E}{64Tl^2} \qquad \textbf{(2.4.2– 3)}$$

where d is the diameter of the string.

From equation **(2.4.2 – 3)**, it can be deduced that the inharmonicity B related to the bending stiffness of the string, is greater in the case of short strings for a given radius and tension as opposed to long strings for the same radius and tension. Also, the

inharmonicity *B* of the string increases sharply with its radius and is more noticeable in the case of the bass strings since their radius is larger as opposed to the strings in the high frequency register. A way of reducing the inharmonicity effect in the low frequency register is by making the string less stiff, by means of one or two layers of wrapped strings as opposed to using solid strings.

For a piano string, which is displaced a distance *y* at the position *x*, the equation governing the motion of the piano string is given by (Fletcher, 1964)

$$-EAK^2\left(\frac{\delta^4 y}{\delta x^4}\right)+T\left(\frac{\delta^2 y}{\delta x^2}\right)=\rho\left(\frac{\delta^2 y}{\delta t^2}\right) \qquad \textbf{(2.4.2 – 4)}$$

where *y* takes the form

$$y = Ce^{2\pi kx}e^{-j2\pi ft} \qquad \textbf{(2.4.2 – 5)}$$

where *C* and *k* are constants determined from the boundary and initial conditions.

Also, the terms $T\left(\frac{\delta^2 y}{\delta x^2}\right)$ and $-EAK^2\left(\frac{\delta^4 y}{\delta x^4}\right)$ represent the restoring force due to tension *T* and the restoring force due to the elastic stiffness respectively.

The general solution of equation **(2.4.2 – 4)** is

$$y = e^{-j2\pi ft}\left[A_1\cosh(2\pi k_1 x)+A_3\cos(2\pi k_2 x)+A_2\sin(2\pi k_1 x)+A_4\sin(2\pi k_2 x)\right] \qquad \textbf{(2.4.2 – 6)}$$

where A_1, A_2, A_3 and A_4 represent the amplitudes of the general solution and four possible values of *k* for any possible frequency *f* can be defined as follows

$$k = \pm k_1 \text{ and } k = \pm jk_2$$

The general solution of equation **(2.4.2 – 6)** can be solved for different boundary conditions, such as pinned-pinned or clamped-clamped conditions.

Assuming that the origin of the x-axis is at the centre of the piano string and that the two string ends are defined at $x=\frac{l}{2}$ and $x=-\frac{l}{2}$ respectively, in the case of pinned-pinned boundary conditions

$$y = \frac{\delta^2 y}{\delta x^2} = 0 \text{ at } x = \frac{l}{2} \text{ and } x = -\frac{l}{2}$$

and for zero values of A_2, A_4 (even functions), the boundary conditions will fit if A_1 is also zero and $\cos \pi k l = 0$. The following formula can be obtained for each value *k* (Fletcher, 1964; Nishiguchi, 2004; Ortiz-Berenguer *et al.*, 2005)

$$f_n = n f_0 \sqrt{1 + n^2 B} \qquad \textbf{(2.4.2 – 7)}$$

Equation **(2.4.2 – 7)** describes in a very simple form the real frequencies of a piano string, where its ends are fixed, but not clamped (Ortiz-Berenguer *et al.*, 2005). In particular, it can be seen that the harmonics of a piano note do not manifest themselves at exact multiples of the fundamental frequency of the string due to the additional inharmonicity parameter *B* associated with the string.

The value of inharmonicity *B* in the case of a real piano typically may vary between 10^{-4} and 10^{-2} (Fletcher, 1964). It is also worth noting that the inharmonicity value will be different for different pianos due to the different selection of strings and level of craftsmanship.

Equation **(2.4.2 – 7)**, which represents the *exact* solution, it can also by approximated using Taylor's expansion series as follows

$$f_n = n f_0 \left(1 + n^2 B\right)^{\frac{1}{2}} \approx n f_0 \left[1 + \frac{n^2 B}{2} - \frac{n^4 B^2}{8} + \ldots\right]$$

If $n^2 B << 1$, then only the first two terms of the binomial expansion may be used as an approximation, in which case

$$f_{n,approximate} = n f_0 \left(1 + \frac{n^2 B}{2}\right) \qquad \textbf{(2.4.2 – 8)}$$

The percentage error derived from this approximation as a function of the harmonic order for typical values of the inharmonicity *B* can be calculated as

$$\%\ \ Error = \frac{f_n - f_{n,approximate}}{f_n} 100 \qquad \textbf{(2.4.2 – 9)}$$

By using equation (**2.4.2 – 9**), it can be calculated that for relatively high values of inharmonicity B, i.e. 10^{-2}, the percentage error can be over 25% for predicting the frequency of the 20^{th} harmonic. Therefore, if one was interested in representing the signal model accurately, even for the first 10 to 20 harmonics, then the exact solution should be used rather than the approximate solution of Taylor's expansion series.

Another interesting feature is that the piano strings held by the bridge are not fixed and their behaviour also depends on the mechanical impedance of the soundboard, hence resulting in a higher or lower vibrating frequency of a piano string than that for fixed ends (Ortiz-Berenguer *et al.*, 2005). In particular, if the mechanical impedance of the soundboard has a positive imaginary value, i.e. mass-like impedance, then the resulting frequency would be slightly higher than expected. On the other hand, in the case where the mechanical impedance of the soundboard has a negative imaginary value, i.e. compliance-like impedance, then the resulting frequency will be lower than expected.

A small amount of inharmonicity is desirable in pianos. For example, synthesized piano notes having harmonics at exact multiples of the fundamental frequency tend to lack subjective "warmth" (Rossing, 1990).

Finally, Appendix B provides a short description of the tuning process based on the 'circle of fifths'. However, in the particular case of the piano, these musical intervals should be 'stretched' further and not tuned to their "tempered" values in order to minimise beating between notes due to the presence of inharmonicity in the piano strings.

2.4.3 Double decay rate

When a piano key is depressed and a sound is heard, a careful examination of the time-history of that note shows a two-stage decay rate; the first stage has a rapid roll-off, whilst the second decay rate has a much slower roll-off resulting into what is called the "aftersound" of a piano note (Weinreich, 1977). This is illustrated in Figure

2.3 through a simulation as will be presented later in chapter 4[1]. In Figure 2.3, the natural logarithm of the instantaneous amplitude of a typical piano note as a function of time, which was recorded at 44.1 kHz using samples from the "Gigastudio" library, is shown:

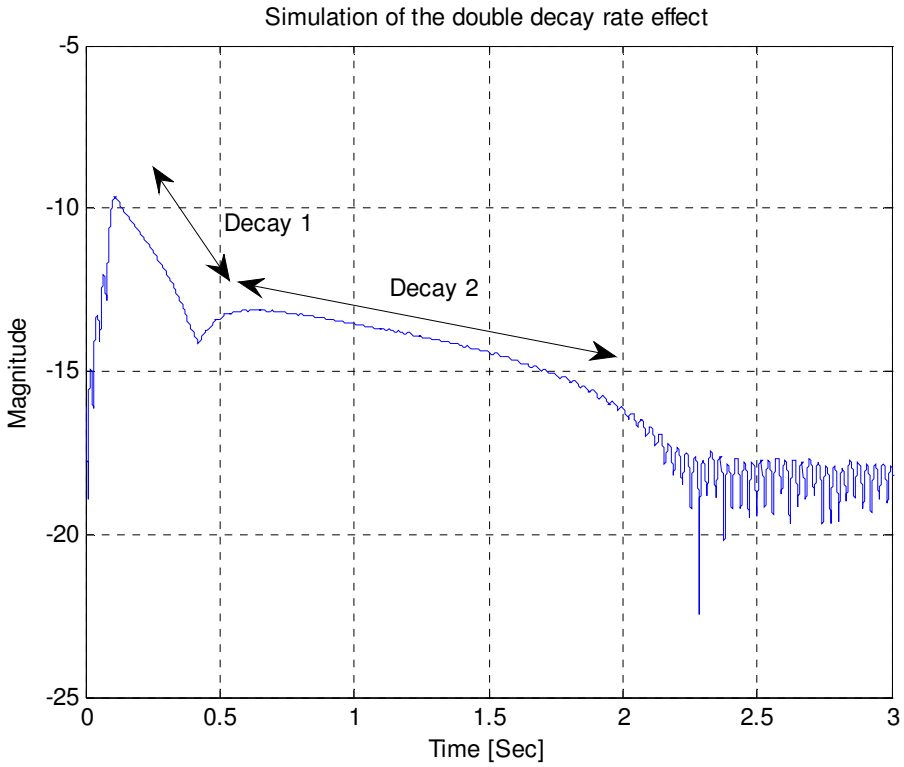


Figure 2.3 – Modelling of the characteristic double decay rate effect

There are mainly two factors believed to be associated with the double decay rate or compound decay characteristic of the sound of a piano note; the polarisation of the strings and, mainly, the coupling effect between different strings comprising a single note.

2.4.3.1 Polarisation of piano strings

An experiment was carried out (Weinreich, 1977) and the sound pressure level was plotted when a sound was produced by a single string of a piano note. The piano

[1] Note also that the "dip" at around 0.35 sec in this particular example is explained in section 4.2.3.1 of chapter 4, where the characteristics of the double decay rate are discussed in more detail.

string was freely vibrating at its fundamental frequency, as measured by a microphone. A two-stage or compound decay was revealed; in the first stage the sound pressure level starts at a high level and decays quickly afterwards, whilst in the second stage, the sound pressure level takes over at a lower level and decays more slowly.

The same phenomenon was observed by using a vibration pickup, where two electrodes measured the string at an angle of 45 degrees to the vertical and horizontal directions (Weinreich, 1977). From this experiment, it was deduced that the vertical motion of the string is associated with the fast decay, whilst the slower decay is associated with the horizontal motion of the string. Hence, it is believed that the two-stage decay phenomenon is closely associated with the piano string polarisation, where the superposition of two different decay rates produces the overall compound effect (Weinreich, 1977).

In simple physical terms, energy is transferred from the string to the soundboard via the bridge in both the vertical and horizontal polarisations, despite the primary excitation of the piano hammer is in the vertical direction. Also, there is a moment excitation at the bridge, which may further contribute to the transfer of energy in both directions. Therefore, since the soundboard is relatively compliant to out-of-plane motion, energy will be transferred from the vertical motion of the string to the vertical motion of the soundboard at a high rate resulting in a fast decay. On the other hand, the soundboard is stiffer in the horizontal direction and energy is transferred at a much lower rate resulting in a lower decay rate.

In terms of modelling the polarisation of the piano strings, a model has been suggested in which forces and velocities in the dynamical system correspond to currents and voltages as part of an equivalent electrical circuit analogy (Naganuma *et al.*, 2004). In this method, a string is modelled in both the vertical and horizontal directions when coupled with a soundboard. The soundboard is modelled as a spring and a dashpot connected in series, whilst the horizontal and vertical directions of the string are also modelled as masses connected to a spring and a dashpot. Coupling is modelled as an idealised transformer. This model can be used to explain the slight different frequencies associated with the vertical and horizontal polarisation of the string (Tanaka *et al.*, 1999).

Other modelling techniques have shown the importance of coupling the vertical and the horizontal motion of a string at the bridge with the soundboard (Nishiguchi *et al.*, 2003 and 2004). From finite element simulations of a string, it has been demonstrated that the motion of the string is significantly different to that of the soundboard. In this model, a pair of orthogonally oriented springs and a rotational angle were employed for the two different directions of the motion of the string.

2.4.3.2 Coupling of piano strings

The notes of a piano are normally made using a pair or a triplet of strings. These coupled strings are slightly mistuned from each other by much less than 1 Hz. It has been shown that the mistuning of a group of strings comprising a piano note affects the decay characteristics of the produced sound (Weinreich, 1977). For instance, when two or three strings of a group are struck, they are initially in-phase and so all will force the bridge up or down at the same time. This will result in a fast decay of the energy. However as the note continues to decay, the phase relationships change since the strings are slightly mistuned from each other and they are no longer working together to move the bridge. Ultimately, according to Weinreich, the strings will be out-of-phase and the rate of decay of energy will be lower resulting in a slower decay. However, this is not a complete description because of the additional presence of moment excitation in a coupled dynamic system. Note that as the mistuning increases, the phase difference between the strings also increases and for large phase differences, beats can be heard. A piano tuner can adjust the tuning between a group of strings in order to alter the characteristics of the double decay sound of a piano note.

Other authors (Hundley *at al.*, 1978) showed that there is a change in rate of energy transfer from the multiple-string source to the bridge during the transition from an initial in-phase condition to a later out-of-phase condition.

It has also been shown (Nakamura, 1988) that the coupling between a pair of strings is dependent on two factors, i.e. the degree of mistuning and the ratio of soundboard impedance to string impedance.

Another factor that contributes to the double decay effect is any hammer imperfections that may be present in the piano (Weinreich, 1977). Hammer imperfections can result in string amplitudes that are not absolutely equal. For instance, suppose that initially the struck strings are in phase but with the first string having larger amplitude than the second one due to a hammer irregularity. The motion of the strings start to decay and when the amplitude of the second string with the lower initial amplitude approaches zero, the bridge continues to move because it is still being forced to do so by the first string with the higher initial amplitude. Subsequently, the second string does not reach zero amplitude but it actually starts to build up vibration in the opposite phase by absorbing energy from the bridge. Ultimately, the motions are anti-symmetric giving rise to a compound decay effect.

Modelling of coupled piano strings can be based on the principles of digital waveguides. In these methods, a digital waveguide model simulates the wave propagation equation for a string when bounded in a medium called the resonator. Aramaki *et al.* (2001) modelled the vibration of two coupled strings using digital waveguides only for vertical motions and carried out a parameter estimation of the amplitude, frequencies and damping factors of the strings. This type of modelling is discussed in more detail in chapter 3.

Finally, Bensa *et al.* (2003) described the problem of coupled piano strings as an initial-boundary value problem permitting stable finite difference schemes. The model can then be related to a digital waveguide model. An experiment was also carried out by estimating parameters for a grand piano using a laser Doppler vibrometer. In this particular method, the stiffness parameter of the strings was calculated along with the fundamental frequencies and the inharmonicity factor. The time evolution of amplitudes of the first six harmonics of a piano note was also deduced where the double decay rate and beating effect were evident.

2.5 Conclusions

This chapter has provided a literature overview of the methods of monophonic and polyphonic music transcription, as well as presented the unique acoustic features associated with the sound generation mechanism of the piano.

Early approaches utilised simple frequency analysis techniques with successful results in the case of monophonic performances, whilst more recent neural network approaches have had mixed results for transcribing monophonic and polyphonic performances. Similarly, some evidence has shown that Bayesian approaches have been successful in transcribing monophonic performances and with some success in transcribing polyphonic performances. Blind separation methods in the form of ICA or sparse coding have been gaining more ground over recent years especially in the case of transcribing polyphonic music performances.

The drawback with most of the above methods, apart from the blind separation methods where there is no explicit representation of the signal model, is the simplistic representation of the signal model, especially in the case of the piano in which the sound generation mechanism is non-trivial. Hence, a conjecture is considered where a more detailed representation of the signal is required in order to improve the accuracy of the transcription methods. As a result, the interaction of the hammer with the piano string, and the concept of inharmonicity and double decay rate effect have been presented.

It was shown that the inherent bending stiffness in the piano strings leads to the inharmonicity effect in which the harmonics are not exact multiples of the fundamental frequency of a piano note. The unique double decay rate effect found in pianos was attributed to two main factors, i.e. the vertical and horizontal polarisation of the piano string and the multiple, slightly de-tuned from each other, strings coupled through the soundboard. Finally, the dynamic behaviour of coupled strings was briefly discussed through the use of digital waveguide techniques.

Chapter 3

Inharmonicity effect

3.1 Introduction

In this research we investigate two of the most important features that define the piano sound; the inharmonicity and the double decay rate. Potentially, these features could be used as part of a comprehensive signal model for transcribing music performances in the future. This chapter concentrates on the inharmonicity and the double decay rate effect is discussed later in chapter 4.

This chapter presents a simple analytical model of a vibrating piano string in the presence of inharmonicity and discusses the various factors that affect its response, such as the magnitude of the inharmonicity factor, the number of harmonics present in the modelled signal, and the evolution of the modelled signal with time. The effect of different phase and amplitude characteristics in the modelled signal are also discussed.

This chapter also presents novel observations associated with the inharmonicity of piano strings for synthesized and real recordings through the use of the cepstrum. The cepstrum is further employed to describe the separation of the response of the strings from the soundboard and the response of multiple strings is discussed through the use of digital waveguides and the cepstrum. The closure of this chapter naturally leads to chapter 4 discussing further the effect of coupling between oscillators and the modelling of the double decay rate effect.

3.2 Inharmonicity model

3.2.1 Theoretical representation

In the piano instrument, the fundamental and harmonics of a vibrating string are given by equation **(2.4.2 – 7)**

$$f_n = nf_0\sqrt{1+n^2B} \qquad \textbf{(3.2.1 – 1)}$$

where B is the inharmonicity associated with the bending stiffness of the vibrating piano string.

If $B=0$, then $f_n = nf_0$, where f_0 is the fundamental frequency in the absence of inharmonicity and the harmonics are integer multiples of f_0. However, if $B \neq 0$, as in the case of real strings, then the fundamental frequency is no longer f_0 but f_1 (Ortiz-Berenguer *et al.*, 2005)

$$f_1 = f_0\sqrt{1+B}$$

and the harmonics are no longer exact multiples of the fundamental frequency.

Figure 3.1 shows the spacing between the fundamental and first two harmonics in the case of a non-zero value of B, where a form of "frequency dilation" takes place relative to the case of $B=0$:

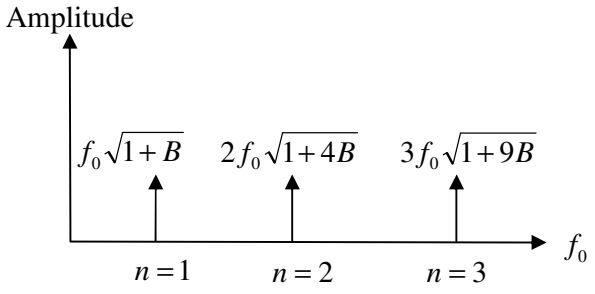


Figure 3.1 – Spacing of non-integer harmonics

Indeed, from Figure 3.1, the spacing between the different harmonics increases with frequency due to the inharmonicity factor *B*.

The response $x(t)$ of a vibrating piano string can be modelled as a simple superposition of N number of harmonics with the fundamental frequency, then based on equation **(3.2.1 – 1)**

$$x(t) = \sum_{n=1}^{N} \alpha_n \sin\left[n2\pi f_0\left(1+n^2B\right)^{0.5}t + \phi_n\right] \qquad \textbf{(3.2.1 – 2)}$$

where α_n and ϕ_n represent the amplitude and phase terms of the n^{th} harmonic respectively. For convenience at this stage, no damping is included.

Alternatively, equation **(2.4.2 – 8)** is recalled, which is the approximate form of equation **(3.2.1 – 1)** for $B << 1$

$$f_{n,approximate} = nf_0\left(1+\frac{n^2B}{2}\right) \qquad \textbf{(3.2.1 – 3)}$$

Equation **(3.2.1 – 2)** can then be approximated as

$$x(t) = \sum_{n=1}^{N} \alpha_n \sin\left[2\pi f_0(n+\frac{n^3B}{2})t + \phi_n\right]$$

$$x(t) = \sum_{n=1}^{N} \alpha_n \left[\sin(2\pi f_0 nt + \phi_n)\cos(\pi f_0 n^3 Bt) + \cos(2\pi f_0 nt + \phi_n)\sin(\pi f_0 n^3 Bt)\right] \qquad \textbf{(3.2.1 – 4)}$$

In an analogy to an "amplitude modulation" effect, the following can be deduced from equation **(3.2.1 – 4)**:

the terms $\sin(2\pi f_0 nt + \phi_n)$ and $\cos(2\pi f_0 nt + \phi_n)$ represent the "carrier" of the signal that are amplitude modulated by $\cos(\pi f_0 n^3 Bt)$ and $\sin(\pi f_0 n^3 Bt)$ terms respectively. These modulations are slowly varying if $n^3 B << 1$.

Equation **(3.2.1 – 4)** represents a "Priestley-type" interpretation in which a signal of the form $x(t)$ is amplitude modulated according to the following relationship (Priestley, 1981)

$$x(t) = \int A(t,\omega)e^{j\omega t}X(\omega)d\omega$$

where $A(t,\omega)$ represents the "modulator" and the term $e^{j\omega t}$ represents the "carrier" of the signal.

3.2.2 Numerical simulations

This section is divided into *five* subsections, in which various factors are discussed that affect the modelled response of a vibrating piano string with the presence of inharmonicity factor *B*. These factors are: the magnitude of the inharmonicity, the number of harmonics present in the modelled signal, the evolution of the modelled signal as a function of time, the phase characteristics in the modelled signal and the effect of different amplitudes.

3.2.2.1 Effect of inharmonicity factor *B*

Equation **(3.2.1 – 2)** is recalled

$$x(t) = \sum_{n=1}^{N} \alpha_n \sin\left[n2\pi f_0 \left(1 + n^2 B\right)^{0.5} t + \phi_n\right]$$

For the purpose of these simulations, it is assumed that α_n amplitudes are constant and that $\phi_n = \pi/2$. Therefore,

$$x(t) = \sum_{n=1}^{N} \alpha_n \cos\left[n2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right]$$

where $x(t)$ is made up of cosine terms and so is an even function of *t*. Note that we only plot the above equation function for $t \geq 0$.

Figure 3.2 is obtained showing the response of a vibrating piano string with and without the presence of inharmonicity *B* for the first 10 terms (fundamental frequency and nine harmonics):

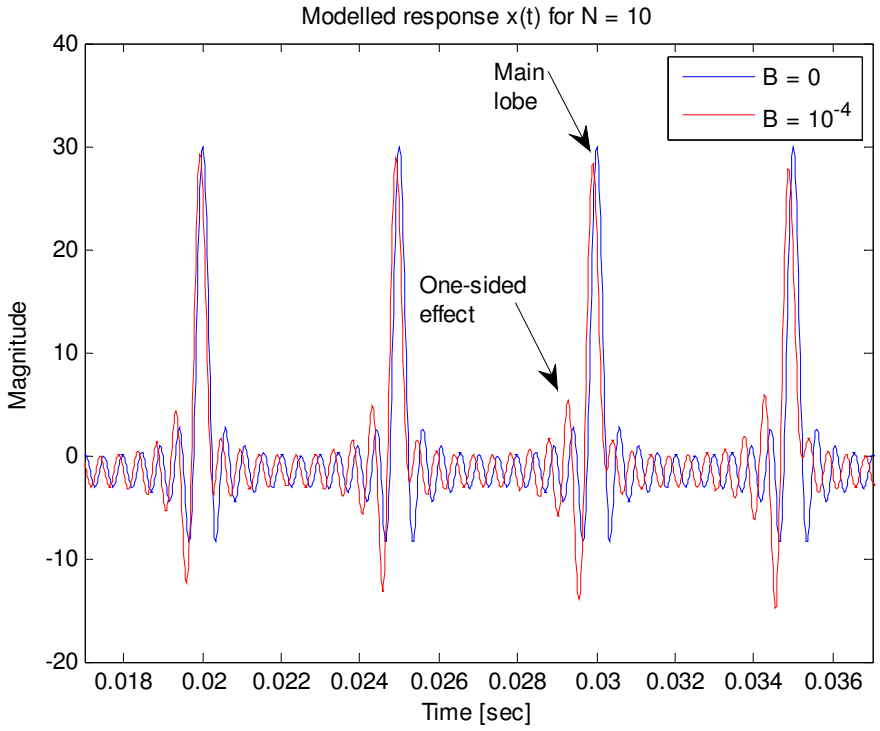


Figure 3.2 – Modelled response $x(t)$ for different values of the inharmonicity factor B, $B = 0$ and $B = 10^{-4}$ for $N = 10$

One-sided effect

From Figure 3.2, it can be seen that the modelled signal with $B = 0$ effectively is the sum of a Fourier series of a periodic function. Note that the spacing between two successive main lobes (or high peaks) is 5 ms corresponding to the fundamental frequency of 200 Hz used in the signal. On the other hand, the modelled response with $B = 10^{-4}$ represents a signal that changes with time and the presence of inharmonicity gives rise to a feature that we refer to as a "one-sided" effect in the modelled signal. For the purpose of this research, the latter name is proposed in relation to the formation of a strong non-symmetrical response relative to the main lobe structure of the signal. In particular, the signal generally retains a repetitive type structure, whereas the structure to the left (before) of a main lobe starts to increase with time. However, this effect is obscured in an actual piano note recording, as will be shown in section 3.2.2.5 for reasons explained later. Therefore, it is important to investigate this effect further since it may be part of the sound reproduction mechanism of a piano note and could perhaps be part of the sound perception mechanism.

Additionally, a pole/zero map of the Laplace transform signal shown in Figure 3.2 is plotted in Figure 3.3 below. This is shown for the case of $N = 10$ and $B = 0$:

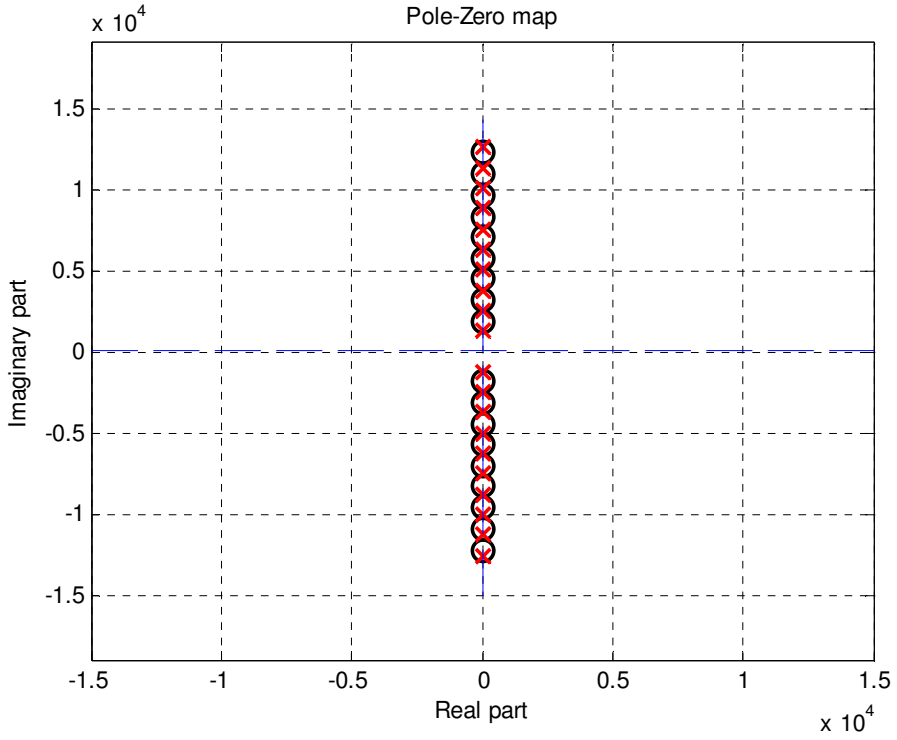


Figure 3.3 – Pole/zero mapping on the s-plane of the modelled response $x(t)$ for $B = 0$ and $N = 10$

From Figure 3.3, it can be seen that the poles and zeros all lie on the imaginary axis as there is no damping in the modelled response.

For further insight, the pole/zero map of the inharmonic signal, i.e. for $B \neq 0$, demonstrates a "distortion" of the location of the poles/zeros that lie on the imaginary axis when compared with the pure harmonic signal, i.e. for $B = 0$. In particular, by increasing the value of B, the poles/zeros move apart on the imaginary axis. A pole/zero map, which demonstrates this effect, is plotted in Figure 3.4 for the case of $N = 10$ and $B = 10^{-2}$:

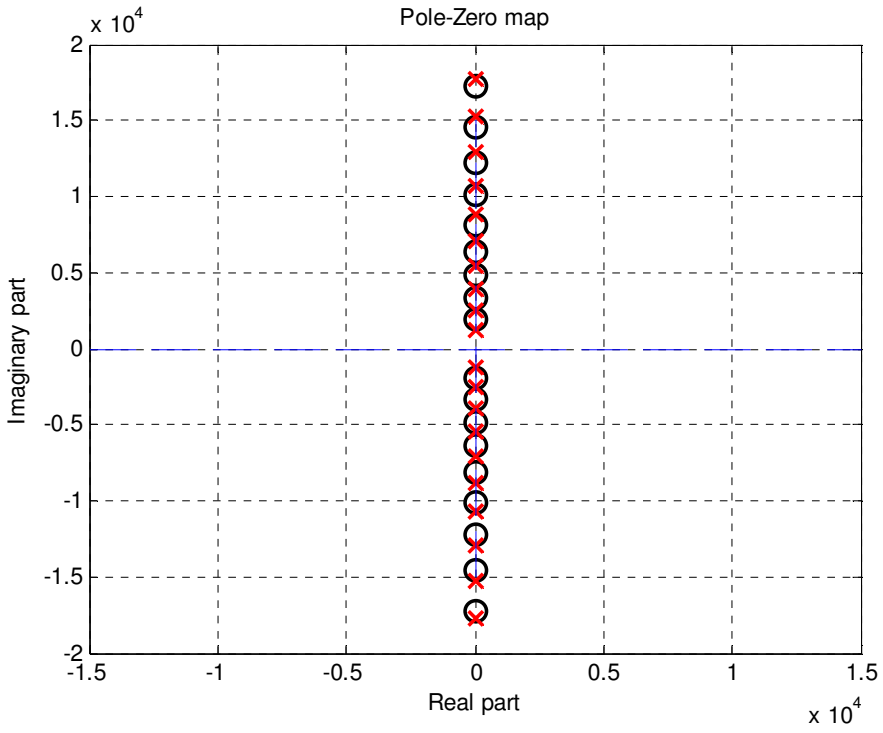


Figure 3.4 – Pole/zero mapping on the s-plane of the modelled response $x(t)$ for $B = 10^{-2}$ and $N = 10$

Also, an interesting feature, which does not correspond to a real physical system, is that for negative values of B, the one-sided lobe structure will shift from the left-hand side to the right-hand side of the main lobe structure as shown in Figure 3.5 below. Note that this is not a physically realisable solution since the bending stiffness, and hence the inharmonicity factor B, in piano strings would always be a non-negative value:

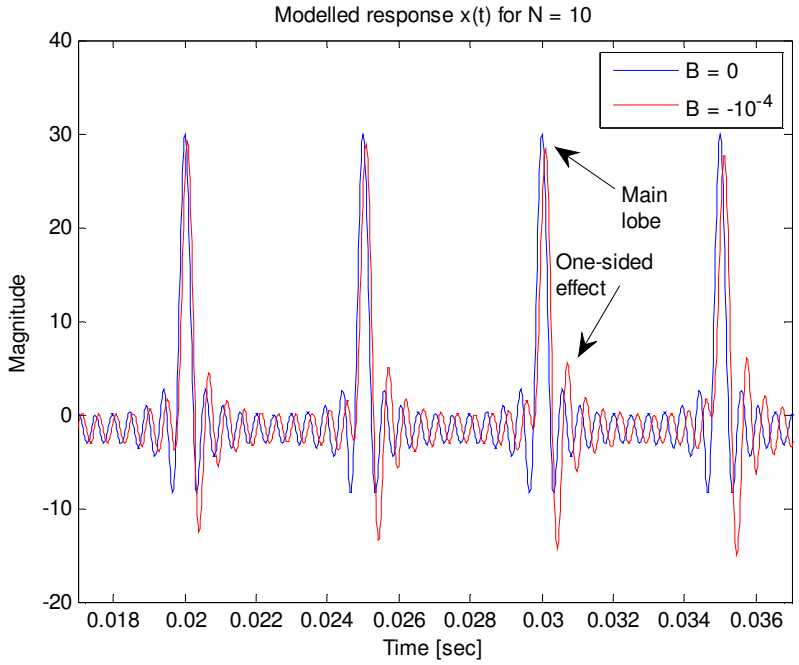


Figure 3.5 – Modelled response $x(t)$ for different values of the inharmonicity factor B,

$$B = 0 \text{ and } B = -10^{-4} \text{ for } N = 10$$

Finally, the effect of increasing the magnitude of inharmonicity factor B is illustrated in Figure 3.6:

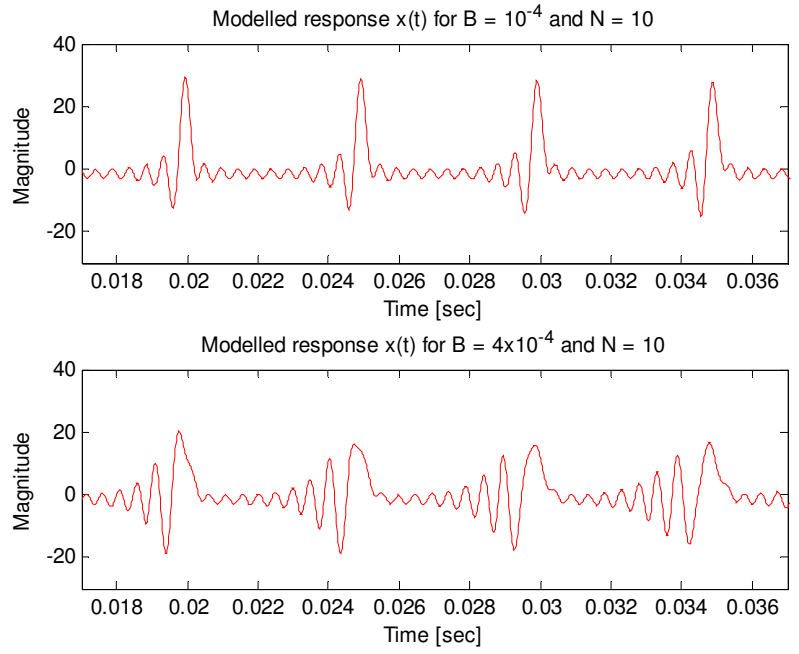


Figure 3.6 – Modelled response $x(t)$ by varying the inharmonicity factor B,

$$B = 10^{-4} \text{ and } B = 4 \times 10^{-4} \text{ for } N = 10$$

In this particular example, the magnitude of the inharmonicity was quadrupled and the effect of one-sided energy was observed in Figure 3.6. From the latter figure, it can be seen that as the magnitude of the inharmonicity factor increases, so as the signal will be modulated further. Hence, the effect of the one-sided energy will become more pronounced for higher values of B used in equation **(3.2.1 – 2)**.

3.2.2.2 Effect of harmonic number *N*

In this subsection, the effect of increasing the number of harmonics N is demonstrated as shown in Figure 3.7. In particular, response $x(t)$ is modelled with 10 and 20 terms respectively for a typical value of inharmonicity, i.e. $B = 10^{-4}$:

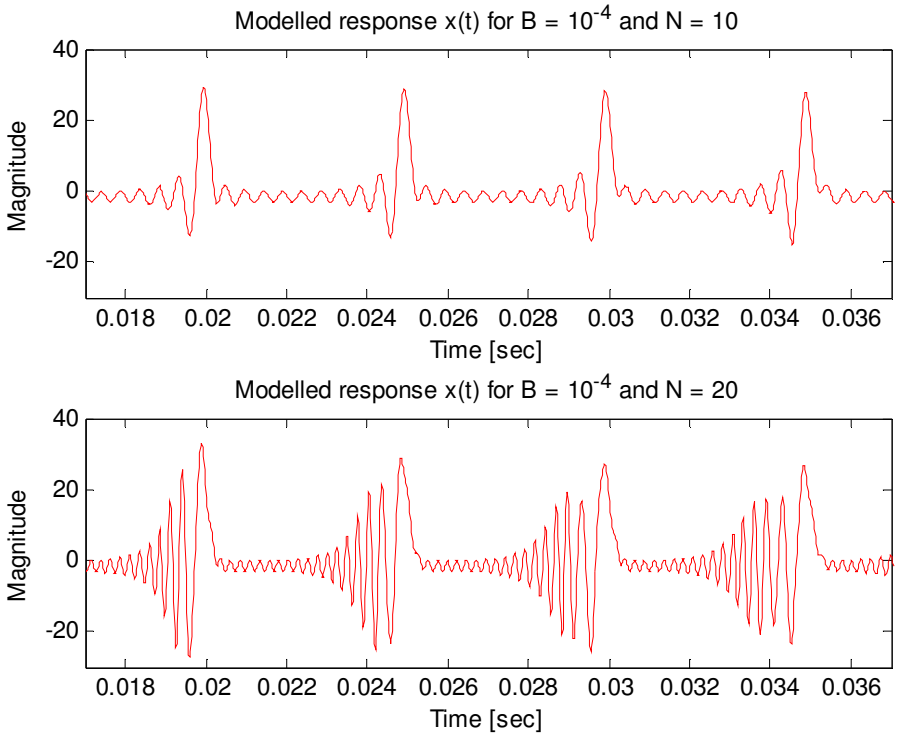


Figure 3.7 – Modelled response $x(t)$ by varying the harmonic number N,

$$N = 10 \text{ and } N = 20 \text{ for } B = 10^{-4}$$

From Figure 3.7, it can be seen that by increasing the number of harmonics N, the effect of one-sided behaviour response will be more pronounced in the modelled signal since the resulting modulation effect will increase as described in equation

(3.2.1 – 2). However, this may be unrealistic in real recordings since all harmonics may not have the same amplitudes. This is discussed in more detail in section 3.2.2.5, where the effect of various amplitudes for the different harmonic terms is demonstrated.

3.2.2.3 Effect of time parameter *t*

Another feature of the modelled signal represented by equation **(3.2.1 – 2)** is that changes in the response of the signal will be dramatic as time increases. So far, we have looked at the response of a single piano string for low-time regions, normally up to about 0.04 seconds. Figure 3.8 depicts the response of $x(t)$ for different times in the signal, for the same number of harmonics and the same inharmonicity factor B:

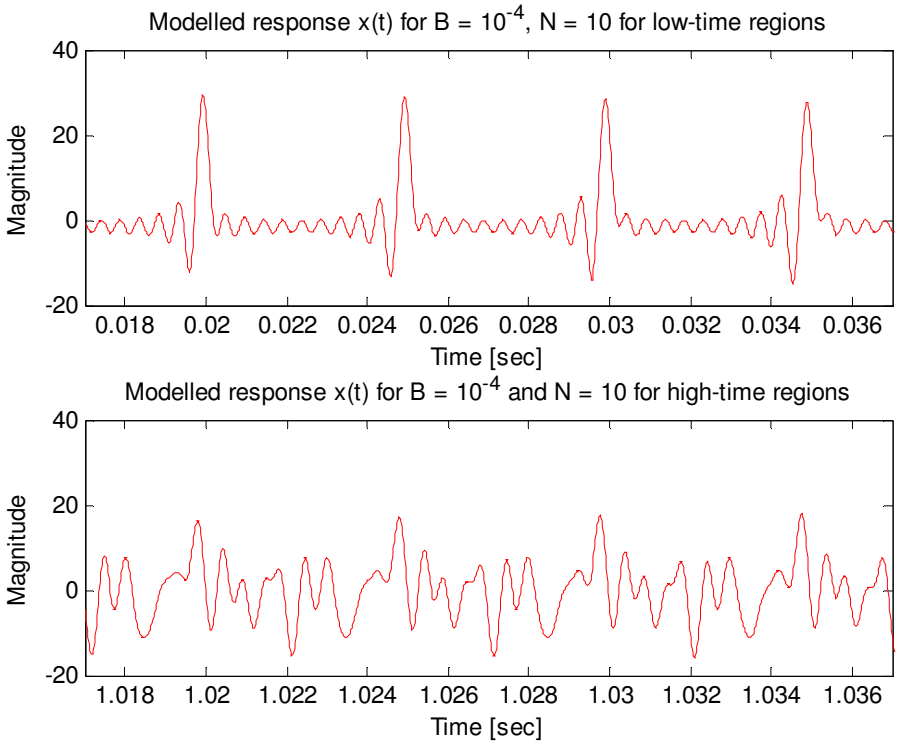


Figure 3.8 – Modelled response $x(t)$ depicted at different times for $N = 10$ and $B = 10^{-4}$

From Figure 3.8, it can be seen that dramatic changes in the response of the modelled signal will result as time increases. This is because by increasing the time parameter variable of equation **(3.2.1 – 2)**, the signal will be modulated further.

Given that the repetitive harmonic structure is a feature of a periodic signal (classical Fourier series), then the question lies on whether an inharmonic signal with non-equally spaced harmonic structure is periodic or not and under what conditions.

Initially, a signal $x(t)$ with two frequencies, f_1 and f_2, are combined as

$$x(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t) \qquad \textbf{(3.2.2.3 – 1)}$$

Hence, under what conditions is the above signal periodic? i.e. one is interested in finding the period T of the signal so that

$$x(t) = x(t+T) \qquad \textbf{(3.2.2.3 – 2)}$$

Let now $t = 0$ for convenience, equation **(3.2.2.3 – 2)** can be written as

$$x(t) = \sin(2\pi f_1 T) + \sin(2\pi f_2 T) = 0 \qquad \textbf{(3.2.2.3 – 3)}$$

Equation **(3.2.2.3 – 3)** is satisfied for

$$2\pi f_1 T = 2n\pi \text{ and } 2\pi f_2 T = 2m\pi$$

where m and n are integers.

Hence,

$$\frac{f_1}{f_2} = \frac{n}{m} \qquad \textbf{(3.2.2.3 – 4)}$$

Case 1

Assume that the ratio of frequencies f_1 and f_2 corresponds to a fraction of two integer numbers, i.e. a rational number. For example, $f_1 = 140$ Hz and $f_2 = 150$ Hz. According to equation **(3.2.2.3 – 4)**, $n = 14$ and $m = 15$. Therefore, the signal is *periodic* with $T = 0.1$ sec.

Case 2

Assume now that $f_1 = \sqrt{20 \times 10^4} \approx 141.4$ Hz and $f_2 = 150$ Hz. According to equation **(3.2.2.3 – 4)**, $n \approx 1414$ and $m = 1500$. So, the period would be $T \approx 10$ sec. However strictly speaking, there are no true integer solutions for m and n that satisfy equation **(3.2.2.3 – 4)**, hence, the signal does not have a period, i.e. it is *non-periodic*.

This simple argument can be extended to a sum of multiple frequencies present in the modelled signal, i.e.

$$x(t) = \sum_{n=1}^{N} \sin(2\pi f_n t)$$

where *each* frequency ratio should be rational. So, extending this to account for the inharmonicity factor *B*, each ratio

$$\frac{n\sqrt{1+n^2B}}{m\sqrt{1+m^2B}}$$

should also correspond to a rational number for all pairs m, n.

If the above conditions are met, then the signal is periodic. However, this is a condition, which is unlikely to be met, so for all practical purposes the modelled signal is considered as non-periodic.

3.2.2.4 Effect of phase

Up to now, we have assumed a constant phase term, $\phi_n = \pi/2$, in equation **(3.2.1 – 2)**, which would correspond to cosine components (even functions). However, in general, ϕ_n may differ from this.

Let us therefore consider how the choice of ϕ_n affects the shape of the signal.

For $\phi_n = \pi/2$ and $\phi_n = 0$, Figures 3.9 and 3.10 are obtained:

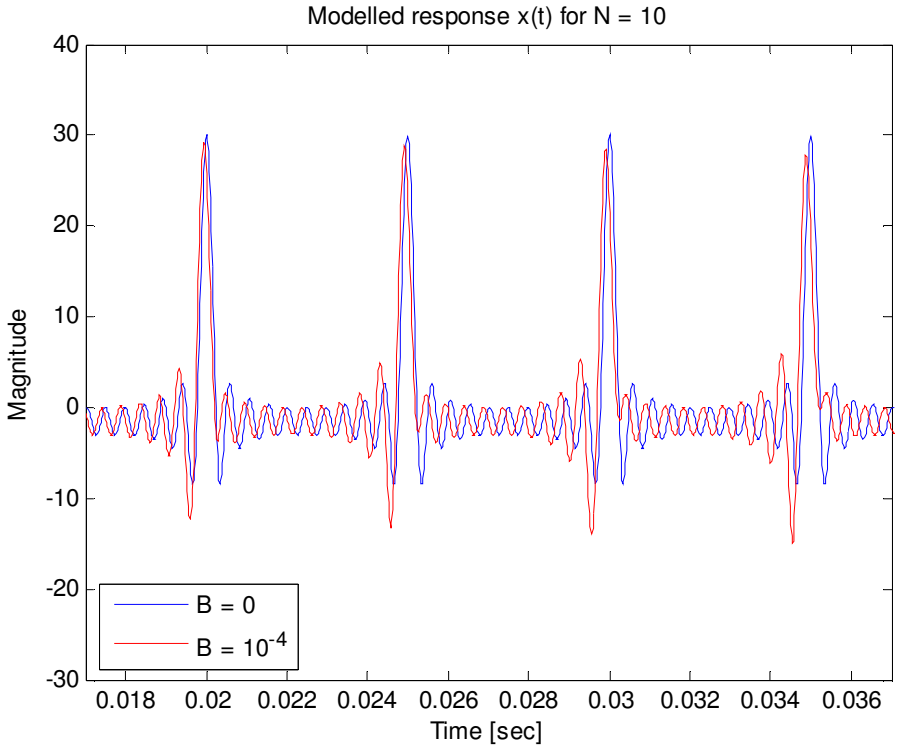


Figure 3.9 – Modelled response $x(t)$ *for* $N = 10$ *and* $\phi_n = \pi/2$

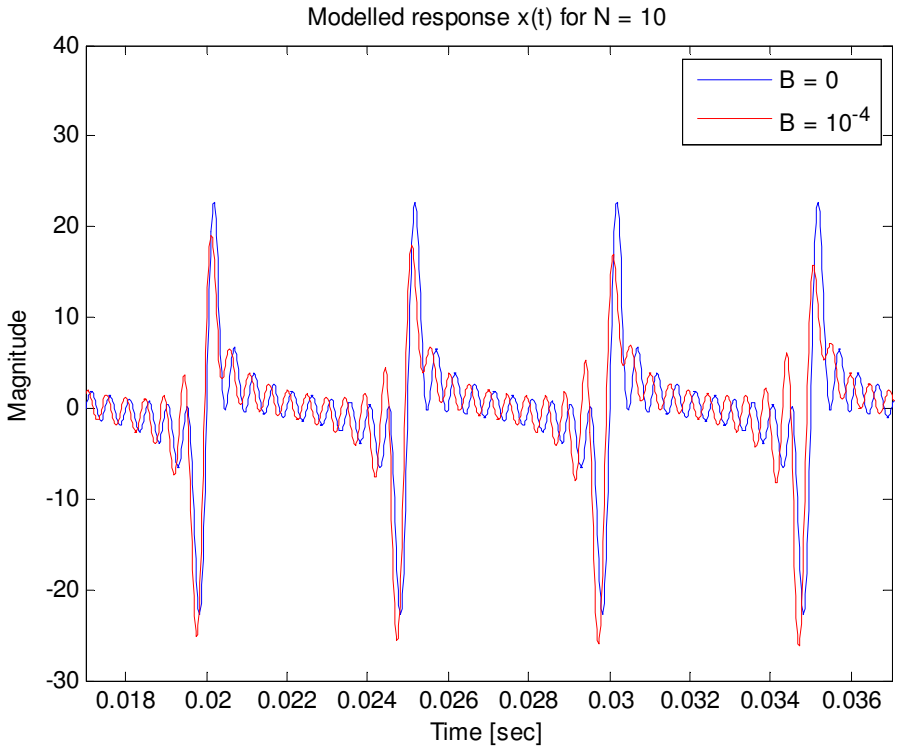


Figure 3.10 – Modelled response $x(t)$ *for* $N = 10$ *and* $\phi_n = 0$

Figures 3.9 and 3.10 represent the cosine and sine components of the modelled signal $x(t)$ and the phase relationship between all the N terms is *fixed (or constant)*. This means that the one-sided effect is related to the non-integer spacing of the harmonics relative to the fundamental frequency and not to the phase of the modelled components. However, note that a positional shift of the inharmonic signal relative to the pure harmonic signal is observed.

Also, equation **(3.2.1 – 2)** is simulated for the cosine terms with a randomly chosen phase term, which may vary between 0 and π/4, i.e. $\phi_{n,rad} \in [0, \pi/4]$

$$x(t) = \sum_{n=1}^{N} \alpha_n \cos\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t + \phi_{n,rad}\right]$$

This signal is shown in Figure 3.11, where the one-sided effect can still be observed:

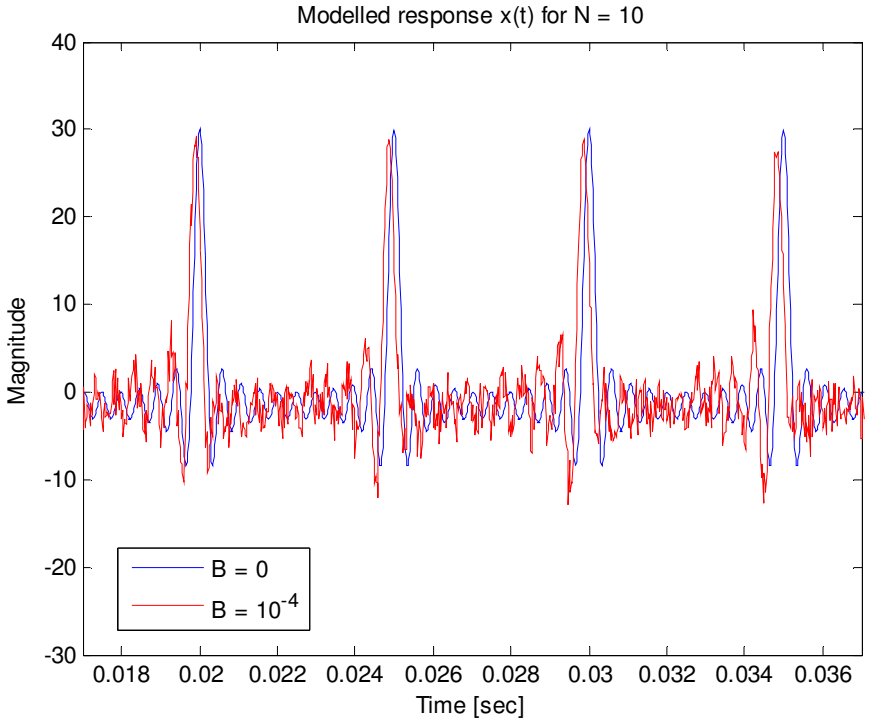


Figure 3.11 – Modelled response $x(t)$ for $N = 10$ and $\phi_{n,rad} \in [0, \pi/4]$

From Figure 3.11, it can be seen that by randomising the phase by a small amount, the one-sided effect is still evident. However, note that if the phase term is completely randomised, i.e. $\phi_{n,rad} \in [0, 2\pi]$, the signal becomes random and no such one-sided structure can then be seen.

3.2.2.5 Effect of amplitude

So far, numerical simulations of equation **(3.2.1 – 2)** were based on the assumption that the harmonic amplitudes α_n are constant, so in this section the amplitudes of the harmonics are varied uniformly and non-uniformly as described below.

Uniformly varying amplitudes

The amplitudes of the harmonic terms are expressed as a ratio of $1/n$ and a damping term ζ_n is also included. This is to represent a case where the amplitudes of the higher frequencies contribute less than the lower frequencies (Cemgil *et al.*, 2008)

$$x(t) = \sum_{n=1}^{N} \frac{e^{-\zeta_n t}}{n} \sin\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t + \phi_n\right] \qquad \textbf{(3.2.2.5 – 1)}$$

Note that the phase term is constant, i.e. $\phi_n = \pi/2$. Figure 3.12 is obtained:

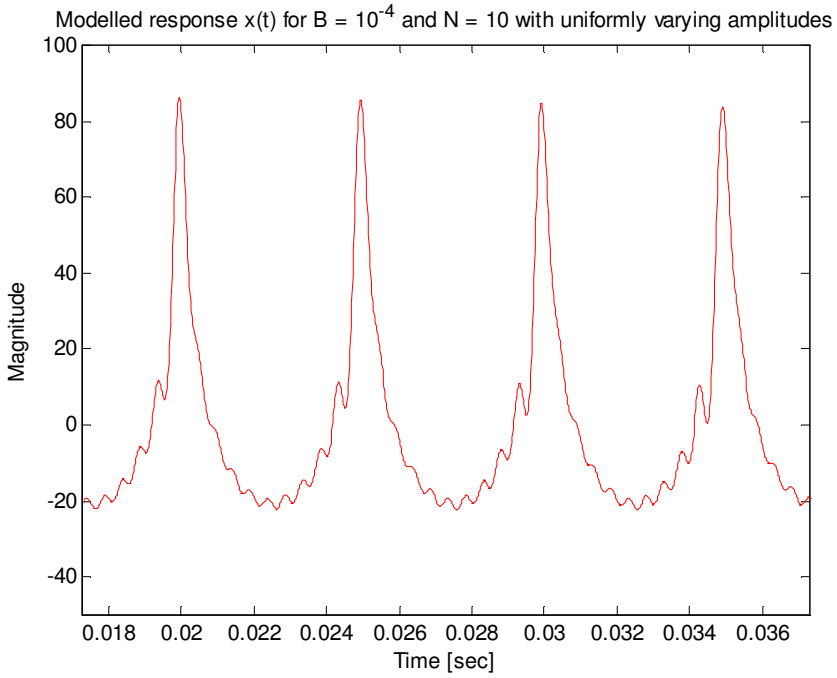


Figure 3.12 – Modelled response $x(t)$ with uniformly varying amplitudes using equation **(3.2.2.5 – 1)** *for $N = 10$ and $B = 10^{-4}$*

From Figure 3.12, it can be seen that the varying amplitudes of the form $1/n$ have a "low pass" effect on the modelled signal, where the high frequency modulation effect associated with the one-sided energy is partially filtered out, but its main asymmetric one-sided structure is yet retained.

Non-uniformly varying amplitudes

Equation **(3.2.2.5 – 1)** can also be written as

$$x(t) = \sum_{n=1}^{N} e^{-\zeta_n t}\alpha(n)\sin\left[n2\pi f_0\left(1+n^2B\right)^{0.5}t+\phi_n\right] \qquad \textbf{(3.2.2.5 – 2)}$$

where the amplitudes $\alpha(n)$ can now vary non-uniformly as in the case of a real signal and $\phi_n = \pi/2$ as before.

For this particular example, the amplitudes of the harmonics where chosen from a real piano note recording C_3 sampled at 44.1 kHz using the "Gigastudio" sample library. The modelled response $x(t)$ is shown in Figure 3.13 by using equation **(3.2.2.5 – 2)**:

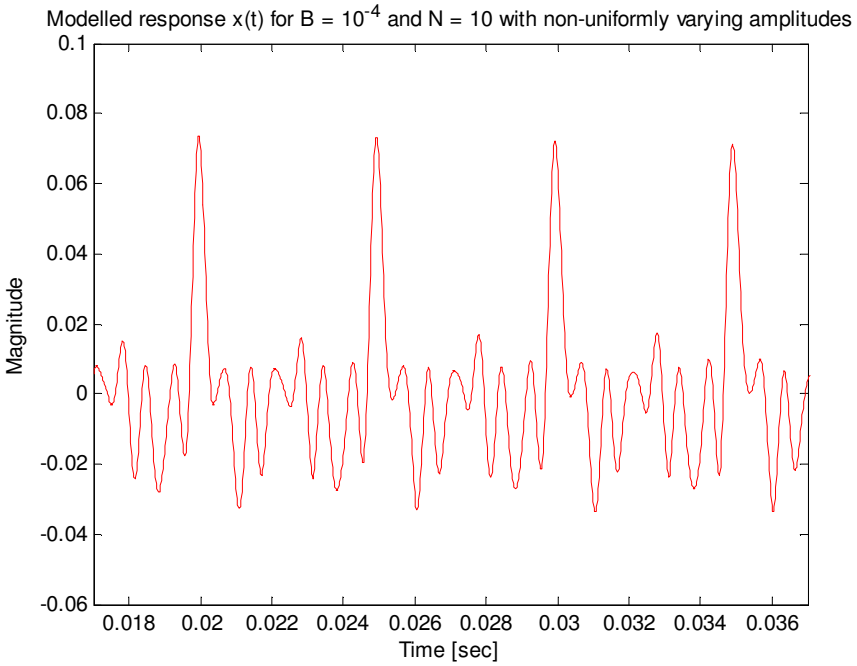


Figure 3.13 – Modelled response $x(t)$ with non-uniformly varying amplitudes using equation **(3.2.2.5 – 2)** *for $N = 10$ and $B = 10^{-4}$*

From Figure 3.13, it can be seen that for non-uniformly varying amplitudes, the one-sided effect is *obscured*.

The modelled response $x(t)$ of equation **(3.2.2.5 – 2)** is also plotted for higher times, where the one-sided effect is still obscured, although there is some indication of a temporal variation/modulation relative to the main lobe structure of the modelled signal. This is shown in Figure 3.14:

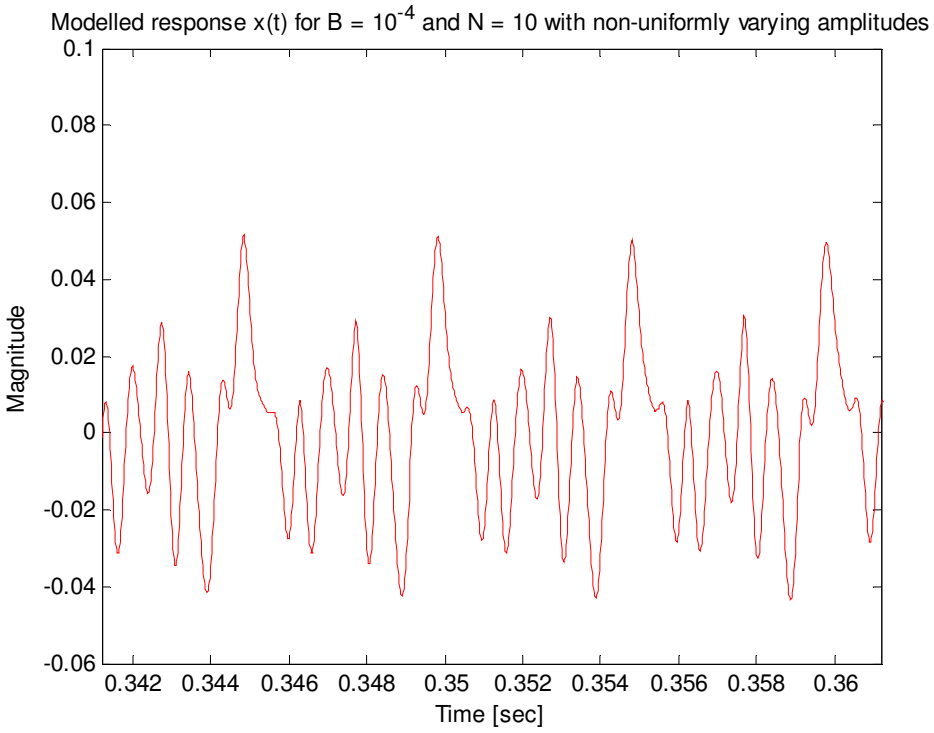


Figure 3.14 – Modelled response $x(t)$ with non-uniformly varying amplitudes using equation **(3.2.2.5 – 2)** *for* $N = 10$ *and* $B = 10^{-4}$

Finally, Figures 3.15 and 3.16 below compare the time histories of a modelled piano note C_3 using equation **(3.2.2.5 – 2)** and a real recording of that note (band-passed for the first 10 terms) over a typically long time window, i.e. for. $0.65 \le t \le 0.72$:

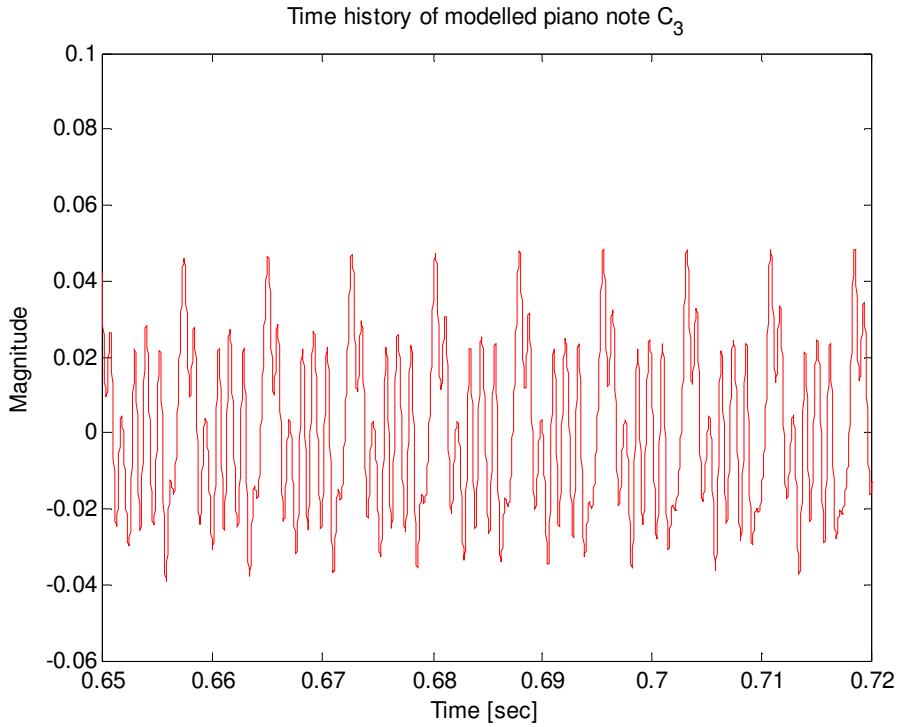


Figure 3.15 – Time history of modelled piano note C_3

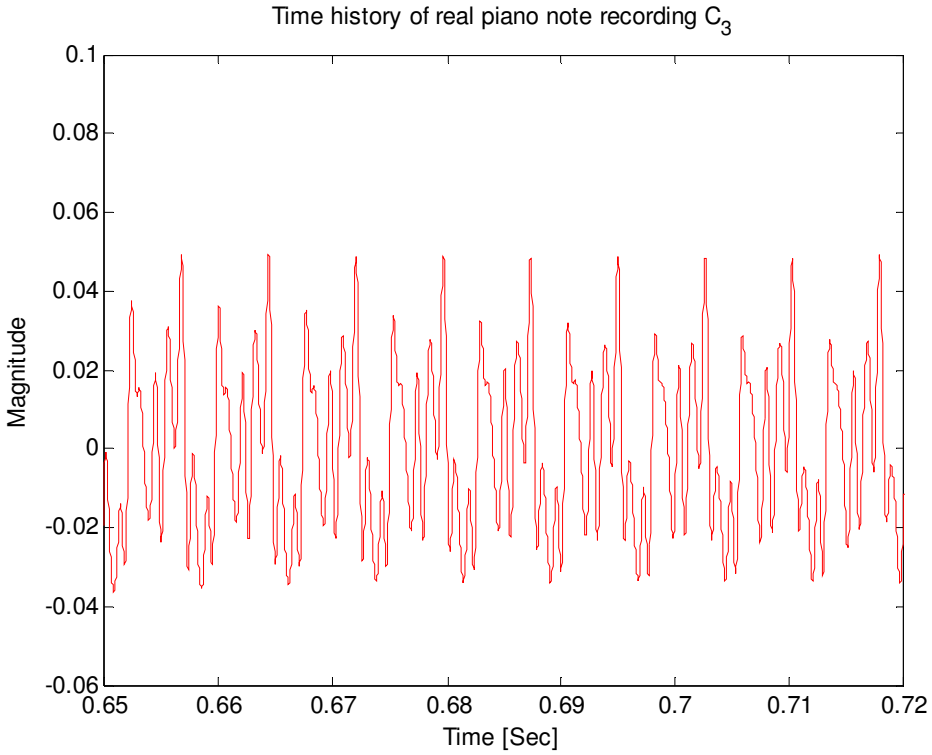


Figure 3.16 – Time history of real piano note recording C_3

From Figures 3.15 and 3.16, it can be seen that the one-sided effect is obscured since both signal representations have non-uniformly varying amplitudes, as discussed before. This perhaps explains why the one-sided effect is not apparent when observing the real time histories of piano note recordings. Also, there are differences in the temporal structure between the modelled and the real piano note recording meaning that our model does not perfectly represent the true signal. Finally, from a perception point of view, the one-sided effect may not be perceived due to the phase insensitivity of the hearing mechanism.

3.3 Cepstrum analysis

In this section, the cepstrum is introduced in order to further unravel the structure of piano sounds. So far, one has simply considered only the response $x(t)$ of a vibrating string with or without inharmonicity, but in reality the string is also coupled directly with the soundboard of the piano, which ultimately radiates the reproduced sound. Hence, it is in the direct interests of this research to further formulate the coupling of the vibrating piano string with the main mechanism of sound radiation, which is the soundboard.

3.3.1 Concept

One way of analysing compound signals is the cepstrum. In principle, the cepstrum is a "homomorphic" operator by which convolved signals in the time domain can be converted to additions. In this way, two different components of a composite signal can be separated out in the "pseudo-time" domain and can be analysed on an individual basis.

The mathematical expression of the *complex cepstrum*, $\hat{x}(t)$, of a signal is defined as follows

$$\hat{x}(t) = \int_{-\infty}^{+\infty} \ln[X(f)] e^{j2\pi ft} df \qquad \textbf{(3.3.1 – 1)}$$

So, the complex cepstrum of a signal is defined as the inverse Fourier transform of the natural logarithm of the Fourier transform of signal $x(t)$. Note that the complex cepstrum retains the phase structure of the signal.

In order to appreciate the merits of using the cepstrum of a signal for analysis, one has to understand the basic mechanism of homomorphic processing. A homomorphic process is a method by which a composite signal can be decomposed into a series of simpler operations. For example in the case of the cepstrum, suppose that a composite signal $s_{tot}(t)$ consists of two simpler signals $s_1(t)$ and $s_2(t)$, which are convolved in the time domain

$$s_{tot}(t) = s_1(t) * s_2(t) \qquad \textbf{(3.3.1 – 2)}$$

Equation **(3.3.1 – 2)** in the frequency domain may be written as

$$S_{tot}(f) = S_1(f)S_2(f) \qquad \textbf{(3.3.1 – 3)}$$

By then taking the logarithms of either side of equation **(3.3.1 – 3**), the product of the two signals in the frequency domain becomes an addition

$$\ln[S_{tot}(f)] = \ln[S_1(f)] + \ln[S_2(f)] \qquad \textbf{(3.3.1 – 4)}$$

By applying the inverse Fourier transforms on equation **(3.3.1 – 4)**, the signals may now be separated, although there is no guarantee for this, in the inverse-log-frequency or the pseudo-time domain and can be analysed on a separate basis

$$F^{-1}\{\ln[S_{tot}(f)]\} = F^{-1}\{\ln[S_1(f)]\} + F^{-1}\{\ln[S_2(f)]\}$$

$$\hat{s}_{tot}(t) = \hat{s}_1(t) + \hat{s}_2(t) \qquad \textbf{(3.3.1 – 5)}$$

where $\hat{s}_1(t)$, $\hat{s}_2(t)$ and $\hat{s}_{tot}(t)$ are the complex cepstra of signals $s_1(t)$, $s_2(t)$ and $s_{tot}(t)$ respectively.

3.3.2 Application to piano signals

The sound of a piano note can be regarded as the convolution of the output from the vibrating strings with the soundboard and bridge response.

In the simple case, the sound of a single piano note is produced by the hammer exciting the string, which is then amplified by the soundboard. Figure 3.17 shows a schematic layout of the sound generation mechanism of a single piano note:

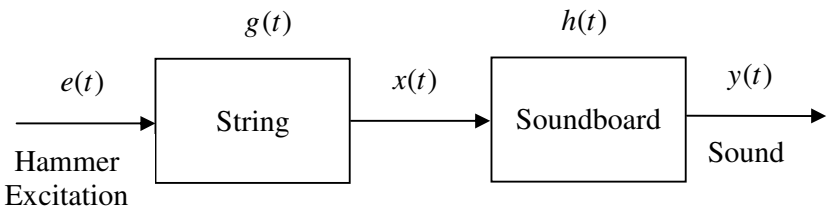


Figure 3.17 – Schematic layout of sound generation mechanism of a single piano note

The following hypothesis is made in the case of the piano instrument (Karatsovis *et al.*, 2006): the soundboard has a broad spectral response (Giordano, 1997 and 1998) and consequently a cepstrum that predominantly contains its information in the low "quefrencies" (or low-time regions). This hypothesis is supported by experimental work on a real grand piano and described later in chapter 4 (ref: Figure 4.15). On the other hand, the vibrating strings give rise to a series of impulses in the pseudo-time domain due to the relative periodicity of the harmonics associated with the vibrating length of the strings. These impulses will manifest themselves in the higher quefrencies (or high-time regions) and the vibrating strings will therefore be well separated from cepstral coefficients associated with the soundboard. The above hypothesis is demonstrated mathematically in section 3.3.3.

The resulting piano sound $y(t)$ is the convolved impulse response of the soundboard, $h(t)$, with the response of a vibrating piano string, $x(t)$

$$y(t) = h(t) * x(t) \qquad \textbf{(3.3.2 – 1)}$$

Also, the resulting response of the vibrating piano string $x(t)$ is the convolution of the impulse response of the string, $g(t)$, with the hammer excitation $e(t)$

$$x(t) = g(t) * e(t) \qquad \textbf{(3.3.2 – 2)}$$

Equation **(3.3.2 – 1)** due to **(3.3.2 – 2)** becomes

$$y(t) = h(t) * g(t) * e(t) \qquad \textbf{(3.3.2 – 3)}$$

By taking the Fourier transforms on either side of equation **(3.3.2 - 3)**

$$Y(f) = H(f)G(f)E(f) \qquad \textbf{(3.3.2 – 4)}$$

By taking the natural logarithms on either side of equation **(3.3.2 – 4)**

$$\ln Y(f) = \ln H(f) + \ln G(f) + \ln E(f) \qquad \textbf{(3.3.2 – 5)}$$

Finally, by taking the inverse Fourier transforms of equation **(3.3.2 – 5)**, one obtains the following complex cepstrum representation as follows

$$\hat{y}(t) = \hat{h}(t) + \hat{g}(t) + \hat{e}(t) \qquad \textbf{(3.3.2 – 6)}$$

where $\hat{y}(t)$, $\hat{h}(t)$, $\hat{g}(t)$, $\hat{e}(t)$ are the complex cepstra of the resulting piano sound, the soundboard response, the piano string impulse response and the excitation force respectively.

As one can observe from equation **(3.3.2 – 6)**, the resulting piano sound may be divided into separate components with the help of the complex cepstrum, so they can be treated separately for further analysis.

As a result, cepstrum analysis can be used in complicated signals, which are rich in harmonics, in order to extract periodicities associated with them, and could ultimately be used to decouple the response of the vibrating strings from the response of the soundboard (Karatsovis *et al.*, 2006). The character of the cepstrum of real piano notes can also be used to describe the phenomenon of inharmonicity associated with the vibrating strings shown earlier in the numerical simulations of section 3.2.2.

3.3.3 Analytical formulations

It is convenient to carry out any formulations in discrete time. The cepstrum $\hat{x}(n)$ can therefore be defined for sequence $x(n)$ as

$$\hat{x}(n) = Z^{-1}\{\ln[X(z)]\}$$

Assume now that the string response is periodic, i.e. there is no inharmonicity, and the excitation of a single string results in its fundamental and exact harmonic frequencies. The single string response $p(n)$ in the discrete time with period n_0 is

$$p(n) = p(n + kn_0)$$

where k is an integer.

Let us also assume that there is an exponential decay $m(n)$ uniform for all components, so the modulated response of the string $x(n)$ is

$$x(n) = m(n)\,p(n) \qquad \textbf{(3.3.3 – 1)}$$

We will model the period signal $p(n)$ as a delta train *convolved* with an n_0 length sequence $p_0(n)$, so equation **(3.3.3 – 1)** becomes

$$x(n) = m(n)\,p_0(n) * \sum_k \delta(n - kn_0) \qquad \textbf{(3.3.3 – 2)}$$

Assuming that $m(n)$ is sufficiently slowly varying over one period, then the delta train is modulated directly by $m(n)$, hence in the complex frequency domain, equation **(3.3.3 – 2)** becomes

$$X(z) \approx P_0(z)\mathrm{Z}\left[m(n)\sum_k \delta(n - kn_0) \right]$$

$$X(z) = P_0(z)\sum_n m(n)\sum_k \delta(n - kn_0)z^{-n}$$

$$X(z) = P_0(z)\sum_k m(kn_0)z^{-kn_0} \qquad \textbf{(3.3.3 – 3)}$$

Assuming $m(n)$ is an exponential decay of the form $m(n) = a^n$, equation **(3.3.3 – 3)** becomes

$$X(z) = P_0(z)\sum_k \left(a^{n_0} z^{-n_0}\right)^k$$

$$X(z) = P_0(z)\frac{1}{\left(1 - a^{n_0} z^{-n_0}\right)} \qquad \textbf{(3.3.3 – 4)}$$

In the cepstrum domain, equation **(3.3.3 – 4)** becomes

$$\hat{x}(n) = \hat{p}_0(n) + \mathrm{Z}^{-1}\left\{ \ln \frac{1}{\left(1 - a^{n_0} z^{-n_0}\right)} \right\}$$

$$\hat{x}(n) = \hat{p}_0(n) + a^{n_0}\delta(n-n_0) + \frac{a^{2n_0}}{2}\delta(n-2n_0) + \frac{a^{3n_0}}{3}\delta(n-3n_0) + \ldots$$

Therefore, the cepstrum of the overall response $\hat{y}(n)$ of equation **(3.3.2 – 1)** will lead to equation **(3.3.3 – 5)** as follows

$$\hat{y}(n) = \hat{h}(n) + \hat{p}_0(n) + a^{n_0}\delta(n-n_0) + \frac{a^{2n_0}}{2}\delta(n-2n_0) + \frac{a^{3n_0}}{3}\delta(n-3n_0) + \ldots \textbf{(3.3.3 – 5)}$$

It has therefore been shown analytically from equation **(3.3.3 – 5)** that the cepstrum of the response $\hat{y}(n)$ approximately consists of the cepstrum of the soundboard $\hat{h}(n)$, the cepstrum of the periodic shape $\hat{p}_0(n)$ and a train of amplitude-decaying spikes $\hat{i}(n)$ of the form

$$\hat{i}(n) = a^{n_0}\delta(n-n_0) + \frac{a^{2n_0}}{2}\delta(n-2n_0) + \frac{a^{3n_0}}{3}\delta(n-3n_0) + \ldots \quad \textbf{(3.3.3 – 6)}$$

The above formulation demonstrates the original hypothesis (Karatsovis *et al.*, 2006).

However, normally in computational simulations, as will be shown in section 3.3.4 later, there is always the effect of windowing $w(n)$ on the total response $y(n)$ that has not yet been accounted for. This can be realised through

$$y_w(n) = w(n)y(n) \qquad \textbf{(3.3.3 – 7)}$$

where

$$w(n) = 1 \quad 0 \le n \le M-1$$
$$w(n) = 0 \qquad else$$

where M is the data length.

Equation **(3.3.3 – 7)** can now be written as

$$y_w(n) = w(n)\left[h(n) * \{m(n)\,p(n)\}\right]$$

The windowed train of spikes $i_w(n)$ of the form $i_w(n) = w(n)i(n)$ is written as a finite sum

$$i_w(n) = \sum_{k=0}^{P} \delta(n - kn_0) \qquad \textbf{(3.3.3 – 8)}$$

where $Pn_0 = M - 1$

By taking transforms in the complex frequency domain, equation **(3.3.3 – 8)** becomes

$$I_w(z) = \sum_{k=0}^{P} z^{-kn_0}$$

$$I_w(z) = \frac{1 - z^{-(M-1)}}{1 - z^{-n_0}} \qquad \textbf{(3.3.3 – 9)}$$

By taking natural logarithms on either side of equation **(3.3.3 – 9)**, one obtains

$$\ln[I_w(z)] = \left[z^{-n_0} + \frac{z^{-2n_0}}{2} + \frac{z^{-3n_0}}{3} + \ldots \right] - \left[z^{-(M-1)} + \frac{z^{-2(M-1)}}{2} + \frac{z^{-3(M-1)}}{3} \right] \qquad \textbf{(3.3.3 – 10)}$$

By taking the frequency domain version of equation **(3.3.3 – 10)** and by setting $z = e^{j\omega\Delta}$, one obtains the following equation

$$\ln[I_w(z)] = \left[e^{-j\omega n_0\Delta} + \frac{e^{-2j\omega n_0\Delta}}{2} + \frac{e^{-3j\omega n_0\Delta}}{3} + \ldots \right] - \left[e^{-j\omega\Delta(M-1)} + \frac{e^{-2j\omega\Delta(M-1)}}{2} + \frac{e^{-3j\omega\Delta(M-1)}}{3} + \ldots \right] \qquad \textbf{(3.3.3 – 11)}$$

The finite inverse transform of equation **(3.3.3 – 11)** for *each* term would be of the form

$$\int e^{-j\omega q} e^{j\omega} d\omega \approx k\delta(t - q)$$

where $q = n_0, 2n_0, 3n_0, \ldots$ and $k = 1, \frac{1}{2}, \frac{1}{3}, \ldots$

Ultimately, the windowed train spikes in the pseudo-time domain will be of the following form (set also $\Delta = 1$ for convenience)

$$\hat{i}_w(t) = w(t - n_0) + \frac{1}{2} w(t - 2n_0) + \frac{1}{3} w(t - 3n_0) + \ldots \qquad \textbf{(3.3.3 – 12)}$$

From equation **(3.3.3 – 12)**, one can deduce that the amplitude of spikes is diminishing and is symmetric of the form $\frac{\sin(x)}{x}$.

However, in the case of a vibrating piano string with inherent inharmonicity B, the above analysis is *not* valid since

$$p(n) \neq p(n + kn_0)$$

One may now recall our simple model of a vibrating piano string with inherent inharmonicity as described in equation **(3.2.1 – 2)**

$$x(t) = \sum_{n=1}^{N} \sin\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right]$$

where n in the above equation reflects the harmonic number.

Equation **(3.2.1 – 2)** may be written in the discrete domain as

$$x(m\Delta) = \sum_{n=1}^{N} \sin[\Omega_n m\Delta] \qquad \textbf{(3.3.3 – 13)}$$

where $\Omega_n = n 2\pi f_0 \left(1 + n^2 B\right)^{0.5}$. The Z-transform of *each* term of equation **(3.3.3 – 13)** can be calculated

$$Z\left[\sin\left(\Omega_n m\Delta\right)\right] = \frac{z^{-1}\sin(\Omega_m \Delta)}{z^{-2} - 2z^{-1}\cos(\Omega_m \Delta) + 1} \qquad \textbf{(3.3.3 – 14)}$$

However, equation **(3.3.3 – 14)** is only the Z-transform of one term only, so one needs to compute the Z-transform of all the terms of equation **(3.3.3 – 13)** and then compute the natural logarithm of the total sum

$$\ln X(z) = \ln\left[\sum_{n=1}^{N} \frac{z^{-1}\sin(\Omega_m \Delta)}{z^{-2} - 2z^{-1}\cos(\Omega_m \Delta) + 1}\right] \qquad \textbf{(3.3.3 – 15)}$$

It has not been possible to compute analytically equation **(3.3.3 – 15)**, although computer simulations were carried out in the next section as a means of investigating the resulting features of a signal with inherent inharmonicity.

3.3.4 Computational demonstrations

In this section, the concept of cepstrum is used to further unravel the structure of a real piano note recording with inherent inharmonicity and the structure of a synthesized periodic signal with and without inharmonicity present. These computational demonstrations naturally follow from the analytical formulations earlier on in order to reveal features of the signals in the time domain, where the sound reproduction mechanism actually takes place.

In particular, the use of the *real (or power) cepstrum*, $\hat{x}_{real}(t)$, is used for the computational demonstrations in this section where

$$\hat{x}_{real}(t) = \int_{-\infty}^{+\infty} \ln\left[\left|X(f)\right|\right] e^{j2\pi ft} df \qquad \textbf{(3.3.4 – 1)}$$

So, the real cepstrum of a signal is defined as the inverse Fourier transform of the natural logarithm of the magnitude of the Fourier transform of the signal.

A real single piano note C_3 was recorded at 44.1 kHz and analysed in terms of its power spectrum and real cepstrum. Note that the piano note was analysed using a single FFT over the whole 10 second recording using the "Gigastudio" sample library. The results are illustrated in Figure 3.18:

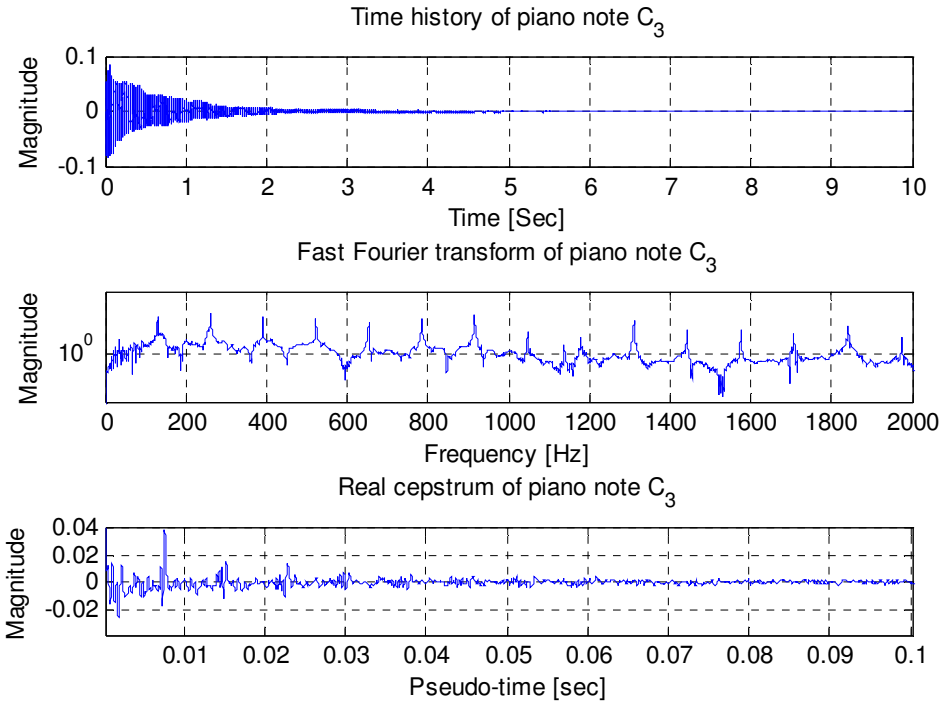


Figure 3.18 – FFT and cepstrum analysis of a real piano note C_3

It can be seen from the FFT of the recorded signal of Figure 3.18 that the fundamental frequency of piano note C_3 is 130.5 Hz and its harmonics are non-integer multiples of this value due to the inharmonicity factor inherent in the real piano note recording.

Additionally, from these measurements, one can identify that there are amplitude decaying impulses every 7.6 ms in the real cepstrum of Figure 3.18. These impulses in the cepstral domain are related to the fundamental frequency of piano note C_3 ($\approx$1/130.5).

We now zoom further around the first 7.6 ms of the pseudo-time domain of the cepstrum as shown in Figure 3.19:

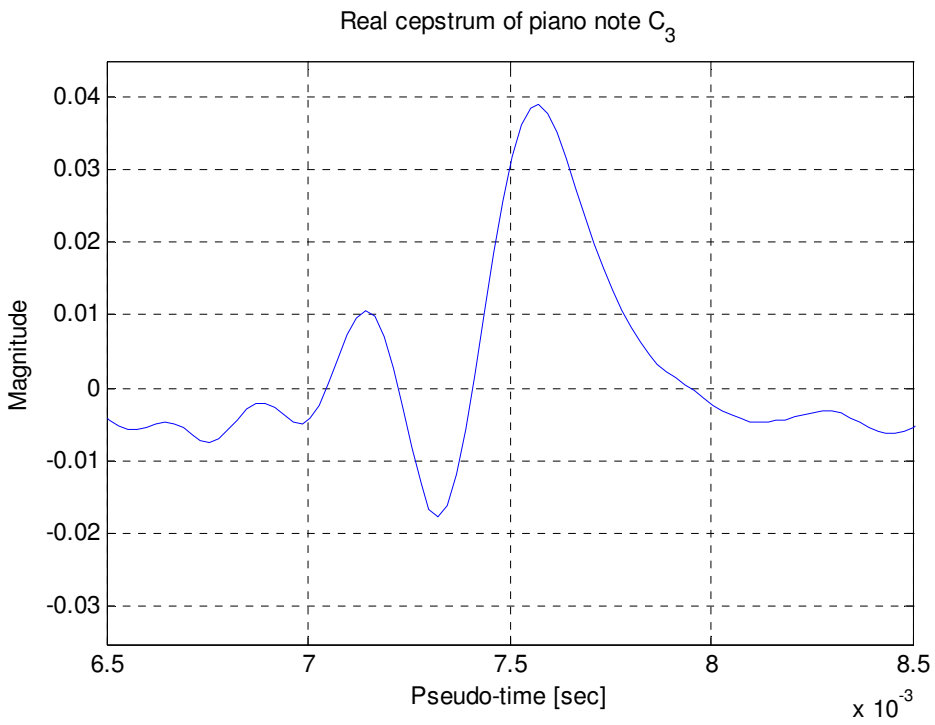


Figure 3.19 – Expanded view of the real cepstrum of Figure 3.18

From Figure 3.19, it can be observed that there is some indication of a one-sided effect in relation to the main lobe of the signal similar to the effect observed in section 3.2.2. The one-sided effect is attributed to the inharmonicity, i.e. the non-equal spacing of the harmonic frequencies as discussed earlier.

The effect of one-sided energy manifests itself more clearly by constructing an artificial signal by replacing the peak values of the power spectrum with delta functions of constant magnitude and replacing the noise floor of the signal by a constant value. The real cepstrum of this artificial signal can then be computed. The new synthesized signal is marked in red in Figure 3.20:

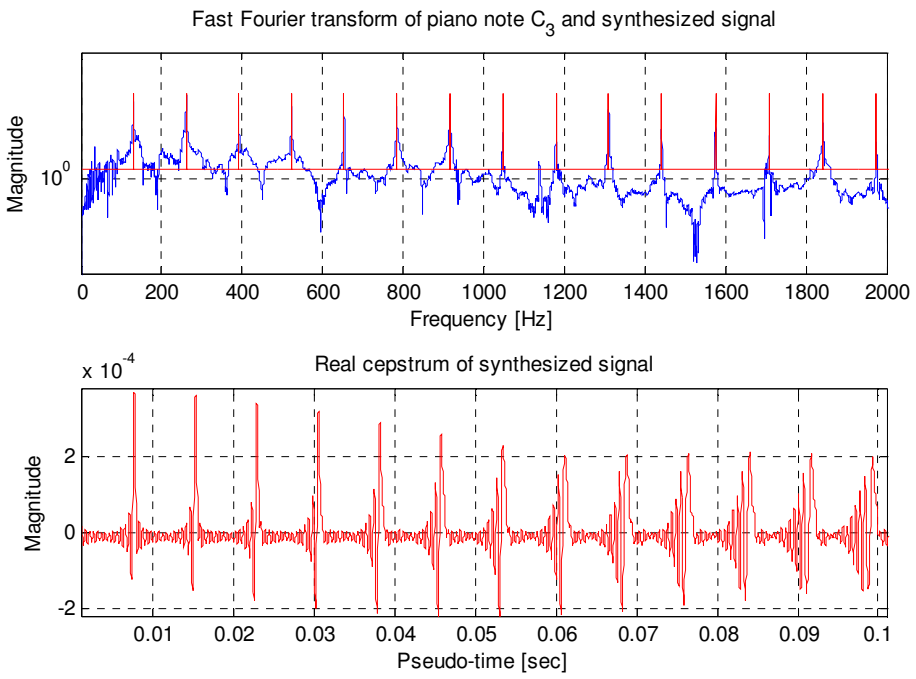


Figure 3.20 – FFT and real cepstrum of synthesized signal

By zooming in on two typical peaks of the synthesized signal in the pseudo-time domain, Figure 3.21 is obtained:

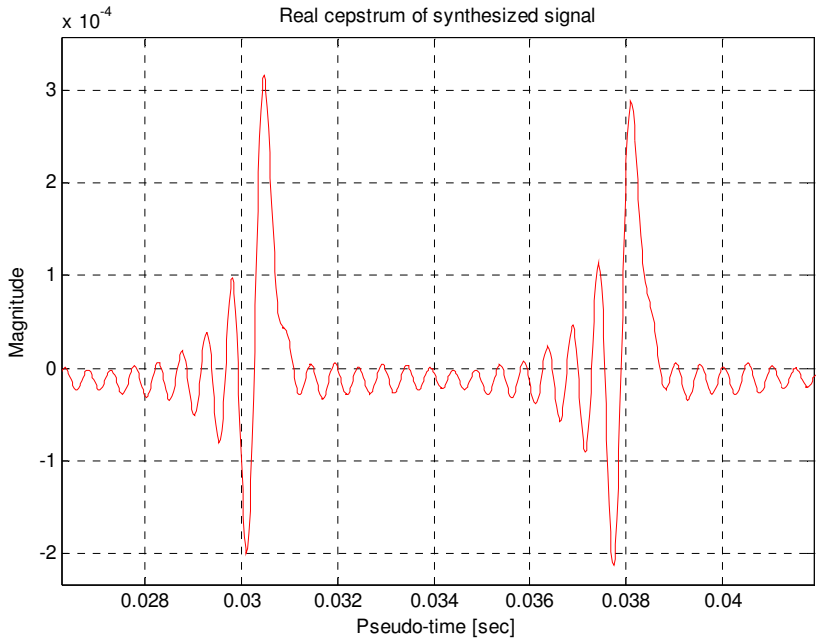


Figure 3.21 – Expanded view of the real cepstrum of Figure 3.20

From Figure 3.21, a clear time-evolving one-sided effect can be seen as also demonstrated in section 3.2.2.

Now, by replacing the peaks of the power spectrum of the original signal with a series of delta functions of constant magnitude, but this time being spaced at exact multiplies of the fundamental frequency, the inharmonicity is "removed" from the original signal. The real cepstrum of this new signal can then be computed and is shown in Figure 3.22:

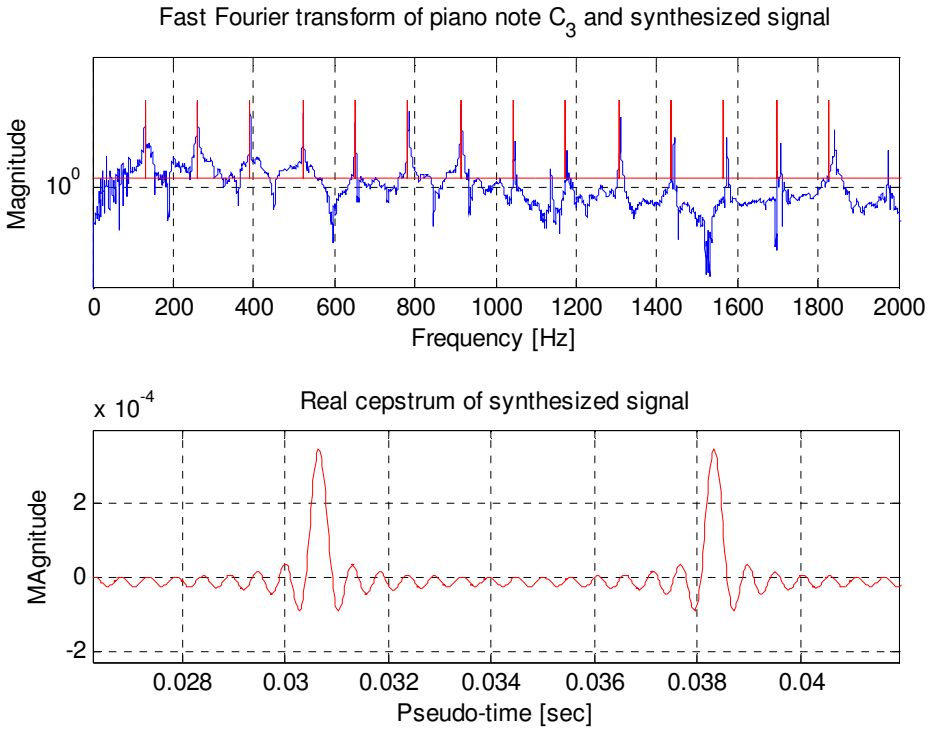


Figure 3.22 – FFT and real cepstrum of new synthesized signal

From Figure 3.22, it can be seen that there is only symmetric energy spreading through windowing and the one-sided effect is absent.

As a final concluding remark, the above analytical and computational investigations support the deduction that the energy in the high-time regions is believed to be associated with the response of the vibrating string of a piano, manifesting as a series of impulses in both the time and pseudo-time domains, whereas the dynamic response of the soundboard features in the low-time regions of the pseudo-time domain due to its broad dynamic response. Indeed, the response of the soundboard is broadband

above around 200 Hz and can be modelled as a linear filter having the characteristics described in the papers by Giordano (1997 and 1998). In particular, the soundboard can be modelled as a lumped parameter model with well defined modes of vibration up to 200 Hz, whereas its behaviour exhibits a multi-modal or broadband behaviour above that frequency according to Moore *et al.* (2006). This argument is further supported by experimental work on a grand piano described in chapter 4.

3.4 Response of multiple piano strings

So far, only the response of a single piano string when coupled with the soundboard has been investigated. As mentioned earlier, the middle and higher piano notes would tend to group a pair or a triplet of strings to a single piano note. The coupling of the strings of the same note would therefore occur through the bridge.

The coupling mechanism of multiple strings (or oscillators) is discussed in more detail in chapter 4, where analytical and computational formulations are presented as part of the unique double decay rate effect in the piano instrument. In this section, the coupling mechanism is investigated through the use of digital waveguides and the properties of the cepstrum.

3.4.1 Digital waveguide representation

In a simple digital waveguide representation, a vibrating piano string is modelled in such a way so that filter *K* represents the dissipation (modulus) and the dispersion (phase) phenomena and filter *D* represents the propagation time for the string vibration reflecting a pure delay. *E* and *S* are the hammer excitation and response of the modelled string respectively (Aramaki *et al.*, 2001)

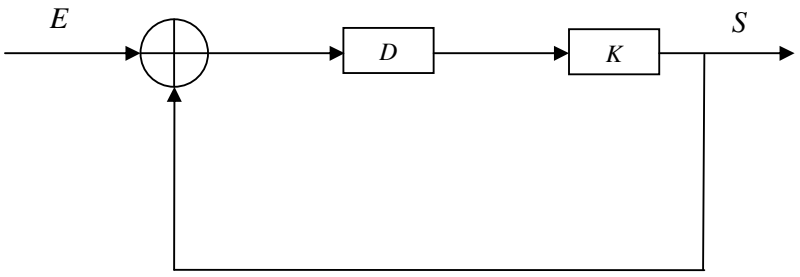


Figure 3.23 – Digital waveguide model representation of a single piano string

The transfer function of such a model is

$$G(\omega) = \frac{S(\omega)}{E(\omega)} = \frac{|K(\omega)|e^{-j(\omega D - \phi(\omega))}}{1 - |K(\omega)|e^{-j(\omega D - \phi(\omega))}}$$

where $\phi(s)$ is the Laplace transform of the phase and the above model can further be extended to two coupled strings in this research work, where C_1 and C_2 represent the coupling factors at the attachment points between the bridge and the two strings as shown in Figure 3.24:

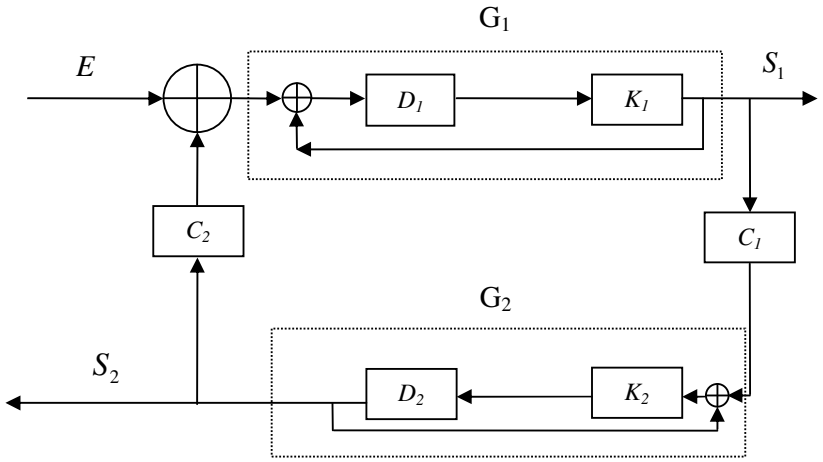


Figure 3.24 – Digital waveguide model representation of two coupled piano strings

where $\left\{ \begin{matrix} G_1 = \dfrac{D_1 K_1}{1 - D_1 K_1} \\ G_2 = \dfrac{D_2 K_2}{1 - D_2 K_2} \end{matrix} \right\}$.

From Figure 3.24, one may obtain the following relationships for the response of strings s_1 and s_2 respectively

$$\left\{ \begin{matrix} S_1 = G_1 \left[E + C_2 S_2 \right] \\ S_2 = G_2 C_1 S_1 \end{matrix} \right\} \qquad \textbf{(3.4.1 – 2)}$$

The response S_1 can be written as

$$S_1 = G_1 \left[E + C_2 C_1 G_2 S_1 \right]$$

$$S_1 = \frac{G_1 E}{1 - C_1 C_2 G_1 G_2} \qquad \textbf{(3.4.1 – 3)}$$

Also, the response S_2 due equation **(3.4.1 – 3)** can be written as

$$S_2 = \frac{C_1 G_1 G_2 E}{1 - C_1 C_2 G_1 G_2} \qquad \textbf{(3.4.1 – 4)}$$

Let us now assume that the total response, S_{tot}, of the two coupled strings is a linear sum of the motion of the two strings with constant coefficients a_1 and a_2 respectively

$$S_{tot} = a_1 S_1 + a_2 S_2 \qquad \textbf{(3.4.1 – 5)}$$

Equation **(3.4.1 – 5)** due to **(3.4.1 – 3)** and **(3.4.1 – 4)** becomes

$$S_{tot} = \frac{G_1 [a_1 + a_2 C_1 G_2] E}{1 - C_1 C_2 G_1 G_2} \qquad \textbf{(3.4.1 – 6)}$$

Equation **(3.4.1 – 6)** represents the total resulting response of two piano strings coupled through the bridge.

3.4.2 Case studies

One utilises the homomorphic properties of the cepstrum of the total signal, S_{tot}, by first taking the natural logarithms on either side of equation **(3.4.1 – 6)**

$$\ln S_{tot} = \ln G_1 + \ln[a_1 + a_2 C_1 G_2] + \ln E - \ln[1 - C_1 C_2 G_1 G_2] \qquad \textbf{(3.4.2 – 1)}$$

The above equation consists of three separate terms:

the first term, "$\ln G_1$" is simply related to response S_1 of string 1.

the second term, "$\ln[a_1 + a_2 C_1 G_2]$", may be approximated with a Taylor's series expansion as follows

$$\ln\left[a_1 \left(1 + \frac{a_2}{a_1} C_1 G_2 \right) \right] = \ln a_1 + \ln\left[1 + \frac{a_2}{a_1} C_1 G_2 \right] \approx \ln a_1 + \frac{a_2}{a_1} C_1 G_2 \qquad \textbf{(3.4.2 – 2)}$$

providing there is "weak" coupling only, i.e. $C_1 << 1$.

The third term of equation **(3.4.2 – 1)**, "$\ln[1 - C_1C_2G_1G_2]$", may also be approximated with a Taylor's series expansion

$$\ln[1 - C_1C_2G_1G_2] \approx -C_1C_2G_1G_2 \qquad \textbf{(3.4.2 – 3)}$$

providing there is also "weak" coupling, i.e. $C_1, C_2 << 1$.

Equation **(3.4.2 – 1)** due to **(3.4.2 – 2)** and **(3.4.1 – 3)** can now be written as

$$\ln S_{tot} = \ln G_1 + \ln a_1 + \frac{a_2}{a_1} C_1G_2 + \ln E + C_1C_2G_1G_2 \qquad \textbf{(3.4.2 – 5)}$$

Two cases are explored below.

Case 1

In the previous sections of this chapter, one has modelled the response of a single vibrating string with the presence of inharmonicity. So, if there is no coupling, i.e. $C_1 = C_2 = 0$ and there is no response from the second string, i.e. $G_2 = 0$, then equation **(3.4.2 – 5)** can be reduced to

$$\ln S_{tot} = \ln G_1 + \ln a_1 + \ln E \qquad \textbf{(3.4.2 – 6)}$$

By taking the inverse Fourier transforms on either side of equation **(3.4.2 – 5)**, one obtains

$$F^{-1}[\ln S_{tot}] = F^{-1}[\ln G_1] + F^{-1}[\ln a_1] + F^{-1}[\ln E]$$

$$\hat{S}_{tot} = \hat{G}_1 + \hat{a}_1 + \hat{E} \qquad \textbf{(3.4.2 – 7)}$$

Equation **(3.4.2 – 7)** represents the response G_1 of a single vibrating string with inharmonicity in the pseudo-time domain, where a series of repetitive impulses will manifest in the higher-time regions, as discussed previously. The cepstrum of the hammer excitation, $\hat{E}$, simply denotes an impulse and $\hat{a}_1$ has no real contribution to the total response.

Case 2

In the more complex case, where there is weak coupling and both of the strings vibrate, then the total resulting response of the cepstrum, $\hat{S}_{tot}$, is different. By taking the inverse Fourier transforms on either side of equation **(3.4.2 – 5)**, one obtains

$$\hat{S}_{tot} = \hat{G}_1 + \hat{E} + \hat{a}_1 + F^{-1}\left[\frac{a_2}{a_1} C_1 G_2\right] + F^{-1}\left[C_1 C_2 G_1 G_2\right] \qquad \textbf{(3.4.2 – 8)}$$

From the resulting total response of the two weakly coupled strings, as described in equation **(3.4.2 – 8)**, one can deduce the following contributions for each term individually:

the terms $\hat{G}_1$ and $\hat{E}$ are simply the cepstra of response S_1 and hammer excitation E, respectively, as discussed before. In other words, $\hat{E}$ is simply a pulse in the pseudo-time domain, whereas $\hat{G}_1$ represents the response of string S_1. Also, $\hat{a}_1$ has no real contribution to the total response since it merely represents the cepstrum of a constant coefficient.

the term $F^{-1}\left[\frac{a_2}{a_1} C_1 G_2\right]$ represents the response of the second string modified by the coupling term C_1 and the constant coefficients a_1 and a_2.

Finally, the term $F^{-1}\left[C_1 C_2 G_1 G_2\right]$ represents the convolved responses of the two strings, S_1 and S_2, modified by both coupling terms C_1 and C_2.

3.5 Conclusions

In this chapter, the effect of inharmonicity has been presented through a set of novel demonstrations and analytical formulations. The inharmonicity may be one of the key features of further unravelling the problem of music transcription and might provide an additional important parameter in the development of a more comprehensive signal model.

A simple model for a single vibrating piano string with inherent inharmonicity was proposed. Through the numerical simulations of such a signal, one showed how the response changed by varying five factors; the magnitude of the inharmonicity factor B, the number of harmonics N present in the modelled signal, the time parameter t, the phase characteristics and the amplitude of harmonics. From these investigations, it was shown that there is a so-called "one-sided" effect, as a direct result of the non-integer spacing of the harmonic components (relative to the fundamental frequency) with uniformly varying amplitudes added to the modelled signal. However, the effect of one-sided energy was shown that it is actually obscured in real piano note recordings, since the amplitudes of the harmonic terms vary non-uniformly. Also, the phase characteristics of the harmonics of the signal in real recordings may be totally random and hence may further obscure the one-sided effect associated with the inharmonicity. As a result, such an effect may not be perceived by the human auditory system.

The cepstrum has also been utilised to reveal a similar, strong, one-sided energy effect. Indeed, real single piano recordings and synthesized signals were analysed and there was evidence that one-sided energy occurs in relation to the main lobes, when a train of spikes are non-equally spaced in the frequency domain. On the other hand, a synthesized signal with equally-spaced spikes showed no such effect.

As far as the analytical formulations are concerned using the cepstrum, it was possible to derive representations for the resulting sound of a piano note without the inharmonicity factor (periodic signal) into separate components (response of string, and soundboard). The difficulties of deriving similar analytical representations of the modelled signal with the inharmonicity were also discussed. In any case, the use of the cepstrum has enabled us to substantiate an earlier hypothesis (Karatsovis *et*

al., 2006), where the soundboard would have a broad spectral response and consequently a cepstrum that predominantly contains its information in the low "quefrencies" (or low-time regions), whereas a vibrating piano string would give rise to a series of impulses in the pseudo-time domain manifesting themselves in the higher quefrencies (or high-time regions). Note also that response of the vibrating string will be well separated from cepstral coefficients associated with the soundboard due to the homomorphic properties of the cepstrum.

Finally, the response of multiple piano strings was discussed through the use of digital waveguides and analysis was carried out for two weakly coupled vibrating piano strings using the cepstrum. This leads naturally to the next big topic of this research work, which is the effect of the coupling mechanism between oscillators and the modelling of the double decay rate effect.

Chapter 4

Double decay rate effect

4.1 Introduction

The double decay rate effect is the second feature that is discussed in this research since its use could potentially aid further in understanding the sound generation mechanism of the piano.

This chapter initially presents a simple analytical model of the double decay rate that can be observed in piano recordings through the properties of coupled oscillators. The model is developed for the response of two masses connected through linear coupling. Each mass can either represent the vertical (or horizontal) motion of one string coupled with the response of a second string through the bridge, or the modelled masses can represent the coupled vertical and horizontal motions of one string only. Analogies of this general model of coupled oscillators are drawn from the formulation of digital waveguides as shown in chapter 3. Furthermore, numerical simulations are carried out in attempt to replicate the double decay rate effect found in real piano recordings in the case of identical and mistuned oscillators.

A physical model is also developed to describe the dynamics of more than two strings coupled through the bridge of a piano. This type of physical modelling is based on a frequency domain mobility-based method providing an extension to the general model of coupled oscillators.

Finally, the coupling of the piano strings with the bridge and the soundboard is further investigated by carrying out vibration measurements on a real grand piano.

4.2 Coupled oscillators

As mentioned and illustrated earlier in section 2.4.3, a careful examination of the time history (shown at a particular frequency) of a piano note recording reveals a two-stage decay rate; the first stage has a rapid roll-off, whilst the second decay rate has a much slower roll-off resulting into what is called the "aftersound" of a piano note (Weinreich, 1977). This compound phenomenon is known as the double decay rate effect. There are mainly two factors believed to be associated with the double decay rate of the sound of a piano note; first, the polarisation of the strings (vertical and horizontal motion) and second, the coupling effect between different strings of the same piano note. The model developed in this chapter can ultimately be used to describe both of these factors.

4.2.1 The general model

Single piano notes mainly comprise two or three strings grouped together. The complete response of three masses, m_1, m_2 and m_3, representing three vibrating strings in the vertical and horizontal motions (two polarisations) is shown in Figure 4.1:

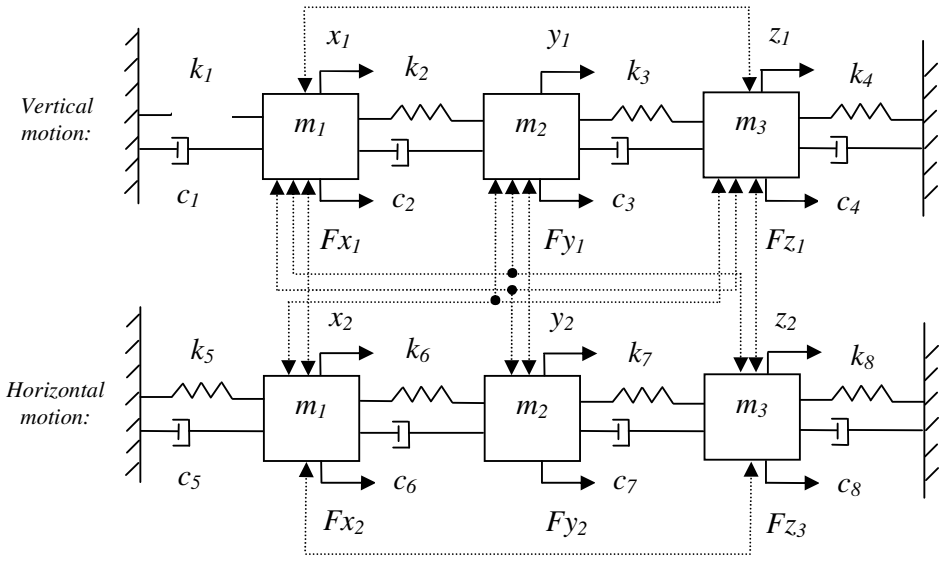


Figure 4.1 – Coupled six-degree-of-freedom-system with motion represented both in the vertical and horizontal directions

where m_1, m_2 and m_3 are the same for all three strings, and the stiffness terms may vary between the three strings: k_1, k_2, k_3 and k_4 represent the stiffness in the vertical polarisation, k_5, k_6, k_7 and k_8 represent the stiffness in the horizontal polarisation, also c_1, c_2, c_3 and c_4 represent the damping in the vertical polarisation, c_5, c_6, c_7 and c_8 represent the damping in the horizontal polarisation, x_1, y_1, z_1 represent the displacement of masses m_1, m_2 and m_3 in the vertical polarisation, x_2, y_2, z_2 represent the displacement of masses m_1, m_2 and m_3 in the horizontal polarisation, F_{x1}, F_{y1}, F_{z1} represent the forces acting on masses m_1, m_2 and m_3 in the vertical polarisation and F_{x2}, F_{y2}, F_{z2} represent the forces acting on masses m_1, m_2 and m_3 in the horizontal polarisation.

The additional dashed arrows represent the coupling of each mass in any polarisation with every other mass in any polarisation. The coupling between two masses (or oscillators) can normally be expressed with a stiffness and a damping parameter.

The above full model is appropriate to be solved either analytically or numerically. However, in our case the objective is to derive simpler relationships of the coupled motion of the strings in order to understand the mechanism of coupling and the double decay rate effect.

Let us now consider a simple two-degree-of-freedom system dynamic model as shown in Figure 4.2. In this particular example, each mass can either represent the vertical (or horizontal) motion of one string coupled with the response of a second string through the bridge, or the modelled masses can represent the coupled vertical and horizontal motions of one string only. The two masses, mass m_1 and m_2, are connected through coupling of a spring with stiffness k_c and a dashpot with viscous damping c_c representing the properties of the bridge. Displacement responses x, y represent the displacement for each mass due to forces F_1 and F_2 respectively:

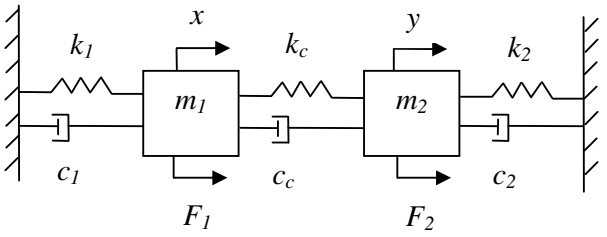


Figure 4.2 – Coupled two-degree-of-freedom-system

The equation of motion for mass m_1 is

$$m_1\ddot{x} + c_1\dot{x} + k_1x + c_c\dot{x} + k_cx = c_c\dot{y} + k_cy + F_1 \qquad \textbf{(4.2.1 – 1)}$$

Defining the following parameters

$$\omega_1^2 = \frac{k_1}{m_1},\ \omega_3^2 = \frac{k_c}{m_1},\ 2\zeta_1\omega_1 = \frac{c_1}{m_1},\ 2\zeta_3\omega_3 = \frac{c_c}{m_1} \qquad \textbf{(4.2.1 – 2)}$$

By taking Laplace transforms, and using the substitutions given in **(4.2.1 – 2)**, equation **(4.2.1 – 1)** becomes

$$\left[\left(s^2 + 2\zeta_1\omega_1 s + \omega_1^2\right) + \left(2\zeta_3\omega_3 s + \omega_3^2\right)\right]X(s) = \left[2\zeta_3\omega_3 s + \omega_3^2\right]Y(s) + \frac{F_1(s)}{m_1} \qquad \textbf{(4.2.1 – 3)}$$

Similarly, for mass m_2, the equation of motion is

$$m_2\ddot{y} + c_2\dot{y} + k_2y + c_c\dot{y} + k_cy = c_c\dot{x} + k_cx + F_2 \qquad \textbf{(4.2.1 – 4)}$$

Hence,

$$\left[\left(s^2 + 2\zeta_2\omega_2 s + \omega_2^2\right) + \left(2\zeta_4\omega_4 s + \omega_4^2\right)\right]Y(s) = \left[2\zeta_4\omega_4 s + \omega_4^2\right]X(s) + \frac{F_2(s)}{m_2} \qquad \textbf{(4.2.1 – 5)}$$

Now introduce the following substitutions

$$\begin{Bmatrix} \alpha(s) = (s^2 + 2\zeta_1\omega_1 s + \omega_1^2) + (2\zeta_3\omega_3 s + \omega_3^2) \\ \beta(s) = (2\zeta_3\omega_3 s + \omega_3^2) \\ \gamma(s) = (s^2 + 2\zeta_2\omega_2 s + \omega_2^2) + (2\zeta_4\omega_4 s + \omega_4^2) \\ \delta(s) = (2\zeta_4\omega_4 s + \omega_4^2) \\ \overline{F}_1(s) = \dfrac{F_1(s)}{m_1}, \overline{F}_2(s) = \dfrac{F_2(s)}{m_2} \end{Bmatrix} \qquad \textbf{(4.2.1 – 6)}$$

Using system of equations **(4.2.1 – 6)**, equations **(4.2.1 – 3)** and **(4.2.1 – 5)** can be written as

$$\begin{Bmatrix} X(s) = \dfrac{1}{\Delta(s)}\left[\gamma(s)\overline{F}_1(s) + \beta(s)\overline{F}_2\right] \\ Y(s) = \dfrac{1}{\Delta(s)}\left[\delta(s)\overline{F}_1(s) + \alpha(s)\overline{F}_2\right] \end{Bmatrix} \qquad \textbf{(4.2.1 – 7)}$$

where $\Delta(s) = \alpha(s)\gamma(s) - \beta(s)\delta(s)$. Recalling equation **(C – 1)** of Appendix C, the coefficients of the 4th order polynomial, which represent the polynomial $\Delta(s)$, are

$$A_1 = 1$$

$$A_2 = 2\zeta_2\omega_2 + 2\zeta_4\omega_4 + 2\zeta_1\omega_1 + 2\zeta_3\omega_3$$

$$A_3 = \omega_2^2 + \omega_4^2 + 4\zeta_1\omega_1\zeta_2\omega_2 + 4\zeta_1\omega_1\zeta_4\omega_4 + \omega_1^2 + 4\zeta_3\omega_3\zeta_2\omega_2 + 4\zeta_3\omega_3\zeta_4\omega_4 + \omega_3^2 - 4\zeta_3\omega_3\zeta_4\omega_4$$

$$A_4 = 2\zeta_1\omega_1\omega_2^2 + 2\zeta_1\omega_1\omega_4^2 + 2\zeta_2\omega_2\omega_1^2 + 2\zeta_4\omega_4\omega_1^2 + 2\zeta_3\omega_3\omega_2^2 + 2\zeta_3\omega_3\omega_4^2 \ldots$$

$$\ldots + 2\zeta_2\omega_2\omega_3^2 + 2\zeta_4\omega_4\omega_3^2 - 2\zeta_3\omega_3\omega_4^2 - 2\omega_3^2\zeta_4\omega_4$$

$$A_5 = \omega_1^2\omega_2^2 + \omega_1^2\omega_4^2 + \omega_3^2\omega_2^2$$

The roots of $\Delta(s)$ can now be used to calculate the damping and natural frequencies of the coupled oscillators, and the numerators in **(4.2.1 – 7)** reflect the "participation" of each mode in the response of $x(t)$ and $y(t)$. By taking the inverse Laplace transform of system of equations **(4.2.1 – 7)**, one can now obtain the responses, $x(t)$ and $y(t)$, of the two coupled masses respectively

$$x(t) = g(t) * \left\{L^{-1}\left(\gamma(s)\right) * L^{-1}\left(\overline{F}_1\right) + L^{-1}\left(\beta(s)\right) * L^{-1}\left(\overline{F}_2\right)\right\} \qquad \textbf{(4.2.1 – 8)}$$

$$y(t) = g(t) * \left\{L^{-1}\left(\delta(s)\right) * L^{-1}\left(\overline{F}_1\right) + L^{-1}\left(\alpha(s)\right) * L^{-1}\left(\overline{F}_2\right)\right\} \qquad \textbf{(4.2.1 – 9)}$$

where $g(t) = L^{-1}\left[\frac{1}{\Delta(s)}\right]$, i.e. this is the impulse response corresponding to the transfer function of $\frac{1}{\Delta(s)}$. In particular,

$$g(t) = L^{-1}\left\{\frac{1}{\left(s^2 + 2\zeta_a\omega_a s + \omega_a^2\right)\left(s^2 + 2\zeta_b\omega_b s + \omega_b^2\right)}\right\} \qquad \textbf{(4.2.1 – 10)}$$

where ω_a, ω_b are the roots of $\Delta(s)$ that can be found as per equation **(C – 1)** and ζ_a, ζ_b are the damping ratios for each mode of the dynamic system that control the different decay rates. So, by using partial fractions, one can use equation **(4.2.1 – 10)** to describe the two different modes of the coupled dynamic system as follows

$$g_1(t) = L^{-1}\left\{\frac{As + B}{(s^2 + 2\zeta_a\omega_a s + \omega_a^2)}\right\} \text{ is mode 1 of the coupled system}$$

and

$$g_2(t) = L^{-1}\left\{\frac{Cs + D}{s^2 + 2\zeta_b\omega_b s + \omega_b^2}\right\} \text{ is mode 2 of the coupled system}$$

Alternatively to equations **(4.2.1 – 8)** and **(4.2.1 – 9)**, the total forced responses $x(t)$ and $y(t)$ can be found by directly solving for the coupled second order differential equations **(4.2.1 – 1)** and **(4.2.1 – 4)** directly. Specifically, the following can be written

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} + \begin{bmatrix} c_1+c_c & -c_c \\ -c_c & c_2+c_c \end{bmatrix}\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} + \begin{bmatrix} k_1+k_c & -k_c \\ -k_c & k_2+k_c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

A set of reduced ordinary differential equations can therefore be used to numerically solve for $x(t)$, $\dot{x}(t)$, $y(t)$ and $\dot{y}(t)$, as shown in equation **(4.2.1 – 11)** in state space form

$$\frac{d}{dt}\begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -(k_1+k_c)/m_1 & -(c_1+c_c)/m_1 & k_c/m_1 & c_c/m_1 \\ 0 & 0 & 0 & 1 \\ k_c/m_2 & c_c/m_2 & -(k_2+k_c)/m_2 & -(c_2+c_c)/m_2 \end{bmatrix}\begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix} - \begin{bmatrix} 0 \\ F_1 \\ 0 \\ F_2 \end{bmatrix} \qquad \textbf{(4.2.1 – 11)}$$

This formulation will be used later to demonstrate the effect of the double decay rate in the case of coupled oscillators in the time domain.

4.2.2 Digital waveguide model analogy

As mentioned in the previous section, the general model of coupled oscillators, can either be used to model the vertical (or horizontal) motion of one string coupled with the response of a second string through the bridge, or can be used to model the coupled vertical and horizontal motions of one string only.

This section draws analogies with the digital waveguide model, as described in section 3.4, for the response of one string, coupled with another, due to a single force excitation.

One recalls equation **(4.2.1 – 3)**, which is the equation of motion of mass m_1 (for example representing the response of one string for a single polarisation)

$$\left[\left(s^2 + 2\zeta_1\omega_1 s + \omega_1^2\right) + \left(2\zeta_3\omega_3 s + \omega_3^2\right)\right]X(s) = \left[2\zeta_3\omega_3 s + \omega_3^2\right]Y(s) + \frac{F_1(s)}{m_1}$$

Recalling **(4.2.1 – 5)**, which is the equation of motion of mass m_2 (for example representing the response of a second string in the same polarisation)

$$\left[\left(s^2 + 2\zeta_2\omega_2 s + \omega_2^2\right) + \left(2\zeta_4\omega_4 s + \omega_4^2\right)\right]Y(s) = \left[2\zeta_4\omega_4 s + \omega_4^2\right]X(s) = \frac{F_2(s)}{m_2}$$

Introducing the following substitutes

$$\left\{\begin{array}{c} H_x(s) = s^2 + 2\zeta_1\omega_1 s + \omega_1^2 \\ H_y(s) = s^2 + 2\zeta_2\omega_2 s + \omega_2^2 \\ C_x(s) = 2\zeta_3\omega_3 s + \omega_3^2 \\ C_y(s) = 2\zeta_4\omega_4 s + \omega_4^2 \end{array}\right\} \qquad \textbf{(4.2.2 – 1)}$$

Equations **(4.2.1 – 3)** and **(4.2.1 – 5)** can be written as

$$X(s) = \frac{C_x(s)}{H_x(s) + C_x(s)} Y(s) + \frac{F_1(s)}{H_x(s) + C_x(s)} \qquad \textbf{(4.2.2 – 2)}$$

$$Y(s) = \frac{C_y(s)}{H_y(s) + C_y(s)} X(s) + \frac{F_2(s)}{H_y(s) + C_y(s)} \qquad \textbf{(4.2.2 – 3)}$$

Since there is only one force exciting one of the strings, one sets $F_2 = 0$. Hence, the response of one string, as described in **(4.2.2 – 2)**, can be written

$$X(s) = \frac{H_y(s) + C_y(s)}{H_x(S)H_y(s) + C_x(s)H_y(s) + C_y(s)H_x(s)} F_1(s) \qquad \textbf{(4.2.2 – 4)}$$

Recalling equation **(3.4.1 – 3)** representing the response, S_1 of one coupled string in terms of an equivalent digital waveguide representation

$$S_1(s) = \frac{G_1(s)E(s)}{1 - C_1(s)C_2(s)G_1(s)G_2(s)} \qquad \textbf{(4.2.2 – 5)}$$

where $E(s) = F_1(s)$.

Introducing the following substitutions

$$\left\{ \begin{array}{c} G_1(s) = \dfrac{1}{H_x(s) + C_x(s)} \\ G_2(s) = \dfrac{1}{H_y(s) + C_y(s)} \\ C_1(s) = C_x(s), C_2(s) = C_y(s) \end{array} \right\} \qquad \textbf{(4.2.2 – 6)}$$

Equation **(4.2.2 – 5)** becomes

$$S_1(s) = \frac{H_y(s) + C_y(s)}{H_x(s)H_y(s) + C_x(s)H_y(s) + C_y(s)H_x(s)} E(s) \qquad \textbf{(4.2.2 – 7)}$$

Equations **(4.2.2 – 4)** and **(4.2.2 – 7)** are therefore identical, so the simple coupled model of two strings is equivalent to the digital waveguide representation as discussed in section 3.4.

4.2.2.1 Identical oscillators

In the case of two identical oscillators, one writes

$$H_x(s) = H_y(s) = H(s) \text{ and } C_x(s) = C_y(s) = C(s) \qquad \textbf{(4.2.2.1 – 1)}$$

Therefore, equation **(4.2.2 – 7)** becomes

$$S_1(s) = \frac{H(s) + C(s)}{H^2(s) + 2C(s)H(s)} E(s) \qquad \textbf{(4.2.2.1 – 2)}$$

The roots of the denominator of equation **(4.2.2.1 – 2)** define the dynamics of the coupled system, i.e.

$$H(s)\big(H(s) + 2C(s)\big) = 0 \qquad \textbf{(4.2.2.1 – 3)}$$

Equation **(4.2.2.1 – 3)** can be factorised into two terms

$$H(s) = 0 \text{ and } H(s) + 2C(s) = 0$$

This shows that the coupling of two identical oscillators will result in two modes; one of which is identical to the *uncoupled* mode.

Assuming for convenience that $E(s) = 1$ (unit impulse force) and by the taking inverse Laplace transforms of equation **(4.2.2.1 – 2)**, one obtains

$$s_1(t) = L^{-1}\left[\frac{1}{H(s) + 2C(s)} + \frac{C(s)}{H(s)}\frac{1}{H(s) + 2C(s)}\right] \qquad \textbf{(4.2.2.1 – 3)}$$

where the term

$$r_1(t) = L^{-1}\left[\frac{C(s)}{H(s)}\right] = L^{-1}\left[\frac{N_1(s)}{(s^2 + 2\zeta\omega_0 s + \omega_0^2)}\right]$$

represents one mode in the response of $s_1(t)$, which has the same properties as that of the uncoupled mode.

Further, the term

$$r_2(t) = L^{-1}\left[\frac{1}{H(s)+2C(s)}\right] = L^{-1}\left[\frac{N_2(s)}{(s^2+2\zeta_c\omega_c s+\omega_c^2)}\right]$$

represents the second coupled mode in the response of $s_1(t)$.

This can also be demonstrated with a simple example of two identical oscillators with the same uncoupled damping

$$\omega_1 = \omega_2 = 173.2 \text{ rad/s with } c_1 = c_2 = 0.1 \text{ Ns/m}$$

and the coupling parameters of the bridge are $k_c = 100$ N/m and $c_c = 1$ Ns/m.

Solving for equation **(C - 1)**, one can obtain the coupled frequencies, $\omega_{1,c}, \omega_{2,c}$ and damping ratios, $\zeta_{1,c}, \zeta_{2,c}$ of the dynamic system ("*c*" subscript denotes the coupled parameter). The solutions are two quadratic pairs represented by solution *T*

$$T = \begin{Bmatrix} -10.5 \pm 178.6i \\ -0.5 \pm 173.2i \end{Bmatrix} \qquad \textbf{(4.2.2.1 – 4)}$$

Hence, from the two solution pairs of equation **(4.2.2.1 – 4)**, one obtains respectively

$$\begin{Bmatrix} \zeta_{1,c}\omega_{1,c} = 10.5 \\ \omega_{1,c}\sqrt{1-\zeta_{1,c}^2} = 178.6 \end{Bmatrix} \text{ and } \begin{Bmatrix} \zeta_{2,c}\omega_{2,c} = 0.5 \\ \omega_{2,c}\sqrt{1-\zeta_{2,c}^2} = 173.2 \end{Bmatrix} \qquad \textbf{(4.2.2.1 – 5)}$$

Solving for the system of equations **(4.2.2.1 – 5)**, one obtains the properties of the two modes of the system

$$\omega_{1,c} = 178.9 \text{ rad/ with } \zeta_{1,c} = 0.06$$

$$\omega_{2,c} = 173.2 \text{ rad/s with } \zeta_{2,c} = 0.003$$

$$\text{Ratio of damping terms between the two coupled modes: } \frac{\zeta_{1,c}}{\zeta_{2,c}} = 20$$

Therefore, despite the coupling between the two strings, one of the resulting coupled modes will be identical to that of the uncoupled mode. In physical terms, there is

symmetry in the system and the two strings will move together in-phase through the coupling of the bridge, which acts as rigid link.

Finally, the ratio of damping terms between the two coupled modes is 20 times larger than that of the uncoupled modes, which means that there are two very different decay rates. This may ultimately result in the characteristic double decay rate of the reproduced piano sound, when both of these modes are appropriately excited.

4.2.2.2 Mistuned oscillators

In the case of two "mistuned" oscillators, one introduces a small change $\Sigma(s)$ in the system response. The word "mistuning" is used when the fundamental frequencies of two strings may vary by a small amount. One can write

$$H_y(s) = H_x(s) + \Sigma(s)$$

Note that the coupling in both masses is identical

$$C_x(s) = C_y(s) = C(s)$$

Equation **(4.2.2 – 7)** becomes

$$S_1(s) = \frac{H_x(s) + \Sigma(s) + C(s)}{H_x(s)[H_x(s) + \Sigma(s)] + C(s)[2H_x(s) + \Sigma(s)]} E(s)$$

$$S_1(s) = \frac{H_x(s) + \Sigma(s) + C(s)}{[H_x(s) + \Sigma(s)][H_x(s) + C(s)] + C(s)H_x(s)} E(s) \qquad \textbf{(4.2.2.2 – 1)}$$

Assuming for convenience that $E(s) = 1$ (unit impulse force) and by taking inverse Laplace transforms of equation **(4.2.2.2 – 1)**

$$S_1(t) = L^{-1}\left[\frac{1}{[H_x(s) + \Sigma][H_x(s) + C(s)] + C(s)H_x(s)}[H_x(s) + \Sigma + C(s)]\right] \textbf{(4.2.2.2 – 2)}$$

From the above formulation, it is *not* obvious analytically which are the modes of the coupled dynamical system and how they interact with each other, so a numerical example is presented later in section 4.2.3.2.

In a simple example of two mistuned oscillators, one assumes that the oscillators differ by about 1 rad/s (or about 0.2 Hz)

$$\omega_1 = 172.2 \text{ rad/s and } \omega_1 = 173.2 \text{ rad/s with } c_1 = c_2 = 0.1 \text{ Ns/m}$$

As before, the coupling parameters of the bridge are $k_c = 100$ N/m and $c_c = 1$ Ns/m.

Solving equation **(C – 1)**, one obtains the properties of the two modes of the system

$$\omega_{1,c} = 178.4 \text{ rad/ with } \zeta_{1,c} = 0.06$$

$$\omega_{2,c} = 172.7 \text{ rad/s with } \zeta_{2,c} = 0.003$$

$$\text{Ratio of damping terms between the two coupled modes: } \frac{\zeta_{1,c}}{\zeta_{2,c}} = 20$$

Hence, in the case of the mistuned oscillators, both of the natural frequencies of the coupled modes will differ from the values for the uncoupled system. The ratio of damping terms between the two coupled modes is again 20 times larger than that of the uncoupled modes, which means that there are two different decay rates as before.

4.2.3 Double decay rate demonstrations

In this section, one illustrates the effect of double decay rate, as captured in a typical piano note recording, using the Hilbert transform. Further numerical simulations are carried out in an attempt to replicate this effect in the case of mistuned coupled oscillators representing real piano strings.

4.2.3.1 Effect in a real piano recording

In this section, the double decay of a typical piano note recording is observed through the use of the Hilbert transform. The Hilbert transform, $\hat{f}(t)$, can be used to generate the analytical representation of a signal, $f(t)$

$$\hat{f}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} d\tau \qquad \textbf{(4.2.3.1 – 1)}$$

The analytic form of a signal is complex valued and the "instantaneous" amplitude of the signal can be defined as the amplitude of the analytic signal and can be used to calculate the decay rates in the time domain.

Figure 4.3 shows the evaluated instantaneous amplitude of a real piano note C_3, which was sampled at 44.1 kHz, at the fundamental and first two harmonic frequencies using recordings from the "Gigastudio" sample library. The acoustic signal was passed through band pass filters centered at the fundamental and the first two harmonic frequencies respectively with a selected 40 Hz bandwidth. Then, the instantaneous amplitudes were obtaining by evaluating the Hilbert transforms for each of these filtered signals, as shown on a natural log scale:

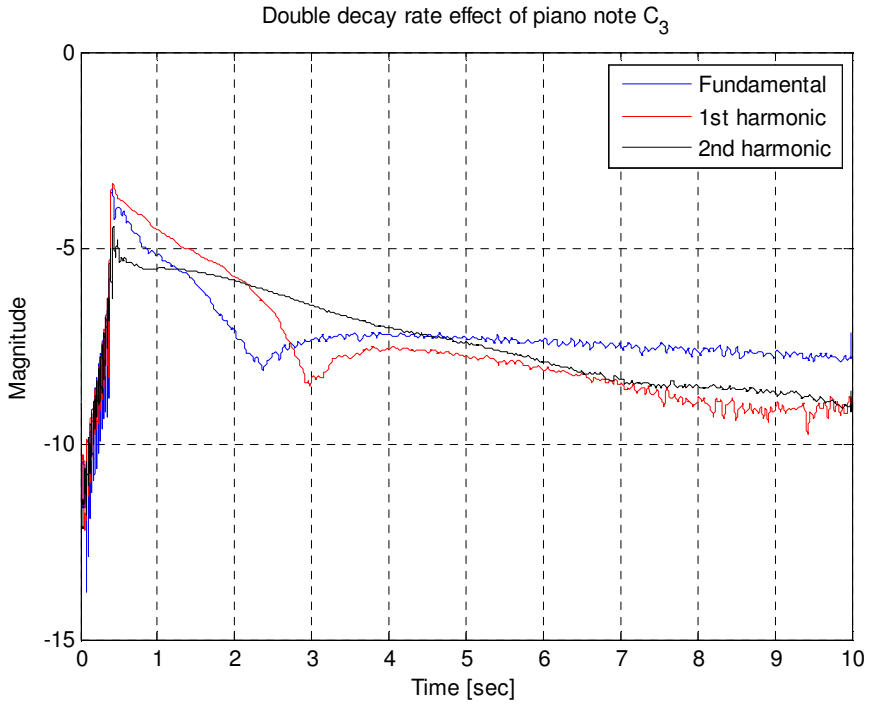


Figure 4.3 – Double decay rates of piano note C_3

From Figure 4.3, it can be seen that for the fundamental and first harmonic frequency there is a relatively fast decay up to about 2.2 s and 3 s respectively, followed by a slower decay. However, in the case of the second harmonic, the first decay rate may only seem to last about 0.2 s and then it is followed by a much longer second decay rate. Therefore, it can be deduced that the double decay rate characteristics will be different for the different frequency terms (fundamental and harmonics). In addition, it is not understood how the decaying rates of the different frequency terms contribute to the way the overall note is perceived (Weinreich, 1977).

Finally, the "dips" at the point where the second decay rate takes over from the first decay rate of each frequency can be attributed to the local destructive interference between the two different modes of the acoustic response forming a local minimum (Weinreich, 1977). Mathematically, it can be proved that the resulting "dip" may be a function of the mistuning relationship between two oscillators. A simple example is illustrated below.

To demonstrate this, one considers two coupled oscillators with frequencies ω_1 and ω_2, where $\omega_2 = \omega_1 + \Delta\omega$ ($\Delta\omega$ representing the mistuning parameter). Consider also that the amplitudes and damping factors of the two oscillators are identical. The resulting response $s(t)$ would be

$$s(t) = \sin(\omega_1 t) + \sin(\omega_1 + \Delta\omega)t \qquad \textbf{(4.2.3.1 – 1)}$$

Alternatively, equation **(4.2.3.1 – 1)** can be written as

$$s(t) = \sin(\omega_1 t)\left(1 + \cos(\Delta\omega t)\right) + \cos(\omega_1 t)\sin(\Delta\omega t) \qquad \textbf{(4.2.3.1 – 2)}$$

Equation **(4.2.3.1 – 2)** is zero provided

$$\Delta\omega t = \pi$$

Therefore, a "dip" will form in the time domain response when

$$t = \frac{\pi}{\Delta\omega}$$

From Figure 4.3, it can be deduced that a dip at about 2 sec would be associated with a $\pi/2$ mistuning relationship between two coupled oscillators.

4.2.3.2 Simulations

This section presents numerical simulations associated with identical and mistuned oscillators with the same physical properties as described in sections 4.2.2.1 and 4.2.2.2 respectively. The objective of these investigations is to deduce which factors promote the appearance of the double decay rate. Solving for the reduced ordinary differential equations, as described in **(4.2.1 – 11)**, one can therefore obtain the separate decay rates associated with coupled masses, m_1 and m_2, in the time domain.

Initially, the masses are excited with equal amplitudes and *in-phase* forces. These mistuned oscillators have the same physical properties as presented in section 4.2.2.2. The forcing characteristics are similar to half-sine pulses representing the excitation of the hammer acting on the strings. Figure 4.4 shows the response of the coupled masses in the case of *mistuned* oscillators on a natural log scale:

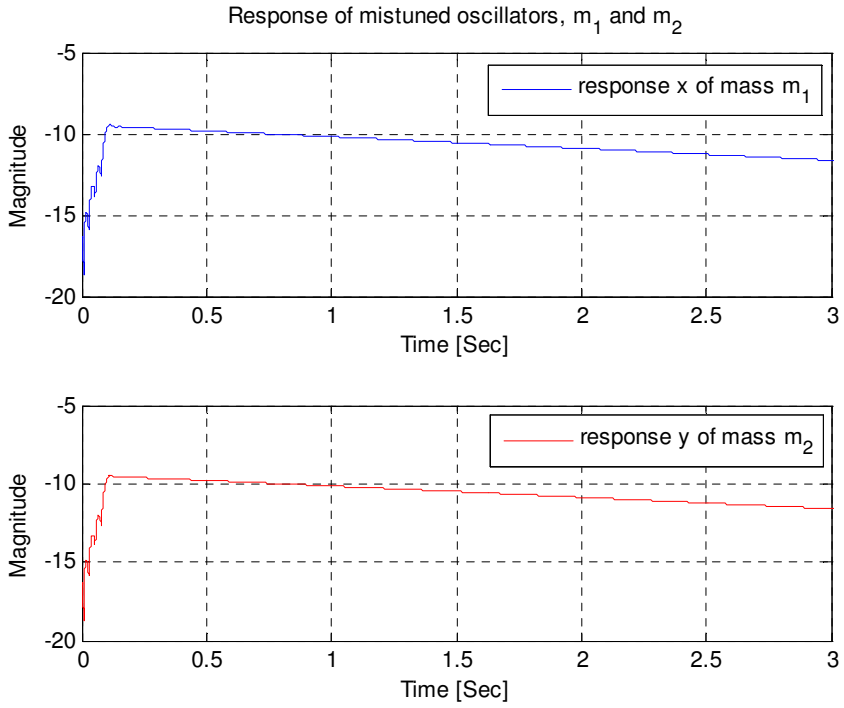


Figure 4.4 – Decay rates of mistuned coupled oscillators when both masses are excited with equal amplitude and in-phase forces

From Figure 4.4, it can be seen that there is no clearly identifiable double decay effect despite the ratio of the damping terms of the coupled modes is 20 times higher than the uncoupled one as discussed in section 4.2.2.2. However, as mentioned in section 2, an out-of-phase relationship between two coupled strings may promote the appearance of double-decay rate (Weinreich, 1977; Hundley *at al.*, 1978). In order to induce a phase difference in our simple model, the two mistuned oscillators are excited with equal amplitude, but out-of-phase forces (simplification of the true physics of the real dynamical system). Figure 4.5 is obtained:

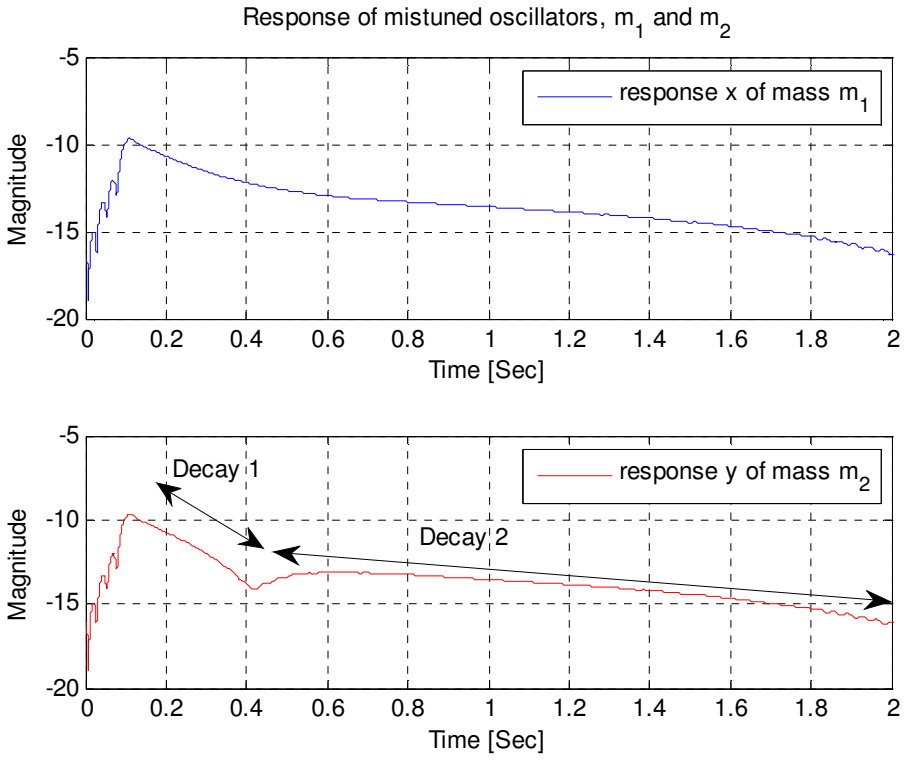


Figure 4.5 – Decay rates of mistuned coupled oscillators when both masses are excited with equal amplitude and out-of-phase forces

From Figure 4.5, a clear formation of the double decay rate effect is observed for mass m_2 (the one with the opposite force excitation applied to). The latter figure resembles the effect observed in the case of a real piano recording as shown earlier in Figure 4.3. Indeed, from Figure 4.5, it can be seen from the response of mass m_2 that there is a fast decay rate up to about 0.42 s followed by a second, much slower, decay rate lasting over a period of 1.5 s in total.

Note also that the resulting double decay rate effect will be dependant upon the "participation" of each mode, which in turn is dependent on the initial conditions and boundary conditions of the model.

Additionally, Figure 4.6 is also obtained showing the response of masses m_1 and m_2 in the case of *identical* oscillators having the same physical properties as presented in section 4.2.2.1 earlier. Note that the resulting responses of mass m_1 and m_2 are almost identical as shown in Figure 4.6:

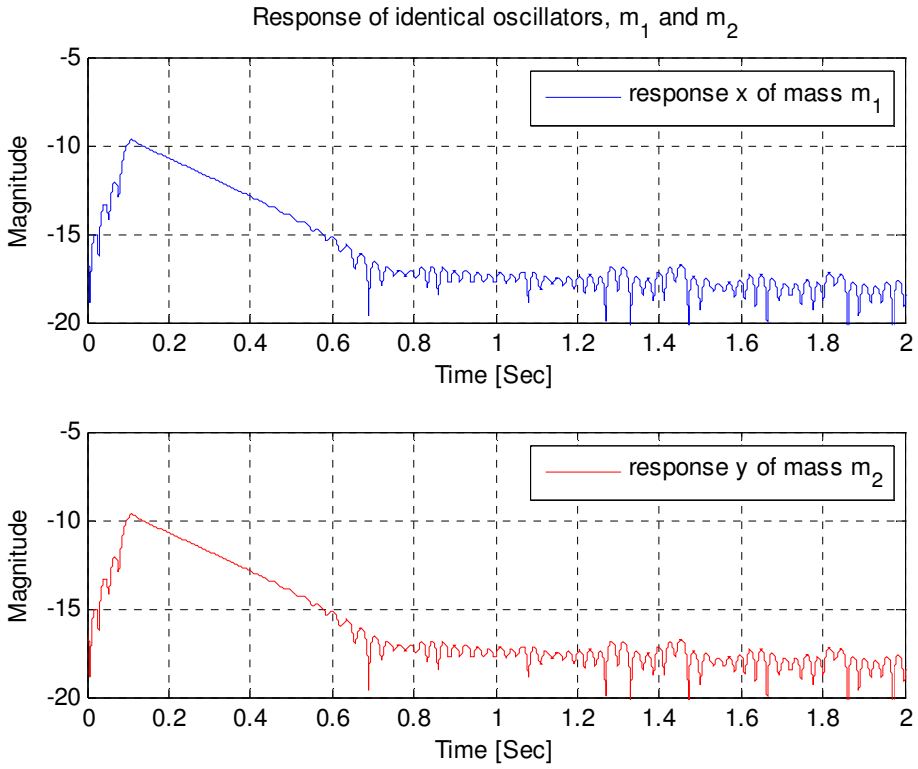


Figure 4.6 – Decay rates of identical coupled oscillators when both masses are excited with equal amplitude and out-of-phase forces

From Figure 4.6, it can be seen that despite the out-of-phase excitation forces acting on the identical oscillators, there seems to be only a single identifiable decay rate, which lasts about 0.7 s and after that the responses dip into the simulated noise floor associated with the numerical estimation process.

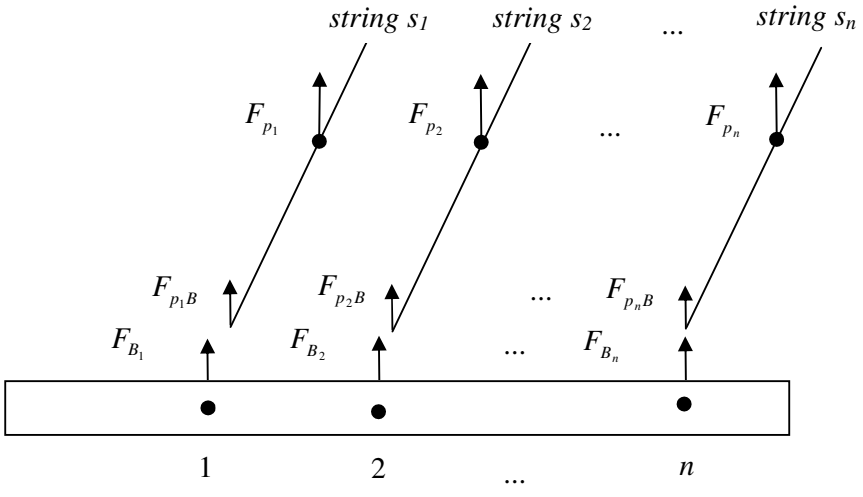
From the above investigations, it can therefore be deduced that both the phase difference between two coupled oscillators and the relative mistuning between the two can promote the clear appearance of a double decay rate effect.

4.4 Physical modelling of multiple strings

This section is used to describe a physical model representing the dynamics of *N* strings coupled with the bridge of a piano. This type of physical modelling effectively represents a further extension of the generalised model of coupled oscillators and is a frequency domain mobility-based method. The following analysis is partially based on the modelling of coupled strings connected to a vibrating body in which the body is represented by a beam clamped at both ends (Carrou *et al.*, 2004). Note that only the vertical motion of the strings and the bridge are modelled here, although the physical model can be extended for the horizontal motion too.

4.4.1 General model of *N* number of strings

A frequency domain model with *N* strings, which are attached to a pinned-pinned beam representing the bridge, is discussed using a mobility-based approach. Figure 4.7 describes such a coupled dynamical system:



and in terms of velocities:

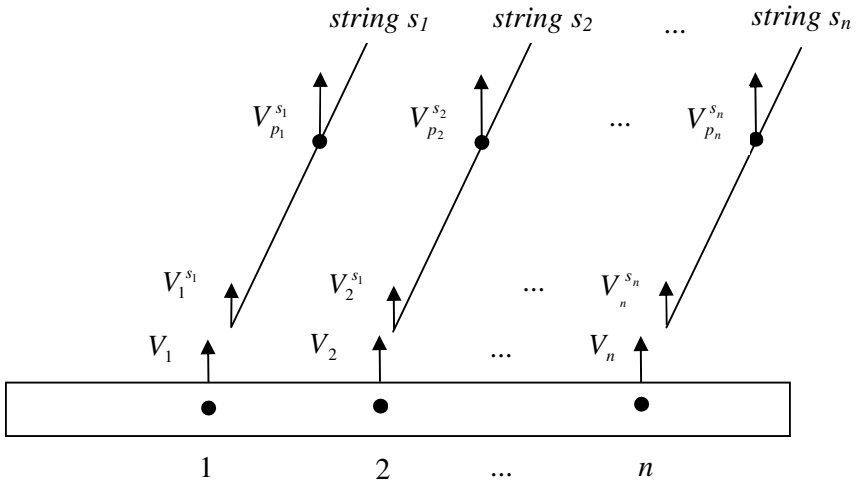


Figure 4.7 – Diagram of forces and velocities in the case of N number of vibrating piano strings when coupled with a finite length beam

F_{p_1}, F_{p_2} and F_{p_n} are the external forces applied at the strings, F_{p_1B}, F_{p_2B} and F_{p_nB} are the internal forces acting on the strings at the attachment points with the beam, whilst F_{B_1} and F_{B_2} and F_{B_n} are the internal forces acting on the beam. $V_{p_1}^{s_1}$, $V_{p_2}^{s_2}$ and $V_{p_n}^{s_n}$ are the velocities of the strings due to F_{p_1}, F_{p_2} and F_{p_n}. Finally, $V_1^{s_1}$, $V_2^{s_2}$ and $V_n^{s_3}$ are the common velocities with V_1, V_2 and V_n acting on the strings and the beam.

The general formulation of the mobility response Y (ratio of velocity over force) for *N* strings can be expressed as follows. Note that the subscripts of $Y_{x_1x_2}^{s_n}$ refer to the mobility between points x_1 and x_2 and the superscript refers to a particular string s_n of the modelled physical system

$$\begin{bmatrix} Y_{11}^{s_1}+Y_{11} & Y_{12} & \dots & \dots & Y_{1n} \\ Y_{21} & Y_{22}^{s_2}+Y_{22} & \dots & \dots & Y_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & Y_{n2} & \dots & \dots & Y_{nn}^{s_n}+Y_{nn} \end{bmatrix} \cdot \begin{bmatrix} F_{p_1B} \\ F_{p_2B} \\ \dots \\ \dots \\ F_{p_nB} \end{bmatrix} = \begin{bmatrix} -Y_{1p_1}^{s_1}F_{p_1} \\ -Y_{2p_2}^{s_2}F_{p_2} \\ \dots \\ \dots \\ -Y_{np_n}^{s_n}F_{p_n} \end{bmatrix}$$

So,

$$\begin{bmatrix} F_{p_1B} \\ F_{p_2B} \\ \cdots \\ \cdots \\ F_{p_nB} \end{bmatrix} = \begin{bmatrix} Y_{11}^{s_1} + Y_{11} & Y_{12} & \cdots & \cdots & Y_{1n} \\ Y_{21} & Y_{22}^{s_2} + Y_{22} & \cdots & \cdots & Y_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Y_{n1} & Y_{n2} & \cdots & \cdots & Y_{nn}^{s_n} + Y_{nn} \end{bmatrix}^{-1} \cdot \begin{bmatrix} -Y_{1p_1}^{s_1} F_{p_1} \\ -Y_{2p_2}^{s_2} F_{p_2} \\ \cdots \\ \cdots \\ -Y_{np_n}^{s_n} F_{p_n} \end{bmatrix}$$

One can therefore obtain the velocity response of any string s_n at any point *x* along its length according to

$$V_x^{s_n} = F_{p_nB} Y_{nx}^{s_n} + F_{p_n} Y_{p_nx}^{s_n} \qquad \textbf{(4.4.1 – 1)}$$

The above generalisation is important since the piano consists in excess of 200 strings mainly grouped in pairs or triplet of strings for the 88 piano notes. The following section investigates the response of a pair of coupled strings in the frequency domain.

4.4.2 Response of a pair of strings

4.4.2.1 The model

Normally, in the second octave from the lower octave register of the piano and above, pairs of strings are used for the notes, whilst for higher octaves, triplets of strings are used to make up the piano notes.

From equation **(4.4.1 – 1)** and in the case of a pair of strings, one can obtain the response of strings s_1 and s_2 at any point *x* along their length as follows

$$V_x^{s_1} = F_{A_1B} Y_{1x}^{s_1} + F_{A_1} Y_{A_1x}^{s_1} \qquad \textbf{(4.4.2.1 – 1)}$$

$$V_x^{s_2} = F_{A_2B} Y_{2x}^{s_2} + F_{A_2} Y_{A_2x}^{s_2} \qquad \textbf{(4.4.2.1 – 2)}$$

where F_{A_1} and F_{A_2} can take any amplitude and phase form, for instance unity or zero and are applied at points $p_1 = A_1$ and $p_2 = A_2$ of the two strings.

$V_x^{s_1}$ and $V_x^{s_2}$ should also exhibit coupled resonances due to both strings and also due to the beam dynamics.

Note that an alternative modelling approach is presented in Appendix D by working out the boundary conditions for the respective equations of motion of the beam and the two strings.

In the simple case of a single string s attached to the beam, one needs to calculate the following point and/or transfer mobilities

$$Y_{x_1x_2}^{s} = j\omega\sum_{n=1}^{\infty}\frac{\varphi_{ns}(y_i)\varphi_{ns}(y_j)}{\rho_s S_s l_s\left(\omega_{ns}^2(1+j\eta_s)-\omega^2\right)} \qquad \textbf{(4.4.2.1 – 3)}$$

where ρ_s, S_s, l_s and η_s denote the density, the cross-sectional surface area, the length and the loss factor of the material of the string respectively. n denotes the mode shape number of mode $\varphi_{ns}(y)$ of the string and the modes are given (Gardonio and Brennan, 2004)

$$\varphi_{ns}(y) = \sqrt{2}\sin k_{ns}y \text{, where } n = 1, 2\ldots \text{ and } k_{ns} = \frac{n\pi}{l_s}$$

The natural frequencies of string s (assuming no bending stiffness or inharmonicity factor B) can also be found from

$$\omega_{ns} = \frac{n\pi c_b}{l_s}$$

where $c_b = \sqrt{\frac{T}{\rho}}$, ρ is the mass per unit length of the string and T is the tension of the string.

Note also that principle of reciprocity implies that $Y_{x_1x_2}^{s} = Y_{x_2x_1}^{s}$.

4.4.2.2 Numerical simulations

The dynamical behaviour of two strings attached to a finite length beam can be simulated numerically using equations **(4.4.2.1 – 1)** to **(4.4.2.1 – 3)**. Suppose that the strings are slightly mistuned by about 2 percent (= 0.2 Hz) as shown before in section

4.2.2.2 but their natural frequencies are set higher, i.e. their natural frequencies are 224.4 Hz and 224.6 Hz, as in the case of a real piano, and suppose that only string s_1 is excited by a unit input force. Figure 4.8 below shows the resonance frequencies of the coupled dynamical system for the forced response including that of the finite length beam at 726.1 Hz.

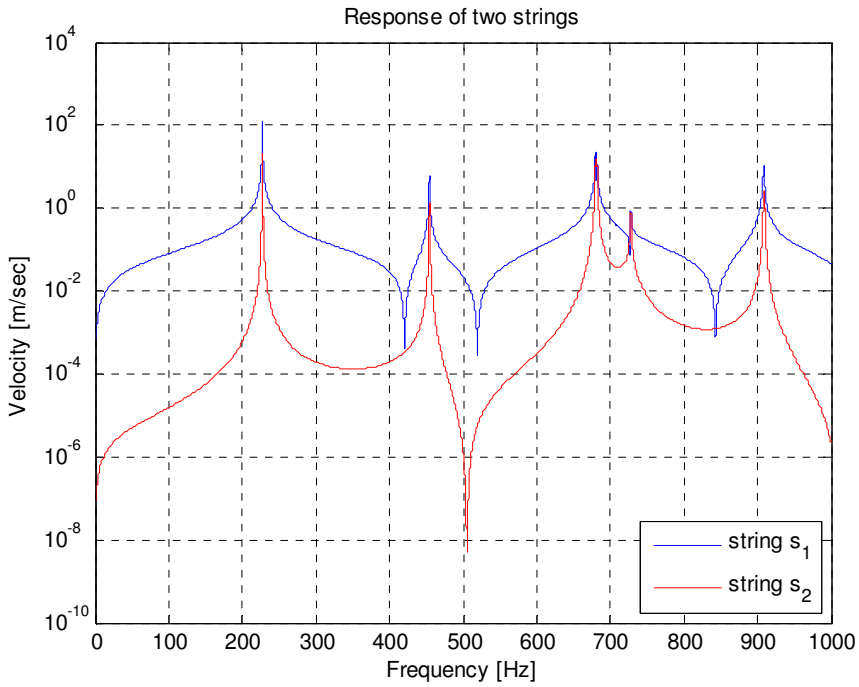


Figure 4.8 – Frequency response of the coupled dynamic system at application point A_1 of force F_{A_1} on string s_1 and response of string s_2 at point A_2

Consider now the power input into the two strings and the beam by exciting string s_1 only. The time-average power input into strings s_1 and s_2 are respectively

$$P_{s_1} = \frac{1}{2}\mathrm{Re}\left\{F_{A_1} V_{A_1}^{s_1 *}\right\} \text{ and } P_{s_2} = \frac{1}{2}\mathrm{Re}\left\{F_{A_2 B} V_2^{s_2 *}\right\} \qquad \textbf{(4.4.2.2 – 1)}$$

where $V_{A_1}^{s_1 *}$ is the complex conjugate of velocity of string s_1 at point A_1 of the applied force F_{A_1}, and $V_2^{s_2 *}$ is the complex conjugate velocity of string s_2 at the attachment point with the beam when $F_{A_2} = 0$

$$V_{A_1}^{s_1 *} = F_{A_1 B}^{*} Y_{1A_1}^{s_1 *} + F_{A_1}^{*} Y_{A_1 A_1}^{s_1 *} \text{ and } V_2^{s_2 *} = F_{A_2 B}^{*} Y_{22}^{s_2 *}$$

The time-average power into the beam is

$$P_{beam} = \frac{1}{2}\mathrm{Re}\left\{F_{B_1}V_1^* + F_{B_2}V_2^*\right\} \qquad \textbf{(4.4.2.2 – 2)}$$

where V_1^* and V_2^* are the complex conjugate velocities of the beam at the attachment points with strings s_1 and s_2

$$\begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = \begin{bmatrix} F_{B_1}^* \\ F_{B_2}^* \end{bmatrix}\begin{bmatrix} Y_{11}^* & Y_{12}^* \\ Y_{21}^* & Y_{22}^* \end{bmatrix}$$

Using equations **(4.4.2.2 – 1)** and **(4.4.2.2 – 2)**, Figure 4.9 can be obtained that shows the power into the driven string s_1 and also the power transferred into the beam and string s_2:

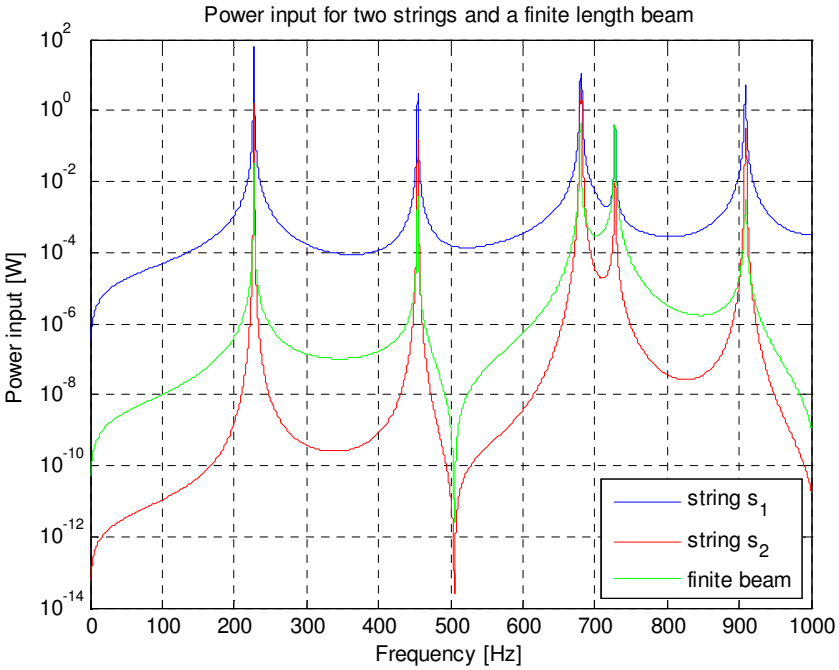


Figure 4.9 - Power input for two coupled strings and a finite length beam

From Figure 4.9, it can be seen that most of the power input dissipates through the material damping of the string and relatively less will be transmitted through the coupling to the beam and the other string. However, the above model does not account for acoustic radiation losses. The acoustic radiation will primarily be related to the soundboard, which is discussed in the next section.

4.5 Experiments on the coupling of strings with the soundboard

So far, the importance of mistuning between strings of the same piano note and the double decay rate effect has been investigated. In this section, vibration measurements on a real grand piano are carried out in order to further investigate the coupling of the piano string with the bridge and the soundboard in the vertical and horizontal directions through the measurement of transfer mobilities. Also, in this section, one attempts to approximately correlate measured point mobilities of the ribbed soundboard and bridge with known theoretical formulations for such built up structures.

4.5.1 Transfer mobilities

Measurements were carried out at various locations on the bridge and the soundboard of a 'Kawai' baby grand piano using a PCB instrumented hammer, Type 086C03, and a PCB accelerometer connected to a high-end data acquisition system (B&K, Type B-frame) using the B&K PULSE software. Transfer mobilities between the bridge and the soundboard were measured to investigate the energy transfer between a vibrating piano string and a soundboard in the horizontal and vertical directions.

The excitation at the bridge represented the direct excitation of a piano string acting on the bridge and the response was initially measured vertically on the soundboard with the accelerometer. However, since it was not possible to measure the response of the soundboard in the horizontal direction due to space constraints, the excitation was provided horizontally by the hammer and the response was again measured vertically at the same point on the soundboard. The methods of excitation and response measurement are illustrated in Figure 4.10:

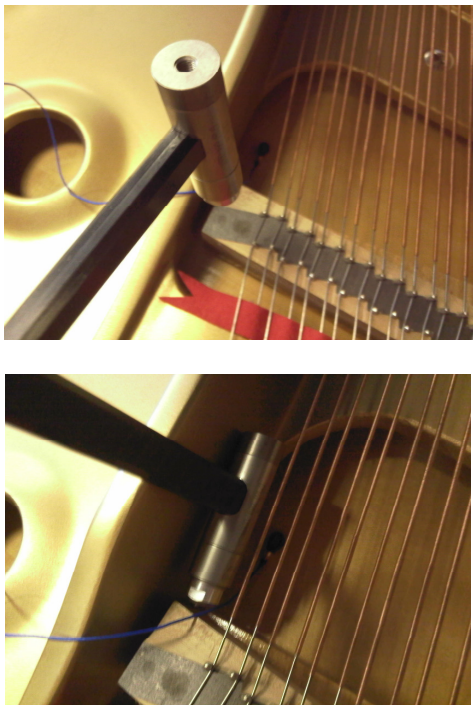


Figure 4.10 – Transfer function measurements between bridge and soundboard for vertical and horizontal hammer excitations. Response is always measured vertically on the soundboard with an accelerometer.

Figure 4.11 below shows a typical transfer function measurement between the bridge and the soundboard for the horizontal and vertical hammer excitations:

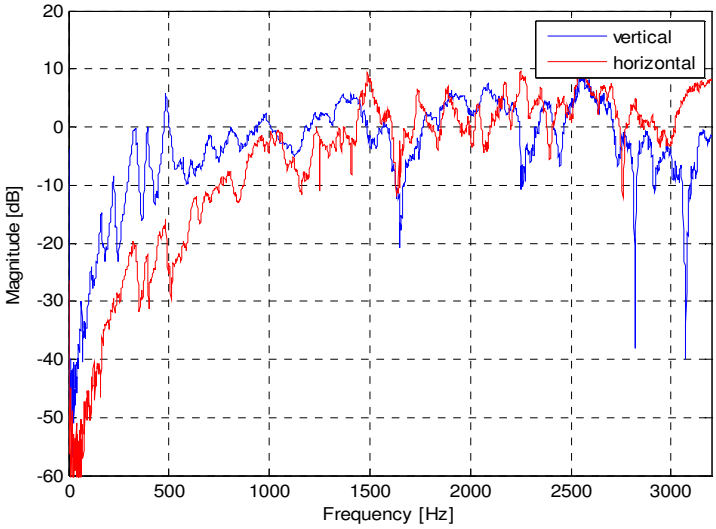


Figure 4.11 – Transfer function measurement for the vertical and horizontal hammer excitations. Response is always measured vertically on the soundboard with an accelerometer.

From Figure 4.11, it can be seen that the measured vertical response on the soundboard, due to a horizontal excitation, can be up to 20 dB lower than its response due to a vertical excitation for frequencies up to about 1,000 Hz. This interesting finding suggests that an excitation in one direction might induce a small response in the other direction too. As a result, one may suggest that the response of a piano string in one direction will be cross-coupled with response of the soundboard in the other direction.

4.5.2 Point mobilities

Point mobilities at the soundboard and bridge were also carried out in the vertical direction. A typical such measurement on the bridge is shown in Figure 4.12:

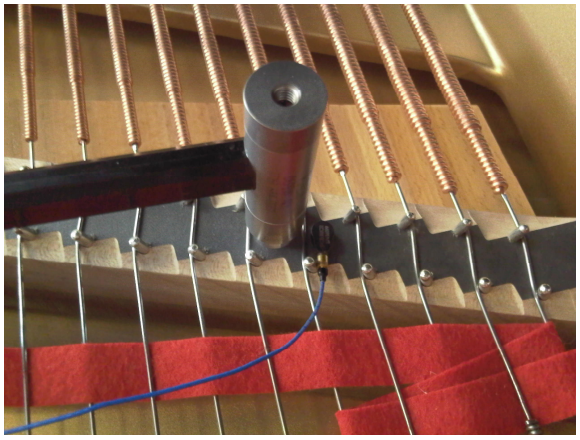


Figure 4.12 – Typical point mobility measurement at the bridge of the piano

Figures 4.13 and 4.14 below show typical point mobility measurements at the bridge and the soundboard respectively:

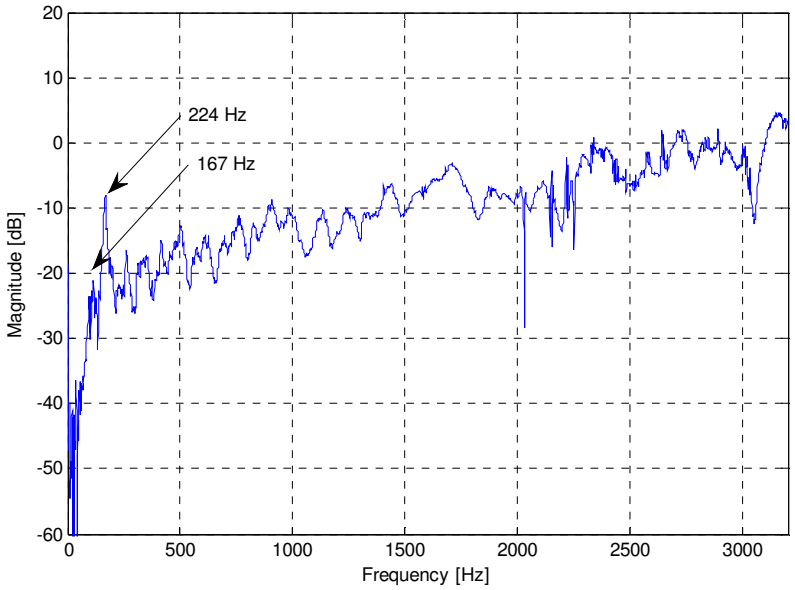


Figure 4.13 – Typical measured point mobility at the bridge close to piano note D_3. First two resonances at 167 Hz and 224 Hz in the dynamical system are indicated.

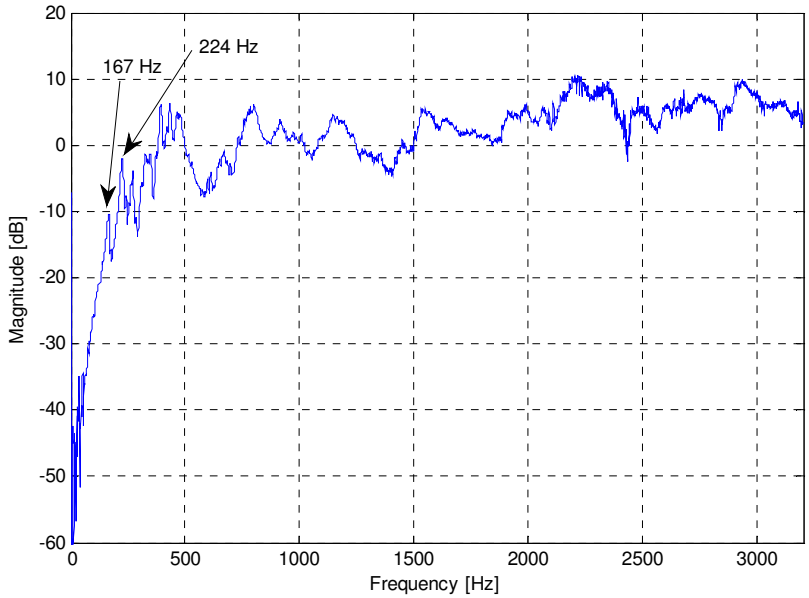


Figure 4.14 – Typical measured point mobility on the soundboard. First two resonances at 167 Hz and 224 Hz in the dynamical system are indicated.

From Figures 4.13 and 4.14, it can be seen that the first resonance frequency, measured either at the bridge or the soundboard, is centered at 167 Hz and it is in this view that this resonance relates to the first resonance frequency of the soundboard.

Note that Giordano (1997 and 1998) previously measured the first resonance frequency at 100 Hz for a larger grand piano, therefore the higher resonance frequency is a consequence of the smaller size of the piano. Also, the fact that the first resonance associated with the soundboard can be measured directly both on the bridge and the soundboard, it implies that the bridge and the soundboard are strongly coupled.

In order to understand further the response of such a complicated coupled dynamical system, one can attempt to correlate the measured point mobilities of the ribbed soundboard and bridge with known theoretical formulations for such built up structures. This is undertaken in the following section.

4.5.3 Theoretical considerations

The power P that enters a homogeneous dynamical system, such as a plate or a beam of mass m, which is excited by a point force with frequency band $\Delta\omega$ is proportional to the modal density $n_d(\omega)$ and the mean-squared force F^2 in that frequency band (ESDU 04010, 2004)

$$P = F^2(\pi/2m)n_d(\omega)$$

$$P = F^2 \mathrm{Re}(Y_\infty)$$

where $\mathrm{Re}(Y_\infty)$ is the real part of the point mobility of an infinite system averaged over the frequency band and space and is independent of the boundary conditions. Note that for a homogeneous structure, such as a finite isotropic plate, the point mobility can be approximated by that of the infinite structure.

In the case of a real piano, the mechanism of sound radiation is related to the dynamic behaviour of the bridge coupled with the ribbed soundboard.

The theoretical point mobility of an infinite beam for flexural wave motion under a force excitation is

$$Y_{\infty} = \frac{(1-j)k_b}{4\rho S\omega} \qquad \textbf{(4.5.3 – 1)}$$

where the complex wavenumber $k_b = \omega^{0.5}\left(\rho S/EI\right)^{0.25}$, and ρ, S, E and I are the density, cross-sectional area, Young's modulus of elasticity and second moment of area of the structure, respectively.

On the other hand, for longitudinal wave motion, the point mobility of an infinite beam is constant and independent of frequency

$$Y_{\infty} = \frac{1}{2S\sqrt{E\rho}} \qquad \textbf{(4.5.3 – 2)}$$

Similarly, the point mobility of an infinite plate due to flexural excitation by a point force is independent of frequency and is of the form

$$Y_{\infty} = \frac{1}{8\sqrt{D\rho h}} \qquad \textbf{(4.5.3 – 3)}$$

where h is the thickness of the plate, $D = \frac{Eh^3}{12(1-\nu^2)}$ and ν is the Poisson's ratio.

However, in the case of the piano, motion will be complicated due to the coupling of the glued bridge on the soundboard and its ribs. An approximate response, divided into two "regions" can be obtained by considering the effect of a "grillage of beams" on a plate as described by Pinnington (1988). This is demonstrated in the measured data in Figure 4.15, where the fluctuating measured pointed mobility on the soundboard, as shown in Figure 4.14, may be approximated with these two regions (regions 1 and 2):

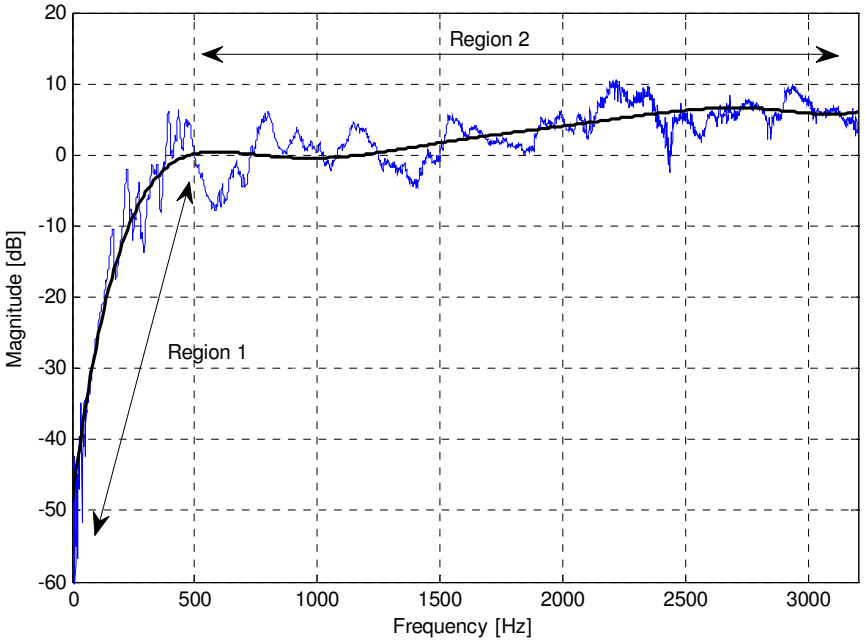


Figure 4.15 – Typical measured point mobility on the soundboard with two approximate regions (fitted regions of Figure 4.14)

In the case of the piano, the beams are the stiffeners of the soundboard and for simplicity one may also assume that the soundboard and the glued bridge are acting as one system instead of being treated separately. According to the theory of such built up structures, at low frequencies, up to roughly the first flexural resonance of the soundboard, the system will act as a spring of stiffness k, so the point mobility will sharply increase over that frequency range, as seen in Region 1 of Figure 4.15. However, for higher frequencies, the point mobility will reach roughly a constant value (Region 2) and may be approximated by that of an infinite plate as described earlier in equation **(4.5.3 – 3)**. In the fact, the response of the soundboard at high frequencies will be broad, as discussed in chapter 3 and by other authors (Giordano 1997 and 1998), resulting in a multi-modal, Statistical Energy Analysis (SEA) type response. This further justifies that the cepstrum of the response of the soundboard would appear in the low-time regions, as originally speculated in chapter 3 (Karatsovis *et al.*, 2006).

4.6 Conclusions

This chapter presents a simple analytical model for describing the effect of the double decay rate as can be observed in a real piano recording. The model was developed upon the response of two masses connected through linear coupling. In this model, each mass can either represent the vertical (or horizontal) motion of one string coupled with the response of a second string through the bridge, or the modelled masses can represent the coupled vertical and horizontal motions of one string only.

It was shown that the formulation of the general model of coupled oscillators is identical to the formulation of digital waveguides as shown in chapter 3. Also, from numerical simulations in the time domain, it was possible to clearly replicate the double decay rate effect in the case of mistuned oscillators, when there is an out-of-phase relationship between the two. From these investigations, one deduced that the phase difference between the two oscillators and the relative mistuning between the two can promote the appearance of a double decay rate effect.

A physical model was also developed to describe the dynamics of multiple strings as an extension to the general model of coupled oscillators. The example of two coupled strings, slightly mistuned, of a piano note and coupled with a finite length beam was also illustrated. Furthermore, a power flow analysis was carried out by exciting one string and evaluating the power that goes into the other string and the beam. From this investigation, it was shown that most of the power input dissipates through the material damping of the string and subsequently less will be transmitted through the coupling to the beam and the other string.

Measurements were also carried out on a real baby grand piano in order to further understand the coupling mechanism between the soundboard, the bridge and the string. Transfer mobility measurements were carried out in order to deduce how the energy is transferred in the vertical and horizontal directions. It was found that an excitation in one direction might also induce a small response in the other direction. As a result, one may suggest that the response of a piano string in one direction will be cross-coupled with response of the soundboard in the other direction.

Finally, point mobility measurements were carried out on the bridge and the soundboard in order to deduce the response of such a complicated dynamical system. From these investigations, it was found that the first resonance of the soundboard can be measured both on the bridge and the soundboard since they are coupled. These dynamics of such a system were further explained through the use of theoretical mobilities. Indeed, such a dynamical system may be approximated with the response of a grillage of beams mounted on plate. Two important frequency regions exist; for low frequencies, the system will act as a spring of a given stiffness and for higher frequencies, the point mobility will reach a constant average value approximating the point mobility of an infinite plate. At high frequencies, a multi-modal behaviour of the soundboard will result in a broad, SEA-type behaviour further justifying our original hypothesis that the soundboard appears in the low-time regions of the cepstrum as outlined in chapter 3.

Chapter 5

Proposed signal model in a Bayesian formalism

5.1 Introduction

The main themes of this research have been the analytical formulation and computational demonstrations of two of the most important features that characterise the piano sound; the inharmonicity and the double decay rate effect.

This chapter initially introduces a very basic form of modelling a piano note, as part of an early prototype model, which is described in section 5.2. The remainder of this chapter presents the pitch estimation of simulated and real monophonic and polyphonic piano note recordings using a more comprehensive model and a more elaborate algorithm implementation, which could ultimately be used in future music transcription methods.

In the past, most efforts in the development of music transcription methods for monophonic and polyphonic music performances had been focused on the analysis and interpretation of recordings without necessarily considering the complex nature of the modelled sound. Such methods tend to employ a simplistic representation of the signal. On the other hand, in this research, a potentially more detailed signal model representation of the piano sound, in particular, is proposed, so that it may ultimately form the basis of an improved transcription method in the future.

In this chapter, a signal model, based mainly on the properties of the inharmonicity effect is proposed, which can be used in a probabilistic Bayesian formalism framework, similar to the one adopted earlier by Godsill and Davy (2002). A monophonic model representing single notes and a polyphonic signal model representing a multitude of notes, such as music intervals, are proposed. The advantage of the proposed models is that their dimensionality may be reduced when compared to the existing models by Godsill and Davy, hence potentially rendering them less computationally expensive.

The potential of the proposed models are illustrated with the transcription of simple examples of real monophonic and polyphonic piano recordings by implementing the M-H algorithm and Gibbs sampler for directly estimating the fundamental frequency and inharmonicity factor of each individual piano note.

The proposed signal models and recordings used are deliberately chosen to be limited (relative to a full music piece) to gain confidence in the methodology. A full transcription method is not within this scope.

5.2 Early prototype model

In this section, a model of the sound of a piano note is presented which does not include the effect of the inharmonicity or the double decay rate as discussed in chapters 3 and 4, respectively. The harmonic components are assumed to be integer multiples of the fundamental frequency and damping is not considered in this model.

The use of the M-H algorithm in conjunction with a Bayesian formalism framework can be used as a method to estimate the frequency of components in signals with random Gaussian noise. In later sections of this chapter, the implementation of the M-H algorithm for multi-variate parameter estimation, e.g. frequency and inharmonicity, will be discussed in more depth and detail.

The M-H algorithm has its origins back in the '50s and was proposed by Metropolis *et al.* (1953) and developed subsequently by Alder and Wainwright (1959) in the field of molecular dynamics, where the energy of a molecule may be predicted based *only* on the last collision by forming of what is called a "Markov" chain. The M-H algorithm can be used to model the successive states of a Markov chain, where the final state of the chain is used to estimate the parameter space, e.g. the energy of a molecule in a gas. In this research, the M-H algorithm may be used to obtain the successive states of a parameter space related to the signal model of an automatic music transcription model, namely the fundamental frequency associated with a piano note.

A piano note can be described in its very basic form as a simple superposition of the fundamental frequency and its integer multiple harmonic terms

$$x(t) = \sum_{n=1}^{N} \alpha_n \sin\left[n 2\pi f_0 t + \phi_n\right] \qquad \textbf{(5.2 – 1)}$$

where N is the number of modelled terms (harmonics and fundamental frequency) and the phase term ϕ_n in equation **(5.2 – 1)** can be implemented by including a cosine component

$$x(t) = \sum_{n=1}^{N} \alpha_n \sin\left[n 2\pi f_0 t\right] + \beta_n \cos\left[n 2\pi f_0 t\right] \qquad \textbf{(5.2 – 2)}$$

where α_n and β_n are the amplitude terms of the sine and cosine components.

An explanation of Bayesian formalism is presented in Appendix A.

Therefore, equation **(5.2 – 2)** in terms of the generalised linear model can be written as

$$\mathbf{d} = \mathbf{G}\mathbf{b} + \mathbf{e}$$

where **d** is an $M \times 1$ data matrix (M is the data length of the signal representing only one single frame of the recording), **b** is an $L \times 1$ (or equivalent to $2N \times 1$) matrix containing the amplitudes of the sine and cosine components of the estimated frequency, and **e** is an $M \times 1$ matrix containing random Gaussian noise entries. Finally, **G** is an $M \times L$ (or equivalent to $M \times 2N$) matrix of the basis functions defined by the parameters of a piano note.

Equation **(A – 9)** is recalled, which describes the probability density of f_0 based on a Bayesian description of the signal (Ó Ruanaidh and Fitzgerald 1996)

$$p(\omega|\mathbf{d}) \propto \frac{\left[\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{d}^{\mathrm{T}}\mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{d}\right]^{\frac{L-M}{2}}}{\sqrt{\det(\mathbf{G}^{\mathrm{T}}\mathbf{G})}} \qquad \textbf{(5.2 – 3)}$$

Expression **(5.2 – 3)** describes a t-distribution probability function and will peak at the most probable value of f_0 of the **G** matrix.

The M-H algorithm can be used to produce random sequences of samples from given densities in order to obtain the parameters of a signal. In particular, suppose that X_i is the i^{th} element of a random walk and that the next variate Y_i in the random sequence is produced by simply adding a random perturbation factor ξ_i to X_i of mean zero and standard deviation one

$$Y_i = X_i + \xi_i \qquad \textbf{(5.2 – 4)}$$

In a very basic algorithm implementation, expression **(5.2 – 3)** can be used to draw probability densities for steps of X_i and Y_i representing single frequency estimates for a model with one term N only (or equivalent to $L = 2$)

$$p(X_i) \propto \frac{\left[\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{d}^{\mathrm{T}}\mathbf{G}_{\mathbf{X}_i}(\mathbf{G}_{\mathbf{X}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{X}_i})^{-1}\mathbf{G}_{\mathbf{X}_i}^{\mathrm{T}}\mathbf{d}\right]^{\frac{2-M}{2}}}{\sqrt{\det(\mathbf{G}_{\mathbf{X}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{X}_i})}}, \quad p(Y_i) \propto \frac{\left[\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{d}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i}(\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i})^{-1}\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{d}\right]^{\frac{2-M}{2}}}{\sqrt{\det(\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i})}} \quad \textbf{(5.2 – 5)}$$

where $\mathbf{G}_{\mathbf{X}_i}$ and $\mathbf{G}_{\mathbf{Y}_i}$

$$\mathbf{G}_{\mathbf{X}_i} = \begin{bmatrix} \sin(2\pi X_i t_1) & \cos(2\pi X_i t_1) \\ . & . \\ . & . \\ . & . \\ \sin(2\pi X_i t_M) & \cos(2\pi X_i t_M) \end{bmatrix}, \quad \mathbf{G}_{\mathbf{Y}_i} = \begin{bmatrix} \sin(2\pi Y_i t_1) & \cos(2\pi Y_i t_1) \\ . & . \\ . & . \\ . & . \\ \sin(2\pi Y_i t_M) & \cos(2\pi Y_i t_M) \end{bmatrix}$$

The condition on which Y_i is accepted instead of X_i is defined by the M-H acceptance function Q

$$Q(X_i, Y_i) = \frac{p(Y_i)}{p(X_i)} \quad \textbf{(5.2 – 6)}$$

The magnitude of equation **(5.2 – 6)** is very small, e.g. of the order of over 10^{-300} (smallest number that can be calculated in MATLAB software package), which can cause numerical underflow problems. To avoid this, equation **(5.2 – 6)** is re-written in terms of logarithms, then set of equations **(5.2 – 5)** are combined, and finally the exponential value of the acceptance function is calculated

$$Q(X_i, Y_i) = \exp\left[\frac{(2-M)}{2}\ln[\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{d}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i}(\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i})^{-1}\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{d}] - \ln[\sqrt{\det(\mathbf{G}_{\mathbf{Y}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{Y}_i})}] - \frac{(2-M)}{2}\ln[\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{d}^{\mathrm{T}}\mathbf{G}_{\mathbf{x}_i}(\mathbf{G}_{\mathbf{x}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{x}_i})^{-1}\mathbf{G}_{\mathbf{x}_i}^{\mathrm{T}}\mathbf{d}] - \ln[\sqrt{\det(\mathbf{G}_{\mathbf{x}_i}^{\mathrm{T}}\mathbf{G}_{\mathbf{x}_i})}]\right]$$

Let us now assume ε is a uniform variable drawn over the range [0,1]. If ε<Q, then $X_{i+1} = Y_i$. Otherwise, if ε>Q, then $X_{i+1} = X_i$.

The process is iterative and a sufficiently large number of iterations is required to obtain accurate estimates of the frequency of a signal component.

In the following example, a single piano note C_4[2] is transcribed, where an estimate of the fundamental frequency and the first two harmonics is obtained sequentially. The

[2] The piano is an instrument that covers a large frequency range, typically of more than seven musical octaves, i.e. from A_0 to C_8. Therefore, the subscript denotes the note of a particular octave. Figure B.1 of Appendix B also shows typical fundamental frequencies associated with different notes of a piano.

upper frame of Figure 5.1 shows the time history of the recorded note, whilst the lower frame shows the evolution of the parameter estimates (frequencies of the harmonics and the fundamental frequency) against iteration number. The real piano note was recorded at 44.1 kHz sampling frequency using the "Gigastudio" sample library:

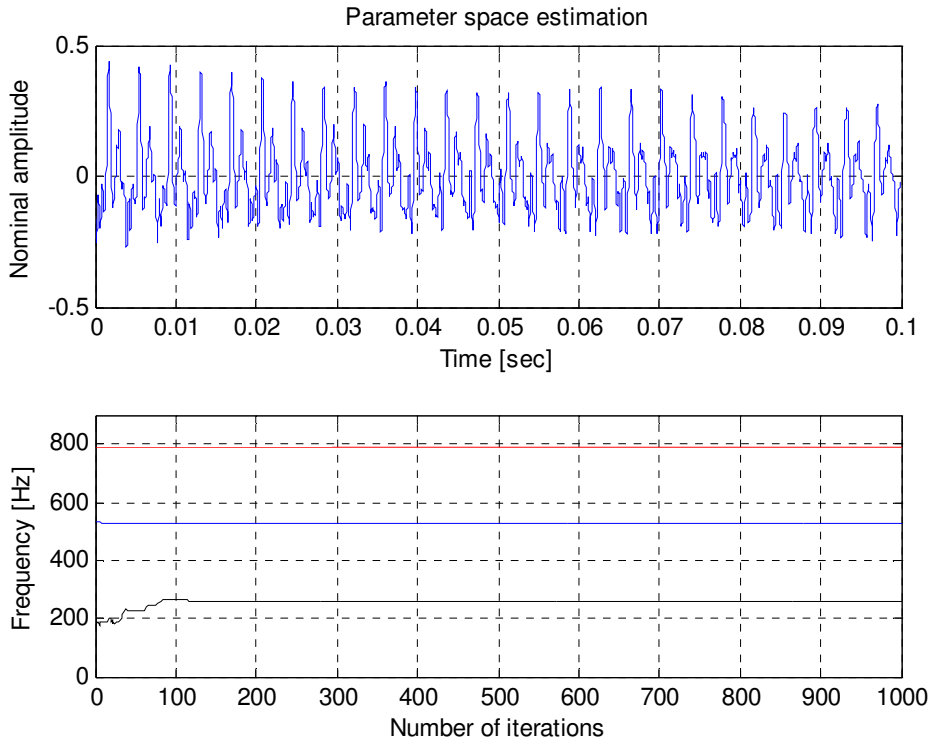


Figure 5.1 – Parameter space estimation for a real piano note recording C_4

As it can be seen from Figure 5.1, the initial frequency estimate was set around 200 Hz and 1000 iterations were employed in order to deduce at which point the M-H algorithm reaches a constant state. Note that the algorithm in this very basic form requires a good initial estimate. The implementation of a more elaborate algorithm with an arbitrary initial value is discussed later in section 5.5.1.

The fundamental, first and second harmonic frequencies of the recorded piano note were estimated as 262 Hz, 527 Hz and 790 Hz respectively.

These estimates were obtained sequentially, i.e. 1000 iterations were employed for the fundamental frequency and 1000 iterations more for each harmonic in turn. In particular, once the first frequency is estimated, then the second frequency estimate is

initialised at twice the first estimate and a further 1000 iterations are used in the estimation process. Finally, the third frequency estimate is initialised at three times the first estimate and a 1000 more iterations are carried out. Therefore, each time, this simple process involves the estimation of a single frequency component for a model with one term only, i.e. for $N = 1$, as discussed before.

A polyphonic example is now considered, e.g. an octave interval of concurrently played notes C_3 and C_4. Note that the octave interval was analysed with the same algorithm as the one used for the single piano note C_4. Figure 5.2 shows the time series data and results of the analysis:

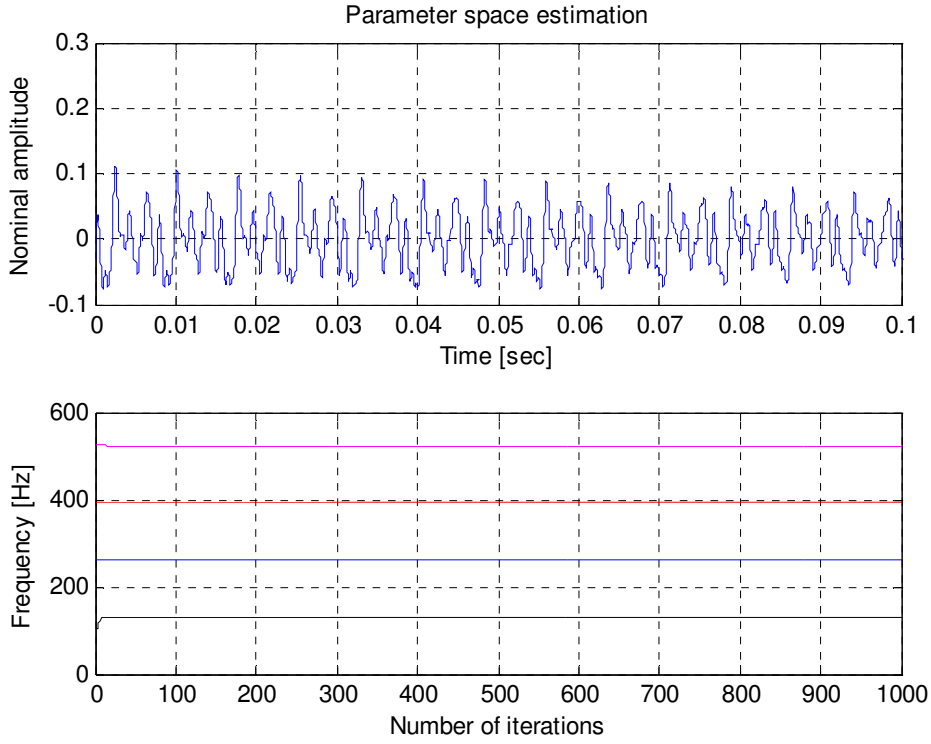


Figure 5.2 – Parameter space estimation for a real recorded octave interval C_3 - C_4

From Figure 5.2, the frequency estimates were sequentially computed as 131 Hz, 261 Hz, 392 Hz and 524 Hz, which could either be harmonics of note C_3 or of note C_4. As a result, one needs to introduce a method for differentiating between multiple notes with commonly shared harmonics. In the next sections of this chapter, the inharmonicity factor and the double decay rate are used to describe monophonic and polyphonic piano note signals. The separation of two different notes comprising a

polyphonic music example can be assisted through the introduction of a different inharmonicity factor for each single piano note. In this way, the overlap between harmonics of different notes is reduced, easing the resolution of problems associated with transcribing music intervals.

5.3 Proposed model parameters

In chapters 3 and 4, the inharmonicity and double decay rate effect of piano notes were discussed, respectively, in order to further unravel the importance of these features mainly from a physical modelling point of view. These features might also form an important part of the auditory perception mechanism in terms of how piano sounds are perceived by humans and what information is extracted (or used) by the listener.

Therefore, the parameters that are used to describe the double decay rate and inharmonicity are incorporated into the proposed signal models of monophonic and polyphonic music. These models form the basis of the Bayesian framework.

5.3.1 Inharmonicity

The inharmonicity effect, as discussed in chapter 3 through analytical and numerical formulations, is the basis of our proposed signal model. As mentioned in section 2.4.2, the response $x(t)$ of a vibrating piano string is a superposition of the fundamental and non-integer harmonic frequencies due to the inharmonicity B found in the piano strings. Equation **(2.4.2 – 7)** is recalled

$$f_n = nf_0\sqrt{1+n^2B}$$

The following conjecture is considered: supposing that each note can be represented by a fundamental frequency f_0, which is associated with the "pitch" of a note and the inharmonicity factor B, then the extraction of these two parameters may help in uniquely characterising either single or multiple notes in a recording. The above assumption forms the core of our proposed signal transcription model.

5.3.2 Double decay rate

The characteristics of the double decay rate were discussed in chapter 4 through analytical and numerical demonstrations using the concept of coupled oscillators.

Since in a music passage the played notes might be either short or long in duration, the following conjecture is considered: the damping of the frequency component amplitudes of a note may represent either the slow decay rate for played notes lasting long in a music performance, or the fast decay rate in the case of a fast music passage. The reason being that fast music passages might only contain the beginning of the played music corresponding to the first (or fast) decay of the note, where the piano hammer has just set the piano string to vibrate, but quickly stopped by the damper mechanism of the piano. On the other hand, longer lasting music passages may be dominated by the second (or slower) decay rate, where the damper mechanism has not yet stopped the vibrating piano string by allowing the sound to develop further the characteristic effect of "aftersound" associated with the second decay rate.

The proposed model may include both possibilities: played notes may last over a long and/or a short period of time. Note that for the purpose of our numerical simulations and the transcription examples of single and multiple notes, the damping factors ζ_n are set to zero. The reason for setting the damping factors ζ_n to zero is for simplicity, but also mainly due to the fact that the data is analysed over a small time window of the order of 100 ms during which the decay of the amplitude is expected to be very small.

Although not shown, the damping factors ζ_n may, alternatively, be treated as nuisance parameters in the Bayesian model and then integrated out.

The above model parameter assumptions lead naturally to the next section, where the inharmonicity B and the double decay rate, through the damping factors ζ_n of the different frequency components, are incorporated into the proposed model based on Bayesian formalism.

5.4 Proposed signal models

5.4.1 Monophonic case

The proposed signal model in the case of single piano notes is discussed in this section.

Equation **(3.2.1 – 2)** is recalled describing the inharmonicity effect in a signal and an additional damping term ζ_n is included. According to our previous conjecture, ζ_n may represent the slow and/or fast decay rate of each modelled frequency

$$x(t) = \sum_{n=1}^{N} \alpha_n e^{-\zeta_n t} \sin\left[n2\pi f_0 \left(1 + n^2 B\right)^{0.5} t + \phi_n\right] \qquad \textbf{(5.4.1 – 1)}$$

The phase term ϕ_n in equation **(5.4.1 – 1)** can be implemented by including a cosine component

$$x(t) = \sum_{n=1}^{N} \alpha_n e^{-\zeta_n t} \sin\left[n2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right] + \beta_n e^{-\zeta_n t} \cos\left[n2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right] \textbf{(5.4.1 – 2)}$$

where α_n and β_n are the amplitude terms of the sine and cosine components, and ζ_n is the decay rate of each harmonic term.

Equation **(5.4.1 – 2)** in terms of the generalised linear model can be written as

$$\mathbf{d} = \mathbf{G}\mathbf{b} + \mathbf{e}$$

where **d** is an $M \times 1$ data matrix.

b is a $L \times 1$ (or equivalent to $2N \times 1$) matrix containing the amplitudes associated with the fundamental frequency and harmonics of the sine and cosine components

$$\mathbf{b} = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \ldots \quad \alpha_N \quad \beta_1 \quad \beta_2 \quad \ldots \quad \ldots \quad \beta_N]^T$$

e is an $M \times 1$ matrix containing random Gaussian noise entries.

Finally, **G** is an $M \times L$ (or equivalent to $M \times 2N$) matrix of the basis functions containing the parameters of a single piano note and can be written as

$$\mathbf{G} = \begin{bmatrix} e^{-\zeta_1 t_1}\sin\left(2\pi f_0(1+B)^{0.5}t_1\right) & e^{-\zeta_2 t_1}\sin\left(4\pi f_0(1+4B)^{0.5}t_1\right) & \ldots & \ldots & e^{-\zeta_N t_1}\sin\left(N2\pi f_0(1+N^2B)^{0.5}t_1\right) \\ e^{-\zeta_1 t_2}\sin\left(2\pi f_0(1+B)^{0.5}t_2\right) & e^{-\zeta_2 t_2}\sin\left(4\pi f_0(1+4B)^{0.5}t_2\right) & \ldots & \ldots & e^{-\zeta_N t_2}\sin\left(N2\pi f_0(1+N^2B)^{0.5}t_2\right) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ e^{-\zeta_1 t_M}\sin\left(2\pi f_0(1+B)^{0.5}t_M\right) & e^{-\zeta_2 t_M}\sin\left(4\pi f_0(1+4B)^{0.5}t_M\right) & \ldots & \ldots & e^{-\zeta_N t_M}\sin\left(N2\pi f_0(1+N^2B)^{0.5}t_M\right) \end{bmatrix}$$

$$\begin{bmatrix} e^{-\zeta_1 t_1}\cos\left(2\pi f_0(1+B)^{0.5}t_1\right) & e^{-\zeta_2 t_1}\cos\left(4\pi f_0(1+4B)^{0.5}t_1\right) & \ldots & \ldots & e^{-\zeta_N t_1}\cos\left(N2\pi f_0(1+N^2B)^{0.5}t_1\right) \\ e^{-\zeta_1 t_2}\cos\left(2\pi f_0(1+B)^{0.5}t_2\right) & e^{-\zeta_2 t_2}\cos\left(4\pi f_0(1+4B)^{0.5}t_2\right) & \ldots & \ldots & e^{-\zeta_N t_2}\cos\left(N2\pi f_0(1+N^2B)^{0.5}t_2\right) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ e^{-\zeta_1 t_M}\cos\left(2\pi f_0(1+B)^{0.5}t_M\right) & e^{-\zeta_2 t_M}\cos\left(4\pi f_0(1+4B)^{0.5}t_M\right) & \ldots & \ldots & e^{-\zeta_N t_M}\cos\left(N\pi f_0(1+N^2B)^{0.5}t_M\right) \end{bmatrix}$$

Consider an existing Bayesian model (Godsill and Davy, 2002; Davy and Godsill, 2002) for a single note

$$x(t) = \sum_{n=1}^{N} \alpha_n \cos[(n+\delta_n)2\pi f_0 t] + \beta_n \sin[(n+\delta_n)2\pi f_0 t] \qquad \textbf{(5.4.1 – 3)}$$

where α_n and β_n are the respective amplitudes of the sine and cosine components, and δ_n is called the "de-tuning" parameter associated with *each* harmonic of a single piano note.

Clearly, the model of a single piano note of equation **(5.4.1 – 3)** has higher dimensionality than the proposed one of equation **(5.4.1 – 2)** since there is a different inharmonicity (or so-called "de-tuning") parameter δ_n for each harmonic component making the estimation of the parameter space potentially more computationally expensive. Contrary, the proposed model presented in equation **(5.4.1 – 2)** involves the estimation of only one inharmonicity parameter B for each single piano note.

5.4.2 Polyphonic case

In the case of multiple piano notes, such as music intervals and chords, our proposed model can further be expanded for an *R* number of concurrent notes.

Equation **(5.4.1 – 2)** can be extended to equation **(5.4.2 – 1)** as follows

$$x(t) = \sum_{r=1}^{R}\sum_{n=1}^{N} \alpha_{n,r} e^{-\zeta_{n,r} t} \sin\left[n2\pi f_{0,r}\left(1+n_r^2 B_r\right)^{0.5} t\right] + \beta_{n,r} e^{-\zeta_{n,r} t} \cos\left[n2\pi f_{0,r}\left(1+n_r^2 B_r\right)^{0.5} t\right] \quad \textbf{(5.4.2 – 1)}$$

where each note r has its own set of parameters.

From a practical implementation point of view, the dimensions of the amplitude matrix **b** would be $R \cdot L \times 1$ (or equivalent to $R \cdot 2N \times 1$) and matrix **G** would be $M \times R \cdot L$ (or equivalent to $M \times R \cdot 2N$) containing the basis functions of the parameters of *R* notes present in the data.

The polyphonic model of Davy and Godsill (2002) can be expressed as

$$x(t) = \sum_{r=1}^{R}\sum_{n=1}^{N} \alpha_{n,r} \cos[(n+\delta_{n,r})2\pi f_{0,r} t] + \beta_{n,r} \sin[(n+\delta_{n,r})2\pi f_{0,r} t] \qquad \textbf{(5.4.2 – 2)}$$

The dimensionality of the model described by equation **(5.4.2 – 2)** is again higher than that of the proposed model described by equation **(5.4.2 – 1)**. Specifically, equation **(5.4.2 – 2)** depends on estimating $N \cdot R$ number of parameters compared to the R number of parameters for our proposed polyphonic model of equation **(5.4.2 – 1)**.

5.5 Algorithm implementation for automatic music transcription

The theoretical foundation of our proposed automatic music transcription method is based on the use of the M-H algorithm and Gibbs sampler for multivariate parameter estimation. The implementation of these algorithms is discussed in the following sections in detail.

5.5.1 Implementation for estimating parameter f_0

Equation **(5.2 – 3)** which represents a "cost function" of a t-distribution function can also be used to obtain estimates of more than one parameters, i.e. f_0 and/or B parameters.

One may now assume a simple example of a modelled signal, as described earlier from equation **(5.4.1 – 2)**

$$x(t) = \sum_{n=1}^{N} \alpha_n e^{-\zeta_n t} \sin\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right] + \beta_n e^{-\zeta_n t} \cos\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right]$$

where the data is generated over 100 ms with $f_0 = 200$ Hz, $N = 5$ and $B = 10^{-4}$.

The logarithm of the cost function, described by equation **(5.2 – 3)**, is plotted in Figure 5.3 for a range of fundemental frequency f_0 estimates, for a known value of inharmonicity B. Additionally, Figure 5.4 shows an expanded region around the peak at 200 Hz of Figure 5.3.

Note that the negative values of the cost function arise because of the use of logarithms in order to avoid issues with numerical underflow:

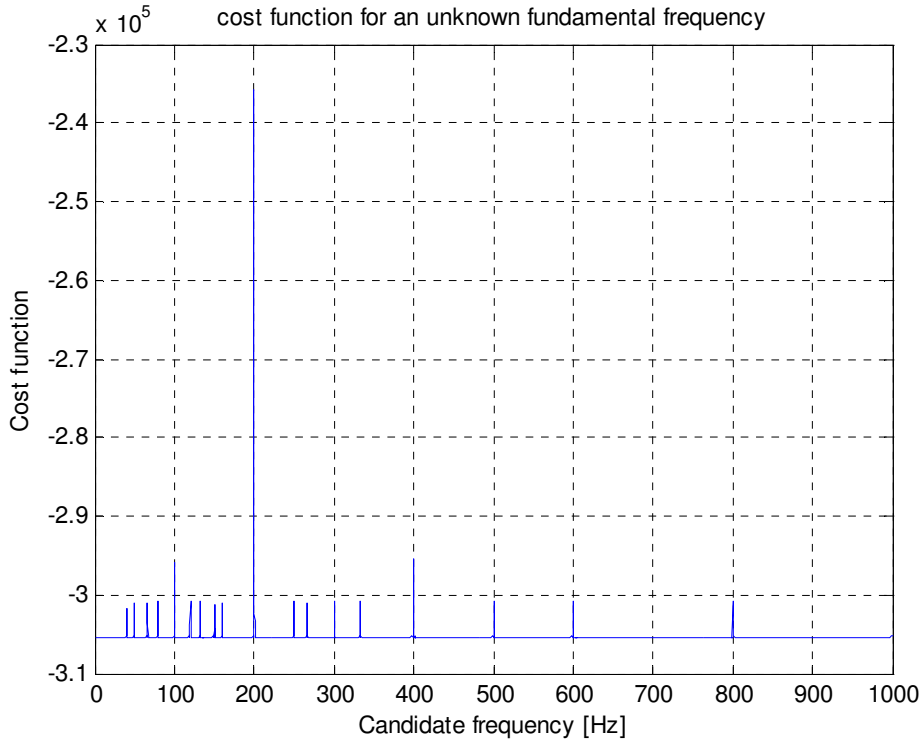


Figure 5.3 – Cost function for a range of fundamental frequencies f_0 given a known value of inharmonicity B

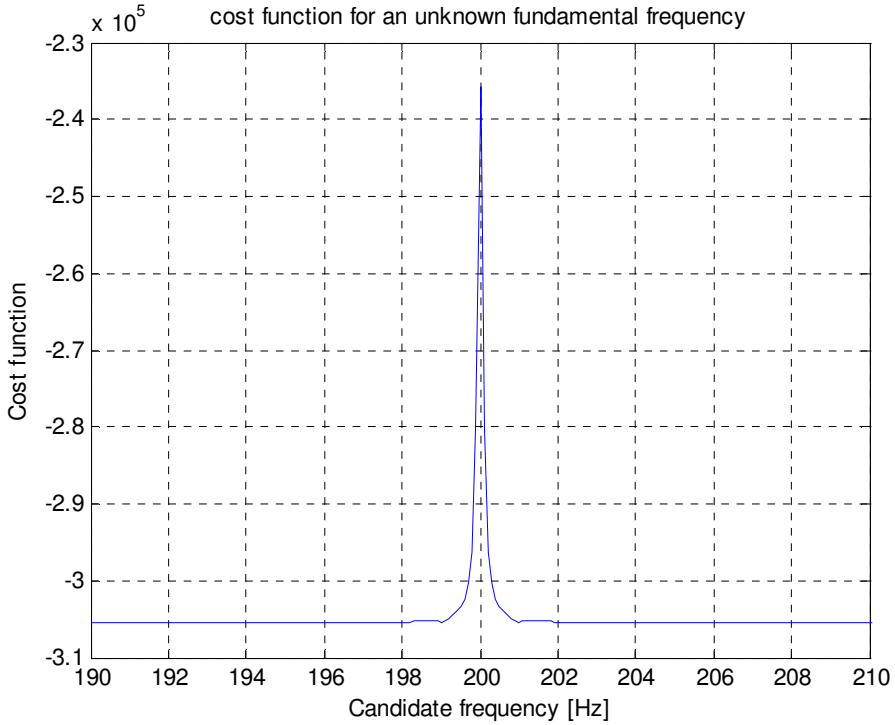


Figure 5.4 – Expanded section view of Figure 5.3

From Figures 5.3, it can be seen that the cost function will have many probable solutions represented as local peaks in the t-distribution function. For instance, solutions will be evident at 200 Hz (fundamental frequency) and any subsequent harmonics, such as at 400 Hz, etc. Also, there will be local optima at multiples and submultiples of the most probable solution (200 Hz), such as at 100 Hz, 150 Hz, etc.

Therefore, one needs to devise an effective algorithm of obtaining only an estimate of the fundamental frequency by ignoring other locally optimal solutions. The following steps have been devised and form the basis of any subsequent implementation of the M-H algorithm for obtaining estimates of the fundamental frequency f_0. The walks of the M-H algorithm are therefore implemented in this particular way, where X and Y represent estimates of f_0:

1_a. For 80% of the time, suppose that X_i is the i^{th} element of a random walk and that the next variate Y_i in the random sequence is produced by simply adding a random perturbation factor ξ_i to X_i of mean zero and standard deviation one

$$Y_i = X_i + \xi_i$$

1_b. For the remaining 20% of the time, the algorithm may randomly follow two different directions:

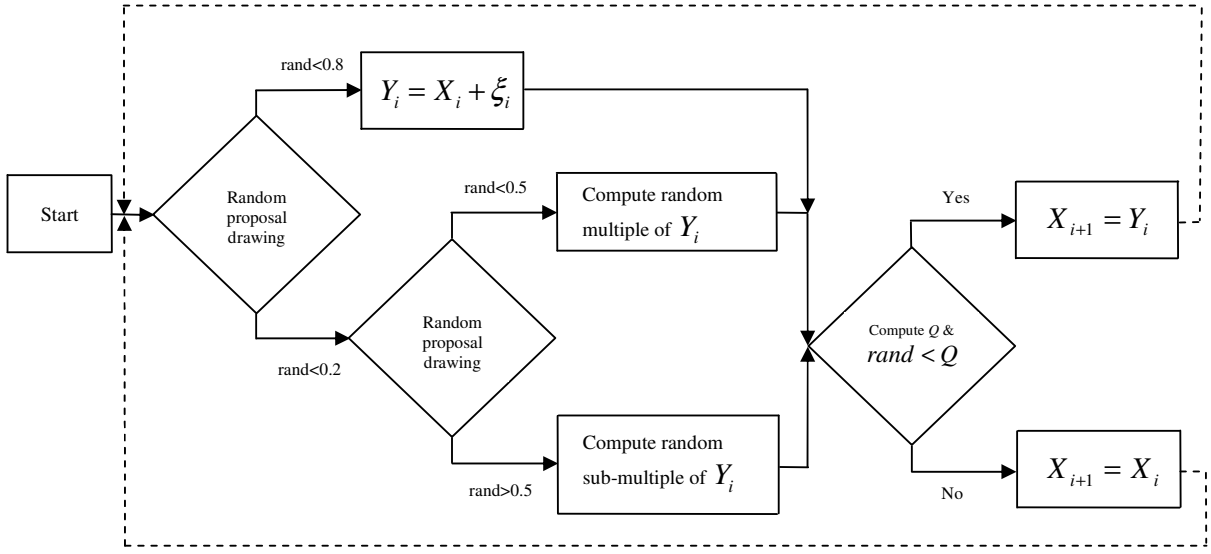
- 50% of the time, the current Y_i estimate is multiplied by a random integer.
- 50% of the time, the current Y_i estimate is divided by a random integer.

The reason behind this decision step is to check whether the estimate is a multiple or a sub-multiple of the fundamental frequency. The additional advantage of this algorithm implementation compared to the original one presented in section 5.2 is that the initial parameter estimate can now be of an *arbitrary* value and hence there is no need to have any prior knowledge about the pitch of the transcribed note.

Note that there are additional checks implemented in the algorithm to stop division if the frequency estimate is below the lowest fundamental frequency of a piano, e.g. below 27.5 Hz, or it stops multiplication if the fundamental is too high, e.g. above 4,186 Hz (typical frequency range of a piano is also shown in Figure B.1 of Appendix B).

2. There are now two possibilities; either the proposed random variate Y_i is accepted, or it is rejected and X_i is repeated. The condition on which Y_i is accepted instead of X_i is defined by the acceptance function Q of equation **(5.2 – 6)**. Let us assume ε is a uniform variable drawn over the range [0,1]. If ε<Q, then $X_{i+1} = Y_i$. Otherwise, if ε>Q, then $X_{i+1} = X_i$.

Steps 1 and 2 are repeated, until the M-H algorithm reaches an equilibrium state, where the estimated fundamental frequency f_0 is almost constant. A flow chart is added here showing a single iteration of the algorithm in its basic form:



5.5.2 Implementation for estimating parameter *B*

Assume now for the same signal that the fundamental frequency f_0 is known, but in this instance the inharmonicity factor B is unknown. The cost function of equation **(5.2 – 3)** for a range of inharmonicity B estimates is shown in Figure 5.5:

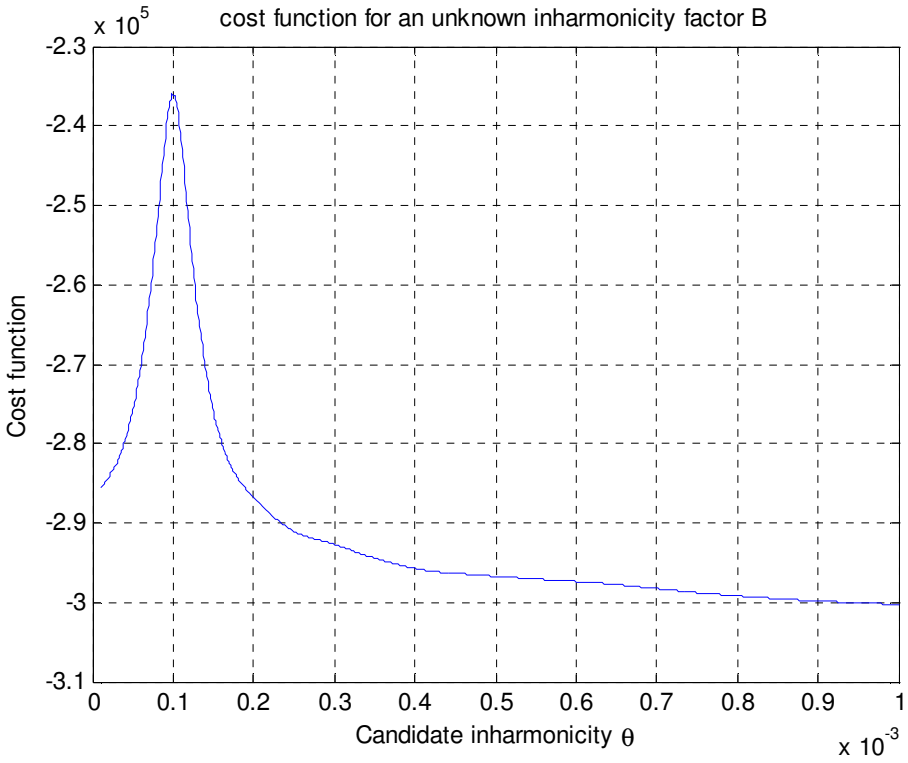


Figure 5.5 – Cost function for a range of inharmonicity factors B given a known value of fundamental frequency f_0

From Figure 5.5, it can be seen that the cost function *only* has a single peak at the correct value, i.e. $B = 10^{-4}$. Therefore, the implementation of the algorithm for calculating the inharmonicity factor B would be much simpler than that of the fundamental frequency f_0.

In particular, the algorithm for calculating the inharmonicity factor B is similar to the one presented in the previous section, although it does not require random multiplications or divisions of the estimates since there is simply only one peak in the cost function as shown in Figure 5.5.

5.5.3 Implementation for simultaneously estimating f_0 and B

The cost function for a range of values of the fundamental frequency f_0 and the inharmonicity factor B is shown in Figure 5.6. In this particular case, the parameter space that needs to be calculated has two dimensions; f_0 and B. The cost functions

shown in Figures 5.3 and 5.5 represent slices through the two-dimensional cost function as illustrated in Figure 5.6:

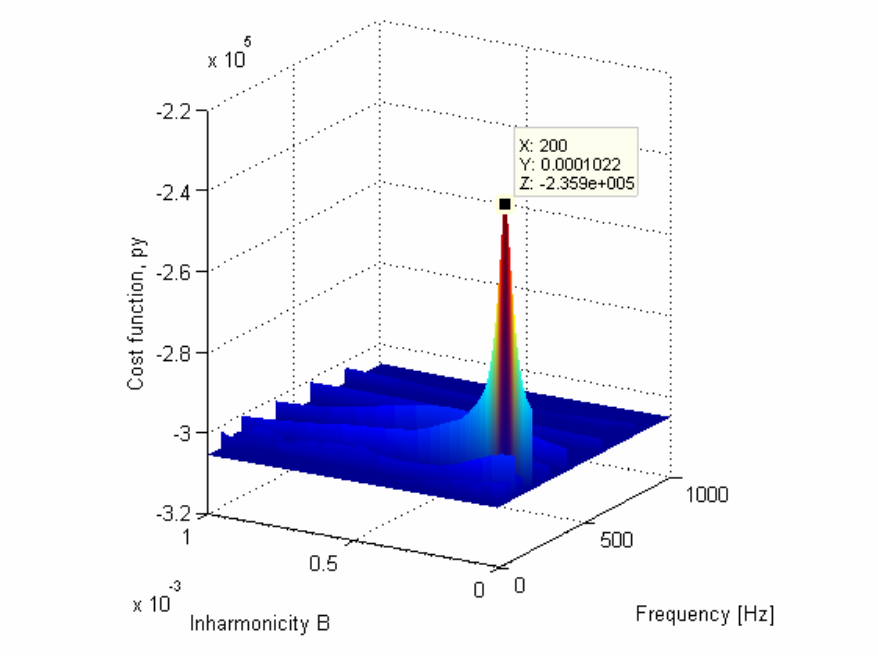


Figure 5.6 – Cost function for a range of fundamental frequency f_0 and inharmonicity B estimates

From Figure 5.6, it can be seen that the cost function peaks at the most probable *pair* of estimates. In this case at $f_0 = 200$ Hz and $B = 10^{-4}$, which correspond to the values used to generate the data.

The challenge now is to implement an algorithm for *multivariate* estimation, i.e. where the estimation of the fundamental frequency f_0 and B takes place in one algorithm. The Gibbs sampler is proposed based on extensive work by Geman and Geman (1984) and Gelfand and Smith (1990) in the field of image restoration processing and statistical data analysis respectively.

The main principle behind the Gibbs sampler of drawing samples from a multivariate density is to break down the problem into one of drawing successive samples from densities of smaller dimensionality (Ó Ruanaidh and Fitzgerald, 1996).

An iterative cyclic pattern can used by assuming a parameter space $\Theta = \{f_0^i, B^i\}$, where the i superscript denotes the i^{th} iteration of the iterative process. The superscript "0" denotes the initial estimates of f_0 and B.

1st iteration:

$$p(f_0|B^0, d) \longrightarrow f_0^1$$

$$p(B|f_0^1, d) \longrightarrow B^1$$

2nd iteration:

$$p(f_0|B^1, d) \longrightarrow f_0^2$$

$$p(B|f_0^2, d) \longrightarrow B^2$$

j^{th} iteration:

$$p(f_0|B^{i-1}, d) \longrightarrow f_0^i$$

$$p(B|f_0^i, d) \longrightarrow B^i$$

Therefore, for each estimate of the fundamental frequency f_0, an estimate of the inharmonicity factor B is drawn, which in turn is used to draw another estimate of f_0 and so forth.

Note that the steps for the Gibbs sampler are based on the implementation of the M-H algorithm for univariate estimates of the fundamental frequency f_0 and inharmonicity factor B as described previously in sections 5.5.1 and 5.5.2 respectively.

5.6 Automatic music transcription examples

Monophonic and polyphonic example cases are presented here using simulated signals and real piano recordings, where M-H algorithm steps are used to compute the successive states of the Gibbs sampler for multivariate parameter estimation as discussed in the previous section.

5.6.1 Simulated monophonic signals

A piano note is simulated based on our proposed model, as described by equation **(5.4.1 – 2)**

$$x(t) = \sum_{n=1}^{N} \alpha_n e^{-\zeta_n t} \sin\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right] + \beta_n e^{-\zeta_n t} \cos\left[n 2\pi f_0 \left(1 + n^2 B\right)^{0.5} t\right]$$

The fundamental frequency is set to 261.6 Hz representing note C_4, as typically shown in Figure B.1 of Appendix B, the inharmonicity factor is typically set to 2.5×10^{-4} representing a typical value for a piano note (Fletcher, 1964), and term *N* is set to 5.

The data is generated over a 100 ms time period. Note that other authors (Godsill and Davy, 2002; Davy and Godsill, 2002) have used smaller "frames" of the order of 20 ms.

By employing the Gibbs sampler, as described in section 5.5.3, simultaneous parameter estimation of the fundamental frequency and inharmonicity factor *B* can be obtained. A typical such run is shown in Figure 5.7:

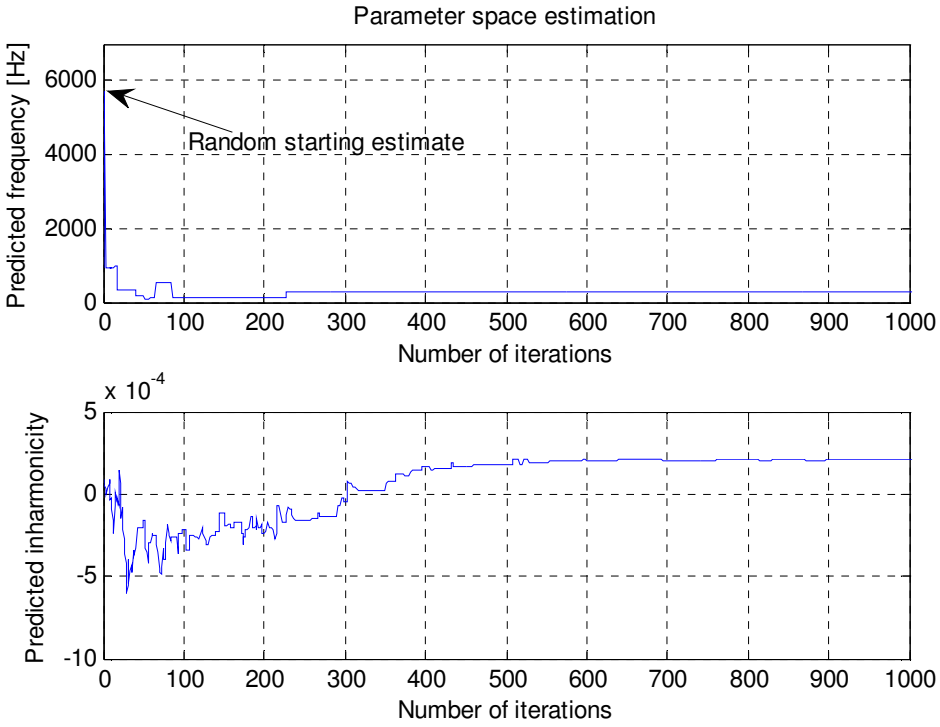


Figure 5.7 – Parameter space estimation for simulated single piano note C_4 based on Gibbs sampler

From Figure 5.7, it can be seen that the estimate of the fundamental frequency is accurately predicted after about 250 iterations at 261.6 Hz. As discussed in section 5.5.1, the advantage of the implemented algorithm is that there is no need to have a prior idea of the "pitch" of the transcribed note. In fact, the starting estimate is chosen *arbitrarily* by the implemented algorithm, which in this case was selected at 5,750 Hz, making the method of frequency estimation applicable to any unknown note.

As far as the estimate of the inharmonicity factor B is concerned, this is also accurately predicted over a longer number of iterations, i.e. just under 1,000. Note that in the case of estimating B, the initial estimate is *not* chosen arbitrarily, but rather is set to the lowest typical value of B, i.e. 10^{-4}. Also, note that the performance of the algorithm may be improved further by discarding negative intermediate estimates for the inharmonicity as shown in Figure 5.7.

Finally, since the steps of the M-H algorithm are drawn randomly, it was found that the results from repeated estimates (about 20 in total) of the inharmonicity factor B on

the same data set exhibit a 12% variation, whereas the variability in the estimation of f_0 is very small of the order of up to 2%.

5.6.2 Real monophonic piano recording signals

In this section, real monophonic piano notes were transcribed with our proposed model as before. Recordings of real piano notes were carried out at 44.1 kHz using samples from the "Gigastudio" sample library.

In this example, a real piano note C_4, with a fundamental frequency of about 261 Hz, is attempted to be transcribed. Figure 5.8 shows the evolution of the estimates for the fundamental frequency and inharmonicity factor against the number of iterations using the Gibbs sampler:

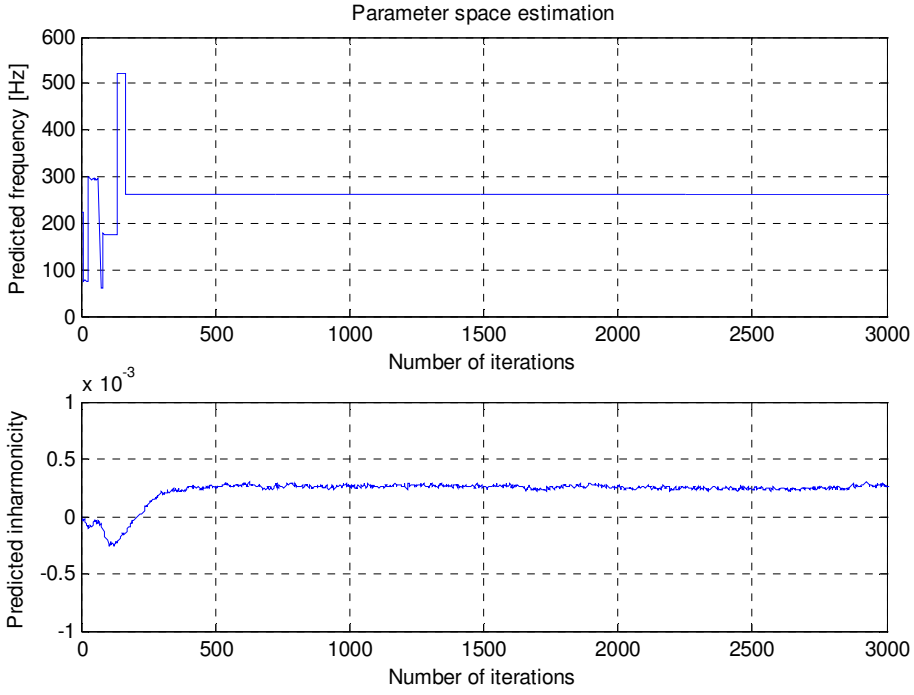


Figure 5.8 – Parameter space estimation for a real piano note recording C_4 based on Gibbs sampler

From Figure 5.8, it can be seen that the estimates of the fundamental frequency and the inharmonicity factor *B* are predicted after about 250 iterations of the algorithm by returning estimates close to 261 Hz and $2.7\text{x}10^{-4}$ respectively.

The procedure is repeated on two more notes, i.e. G_4 with fundamental frequency of about 392 Hz and C_5 with a fundamental frequency of about 523 Hz. Table 1 summarises the results showing the estimated fundamental frequencies and the four harmonics that can be derived from the simple model of inharmonicity

$$f_n = nf_0\sqrt{1+n^2B}$$

where for $n=1$, the fundamental frequency is, strictly speaking, no longer f_0 but f_1 due to the inherent inharmonicity in the strings (Ortiz-Berenguer *et al.*, 2005)

$$f_1 = f_0\sqrt{1+B}$$

Table 5.1 is shown below:

Piano Note	f_1 **[Hz]**	**1st harmonic [Hz]**	**2nd harmonic [Hz]**	**3rd harmonic [Hz]**	**4th harmonic [Hz]**	**Inharmonicity *B***
C_4	**261.4**	523	785	1047.6	1311.1	**$2.7\text{x}10^{-4}$**
G_4	**391.8**	784.3	1177.7	1572.6	1969.5	**$4.3\text{x}10^{-4}$**
C_5	**523.5**	1048.3	1576	2107.8	2645.2	**$8.9\text{x}10^{-4}$**

Table 5.1 –Automatic music transcription of real single piano note recordings using the Gibbs sampler

From Table 5.1, it can be seen that the estimates of the three fundamental frequencies are accurate when compared with typically expected values found in a piano instrument. Also, it can be seen that the inharmonicity factor will increase with the

fundamental frequency or "pitch" of the piano note, as expected (Fletcher, 1964). Indeed, this observation is in line with equation **(2.4.2 – 3)** as discussed earlier

$$B = \frac{\pi^3 d^4 E}{64 T l^2}$$

where for example the inharmonicity is greater in the case of short strings (high frequency notes) for a given radius and tension as opposed to long strings (low frequency notes) for the same radius and tension.

This was a demonstration of a successful transcription of real monophonic piano notes with the implementation of our algorithms.

5.6.3 Simulated polyphonic signals

A music interval is the simplest form of polyphonic music, where two notes are played together. One may recall equation **(5.4.2 – 1)**, which may be used to describe an $R = 2$ number of concurrent notes

$$x(t) = \sum_{r=1}^{2} \sum_{n=1}^{N} \alpha_{n,r} e^{-\zeta_{n,r} t} \sin\left[n 2\pi f_{0,r} \left(1 + n_r^2 B_r\right)^{0.5} t\right] + \beta_{n,r} e^{-\zeta_{n,r} t} \cos\left[n 2\pi f_{0,r} \left(1 + n_r^2 B_r\right)^{0.5} t\right]$$

A typical music interval, C_4 to G_4, is simulated based on the above proposed model. The fundamental frequencies are set to 261.6 Hz for C_4 and 392 Hz for G_4, whereas the inharmonicity factors of each note are set to 2.5×10^{-4} and 4.0×10^{-4}, respectively.

By employing the Gibbs sampler, as described in section 5.5.3, simultaneous successful parameter estimation of the fundamental frequency and inharmonicity factor *B* can be obtained. This is shown in Figure 5.9:

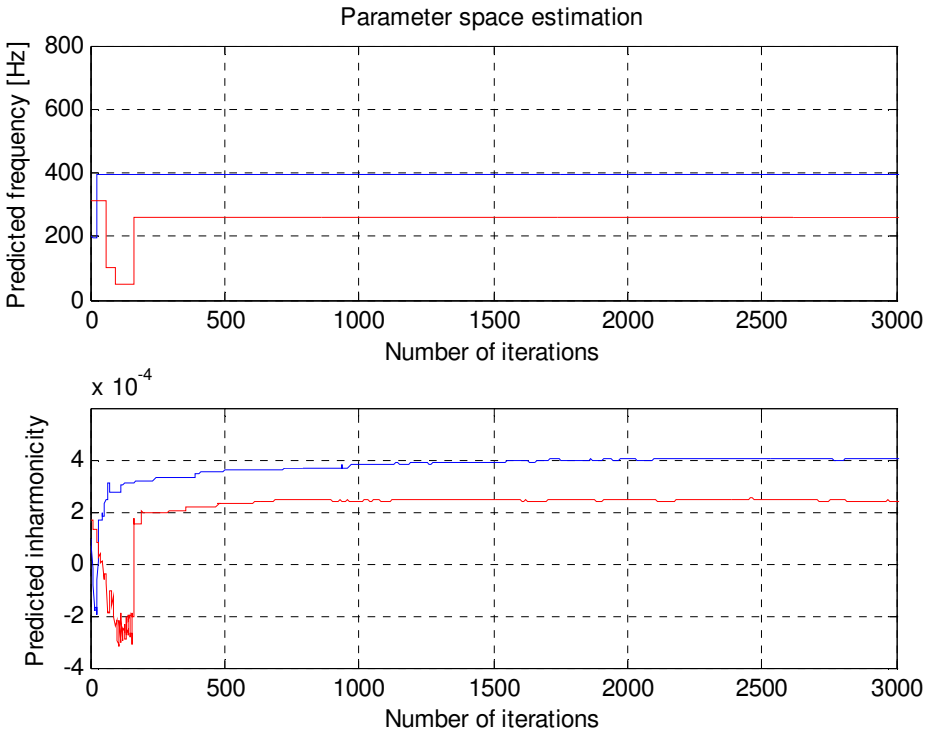


Figure 5.9 – Parameter space estimation for two simulated piano notes based on Gibbs sampler

5.6.4 Real polyphonic piano recording signals

For the purpose of this exercise, music intervals are generated by combing the real monophonic recordings, which represent the individual notes, in a number of music interval combinations. Then, the fundamental frequencies and inharmonicity factors of the constituent notes are estimated.

In particular, it has been shown from the literature overview in chapter 2 that the most difficult music intervals to transcribe are those where the notes have a large number of shared harmonics. These are known as "consonant" intervals as opposed to "dissonant" intervals, where there is very little overlap between the shared harmonics of the involved notes.

The most consonant music intervals are: the "unison" (duplicated note; i.e. a "pseudo-interval") and the "octave" (where the highest note in the interval approximately

shares its fundamental frequency and its entire harmonics with the lowest note). Other consonant intervals exist such as the "perfect fifth" (having a 3:2 approximate integer ratio of harmonics between two notes) and the "perfect fourth" (having a 4:3 approximate integer ratio of harmonics between two notes). Note that the signal discussed in section 5.6.3 was an example of a simulated "perfect fifth" interval.

Figure 5.10 shows the automatic music transcription of a real perfect fifth music interval (C_4 – G_4) for a known number of transcribed notes:

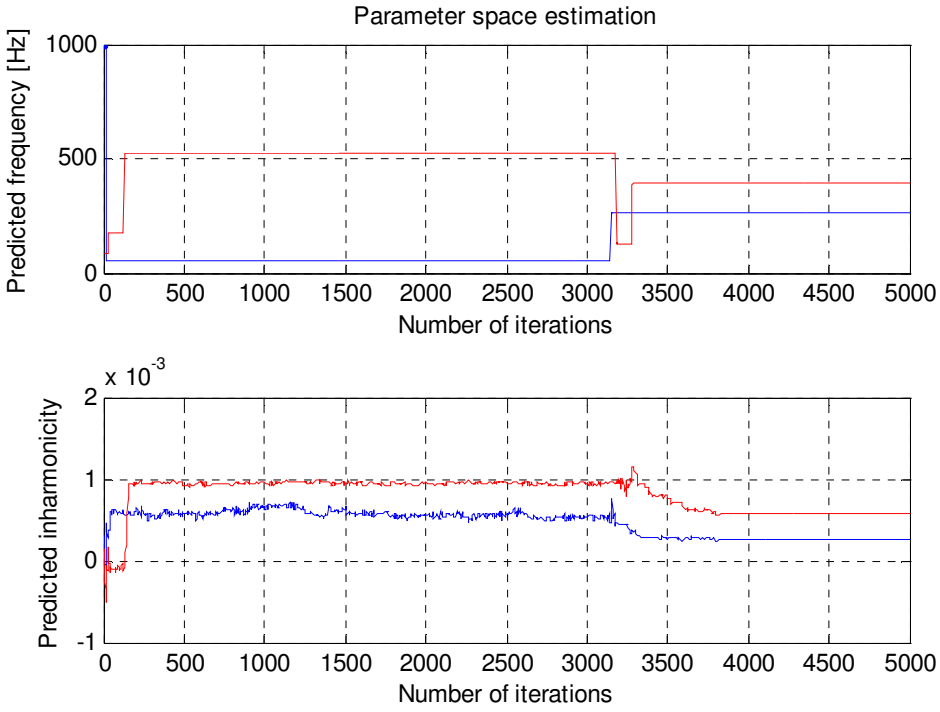


Figure 5.10 – Parameter space estimation for a typical music interval, C_4 – G_4 (perfect fifth), from real piano note recordings based on Gibbs sampler

From Figure 5.10, it can be seen that the estimates of the fundamental frequencies of the above perfect fifth music interval are successfully predicted at around 261 Hz and 392 Hz. These estimates hardly differ from the estimates of the single transcribed notes as shown earlier in Table 5.1.

However, as far as the estimated inharmonicity factors are concerned for the above perfect fifth music interval (C_4 – G_4), parameter B of note C_4 is predicted at around 2.5×10^{-4}, which is 7% different when compared with parameter B of the single

transcribed note C_4 as shown earlier in Table 5.1, i.e. $B = 2.7\text{x}10^{-4}$. Also, parameter B of note G_4 from the same polyphonic transcription is estimated at $5.9\text{x}10^{-4}$, which is 37% different when compared with parameter B of the single transcribed note G_4 as shown earlier in Table 5.1, i.e. $B = 4.3\text{x}10^{-4}$.

In another example of polyphonic transcription, it was found that for the transcribed octave music interval $C_4 – C_5$, parameter B of note C_4 is predicted at around $4.0\text{x}10^{-4}$, which is 48% different when compared with parameter B of the single transcribed note C_4 as shown earlier in Table 5.1, i.e. $B = 2.7\text{x}10^{-4}$. Also, parameter B of note C_5 from the same polyphonic transcription is estimated at $1.1\text{x}10^{-3}$, which is 24% different when compared with parameter B of the single transcribed note C_5 as shown earlier in Table 5.1, i.e. $B = 8.9\text{x}10^{-4}$.

The majority (3 out of 4) of the above discrepancies in the estimation of the inharmonicity factors B of the two music intervals ($C_4 – G_4$ and $C_4 – C_5$) is higher than the expected 12% variation, which is related to the randomness of the estimation process of the inharmonicity factor as discussed in section 5.6.1.

Table 5.2 below summarises the above results of the two polyphonically transcribed music intervals ($C_4 – G_4$ and $C_4 – C_5$) by presenting their estimated fundamental frequencies and inharmonicity factors, along with their respective changes[3] shown in the parentheses, when compared with those in Table 5.1:

Music interval (note 1-note 2)	**Type**	**f_1 of note 1 [Hz]**	**Inharmonicity B of note 1**	**f_1 of note 2 [Hz]**	**Inharmonicity B of note 2**
C_4-G_4	**Perfect fifth**	**261.4 (0% change)**	**$2.5\text{x}10^{-4}$ (7% change)**	**392.1 (-0.08% change)**	**$5.9\text{x}10^{-4}$ (-37% change)**
C_4-C_5	**Octave**	**260.2 (0.5% change)**	**$4.0\text{x}10^{-4}$ (-48% change)**	**522.8 (0.1% change)**	**$1.1\text{x}10^{-3}$ (-24% change)**

Table 5.2 – Automatic music transcription of real polyphonic piano note recordings using the Gibbs sampler

[3] Positive and negative percentile changes indicate a reduction and an increase respectively in the estimates when directly compared with the estimates of single transcribed notes.

The reason for the large discrepancies of the inharmonicity estimates between the monophonically and polyphonically transcribed notes is perhaps due to the fact that in the case of a polyphonic transcription, more parameters are fitted over the same length of data and as a result the variability in the parameter estimation will be larger than in the case of a monophonic transcription. However, more work needs to be carried out in this field to understand, or support, the reason for such a discrepancy.

Note that it has to be said that the estimation of the inharmonicity factor in the case of *simulated* polyphonic signals is very accurate since the model accurately fits the generated data as shown earlier in section 5.6.3.

The question is therefore how important is the accurate estimation of the inharmonicity factor B for calculating the fundamental frequency and the associated harmonics of a real piano note.

Assume now that one is interested in calculating the difference between the frequency estimates in the case of monophonically and polyphonically transcribed notes for the first five N terms.

In this particular case, the value of the inharmonicity factor, in the case of a polyphonic transcription, is chosen to be 48% higher than that of the monophonic transcription. This percentage reflects the largest discrepancy as shown earlier in Table 5.2. Note also that the magnitude of the inharmonicity factor B is of the order of 10^{-4}. The following formula may be used to estimate the percentage difference in the frequency estimates

$$\%\ \ difference = \frac{f_{n,monophonic} - f_{n,polyphonic}}{f_{n,monophonic}} 100 \qquad \textbf{(5.6.4 – 1)}$$

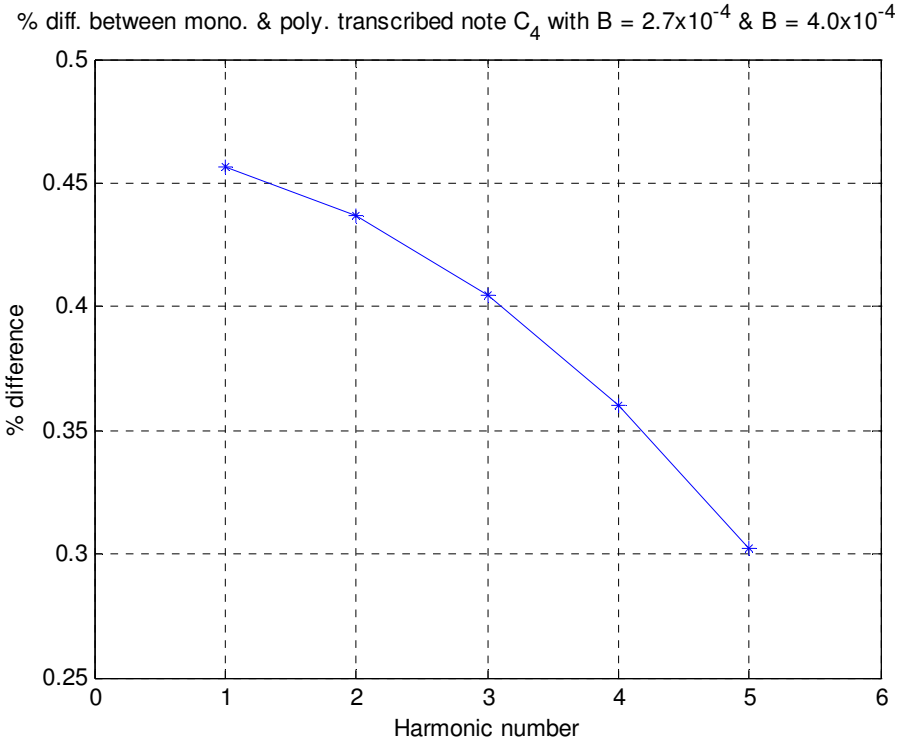


Figure 5.11 – Percentage difference in the frequency estimates between a monophonically and polyphonically transcribed piano note C_4 with a 48% discrepancy in the values of the inharmonicity factor B, which are of the order of ***10^{-4}***

Figure 5.11 shows that the percentage difference of the resulting first five harmonic terms (including the fundamental frequency) between the monophonically and polyphonically transcribed piano note C_4 is less than 0.5%, which can be regarded as negligible.

However in the *hypothetical* case, where the value of the estimated inharmonicity factor *B* is of the order of 10^{-2}, which is typically the highest value of inharmonicity found in a piano (Fletcher, 1964), the difference in estimating the resulting harmonic terms can be of a measurable magnitude. This is shown in Figure 5.12:

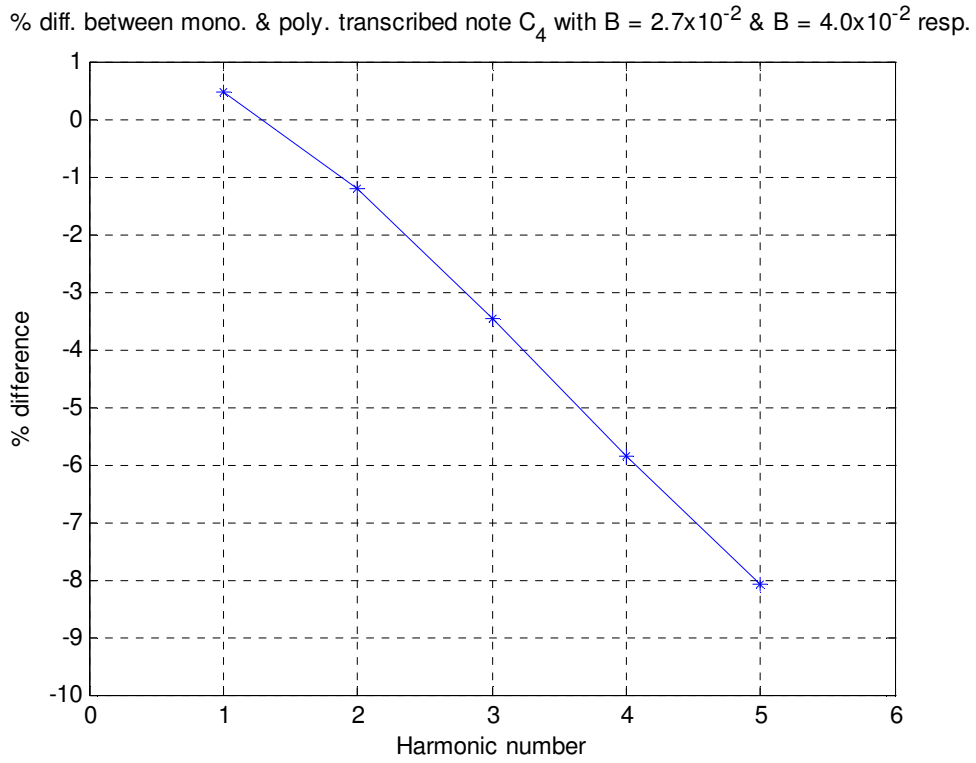


Figure 5.12 – Percentage difference in the frequency estimates between a monophonically and polyphonically transcribed piano note C_4 with a 48% discrepancy in the values of the inharmonicity factor B, which are of the order of ***10^{-2}***

From Figure 5.12, it can be seen that for large values of the inharmonicity factor, i.e. of the order of 10^{-2}, there would be a percentage difference of up to 8% when calculating the first five terms of a polyphonically transcribed piano note.

As an overall remark, it can be deduced that the estimation of the inharmonicity factor from real polyphonic recordings may differ significantly from the estimate of the real monophonic recordings. As a result, for very large values of the inharmonicity factor, a relatively large difference in the estimation of the harmonics terms between the monophonically and polyphonically transcribed notes may result.

As a final remark, the results from the transcription examples presented in this research were promising, but the transcription model is simple and has not been generalised for an unknown number of notes present in a recording. Future work is encouraged to consider the varying model dimensionality.

5.7 Model performance

In this section, the performance characteristics of our algorithm implementation for the parameter space estimation are discussed in relation to two topics.

First, the inharmonicity factor B is estimated for a typical piano note over different time windows of the same recording in order to deduce how well the model fits the data. Second, one of the objectives of this research work has been the development of a signal model, which has smaller dimensionality than existing models in the literature (Godsill and Davy, 2002; Davy and Godsill, 2002). Hence, the potentially reduced computational expense of our proposed signal model is discussed in relation to the model of Godsill and Davy (2002) through simple numerical demonstrations.

5.7.1 Parameter *B* estimation across successive time windows

In order to provide further insight into the accuracy of the proposed model across successive time windows of a typical piano note recording, the monophonic model of a single note was employed for estimating parameter B. For the purpose of this analysis, a 1 sec recording of a piano note C_4 was divided into 10 successive time windows of 100 ms duration each.

Figure 5.13 illustrates the variation in estimating the inharmonicity factor in each different time window of the same recording:

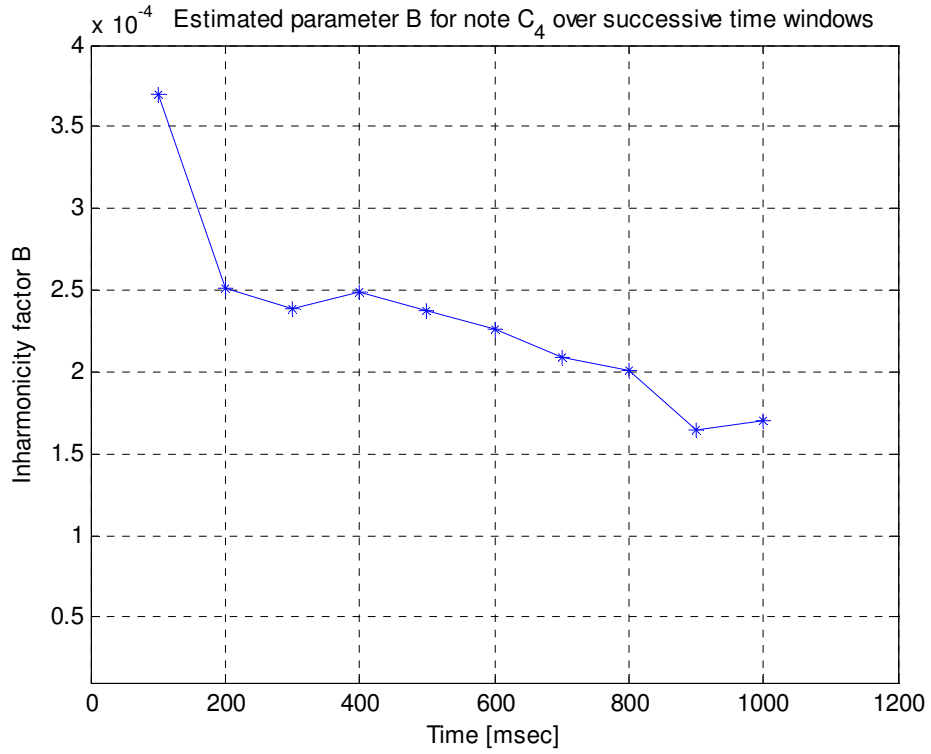


Figure 5.13 – Estimated inharmonicity factor B for a single piano note C_4 across successive time windows of the same recording

From Figure 5.13, it can be seen that the inharmonicity factor *B* estimate in the first frame of the data, between 0 ms and 100 ms, is considerably different than the estimates in time windows above 200 ms. This is expected since the initial attack of the piano note is contained within this frame, whereas our model does not account for this feature.

For higher-time windows, e.g. between 200 ms and 600 ms, the inharmonicity factor does not vary significantly, although above 600 ms it starts to roll off considerably. This might be due to the fact that the time-varying amplitudes of the signal are not accurately represented by the model throughout the recording (potential existence of a non-linear mechanism affecting the amplitudes of the signal).

5.7.2 Computational expense

As it has been discussed earlier, our proposed signal model has smaller dimensionality, since we are only interested in computing R inharmonicity factors for R number of notes, as opposed to the estimation of $N \cdot R$ inharmonicity (or so-called "de-tuning") parameters according to Godsill and Davy (2002).

In this section, we simply demonstrate the potential computational savings with our proposed model when compared to the existing model by Godsill and Davy. In particular, the two different signal models are used within a Gibbs sampler implementation and the parameter estimation space is simply *timed* for a *fixed* number of iterations. Note that the accuracy of the estimates from the two different models is not discussed in this research.

Table 5.3 below shows the additional computational expense from the model of Davy and Godsill as a function of N terms for a fixed number of iterations when transcribing a single piano note:

Additional computational expense [time percentage]	**Number of terms *N***
32%	**3**
47%	**4**
58%	**5**

Table 5.3 – Additional computational expense of model by Godsill and Davy

From Table 5.3, it can be seen that as the number of harmonics increases, so as the additional computational expense increases (between about 11% and 15% for every

additional term)[4]. This is due to the fact that each harmonic term will be associated with a different inharmonicity parameter and therefore there will be additional computational steps in the Gibbs sampler as the number of harmonics increases.

[4] Note that for $N = 1$, the existing model by Godsill and Davy performs marginally better than the proposed one. This is probably due to the fact that the proposed and the existing models are implemented slightly differently within Matlab numerical software.

5.8 Conclusions

Signal models based on Bayesian formalism were proposed for describing the sound of a piano by employing an inharmonicity parameter and decay rates for either single or multiple piano notes.

In particular, the proposed models for the monophonic and polyphonic cases were based on the description of notes with superimposed fundamental and harmonic frequencies including the inharmonicity factor inherently present in the piano strings. The fast and/or slow decay rate of the resulting sound can also be incorporated in the model through the simple use of a damping parameter term depending on whether the transcribed music passage/note is short or long in duration. However, in our analysis the damping parameters were set to zero since the data is analysed over a small time window during which the decay of the amplitude is expected to be very small.

It was demonstrated that the proposed models in this research are potentially more computationally attractive when compared with existing models in the literature (Godsill and Davy, 2002; Davy and Godsill, 2002). Indeed, the latter models would use a different inharmonicity parameter for each harmonic component of each note making the estimation of the parameter space more computationally expensive, whereas our proposed models involve only one inharmonicity parameter for each piano note.

A computational route for calculating the desired parameter space, such as the fundamental frequency and inharmonicity factor of each transcribed piano note, was outlined by utilising the M-H algorithm and Gibbs sampler for multivariate parameter estimation. Examples of monophonic and polyphonic music transcription were also featured and discussed.

The estimation of the inharmonicity factor from real polyphonic recordings may differ from the estimate of real monophonic recordings. This may be due to the fact that in the case of a polyphonic transcription, more parameters are fitted over the same length of data and as a result the variability in the parameter estimation will be larger than in the case of a monophonic transcription. Hence, for very large values of the

inharmonicity factor, a relatively large discrepancy in the estimation of the harmonics terms may result from this.

It has been shown that the estimation of the inharmonicity factor B may vary across successive sections of the same recording. In particular, the biggest discrepancy is found when analysing a section of the recording where the attack of the note is included, i.e. at the beginning of the recording between 0 ms and 100 ms. This is expected since our proposed model does not account for this feature. Further discrepancies may be identified as moving closer to end of the recording. This might be due to the fact that the time-varying amplitudes of the signal are not accurately represented by the model throughout the recording.

The transcription results using the proposed signal models that are presented in this research are encouraging rather than comprehensive. Future studies could explore and compare this approach with other known methods in the literature. Also, the transcription model presented has not been generalised for an unknown number of notes present in a real recording. The model should consider the increased dimensionality of the model through the use of a more generic framework.

Chapter 6

Overall conclusions

6.1 Final remarks

This research has concentrated on the investigation of two important features of the sound reproduction mechanism of the piano instrument through analytical formulations and numerical investigations. First, the inharmonicity associated with the piano strings, which results in the formation of non-integer spaced harmonics relative to the fundamental frequency of the playing note, was discussed. Second, the double decay rate effect was discussed, where the time-varying reproduced sound initially has a fast decay rate characteristic followed by a much slower decay rate (also known as "aftersound").

Since the inharmonicity and double decay rate effect form an integral part of the sound reproduction mechanism of the piano, then one may assume that these features may also need to be incorporated into a signal model to aid in music transcription.

Chapters 1 and 2 provide the introduction to this research work by mainly presenting the existing knowledge on the inharmonicity and double decay rate effect. Also, since the original motivation behind this research work was the development of an automatic music transcription method, a review of the different transcription methods was carried out, mainly in relation to neural network, probabilistic and blind separation methods.

Chapters 3, 4 and 5 provide the original contributions of this research. In particular, in chapters 3 and 4, the inharmonicity and the double decay rate features are discussed through analytical formulations and numerical demonstrations, whilst in chapter 5, signal models for monophonic and polyphonic music performances are proposed based on the latter features. These models are presented in a probabilistic Bayesian framework and transcription examples of simulated and real piano note recordings are demonstrated through the use of the Metropolis-Hastings (M-H) algorithm and Gibbs sampler for multivariate parameter estimation.

6.1.1 Inharmonicity

The following can be concluded from *chapter 3* in relation to the inharmonicity factor B of piano strings:

- The numerical simulations of a vibrating piano string with inharmonicity showed a characteristic "one-sided" effect in the time domain, as a result of the non-integer spacing of the harmonic components of the modelled signal in relation to the fundamental frequency. The latter name was derived from the formation of a strong non-symmetrical response relative to the main lobe temporal structure of the signal, which represents the fundamental frequency of the vibrating string.

- The one-sided effect is not present in the case when the inharmonicity factor is set to zero. The modelled signal with $B = 0$ is simply the sum of a Fourier series of a periodic function. However, in general, the signal with non-zero inharmonicity is not periodic, unless each ratio of the modelled frequencies is a rational number.

- The response of the modelled signal with inharmonicity is dependant on five different parameters: the magnitude of the inharmonicity factor, the number of harmonics present in the modelled signal, the time parameter, the amplitude of harmonics and the phase characteristics. Generally speaking, an increase in any of the five parameters results in a stronger manifestation of the one-sided effect.

- The modelled signal of a vibrating piano string was directly compared with the time history of a real piano note recording. From these investigations, it was shown that the presence of one-sided structure is dependent on how the amplitude of the harmonic terms varies. In the case of non-uniformly varying amplitudes, this effect is obscured both in simulated and real piano note recordings.

- Further analysis work was presented using the cepstrum in order to reveal a one-sided effect due to the inharmonicity factor present in simulated and real piano note recordings. Once again, the non-integer spacing of the manifesting

harmonics was acknowledged to be the main reason for this characteristic effect through numerical demonstrations.

- Analytical formulations were presented using the homomorphic properties of the cepstrum in order to derive representations of the resulting sound, with and without inharmonicity, of a piano note into separate components (response of a vibrating piano string and soundboard). The cepstrum has been used to consider a conjecture (Karatsovis *et al.*, 2006), where the soundboard would have a broad spectral response and consequently a cepstrum that predominantly contains its information in the low "quefrencies" (or low-time regions), whereas a vibrating piano string would give rise to a series of impulses in the pseudo-time domain manifesting themselves in the higher quefrencies (or high-time regions).

- The coupled response of two vibrating piano strings was also discussed through both the use of digital waveguides and the cepstrum. This work naturally led to the next chapter, where the coupling mechanism was discussed in more detail.

- From the analytical and numerical demonstrations in this chapter, it was possible to show the importance of the inharmonicity in the reproduced sound mechanism of a piano note. Therefore, it was decided that this parameter should form an integral part of any proposed signal model in the future.

6.1.2 Double decay rate

The following can be concluded from *chapter 4* in relation to the double decay rate effect of the reproduced sound:

- Coupled oscillators were used to demonstrate the effect of double decay rate as observed in real piano note recordings. It was also shown that the analytical formulation of two coupled oscillators is identical to the analogous formulation of digital waveguides as shown in chapter 3.

- From numerical simulations, it was possible to show that in the case of two identical oscillators, representing a group of two identically-tuned piano strings that may form a piano note, one of the resulting coupled modes will be identical to that of the uncoupled mode. In physical terms, there is symmetry in the system and the two masses will move together in-phase through the coupling of the bridge, which acts as rigid link connection between the two masses. Also, the ratio of the damping terms between the two coupled modes will be larger when compared with that of the uncoupled modes. This may ultimately result in the characteristic double decay rate of the reproduced piano sound, when both of these modes are appropriately excited.

- In the case of mistuned oscillators, it was possible to clearly replicate the double decay rate effect, when there is an out-of-phase relationship between the two, following work by other authors (Weinreich, 1977; Hundley *at al*., 1978). It was therefore deduced that the phase difference between the two oscillators and their relative mistuning can promote the appearance of a double decay rate effect.

- The double decay rate effect is different for the various harmonic terms (fundamental and harmonics). However, it is still not well understood how the decaying rates of the different harmonic terms contribute to the way the overall sound is perceived (Weinreich, 1977).

- There is normally a “dip” in the response of a vibrating piano string at the point where the second (or slower) decay rate takes over from the first (or fast) decay. It was simply shown analytically that the resulting dip would be a direct function of the mistuning relationship between two oscillators.

- A physical model was developed to describe the dynamics of multiple strings coupled with a piano bridge as an extension to the general model of coupled oscillators. Through the use of this model, it was possible to simulate the response of two mistuned strings of a piano note coupled with a finite length beam in the frequency domain.

- Power flow analysis was carried out by exciting one string and evaluating the power that goes into the other string and the beam. It was demonstrated that most of the power input dissipates through the material damping of the string and subsequently less will be transmitted through the coupling to the beam and the other string.

- Measurements were also carried out on a real grand piano in order to understand further the coupling mechanism between the soundboard, the bridge and a piano string. From the measurement of transfer mobilities, it was found that an excitation in one direction might also induce a small response in the other direction. As a result, this might mean that the response of a piano string in one direction may be "cross-coupled" with response of the soundboard in the other direction.

- The dynamics of such a complicated system (piano strings, soundboard and bridge) were further explained through the use of simple theoretical mobilities for such built-up structures. Such a system may be approximated with the response of a grillage of beams mounted on plate, where two important frequency regions exist. At low frequencies, the system will act as a spring of a given stiffness and at higher frequencies, the point mobility will reach a constant average value approximating the point mobility of an infinite plate. Note that at high frequencies, a multi-modal behaviour of the soundboard will result in a broad, Statistical Energy Analysis (SEA) type behaviour. The latter argument further justifies our original hypothesis that the soundboard appears in the low quefrencies or low-time regions of the cepstrum as initially discussed in chapter 3 (Karatsovis *et al.*, 2006). These theoretical observations are backed by measuring the real point mobilities directly on the soundboard and the bridge of a grand piano.

6.1.3 Proposed signal model

The following can be concluded from *chapter 5* in relation to our proposed signal model:

- The proposed signal model may incorporate both the inharmonicity and the double decay rate of piano notes, as described earlier in chapters 3 and 4, based in a Bayesian formalism framework. Using the proposed model, it is possible to account for both monophonic and polyphonic music performances.

- The modelling of the inharmonicity factor forms the core of our proposed model by using a very simple formulation, where the harmonics are non-integer multiples of the fundamental frequency (or pitch) of the piano note.

- The use of the double decay rate can be incorporated in the model through a damping term for the different frequency components of the piano sound. It was proposed that depending on the time duration of the performed notes, the damping factor of the frequency components of a note may either represent the slow decay rate for played notes lasting long in a music performance, or the fast decay rate in the case of fast music passages.

- The rationale for our simple transcription method is based on the following conjecture: supposing that each note can be represented by a single inharmonicity factor B and a fundamental frequency f_0, then the joint extraction of these two parameters might help in uniquely characterising either single or multiple notes in a recording.

- The amplitudes of the sine and cosine components of the proposed Bayesian model are integrated out as nuisance parameters, whereas the damping decay rates of the sine and cosine components are set to zero. This is due to the fact that the data is analysed over a small time window of the order of 100 ms during which the decay of the amplitudes is expected to be very small.

- The computational route for estimating the desired parameter space, such as the fundamental frequency and inharmonicity factor of each transcribed piano note, was outlined by utilising the M-H algorithm and Gibbs sampler for multivariate parameter estimation. Examples of monophonic and polyphonic music transcription were presented and discussed with emphasis on “difficult” music intervals (octaves and perfect fifths), where there is a large number of commonly shared harmonics between two notes.

- The estimation of the inharmonicity factor from real polyphonic recordings may differ significantly from the estimate of real monophonic recordings. This may be due to the fact that in the case of a polyphonic transcription, more parameters are fitted over the same length of data and as a result the variability in the parameter estimation will be larger than in the case of a monophonic transcription. As a result, for large values of the inharmonicity of the order of 10^{-2} (typical highest value according to Fletcher, 1964), a relatively large variation in the estimation of the harmonics terms between the monophonically and polyphonically transcribed notes may result from this.

- The estimation of the inharmonicity factor B may vary across successive time windows of the same recording. This might be due to the fact that our proposed model might not accurately represent the time-varying characteristics of a real note throughout the recording (potentially an existence of a non-linear mechanism in the response). Note also that the biggest discrepancy is observed when analysing a section of the recording where the attack of the note is included. This is expected since our proposed model does not account for this feature.

- It was possible to compare our proposed model in this research with previous models in the literature (Godsill and Davy, 2002; Davy and Godsill, 2002) in terms of their computational expense. Indeed, the above models in the literature require the estimation of a different inharmonicity, or so-called "de-tuning" parameter, for each harmonic component of a single note. Hence, for N number of terms, there will be N de-tuning parameters that need to be computed, whereas, in this research, we propose a model with a single inharmonicity factor for all N terms.

- In general, the transcription examples presented in this research were successful, but the transcription model is simple and has not been generalised for an unknown number of notes present in a recording. However, this generalisation is possible in practice with the implementation of a Reversible Jump Markov Chain Monte Carlo (RJMCMC) method, where the dimensionality of the model may vary (Green, 1995).

6.2 Future work

The following can be proposed for future work following our investigations on the acoustic features of piano sounds:

- This work was focused in describing the unique "imperfections" in piano sounds, which may form the basis of a signal model in a music transcription method. In particular, the inharmonicity, which is associated with the bending stiffness of the piano strings, formed the main core of this research work underlying the importance of an accurate and more comprehensive signal model. This concept could further be extended to other instruments provided other similar imperfections are identified and subsequently modelled. Therefore, despite our proposed signal model is potentially more computationally attractive than existing models in the literature, it is yet less generic since it only covers the piano instrument.

- The results following the automatic music transcription of music intervals seem to be encouraging, although the transcription model presented in this research is simple and has not been generalised for an unknown number of notes present in a recording. This generalisation is possible with the development of an appropriate RJMCMC method, where the dimensionality of the parameter space may vary. In particular, the latter method is a variant of the MCMC method, where a model indicator is introduced and the resulting Markov chain simulations may jump between models of different dimensions by forming samples from the posterior density estimates.

- There seems to be a discrepancy in the computation of the inharmonicity factor between monophonically and polyphonically transcribed notes. This is believed to be associated with the fact that in the case of a polyphonic transcription, more parameters are fitted over the same length of data and as a result the variability in the parameter estimation will be larger than in the case of a monophonic transcription. However, there may be other reasons behind this computational discrepancy, which have not been addressed in this research. More work needs to be carried out in this field to understand, or support, the reason for such a discrepancy.

- Music transcription is the process of converting a live or recorded performance into a written score, where other types of information are required apart from the pitch of the notes, such as the duration, the tempo and the dynamics of the played notes. The transcription task in this research has been restricted to the identification of only two parameters; the fundamental frequency and the inharmonicity factor of single or multiple notes. Therefore, more effort should be placed in the future in order to transcribe other meaningful musical information.

- The current research work has only considered the transcription of a few music intervals between the middle and the high frequency register of the piano. A more extensive frequency range should be considered in future work to cover a broader range of notes. Also, a more comprehensive 'bank' of audio information is proposed in order to test the model for a larger number of different music interval combinations.

- The transcription of 2-note mixture, which is the simplest form of polyphonic music, should further be extended to a higher number of note mixtures, e.g. 4, 6 or even 8-note mixtures to represent more appropriately performances of real polyphonic music pieces.

- The current proposed model does not account for different music instruments that might be playing together at any given time, which is typical in an orchestral piece (this task is also related to instrument classification).

- Other, perhaps more important, non-musical applications may be considered using similar probabilistic Bayesian frameworks of analysis, such as the source separation of signals in hearing aids. Analogously, the principle is the same since in the automatic music transcription one is interested in separating the individual notes of a polyphonic music recording.

Appendix A

Bayesian formalism

It is useful to define data in terms of a linear combination of a basis function with a Gaussian noise component when this is possible. This is known as the general linear model of a signal and can be written in the matrix form

$$\mathbf{d} = \mathbf{G}\mathbf{b} + \mathbf{e}$$

where **d** is an $M \times 1$ matrix containing the data points of the signal, **G** is an $M \times L$ matrix of the basis functions containing the parameters of the signal, **b** is an $L \times 1$ matrix containing any linear coefficient corresponding to each particular column of **G** matrix, and **e** is an $M \times 1$ matrix containing random Gaussian noise components.

The advantage of using the general linear model in signal analysis is that the parameters of a signal can be inferred when combined with common probabilistic methods such as Bayesian analysis, maximum likelihood estimation and MCMC methods.

A method of estimating the parameters of the **G** matrix, consisting of a parameter set of $\boldsymbol{\omega}$, of the model is to develop a method that incorporates any information on the likelihood function and any information on the state of knowledge about the parameters before the data is observed. The latter descriptor is called the prior probability density function and is supposed to represent the user's state of uncertainty about the parameter vector $\boldsymbol{\omega}$ of the signal. So, one could obtain the posterior density function that describes the data after being deduced. This is the basis of Bayesian analysis.

In the case of the general linear model, Bayesian formalism can be summarised into a single expression

$$p(\boldsymbol{\omega}, \mathbf{b}, \sigma | \mathbf{d}) = \frac{p(\mathbf{d} | \boldsymbol{\omega}, \mathbf{b}, \sigma)\, p(\boldsymbol{\omega}, \mathbf{b}, \sigma)}{p(\mathbf{d})} \qquad \textbf{(A – 1)}$$

where $p(\mathbf{d}|\boldsymbol{\omega},\mathbf{b},\sigma)$ is the chosen likelihood function, $p(\boldsymbol{\omega},\mathbf{b},\sigma)$ conveys the prior knowledge about the probable values of the parameters before the data is observed, $p(\mathbf{d})$ is called the evidence and only has a normalising effect, and $p(\boldsymbol{\omega},\mathbf{b},\sigma|\mathbf{d})$ is the joint posterior density of $\boldsymbol{\omega}$, $\mathbf{b}$ and σ given $\mathbf{d}$.

Note that σ denotes the standard deviation of the Gaussian noise component induced into the signal.

The likelihood function can mathematically be defined as follows

$$p(\boldsymbol{\omega},\mathbf{b},\sigma;\mathbf{d}) = p(\mathbf{d}|\boldsymbol{\omega},\mathbf{b},\sigma) \qquad \textbf{(A – 2)}$$

For additive noise, the above equation can be written as

$$p(\mathbf{d}|\boldsymbol{\omega},\mathbf{b},\sigma) = p(\mathbf{e}) \qquad \textbf{(A – 3)}$$

Also, if $\mathbf{e}$ is considered to be composed of many independent identical distributed components M, then by summing them, equation **(A – 3)**, for a Gaussian process, can be written as follows

$$p(\mathbf{d}|\boldsymbol{\omega},\mathbf{b},\sigma) = p(\mathbf{e}) = (2\pi\sigma^2)^{-\frac{M}{2}} \exp\left[-\frac{\sum_{i=1}^{M} e_i^2}{2\sigma^2}\right] \qquad \textbf{(A – 4)}$$

$$p(\mathbf{d}|\boldsymbol{\omega},\mathbf{b},\sigma) = (2\pi\sigma^2)^{-\frac{M}{2}} \exp\left[-\frac{(\mathbf{d}-\mathbf{G}\mathbf{b})^T(\mathbf{d}-\mathbf{G}\mathbf{b})}{2\sigma^2}\right] \qquad \textbf{(A – 5)}$$

From equation **(A – 2)** there are certain parameters that are of no interest to us. In particular, σ and $\mathbf{b}$ are generally of little importance and one aims to remove them from the posterior probability density function to yield information solely on the parameters $\boldsymbol{\omega}$ of the $\mathbf{G}$ matrix. These unwanted parameters are called nuisance parameters and can be integrated out via a marginalisation procedure.

One can integrate the posterior density function with respect to $\mathbf{b}$ and σ, so that only the set of parameters $\boldsymbol{\omega}$ remains

$$p(\boldsymbol{\omega}|\mathbf{d}) = \iint p(\boldsymbol{\omega}, \mathbf{b}, \sigma|\mathbf{d})\mathbf{db}\mathrm{d}\sigma \qquad \textbf{(A – 6)}$$

Assuming also that the probability density function can be written as

$$p(\boldsymbol{\omega}, \mathbf{b}, \sigma) = p(\boldsymbol{\omega})\,p(\mathbf{b})\,p(\sigma) \qquad \textbf{(A – 7)}$$

Equation **(A – 1)** due to equation **(A – 6)** and **(A – 7)** becomes

$$p(\boldsymbol{\omega}|\mathbf{d}) = \iint \frac{p(\mathbf{d}|\boldsymbol{\omega}, \mathbf{b}, \sigma)\,p(\boldsymbol{\omega})\,p(\mathbf{b})\,p(\sigma)}{p(\mathbf{d})}\mathbf{db}\mathrm{d}\sigma \qquad \textbf{(A – 8)}$$

Equation **(A – 8)** due to **(A – 5)** becomes

$$p(\boldsymbol{\omega}|\mathbf{d}) = \frac{p(\boldsymbol{\omega})}{p(\mathbf{d})}\iint (2\pi\sigma^2)^{\frac{-M}{2}}\, e^{\frac{-(\mathbf{d}-\mathbf{Gb})^T(\mathbf{d}-\mathbf{Gb})}{2\sigma^2}}\, p(\mathbf{b})\,p(\sigma)\mathbf{db}d\sigma$$

Solving the above indefinite integral by assigning Jeffrey's prior to σ and uniform priors to **b**, which is an $L\times 1$ matrix, one can obtain the following expression for an estimate of the posterior density of ω (Ó Ruanaidh and Fitzgerald, 1996)

$$p(\omega|\mathbf{d}) \propto \frac{\left[\mathbf{d}^{\mathbf{T}}\mathbf{d} - \mathbf{d}^{\mathbf{T}}\mathbf{G}(\mathbf{G}^{\mathbf{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathbf{T}}\mathbf{d}\right]^{\frac{L-M}{2}}}{\sqrt{\det(\mathbf{G}^{\mathbf{T}}\mathbf{G})}} \qquad \textbf{(A – 9)}$$

The above expression is a student's t-distribution probability function and peaks at the most probable value of ω of the **G** matrix. Note also that the shape of the marginal density is more important than its size and there is no need to find the constant of proportionality in the above equation. The probability density will peak at the most probable value of the parameters of the **G** matrix of the signal irrespective of any multiplicative factor.

The above expression **(A – 9)** can be used to obtain probability densities for a pair of successive states in a "Markov" chain, where the next probable value Y of the desired parameter is dependant upon its previous value X. The acceptance function of whether the next value is accepted or rejected can be defined by the *ratio* of the t-distribution probability densities functions, $p(X)$ and $p(Y)$, as follows

$$Q(X,Y) = \frac{p(Y)}{p(X)}$$

where Q is known as the acceptance function and depending on the returning value of the ratio, the next value is either accepted or rejected. Note that in the case of implementing the Metropolis-Hastings (M-H) algorithm for estimating a parameter space, probable values for the next states of the Marlov chain could randomly be drawn using a sample space with random Gaussian distribution having zero mean and standard deviation one. So, if for example, the ratio Q of the acceptance function is equal to or over unity, then this means that the randomly drawn estimate matches closely the statistical properties of the data (or the signal), therefore its drawn value is accepted and the next state of the Markov chain would move closer to target parameter value of the signal.

Appendix B

Tuning on the equal temperament scale

Nowadays in western music the use of the "equal temperament" scale is universal. This consists of twelve equal semitones, which make up an octave. The frequency ratio of the lower note in the octave with the corresponding higher note in the octave would have a ratio of 1:2. The smallest musical interval k on the equal temperament scale would be

$$k = \sqrt[12]{2} = 1.0595 \qquad \textbf{(B – 1)}$$

This forms the basis of tuning on the equal temperament scale. With such a method, organs can exactly be tuned, whilst in the case of the piano or the harpsichord, where the inharmonicity of the strings plays an important role, the tuned musical intervals should be "stretched" further as it will be discussed later.

So, the general process of tuning to the equal temperament scale is based around the "circle of fifths". A fifth musical interval, or a ratio of 3:2, consists of seven semitones and therefore that would correspond to a ratio of

$$(1.059)^7 = 1.4983 \qquad \textbf{(B – 2)}$$

The tuning of an octave follows the process below:

- C_4, which is middle C[5], is set to, say, 260.97 Hz by using a tuning fork. The higher note of the fifth interval would be G_4, which corresponds to 391.01 Hz (260.97 x 1.059^7). Based on Helmholtz's theory, second order beats will be heard when the frequencies of two tones f_1 and f_2 are similar but not exactly the same. In particular, if $f_2 = \frac{n}{m} f_1 + \delta$, then $m\delta$ beats would

[5] The numeric subscript index refers to the note of an associated octave on a music keyboard. Higher notes on the keyboard will be associated with higher octaves on the keyboard.

occur per second. Therefore, in the case of a mistuned fifth, as in the case of C_4 with G_4

$$\left\{ \begin{array}{l} f_2 = \frac{3}{2} f_1 + 2\delta \\ \\ \frac{f_2}{f_1} = 1.4983 \end{array} \right\} \Leftrightarrow \delta = 0.0017 f_1 \qquad \textbf{(B – 3)}$$

Therefore, $0.0034 f_1$ beats can be heard per second.

So, with $f_1 = 260.97$, the rate of beats would be about 0.89 Hz or 8.9 beats per 10 seconds.

- The next fifth interval would be G_4 to D_5 with a rate of beat of 1.33 beats per second.

- After tuning to D_5, then we tune an octave lower, i.e. for the musical interval of D_4 to A_4.

- The circle of fifths would eventually stop at C_5, having tuned the whole musical scale from C_4 to C_5.

The complete circle of fifths would therefore follow the tuning succession as below:

Fifth Interval	Beat Frequency [Hz]
C_4 - G_4	0.8873
G_4 - D_5	1.3294
D_4 - A_4	0.9960
A_4 - E_5	1.4923
E_4 - B_4	1.1179
B_3 - F_4#	0.8375
F_4# - C_5#	1.2548
C_4# - G_4#	0.9401
G_4# - D_5#	1.4085
D_4# - A_4#	1.0552
A_4# - F_5	1.5810
F_4 - C_5	1.1844

Table B.1 – Tuning intervals and associated beat frequencies

Over the tuning process a metronome can be used to count beats. For example, when counting the beats for the musical interval C_4 to G_4 the metronome can be set to 0.8873 x 60 = 53 ticks per minute.

The tuning of the piano, based on the method as just described, is not satisfactory to the trained ear of a musician, because of the inharmonicity of the strings as mentioned earlier. The tuning process then becomes much more complicated and lies beyond the scope of this work. However as a general rule, when tuning in fifths in a piano, the intervals are "stretched" in order to minimise beating between notes. The similar applies to the octave tuning method in a piano. So, if the higher note on the octave is tuned at exactly twice the fundamental frequency of the lower one, then beating will occur since their harmonics will not be an exact match due to the inharmonicity factor associated with the strings. In order to avoid this, the higher note is tuned to the first harmonic of the lower note (Ortiz-Berenguer *et al.*, 2005). As a result, the octave is "stretched" above its "well-tempered" value but the beating effect is minimised to a level, which is satisfactory providing a well-defined and clear sound.

Finally, a typical frequency range of a piano instrument is shown below, where all the 88 notes are associated with a "pitch" or a fundamental frequency. In the case of the piano instrument, the pitch may typically vary from 27.5 Hz up to about 4,186 Hz as shown in Figure B.1:

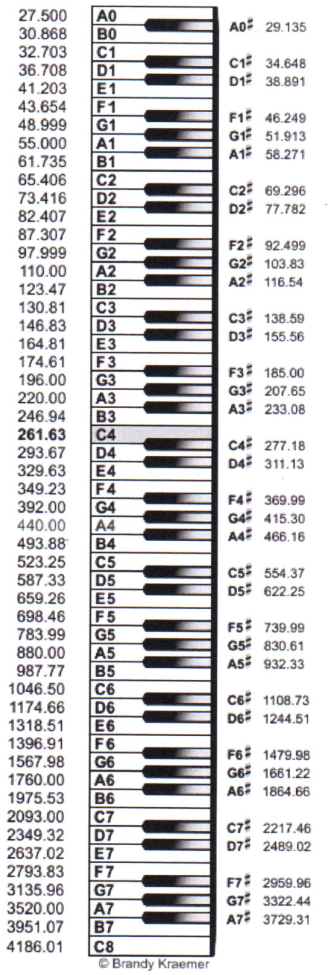


Figure B.1 – Piano keyboard range as a function of the fundamental frequency (reproduced from "about.com" in relation to the scientific pitch notation of the piano)

Appendix C

Roots of quadratic equation

One is interested in obtaining the roots of $\Delta(s)$ as follows

$$\Delta(s) = \alpha(s)\gamma(s) - \beta(s)\delta(s) = 0$$

$$\Leftrightarrow$$

$$\left[(s^2 + 2\zeta_1\omega_1 s + \omega_1^2) + (2\zeta_3\omega_3 s + \omega_3^2)\right]\cdot\left[(s^2 + 2\zeta_2\omega_2 s + \omega_2^2) + (2\zeta_4\omega_4 s + \omega_4^2)\right] - \left[2\zeta_3\omega_3 s + \omega_3^2\right]\cdot\left[2\zeta_4\omega_4 s + \omega_4^2\right] = 0$$

By expanding the above expression and re-arranging the terms in a descending order of a 4th order polynomial of the form

$$A_1 s^4 + A_2 s^3 + A_3 s^2 + A_4 s + A_5 = 0 \qquad \textbf{(C – 1)}$$

One obtains

$$s^4 + \ldots$$

$$\ldots + \left[2\zeta_2\omega_2 + 2\zeta_4\omega_4 + 2\zeta_1\omega_1 + 2\zeta_3\omega_3\right]s^3 + \ldots$$

$$\ldots + \left[\omega_2^2 + \omega_4^2 + 4\zeta_1\omega_1\zeta_2\omega_2 + 4\zeta_1\omega_1\zeta_4\omega_4 + \omega_1^2 + 4\zeta_3\omega_3\zeta_2\omega_2 + 4\zeta_3\omega_3\zeta_4\omega_4 + \omega_3^2 - 4\zeta_3\omega_3\zeta_4\omega_4\right]s^2 + \ldots$$

$$\ldots + \left[2\zeta_1\omega_1\omega_2^2 + 2\zeta_1\omega_1\omega_4^2 + 2\zeta_2\omega_2\omega_1^2 + 2\zeta_4\omega_4\omega_1^2 + 2\zeta_3\omega_3\omega_2^2 + 2\zeta_3\omega_3\omega_4^2 + 2\zeta_2\omega_2\omega_3^2 + 2\zeta_4\omega_4\omega_3^2 - 2\zeta_3\omega_3\omega_4^2 - 2\omega_3^2\zeta_4\omega_4\right]s + \ldots$$

$$\ldots + \left[\omega_1^2\omega_2^2 + \omega_1^2\omega_4^2 + \omega_3^2\omega_2^2\right] = 0$$

Hence, one can solve equation **(C – 1)** numerically to deduce the natural frequencies and damping factors of a coupled dynamical system.

Appendix D

Alternative modelling of a pair of piano strings coupled with bridge

An alternative modelling approach for free vibration is presented here by working out the boundary conditions for the respective equations of motion of the beam and the two strings. The beam is pinned-pinned and the strings are fixed on one end:

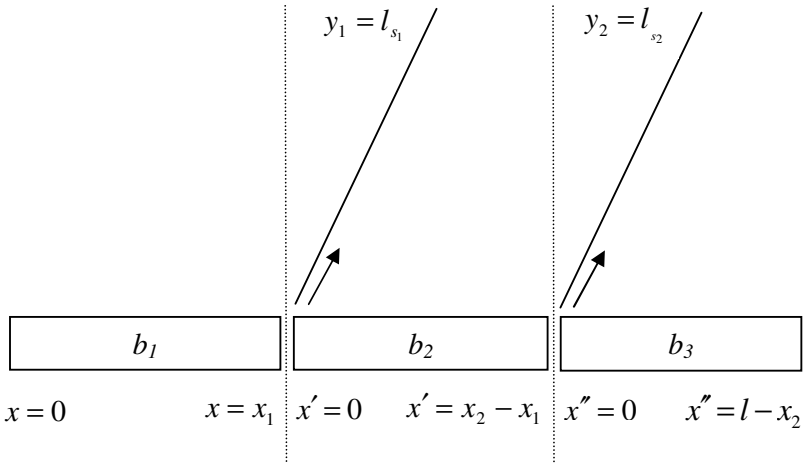


Figure D.1 – Two strings coupled with a beam

The equations of motion for the three-part beam and the strings would be the following:

For the three-part beam

$$W_{b_1}(x) = A_1 e^{j(\omega t + kx)} + A_2 e^{j(\omega t - kx)} + A_3 e^{j\omega t} e^{-kx} + A_4 e^{j\omega t} e^{k(x - x_1)}$$

$$W_{b_2}(x') = B_1 e^{j(\omega t + kx')} + B_2 e^{j(\omega t - kx')} + B_3 e^{j\omega t} e^{-kx'} + B_4 e^{j\omega t} e^{k(x' - x_2 + x_1)}$$

where $x' = x - x_1$

$$W_{b_3}(x'') = C_1 e^{j(\omega t + kx'')} + C_2 e^{j(\omega t - kx'')} + C_3 e^{j\omega t} e^{-kx''} + C_4 e^{j\omega t} e^{k(x'' - \ell + x_2)}$$

where $x'' = x - x_2$

For the two strings

$$W_{s_1}(y_1) = D_1 e^{j(\omega t - k_{s_1} y_1)} + D_2 e^{j(\omega t + k_{s_1} y_1)}$$

$$W_{s_2}(y_2) = E_1 e^{j(\omega t - k_{s_2} y_2)} + E_2 e^{j(\omega t + k_{s_2} y_2)}$$

16 boundary conditions for these equations were identified

$$\rightarrow W_{b_1}\Big|_{x=x_1} = W_{b_2}\Big|_{x'=0} \qquad \textbf{(D – 1)}$$

$$\rightarrow W_{b_2}\Big|_{x'=x_2-x_1} = W_{b_3}\Big|_{x''=0} \qquad \textbf{(D – 2)}$$

$$\rightarrow W_{b_2}\Big|_{x'=0} = W_{s_1}\Big|_{y_1=0} \qquad \textbf{(D – 3)}$$

$$\rightarrow W_{b_3}\Big|_{x''=0} = W_{s_2}\Big|_{y_2=0} \qquad \textbf{(D – 4)}$$

$$\rightarrow \frac{\partial w_{b_1}}{\partial x}\Big|_{x=x_1} = \frac{\partial w_{b_2}}{\partial x'}\Big|_{x'=0} \qquad \textbf{(D – 5)}$$

$$\rightarrow \frac{\partial w_{b_2}}{\partial x'}\Big|_{x'=x_2-x_1} = \frac{\partial w_{b_3}}{\partial x''}\Big|_{x''=0} \qquad \textbf{(D – 6)}$$

$$\rightarrow M_{b_1}\Big|_{x=x_1} = M_{b_2}\Big|_{x'=0} \qquad \textbf{(D – 7)}$$

$$\rightarrow M_{b_2}\Big|_{x'=x_2-x_1} = M_{b_3}\Big|_{x''=0} \qquad \textbf{(D – 8)}$$

$$\rightarrow W_{b_1}\Big|_{x=0} = 0 \qquad \textbf{(D – 9)}$$

$$\rightarrow W_{b_3}\Big|_{x''=l-x_2} = 0 \qquad \textbf{(D – 10)}$$

$$\rightarrow W_{s_1}\Big|_{y_1=l_{s_1}} = 0 \qquad \textbf{(D – 11)}$$

$$\rightarrow W_{s_2}\Big|_{y_2=l_{s_2}} = 0 \qquad \textbf{(D – 12)}$$

$$\rightarrow M_{b_1}\Big|_{x=0} = 0 \qquad \textbf{(D – 13)}$$

$$\rightarrow M_{b_3}\Big|_{x''=l-x_2} = 0 \qquad \textbf{(D – 14)}$$

$$\rightarrow Q_{b_2}\Big|_{x'=0} - Q_{b_1}\Big|_{x=x_1} = T_{s_1}\frac{\partial W_{s_1}}{\partial y_1}\Bigg|_{y_1=0} \qquad \textbf{(D – 15)}$$

$$\rightarrow Q_{b_3}\Big|_{x''=0} - Q_{b_2}\Big|_{x'=x_2-x_1} = T_{s_2}\frac{\partial W_{s_2}}{\partial y_2}\Bigg|_{y_2=0} \qquad \textbf{(D – 16)}$$

Equation **(D – 1)** gives

$$A_1e^{jkx_1} + A_2e^{-jkx_1} + A_3e^{-kx_1} + A_4 - B_1 - B_2 - B_3 - B_4e^{k(-x_2+x_1)} = 0$$

Equation **(D – 2)** gives

$$B_1e^{jk(x_2-x_1)} + B_2e^{-jk(x_2-x_1)} + B_3e^{-k(x_2-x_1)} + B_4 - C_1 - C_2 - C_3 - C_4e^{k(-\ell+x_2)} = 0$$

Equation **(D – 3)** gives

$$B_1 + B_2 + B_3 + B_4e^{k(-x_2+x_1)} - D_1 - D_2 = 0$$

Equation **(D – 4)** gives

$$C_1 + C_2 + C_3 + C_4e^{k(-\ell+x_2)} - E_1 - E_2 = 0$$

Equation **(D – 5)** gives

$$jA_1e^{jkx_1} - jA_2e^{-jkx_1} - A_3e^{-kx_1} + A_4 - jB_1 + jB_2 + B_3 - B_4e^{k(-x_2+x_1)} = 0$$

Equation **(D – 6)** gives

$$jB_1e^{jk(x_2-x_1)} - jB_2e^{-jk(x_2-x_1)} - B_3e^{-k(x_2-x_1)} + B_4 - jC_1 + jC_2 + C_3 - C_4e^{k(-\ell+x_2)} = 0$$

Equation **(D – 7)** gives

$$-A_1 e^{jkx_1} - A_2 e^{-jkx_1} + A_3 e^{-kx_1} + A_4 + B_1 + B_2 - B_3 - B_4 e^{k(-x_2+x_1)}$$

Equation **(D – 8)** gives

$$-B_1 e^{jk(x_2-x_1)} - B_2 e^{-jk(x_2-x_1)} + B_3 e^{-k(x_2-x_1)} + B_4 + C_1 + C_2 - C_3 - C_4 e^{k(-\ell+x_2)} = 0$$

Equation **(D – 9)** gives

$$A_1 + A_2 + A_3 + A_4 e^{-kx_1} = 0$$

Equation **(D – 10)** gives

$$C_1 e^{jk(\ell-x_2)} + C_2 e^{-jk(\ell-x_2)} + C_3 e^{-k(\ell-x_2)} + C_4 = 0$$

Equation **(D – 11)** gives

$$D_1 e^{-jk_{s_1}\ell_{s_1}} + D_2 e^{jk_{s_1}\ell_{s_1}} = 0$$

Equation **(D – 12)** gives

$$E_1 e^{-jk_{s_2}\ell_{s_2}} + E_2 e^{jk_{s_2}\ell_{s_2}} = 0$$

Equation **(D – 13)** gives

$$-A_1 - A_2 + A_3 + A_4 e^{-kx_1} = 0$$

Equation **(D – 14)** gives

$$-C_1 e^{jk(\ell-x_2)} - C_2 e^{-jk(\ell-x_2)} + C_3 e^{-k(\ell-x_2)} + C_4 = 0$$

Equation **(D – 15)** gives

$$EI\left[jk^3 A_1 e^{jkx_1} - jk^3 A_2 e^{-jkx_1} + k^3 A_3 e^{-kx_1} - k^3 A_4 - jk^3 B_1 + jk^3 B_2 - k^3 B_3 + k^3 B_4 e^{k(-x_2+x_1)} + T_{s_1} jk_{s_1} D_1 - T_{s_1} jk_{s_1} D_2\right] = 0$$

Equation **(D – 16)** gives

$$EI\left[-jk^3 C_1 + jk^3 C_2 - k^3 C_3 + k^3 C_4 e^{k(-\ell+x_2)} + jk^3 B_1 e^{jk(x_2-x_1)} - jk^3 B_2 e^{-jk(x_2-x_1)} + k^3 B_3 e^{-k(x_2-x_1)} - k^3 B_4\right] + T_{s_2} jk_{s_2} E_1 - T_{s_2} jk_{s_2} E_2 = 0$$

Therefore, one can now solve for

$$\det(\mathbf{A}) = 0$$

where **A** matrix is formed from equations **(D – 1)** to **(D – 16)**, in order to obtain the natural frequencies of the coupled dynamical system in question.

References

1. Abdallah, S. A. and Plumbley, M. D., "Unsupervised analysis of polyphonic music by sparse coding", IEEE transactions on neural networks, Vol. 17 (1), 179-196 (2006).

2. Alder, B. J. and Wainwright, T. E., "Studies in molecular dynamics I: general method", Journal of Chemical Physics, 31, 459-466 (1959).

3. Alm, J. F. and Walker, J. S., "Time-frequency analysis of musical instruments", SIAM Review, Vol. 44 (3), 457-476 (2002).

4. Andrieu, C. and Doucet, A., "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC", IEEE Transactions on Signal Processing, Vol. 47 (10), 2667-2676 (1999).

5. Aramaki, M., Bensa, J., Daudet. L., Guillemain, P. and Kronland-Martinet, R., "Resynthesis of coupled piano string vibrations based on physical modeling", Journal of New Music Research, Vol. 30 (3), 213-226 (2001).

6. Bello, J. P., Monti, G. and Sandler, M., "Techniques for automatic music transcription", International Symposium on Music Information Retrieval (2000).

7. Bensa, J., Bilbao, S., Kronland-Martinet, R. and Smith III, J. O., "The simulation of piano string vibration: From physical models to finite difference schemes and digital waveguides", J. Acoust. Soc. Am, 114 (2), August (2003).

8. Bensa, J., Bilbao, S., Kronland-Martinet, R. and Smith III, J. O. and Voinier, T., "Computational modelling of stiff piano strings using digital waveguides and finite differences", Acta Acustica United with Acustica, Vol. 91, 289-298 (2005).

9. Blevins, R. D., "Formulas for natural frequency and mode shapes", Litton Educational Publishing, Inc. (1979).

10. Boutillon, X., "Model for piano for piano hammers: Experimental determination and digital simulation", J. Acoust. Soc. Am. 83(2), 746-754, February (1988).

11. Boutillon, X. and Weinreich, G., "Three-dimensional mechanical admittance: Theory and new measurement method applied to the violin bridge", J. Acoust. Soc. Am. 105 (6), 3524-3533, June (1999).

12. Capleton, B., "False beats in coupled piano string unisons", J. Acoust. Soc. Am. 115 (2), 885-892, February (2004).

13. Cemgil, A. T., Godsill, S. J., Peeling, P. and Whiteley, N., "Bayesian statistical methods for audio and music processing", University of Cambridge, Technical Report (2008).

14. Cemgil, A. T., Kappen, B. and Barber, D., "Generative model based polyphonic music transcription", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, October 19-23 (2003).

15. Cheveigné, de A., "Pitch perception models", from the book "Pitch: neural coding and perception", edited by Plack, C. J., Oxenham, A. J., Fay, R. R. and Popper, A. N., published by Springer, New York (2005).

16. Chien, Y.-R. and Jeng, S.-K., "An automatic transcription system with octave detection", IEEE International Conference on Acoustics and Signal Processing (ICASSP), Vol. 2, 1865-1868 (2002).

17. Corradi, R., Fazioli, P., Miccoli, S., Ripamonti, F. and Squicciarini, G., "Experimental modal analysis and structural modelling of a grand piano soundboard to support instrument design", 10th International Conference, RASD, Southampton (2010).

18. Davy, M. and Godsill, S. J., "Bayesian harmonic models for musical signal analysis", University of Cambridge, Technical Report (2002).

19. ESDU: "Mobilities and impedances of structures, part II: compendium of point of infinite structures", March (2004).

20. Fahy, F. and Walker, J. (editors), "Fundamentals of noise and vibration", E & FN Spon, London (1998).

21. Fletcher, H. "Normal vibration frequencies of a stiff piano", J. Acoust. Soc. Am. 36(1), 203-209 (1964).

22. Gardonio, P. and Brennan, M. J., "Mobility and impedance methods in structural vibration", from the book "Advanced applications in acoustics, noise and vibration", edited by Fahy, F. and Walker, J., published by Taylor and Francis (2004).

23. Gelfand, S. and Smith, A. F. M., "Sampling based approaches to calculating marginal densities", Journal of the American Statistical Association, 85, 398-409 (1990).

24. Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Trans. On Pattern Analysis and Machine Intelligence, PAMI-6, 721-741 (1984).

25. Giordano, N., "Simple model of a piano soundboard", J. Acoust. Soc. Am. 102 (2), Pt. 1, 1159-1168 (1997).

26. Giordano, N., "Sound production by a vibrating piano soundboard: experiment", J. Acoust. Soc. Am. 104 (3), Pt. 1, 1648-1653 (1998).

27. Giordano, N., "Mechanical impedance of a piano soundboard", J. Acoust. Soc. Am. 103 (4), 2128-2133 (1998).

28. Giordano, N. and Jiang, M., "Physical modeling of the piano", EURASIP Journal on Applied Signal Processing, Vol. 7, 926-933 (2004).

29. Godsill, S. and Davy, M., "Bayesian on harmonic models for musical pitch estimation and analysis", IEEE conference on Acoustics, Speech and Signal Processing (ICASSP), Orlando, Florida, USA, II-1769 to II-1772 (2002).

30. Green, P., "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", Biometrika, 82, 711-732 (1995).

31. Hundley, T. C., Benioff, H. and Martin, D. W., "Factors contributing to the multiple rate of piano tone decay", J. Acoust. Soc. Am., Vol. 64 (5), 1303-1309 (1978).

32. Karatsovis, C., White, P. R. and Hammond, J. K., "Cepstral analysis of piano notes", Institute of Acoustics Spring Conference: Futures in Acoustics: Today's Research – Tomorrow's Careers, Southampton, UK, 623-632 (2006).

33. Klapuri, A. and Virtanen, T., "Automatic music transcription", from the book "Handbook of Signal Processing in Acoustics", edited by Havelock, D., Kuwano, S. and Vorländer. M., published by Springer (2008).

34. Le Carrou, J.-L., Gautier, F., Dauchez, N. and Gilbert, J., "Modelling of sympathetic string vibrations", Acta Acustica United with Acustica, Vol. 91, 227-288 (2005).

35. Leistikow, R. J., Thonburg, H. D., Smith III, J. O. and Berger, J., "Bayesian identification of closely-spaced chords from single-frame STFT peaks", Proc. Of the 7th Int. Conference on Digital Audio Effects (DAFx'04), Naples, Italy, October 2008, 228-233 (2004).

36. Marolt, M., "Networks of adaptive oscillators for partial tracking and transcription of music recordings", Journal of New Music Research. Vol. 33 (1), 49-59, (2004).

37. Meddis, R. and Hewitt, M. J., "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification", J. Acoust. Soc. Am. Vol. 89 (6), 2866-2882 (1991).

38. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., "Equation of state calculations by fast computing machines". Journal of Chemical Physics, 21, 1087-1092 (1953).

39. Moore, T. R. and Zietlow, A., "Interferometric studies of a piano soundboard", J. Acoust. Soc. Am. 119 (3), 1783-1793 (2006).

40. Nakamura, I., "Fundamental theory and computer simulation concerning the decay characteristics of piano sound", J. Acoust. Soc. Am., Vol. 84 (S1), S135, (1988).

41. Naganuma, D. and Nakamura, I., "Consideration on the two-dimensional vibration of single piano strings", Proc. ISMA 2004, Japan, 223-226 (2004).

42. Nishiguchi, I., "Recent research on the acoustics of pianos", Acoust. Sci. & Tech., Vol. 25 (6), 413-418 (2004).

43. Nishiguchi, I. and Sasaki, M., "Numerical analyses of a piano string by a three dimensional large deformation theory", Proc. ISMA 2004, Nara, Japan, 223-226 (2004).

44. Nishiguchi, I., Sasaki, M. and Kato, K., "Three dimensional motion of a piano string analyzed by FEM", Proc. Spring Meet. Acoust. Soc. Jpn., 831-832 (2003).

45. Ortiz-Berenguer, L. I., Casajús-Quirós. F. J., Blanco-Martin, E. and Ibanez-Cuenca, D., "Modelling of piano sounds using FEM simulation of soundboard vibration", Acoustics Paris Conference, 6251-6256 (2008).

46. Ortiz-Berenguer, L. I., Casajús-Quirós. F. J. and Torres-Guijarro, S., "Multiple piano note identification using a spectral matching method with derived patterns", J. Audio Eng. Soc., Vol. 53 (1/2), 32-43 (2005).

47. Ó Ruanaidh, J. J. K. and Fitzgerald, W. J., "Numerical Bayesian methods applied to signal processing", Springler-Verlag, New York, Inc. (1996).

48. Peeling, P., Li, C.-F. and Godsill, S., "Poisson point process modeling for polyphonic music transcription", J. Acoust. Soc. Am. 121 (4), 168-175, April (2007).

49. Pertusa, A. and Inesta, J. M., "Polyphonic monotimbral music transcription using dynamic networks", Pattern Recognition Letters, Vol. 26, 1809-1818 (2005).

50. Pinnington, R. J., "Approximate mobilities of built up structures", ISVR technical report no. 162 (1988).

51. Poliner, G. E. and Ellis, D. P. W., "A discriminative model for polyphonic piano transcription", EURASIP Journal on Advances in Signal Processing, Article ID 48317, 9 pages (2006).

52. Priestley, M. B., "Spectral series and time series", Academic Press, London, UK (1981).

53. Rossing, T. D., "The science of sound", 2nd edition Addison-Wesley Publishing Company, 287-295 (1990).

54. Smith III, J. O. and Van Duyne, S. A., "Commuted Piano Synthesis", Proceedings of the ICMC-95, 319-926 (1995).

55. Tanaka, H., Mizutani, K. and Nagai, K., "Experimental analysis of two-dimensional vibration of a piano string measured with an optical device", J. Acoust. Soc. Am., Vol. 105 (2), 1181 (1999).

56. Taylor, C., A., "The physics of musical sounds", English Universities Press, London (1965).

57. Triki, M. and Slock, D. T. M., "Periodic signal extraction with global amplitude and phase modulation for music signal decomposition", IEEE Proceedings of ICASSP, 233-236 (2005).

58. Walmsley, P. J., Godsill. S. J. and Rayner, P. J. W., "Multidimensional optimisation of harmonic signals", Proceedings of EUSIPCO (1998).

59. Walmsley, P. J., Godsill, S. J. and Rayner, P. J. W., "Bayesian modelling of harmonic signals for polyphonic music tracking", Cambridge Music Processing Colloquium, 1-6 (1999).

60. Walmsley, P. J., Godsill, S. J. and Rayner, P. J. W., " Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters", IEEE Proceedings of Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, W99-1 to W99-4 (1999).

61. Weinreich, G., "Coupled piano strings", J. Acoust. Soc. Am., Vol. 62 (6), 1474-1484 (1977).

62. Zhou, G. and Giannakis, G., "Polyspectral analysis of mixed processes and coupled harmonics", IEEE Transactions on Information Theory, Vol. 42 (3), 943-958, May (1996).