APPLICATIONS OF WAVELETS TO ANALYSIS OF PIANO TONES

WANG ENBO

(B.Sci and M.Sci, Wuhan University, Wuhan, China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

I am deeply grateful to my supervisor, Prof. Tan B.T.G, for his kind guidance and assistance during the course of my research at National University of Singapore. It has been a great honor and privilege for me to study with him. The supports I received from Prof. Tan in both the signal processing and the computer music has been greatly instrumental in this research effort.

I would also like to thank my family for all the patience and support they have shown during this time.

Table of Contents

Acknowledgements 1

Table of Contents 2

Summary 4

List of Figures 6

List of Tables 10

Chapter 1 Introduction 1

1.1 Musical Acoustics and Computer Music 1
1.2 Review of Computer Music 4

1.2.1 A Brief History 4

1.2.2 Analysis of Musical Sounds 6

1.2.3 Sound Synthesis Techniques 17

1.3 Piano Tones and Their Analysis 20
1.4 The Structure of This Dissertation 27

Chapter 2 Wavelet Fundamentals 29

2.1 General scheme for analyzing a signal 30

2.1.1 Vector space and inner product 30

2.1.2 Orthogonality and orthogonal projections 32

2.2 Wavelets and multiresolution analysis 34

2.2.1 About Wavelet 34

2.2.2 Multiresolution analysis 35

2.2.3 Linking wavelets to filters 37

2.2.4 Fast filter bank implementations of wavelet transform 42

Chapter 3 Waveform Analysis of Piano tones' Onset Transients 48

3.1 Definitions for Onset transients 51
3.2 Measuring Durations of piano onset transients 55

3.2.1 The challenges 55

3.2.2 Wavelet Multiresolution Decomposition by filter banks and 'wavelet crime' 57

3.2.3 Measurement and Analysis 63

Chapter 4 Time-Frequency Analysis of Piano tones 83

4.1 Wavelets Packet Transform and Time-Frequency Plane 84
4.2 The Time-Frequency Planes of Onset Transients by WPT bases 91
4.3 Local cosine bases 105
4.4 Matching pursuit 117

Chapter 5 Reconstructing Waveforms By Wavelet Impulse Synthesis 128

5.1 Wavelet Impulse Synthesis 129
5.2 Effective Approximation And Waveform Reconstruction 138
5.3 A listening test 145

Chapter 6 Determining the inharmonicity coefficients for piano tones 148

6.1 Theoretical Preparation 150

6.1.1 Choice Of Wavelet Bases 151

6.2 Experiments And Results 156

Chapter 7 Conclusions and Suggestions for Future Work 205

7.1 Conclusions 205
7.2 Suggestions for future work 210

References 212

Publication 217

Appendix A 218

Summary

The wavelet analysis has two important advantages over Fourier analysis: localizing ‘unusual’ transient events and disclosing time-frequency information with flexible analysis windows. This dissertation presents the application of wavelet analysis to musical sounds. Among all kinds of attributes of musical sounds, the most basic but also most important attribute might be what is called the tone quality, usually referred to as the timbre. It is the timbre that helps people recognize and identify the distinction between musical instruments when the same note is played at the same loudness on different musical instruments. Besides spectral structures, other factors like the onset transients and inharmonicity may affect the timbre of a musical instrument. The piano is an important western musical instrument and has very short onset transients and significant inharmonicity. Taking piano sounds as the object of study, this dissertation has confirmed the applicability of wavelet analysis to piano tones and has investigated their onset transients and inharmonicity.

Firstly, the ability of wavelets to localize ‘unusual’ transient events is used to estimate the duration of the onset transients of piano tones. A variant wavelet multiresolution analysis was employed for this. After explaining the surprisingly negative dip in the envelope of processed piano waves, we are able to identify the beginning of the onset transients. The duration of the onset transients was therefore obtained by measuring the time between the waveform peak and the identified beginning point.

Secondly, the ability of wavelet analysis to perform time-frequency analysis with flexible windows was adopted to illustrate the distinction in the time-frequency plane between the onset transients and the stationary parts. The analysis of such wavelet time-frequency planes disclosed and verified some of the piano tones' important characteristics.

Thirdly, the reconstruction of piano tones was investigated. Our experiments indicated that only a small number of time-frequency blocks were needed to represent piano tones well. This is due to both the compression capability of the wavelet analysis and the special features of the piano tones. The entire reconstruction process also paves the way for our estimation of inharmonicity coefficients for piano tones.

Finally, most previous studies for estimating the inharmonicity coefficients of piano tones were based on Fourier transform. Little or no works has been based on wavelet transform. Thus in this thesis, an approach based on the wavelet impulse synthesis was designed to estimate the inharmonicity coefficients of piano tones. Each time-frequency block in the plane represented a wave component which is the product of a coefficient with its associated wavelet basis. Each wave component was obtained by wavelet impulse synthesis and classified into a particular partial in terms of a series of analysis frequencies, thus allowing the estimation of the partial's frequency. After eliminating the 'partial shift' effect by a correction process, the combination of fundamental frequency and inharmonicity coefficient was accurately measured. The calculated results agreed closely with the piano' real harmonics obtained by FFT analysis.

List of Figures

Fig 1.1 An individual bandpass filter in phase vocoder.....14
Fig 1.2 Production of piano sounds.....20
Fig 2.1A member vector X in R^3 space.....30
Fig 2.2 The example of a wavelet.....34
Fig 2.3 One level wavelet transform.....45
Fig 2.4 One level inverse wavelet transform.....47
Fig 3.1 A modern standard piano keyboard with the distribution of fundamental frequencies.....49
Fig 3.2 The waveform of a piano tone C4 whose corresponding key is located in the middle of the piano keyboard.....50
Fig 3.3 The waveform of piano tone A0 whose corresponding key is located on the extreme left of the piano keyboard.....52
Fig 3.4 The waveform of piano tone C8 whose corresponding key is located on the extreme right of the piano keyboard.....52
Fig 3.5 The evolving process of the piano tone C4, roughly the initial 1,024 sampled points as the x-axis shows.....53
Fig 3.6 Onset durations of all piano tones in the ideal theoretical situation.....54
Fig 3.7 The arrangement of piano tones in a segment of MUMS CD sound tracks.....56
Fig 3.8 One stage 1-D wavelet transform.....57
Fig 3.9 Multi-level decomposition.....58
Fig 3.10 Multi-level inverse Discrete Wavelet Transform.....59
Fig 3.11 Diagram of multiresolution decomposition.....60
Fig 3.12 Four sine functions with different frequencies at different time.....61
Fig 3.13 The comparison between the original signal and the summation of all subbands in using multiresolution analysis.....61
Fig 3.14 The contents of every subband in the three level multiresolution analysis. From top to bottom, each subband respectively corresponds to d_x^1, d_x^2, d_x^3 and a_x^3.....62
Fig 3.15 The energy envelope of C4 piano tone.....65
Fig 3.16 Scaling and wavelet function of wavelet bases Coiflet 1.....66
Fig 3.17 The waveforms of some subbands in the multiresolution analysis of C4 piano tone.....67
Fig 3.18 Results of the multiresolution analysis for C4 piano tone.....68
Fig 3.19 The measurement of A3 piano tone.....71
Fig 3.20 The measurement of D1 piano tone.....73
Fig 3.21 The measurement of F5 piano tone.....75
Fig 3.22 The measurement of B0 piano tone.....77
Fig 3.23 The measurement of G7 piano tone.....79

Fig 3.24 Onset durations of all piano tones (from A0 to C8) as computed by multiresolution analysis.....81
Fig 4.1 Some T-F planes for an 8 points signal.....87
Fig 4.2 The hierarchy diagram of DWT for 8 points, corresponding to the Fig 4.1(b) (Note: the '+' here does not mean the ordinary plus operation in mathematics. It only means that A3, D3, D2 and D1 together may make up one possible result among the DWT decomposition.).....88
Fig 4.3 The full tree hierarchy diagram of WPT for 8 points, corresponding to the Fig 4.1 (c).....88
Fig 4.4 The time-frequency plane for tone C4 by the wavelet packets transform: onset transients (top) and stationary part (bottom).....94
Fig 4.5 Onset transient (top) and stationary part (bottom) of D7 piano tone 97
Fig 4.6 Onset transient (top) and stationary part (bottom) of E2 piano tone.....98
Fig 4.7 Onset transient (top) and stationary part (bottom) of A3 piano tone.....99
Fig 4.8 Onset transient (top) and stationary part (bottom) of F5 piano tone.....100
Fig 4.9 Onset transient (top) and stationary part (bottom) of B0 piano tone.....101
Fig 4.10 Time-frequency plane of approximately first 50 ms for (a) A0, (b) B0, (c) F5 and (d) C6 piano tone.....104
Fig 4.11 Time-Frequency Partition by local cosine bases (Source: from Mallat [74]).....106
Fig 4.12 The time-frequency plane for C4 by local cosine bases.....109
Fig 4.13 The time-frequency plane for B0 by local cosine bases.....111
Fig 4.14 The time-frequency plane for A2 by local cosine bases.....112
Fig 4.15 The time-frequency plane for G4 by local cosine bases.....113
Fig 4.16 The time-frequency plane for C8 by local cosine bases.....114
Fig 4.17 The time-frequency plane for A7 by local cosine bases.....116
Fig 4.18 box1(t_1, f_1); box2(t_2, f_1); box3(t_1, f_2); box4(t_2, f_2).....117
Fig 4.19 Comparison: the time-frequency plane for tone C4 by wavelet packet (top) and matching pursuit (bottom).....120
Fig 4.20 Comparison: the time-frequency plane for tone D7 by wavelet packets (top) and matching pursuit (bottom).....123
Fig 4.21 Comparison: the time-frequency plane for tone E2 by wavelet packets (top) and matching pursuit (bottom).....124
Fig 4.22 Comparison: the time-frequency plane for tone A3 by wavelet packets (top) and matching pursuit (bottom).....125
Fig 4.23 Comparison: the time-frequency plane for tone F5 by wavelet packets (top) and matching pursuit (bottom).....126
Fig 4.24 Comparison: the time-frequency plane for tone B0 by wavelet packets (top) and matching pursuit (bottom).....127
Fig 5.1 An 8-point 3 level full tree WPT: any coefficient can be uniquely identified by (*d*,*b*,*k*), where $d \equiv depth$, $b \equiv node$, $k \equiv index\ within\ node$ 129
Fig 5.2 The demonstration for zero-nodes' extending or shrinking 131

Fig 5.3 Traditional Wavelet Packet Analysis and Synthesis 133
Fig 5.4 The T-F plane of the onset transient of C4 piano tone 134
Fig 5.5 The T-F block whose coefficient is largest (bottom) and the waveform of the basis this T-F block corresponds (top) 134
Fig 5.6 The T-F block whose coefficient is 2nd largest (bottom) and the waveform of the basis this T-F block corresponds (top) 135
Fig 5.7 The T-F block whose coefficient is 3rd largest (bottom) and the waveform of the basis this T-F block corresponds (top) 135
Fig 5.8 The T-F block whose coefficient is 4th largest (bottom) and the waveform of the basis this T-F block corresponds (top) 136
Fig 5.9 The T-F block whose coefficient is 5th largest (bottom) and the waveform of the basis this T-F block corresponds (top) 136
Fig 5.10 The synthesis by five largest T-F blocks 137
Fig 5.11 Reconstruction of B0 piano tone by 100 most significant T-F blocks.. 139
Fig 5.12 Reconstruction of B0 piano tone by 300 most significant T-F blocks . 139
Fig 5.13 Reconstruction of B0 piano tone by 500 most significant T-F blocks . 140
Fig 5.14 Reconstruction of B0 piano tone by 1000 most significant T-F blocks 140
Fig 5.15 Reconstruction of B0 piano tone by 1500 most significant T-F blocks 141
Fig 5.16 Reconstruction of B0 piano tone by 2000 most significant T-F blocks 141
Fig 5.17 Reconstruction of F1 piano tone by 100 most significant T-F blocks.. 143
Fig 5.18 Reconstruction of F1 piano tone by 500 most significant T-F blocks.. 143
Fig 5.19 Reconstruction of F1 piano tone by 1000 most significant T-F blocks 144
Fig 5.20 Reconstruction of F1 piano tone by 1500 most significant T-F blocks 144
Fig 5.21 Reconstruction of F1 piano tone by 2000 most significant T-F blocks 145
Fig 6.1 Comparison between Daubechies bases (6,1,6) and Battle-Lemarie bases (6,1,6) 152
Fig 6.2 Comparison between Daubechies bases (6,1,2) and Battle-Lemarie bases (6,1,2) 153
Fig 6.3 Comparison between Daubechies bases (6,0,6) and Battle-Lemarie bases (6,0,6) 153
Fig 6.4 Comparison between Daubechies bases (4,1,6) and Battle-Lemarie bases (4,1,6) 155
Fig 6.5 Comparison between Daubechies bases (7,1,6) and Battle-Lemarie bases (7,1,6) 156
Fig 6.6 The result of rough estimation: the expected curve $nF_1\sqrt{1+Bn^2}$ *vs* measured partial frequencies 162
Fig 6.7 Results of the 6-iteration correction process for F1 piano tone 171
Fig 6.8 Our prediction on the F1 piano tone inharmonic frequency structure and its real FFT spectrum 173
Fig 6.9 The assumed harmonic structure of the F1 piano tone and its real FFT spectrum Note the frequency range roughly from 800 Hz to 1200 Hz, and the frequencies around the 1600 Hz 174
Fig 6.10 Reconstruction of a 32768-point tone B0 sample by *m*=1500 most

significant time-frequency blocks 175
Fig 6.11 Results of the 8-iteration correction process for the B0 piano tone 180
Fig 6.12 Our prediction on the B0 piano tone inharmonic frequency structure and its real FFT spectrum.....181
Fig 6.13 The assumed harmonic structure of the B0 piano tone and its real FFT spectrum Note the frequency range roughly from 600 Hz to 800 Hz 182
Fig 6.14 Reconstruction of a 32768-point tone G2 sample by *m*=1500 most significant time-frequency blocks 183
Fig 6.15 Results of the 3-iteration correction process for the G2 piano tone 185
Fig 6.16 Our prediction on the G2 piano tone inharmonic frequency structure and its real FFT spectrum.....186
Fig 6.17 The assumed harmonic structure of the G2 piano tone and its real FFT 187
Fig 6.18 Reconstruction of a 32768-point tone D3# sample by *m*=1500 most significant time-frequency blocks 188
Fig 6.19 Results of the 6-iteration correction process for the D3# piano tone ...191
Fig 6.20 Our prediction on the D3# piano tone inharmonic frequency structure and its real FFT spectrum 192
Fig 6.21 The assumed harmonic structure of D3# piano tone and its real FFT spectrum. Note the frequency range roughly from 1500 Hz to 3000 Hz 193
Fig 6.22 Reconstruction of a 32768-point tone C4 sample by *m*=1500 most significant time-frequency blocks 194
Fig 6.23 Results of the 3-iteration correction process for the C4 piano tone 196
Fig 6.24 Our prediction on the C4 piano tone inharmonic frequency structure and its real FFT spectrum.....197
Fig 6.25 The assumed harmonic structure of C4 piano tone and its real FFT spectrum 197
Fig 6.26 Reconstruction of a 16384-point tone A5 sample by *m*=1500 most significant time-frequency blocks 198
Fig 6.27 Results of the 2-iteration correction process for the A5 piano tone 200
Fig 6.28 Our prediction on A5 piano tone inharmonic frequency structure and its real FFT spectrum 201
Fig 6.29 The assumed harmonic structure of A5 piano tone and its real FFT spectrum 201
Fig 6.30 Estimated inharmonicity coefficients for some piano tones.....202

List of Tables

Table 3.1 Comparison with visual inspection 80
Table 5.1 The results of the listening test for tone B0, where the numbers outside the brackets are the No. of choices by the listener and the number pairs within the bracket show the correction rate expressed by (correct : wrong). For a total of 200 choices (110 original and 90 reconstructed), 108 (62 plus 46) were correct and 92 (48 plus 44) were wrong. 147
Table 6.1 Frequencies of some partials of F1 piano tone after rough estimation 159
Table 6.2 F_1 and B for a F1 piano tone 161
Table 6.3 The first iteration: absolute value operation applied 165
Table 6.4 The first iteration: nothing has been done on negative *B* values in the rough estimate 167
Table 6.5 F_1 and B calculated from rough estimation to the 6^{th} iteration 167
Table 6.6 F_1 and B calculated for the B0 piano tone 176
Table 6.7 F_1 and B calculated for the G2 piano tone 183
Table 6.8 F_1 and B calculated for the D3# piano tone 188
Table 6.9 F_1 and B calculated for the C4 piano tone 194
Table 6.10 F_1 and B calculated for the A5 piano tone 198
Table 6.11 The inharmonicity coefficients estimated by Galembo [64] unit:10^{-6} 203
Table 6.12 The values of some piano tones' inharmonicity coefficients 204

Chapter 1 Introduction

1.1 Musical Acoustics and Computer Music

Musical acoustics, an intrinsically multidisciplinary field, mirrors the convergence of two distinct disciplines, science and music. Such convergence, according to Benade [1] is the meeting place of music, physics and auditory science. In other words, the study of *musical acoustics* has intertwined music with physics. It is this intertwining that promotes music from being an ineffable art of emotional expression to being a sophisticated subject of science research.

For example, scientists represent sounds by waves and attribute the production of a sound to the result of air vibrations. Whenever two or more sound waves with

different pitches (i.e. frequencies) are played at the same time, the amplitudes of such sound waves in the air pressure combine with each other and produce a new sound wave as the consequence of such interaction. Also, any given complicated sound wave can be modelled by many different sine waves of the appropriate frequencies and amplitudes (spectral analysis). Finally, the human hearing system, mainly composed of both the ears and the brain, can usually isolate/decode the variation of the air pressure at the ear "containing" these pitches into separate tones and perceive them as distinctive sounds.

The examples mentioned above from the production of a musical sound to the perception of the sound are all within the coverage of musical acoustics. From these examples, we also can deduce how scientists have translated various aspects of musical sounds into physics research topics.

The history of research into *musical acoustics* can be traced back to ancient Greece when Pythagoras (roughly about 580 BC~500 BC) studied the relation between musical intervals and certain string length ratios. In the following centuries, scientists and musicians who continued to believe that science would supply the basis for the foundations of music have steadily expanded the scope of *musical acoustics*, particularly in the design and manufacture of various musical instruments.

Although the use of science and technology in music is not new, the real surge of interest in the study of *musical acoustics* was indeed triggered by the rapid progress and extensive use of computer systems, dating back to the 1950s. Ever since then, important new developments like digital music, computer spectrum analysis, sound

mixing, etc, have sprung up. Driven by these developments, a variety of music-related products from professional music recording/editing studio equipment to electronic pianos for the domestic consumer have emerged.

Against this background, a new research subfield, *computer music*, gradually came into being, whose range covers physics, psychology, computer science, and mathematics. The emergence of *computer music* is a quantum leap for the marriage of technology and music. Acting as a 'super' musical instrument, a well-designed computer system not only can simulate sounds of any existing musical instruments but also, more importantly, may extend musical timbres beyond those conventional musical instruments, by eliminating the constraints of the physical medium on sound production. That means 'new' and previously unheard musical sounds might be synthesized by a computer and the musical waveform heard by being played through a loudspeaker. This generality of computer synthesis implies an extraordinarily larger sound timbre space, which is an obvious attraction to music composers [2] seeking new sounds.

This raises an essential question on how to realize such a 'super' musical instrument. Generally speaking, the answer could be reduced to 2 inverse but closely interrelated processes: the analysis of a sound and the digital synthesis of the sound. The goal of the analysis process is to overcome the barrier when the required knowledge on the nature of a sound in question is lacking, which is related to the physical and perceptual description of sounds. Only with such necessary knowledge can we effectively instruct a computer system to perform the synthesis of musical

sounds. The means by which sounds are synthesized by different synthesis methods will be introduced in detail in section 1.2.2. Therefore, the twin processes of analysis and synthesis are universal in computer music, where researchers often analyze an acoustic signal in order to extract information about certain aspects of the signal and then use this information to reconstruct the signal by various methods of digital sound synthesis.

1.2 Review of Computer Music

1.2.1 A Brief History

As stated previously, when physics, psychology, computer science, and mathematics are integrated with musical knowledge, scientists, musicians and technicians can work together in *Computer Musi*c.

Nowadays, there are many organizations and companies throughout the world who are engaged in this flourishing and profitable area. But all of these can be attributed to the early work which established a solid foundation for today's commercially successful electronic music industry.

Believing that computers could generate new sounds to meet the exacting requirements of human aural perception, researchers at Bell Telephone Laboratories in Murray Hill began the first experiments in digital synthesis in 1957 when computers were still relatively uncommon and bulky. Their experiments confirmed that computers can effectively synthesize sounds with different pitches and waveforms.

Encouraged by the success of these experiments, Max V. Mathews made further remarkable progress in this pioneering stage of computer music. He invented the influential Music I language, a software environment which could implement sound synthesis algorithms. Based on Music I, the psychologist Newman Guttman created a piece of music called “In a Silver Scale” also in 1957, which only lasted 17 seconds. Subsequently, Bell Laboratories further developed the more ambitious Music II to Music V programs that are now looked upon as the original models for many synthesis programs of today.

Then in the following decades, some scientists like Chowning (1973) [3], Moorer (1977) [4], Horner (1993) [5] and Cardoz [6] developed the sound synthesis technique further through various approaches including modulation synthesis, additive synthesis, multiple wavetable synthesis and physical modeling synthesis respectively.

However, computer composers often want to mix and balance several audio channels that are input into computer devices simultaneously to create a synthesized piece of music. In this sound mixing process, it is often necessary to filter, delay, reverberate or localize the synthesized sounds. These operations fall within the domain of signal processing, which has been described by researchers such as Lansky (1982), Freed (1988), Jaffe (1989) [7-9].

Beside sound synthesis, sound analysis also plays an indispensable role in computer music, not only because such analysis is essential to enable a near perfect reconstruction of a musical sound, but also because such analysis is also essential for

the realization of an intelligent computer which can recognize, understand and respond to what it 'hears'. Such sound analysis includes research on the structure of musical tones and various techniques of spectral analysis. Each of these aspects can be further divided into several separate topics. For instance, research on the structure of musical tones, formant theory, onset transients and inharmonicity, etc are frequently mentioned. To improve spectral analysis techniques, all kinds of mathematical tools ranging from the Fourier transform to the Wavelet transform have been involved. From the next section, we will discuss the details of such sound synthesis/analysis techniques.

1.2.2 Analysis of Musical Sounds

In section 1.2.1, the importance of sound analysis in computer music has been briefly introduced. More omni-faceted accounts of the applications of sound analysis have been summarized by Roads [2] as below:

- Analysis $\rightarrow$ Modification $\rightarrow$ Resynthesis
- Making responsive instruments that "listen" via a microphone to a performer and respond in real time
- Creating sound databases in terms of each sound's acoustic properties.
- Adjusting the frequency response of a sound reinforcement system according to the frequency characteristics of the space.
- Restoring old recordings

- Data compression
- Transcribing sounds into common music notation
- Developing musical theories based on real performance of musical sound rather than just paper scores.

All such applications of sound analysis would pave the way for the further development of computer music, which in turn would promote more diversified applications and thereby lead to more intricate analysis on various attributes of a musical sound. These various attributes may range from straightforward sensations like pitch (a psychological and musical notion whose physical counterpart is frequency) and loudness, to more 'elusive' perceptions such as a sound's brightness, etc.

Nevertheless, among all kinds of such attributes of a musical sound, one of the most basic but also most important attributes is what is called the tone quality, or usually referred to as the timbre, which is determined by the harmonic content of the waveform[10].

1) Timbre of a Tone

According to The American Standards Association (ASA which has been renamed the American National Standards Institute, or ANSI), timbre is defined as "the attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar". Simply put, it is the timbre that helps us to recognize and identify the distinction

between musical instruments. For example, the human ear can easily distinguish a violin sound from a piano sound, even if both musical instruments have played the same note, e.g., the note C4 at the same loudness.

It is interesting to note what factors may affect the timbre.

i. The harmonics of a tone

Musicians and scientists have been long aware that the harmonic structure or spectrum of a tone is made up of a number of distinct frequencies, labeled as the partials. The lowest frequency is called the fundamental frequency which determines the perceived pitch of the tone. The other frequencies are called harmonics or partials, whose frequency values are integer multiples of the fundamental frequency. Some proponents of such harmonic analysis have asserted that the differences in the tone quality depend solely on the presence and strength of the partials [11]. Even though this is not entirely true, most theorists still agree that the spectrum of a tone is the primary determinant of its tone quality.

ii. The formant

As a supplement to the classical theories of harmonic analysis, the formant theory holds that "the characteristic tone quality of an instrument is due to the relative strengthening of whatever partial lies within a fixed or relatively fixed region of the musical scale" [12].

The classical theories which assert that the harmonics or partials are the sole determinant of tone quality in practice are not strictly true. For instance, for the

bassoon, there may be no apparent similarity between the Fourier spectra of different bassoon notes, other than an increase in amplitude of the high-frequency harmonics. But a meticulous comparison of the Fourier spectra for every bassoon note may disclose that a certain frequency region which is consistently emphasized relative to the other harmonics. In contrast with the classical theories that only look at the fixed spectrum of a single tone, the formant theory looks at such frequency ranges or "formants" which are consistently emphasized throughout the instrument's range to produce constancy in the characteristic tone quality of the instrument [13]. Furthermore, the perceived tone quality may also be influenced by the amount of emphasis in the formant region and by the width of the frequency band involved.

iii. The onset transient

The onset transient or the attack transient usually refers to the unique stage of a sound that occurs in its very beginning and generally only lasts for a very short period. If the onset transient of, for example, an oboe tone, is spliced together with the sustained stationary portion of the tone of another instrument such as a violin tone, listeners will often identify the combined tone as an oboe tone, although the main body of the combined tone is from another instrument [14]. Also, playing a piano' tone backwards results in a sound very different from that of a piano. Previous work [15, 16] on the sounds of musical instruments have indicated that each sound's onset transient plays a very important role in helping listeners to discriminate between various instruments. There could be several diverse explanations for this. From an acoustical point of view, during the onset transient, the standing wave has not been

established yet in the instrument. The amplitude fluctuates rapidly and the spectrum differs from that of the steady state, and such unstable behavior during the onset transient may contain more specific information regarding a certain instrument. From the human perception point of view, the human auditory system is more sensitive to a transient event than to static phenomena. Consequently, the subject of onset transients has become of considerable contemporary research interest.

iv. Inharmonicity

So far, we have supposed that a musical tone possesses a harmonic structure. This may be true for most western musical instruments, but many other instruments do produce inharmonic tones. Even in a harmonic tone structure, the so-called harmonics may not exactly follow a perfect harmonic structure. There is always a possibility that a partial could deviate from its expected harmonic position.

In music, inharmonicity is the concept of measuring the degree of deviation by which the frequencies of partials of a tone differ from integer multiples of the fundamental frequency. Inharmonicity is particularly evident in piano sounds because of the piano strings' stiffness and non-rigid terminations. Inharmonicity can have an important effect on the timbre. Podlesak [17] and Moore [18] pointed out pitch shifts due to inharmonicity, although having durations of a few tens of milliseconds, can be discriminated by listeners. In an experiment [19], it was found that synthesized piano notes with no inharmonicity were judged as sounding dull compared to real piano sounds.

2) Spectrum Analysis

As stated before, to synthesize musical sounds, it is important to understand which acoustical properties of a musical instrument sound are relevant to which specific perceptual features. Some relationships can be obviously identified, e.g. amplitudes control the loudness and the fundamental frequency regulates the pitch. Other perceptual features are subject to sound spectra and how they vary with time. For example, "attack impact" is strongly related to spectral characteristics during the first 20-100*ms* corresponding to the rise time of the sound, while the "warmth" of a tone points to spectral characteristics such as inharmonicity.

A straightforward definition of spectrum is a measure of the distribution of signal energy as a function of frequency. From such a distribution, we are able to know the contributions of various frequency components, each corresponding to a certain rate of variation in air pressure in the case of a sound wave. Gauging the balance among these components is the task of spectrum analysis [2].

Since spectral diagrams are capable of yielding significant insights into the microstructure of vocal, instrumental or synthetic sounds, not surprisingly, they are considered as essential tools for scientists and engineers. For instance, through revealing the energy spectrum of instrumental and vocal tones, spectrum analysis can help to identify timbres and separate instruments of different timbres playing simultaneously [20]. However, it was Melville Clark Jr.'s laboratory at MIT that accomplished the first time-varying spectrum analysis and synthesis of musical

sounds by a computer [21, 22]. Various applications or explorations of spectrum analysis were subsequently performed by Beauchamp [23, 24] and Risset and Mathews [25]. Some other pioneer work in spectrum analysis on musical sounds worthy of highlighting here include the work of Strong and Clark [26], who were the first to incorporate listening tests on musical sound synthesis derived from spectral analysis, and also the first to stress the importance of the spectral envelopes of musical instruments.

Fourier analysis, a family of different techniques that are still evolving, may be the most prevalent approach in spectrum analysis. In the following discussion, some typical techniques of Fourier analysis will be briefly introduced. The ideas behind such techniques can be very divergent, but they are all modeled on the basis of the Fourier Transform (FT) or the Short Time Fourier Transform (STFT).

i) Pitch-synchronous analysis [27]

In this approach, the essential part is partitioning a sound's waveform into pseudo periodic segments. The pitch of each pseudo periodic segment is also roughly estimated. The size of the analysis segment is adjusted relative to the estimated pitch period. Then the Fourier transform is applied on every analysis segment as though each of them was periodic. This technique thus generates the sound's spectrum for each time segment.

ii) Heterodyne Filter Analysis [28]

The heterodyne filter approach is especially suitable for resolving the harmonics of a sound. In a prior stage of analysis, the fundamental frequency of

the sound is estimated. The heterodyne filter multiplies the input waveform by an analysis signal (a sine wave or cosine wave). Then the resulting waveform is summed over a short time period to obtain amplitude and phase data. The product of the input signal (an approximate sine wave) with an analysis signal (a pure sine wave having the same phase) should be a waveform riding above the zero-axis (i.e. having positive values) if the frequencies of the two signals match. Otherwise, the result scatters symmetrically around the zero-axis (positive or negative). When this scattered waveform is summed over a short time period it will basically cancel out.

However, the limits of the heterodyne method are also well known. For example, Moorer [29] showed that the heterodyne filter approach is invalid for fast attack periods (less than 50ms) or those sounds whose pitch changes greater than about a quarter tone. Although Beauchamp [30] improved the heterodyne filter to allow it to follow changing frequency trajectories, the heterodyne filter approach is seldom used nowadays and has already been supplanted by other methods.

iii) Short-time Fourier Transform and Phase Vocoder [31-34]

One of the most popular techniques based on the Short-time Fourier transform (STFT) for the analysis/resynthesis of spectra is the phase vocoder, developed by Flanagan and Golden [35] in 1966 at Bell Telephone Laboratories. The phase vocoder can be thought of as passing a windowed input signal through a bank of parallel bandpass filters which spread across the audio bandwidth with

equal intervals. Every filter measures the amplitude and phase of a signal in each frequency band (see Fig 1.1). Through a subsequent operation, these values can be converted into two envelopes: one for the amplitude, and one for the frequency.

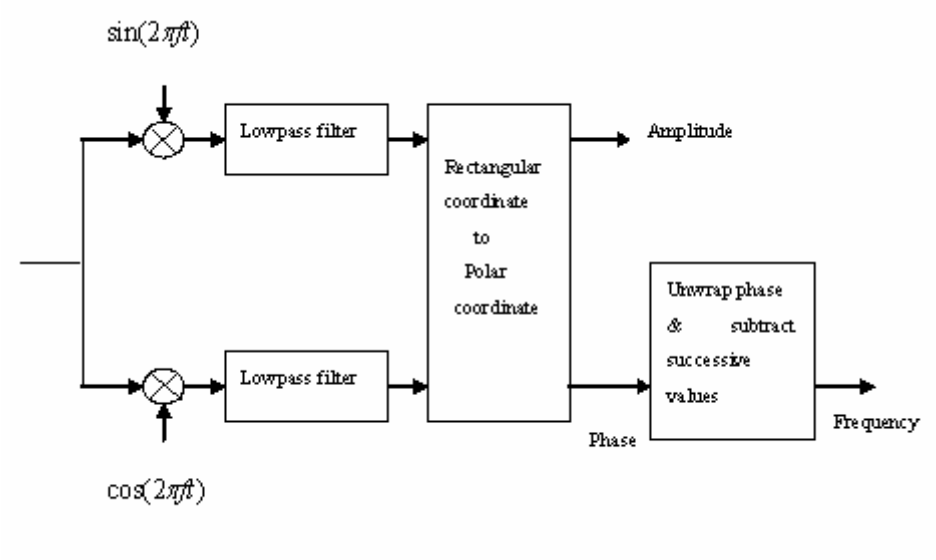


Fig 1.1 An individual bandpass filter in phase vocoder

Moreover, various implementations of the Phase Vocoder provide tools for modifying these envelopes, which make the musical transformations of analyzed sounds possible.

Recently, many implementations of the Phase Vocoder have been improved to follow or track the most prominent peaks in the spectrum over time. Hence they are called Tracking Phase Vocoders (TPV) [36, 37]. Unlike the ordinary phase vocoder, in which the resynthesis frequencies are limited to harmonics of the analysis window, the TPV follows changes in frequencies. The result of peak tracking is a set of amplitude and frequency envelopes that drive a bank of sinusoidal oscillators in the resynthesis stage.

Beside these typical Fourier-based methods, other “non-Fourier” methods (they are

actually extensions of Fourier analysis) have also gained ground in recent years, typically, two of which are Constant-Q Filter Bank analysis and wavelets analysis respectively.

i) The Constant-Q transform

The Constant-Q transform [38] can be thought of as a series of logarithmically spaced filters, with the k-th filter having the central frequency given by

$$f_k = (2^{1/24})^k f_{\min} \qquad (1\text{-}1)$$

where the minimum frequency $f_{\min}$ is an adjustable parameter and can be chosen to be the lowest frequency about which information is desired. The bandwidth Δf_k is defined as the difference between consecutive bin frequencies (central frequencies)

$$\Delta f_k = f_{k+1} - f_k = 2^{1/24} \cdot f_k - f_k = (2^{1/24} - 1) f_k \qquad (1\text{-}2)$$

and thus Q, the ratio of frequency to bandwidth is roughly constant

$$Q = f_k / \Delta f_k = f_k / ((2^{1/24} - 1) f_k) \approx 34 \qquad (1\text{-}3)$$

Distinct from Fourier analysis which divides the spectrum into a set of equally-spaced frequency bins (e.g., for a 1024 point FFT at a sampling rate of 44100 Hz, the bandwidth of a bin is uniformly $\Delta f = (44100/2)/1024 \approx 21.53$ Hz), the Constant-Q transform varies the bandwidth proportionally with the frequency ($\Delta f_k = f_k / 34$). That means, in the constant-Q transform the analysis bands are narrow for low frequencies and wide for high frequencies.

It is also clear from the Equation (1-1) that

$$\log(f_k) = \frac{\log(2)}{24} k + \log(f_{\min}) \qquad (1\text{-}4)$$

Thus, if we translate frequencies into a logarithmic scale (such as corresponds to musical octaves) like Equation (1-4), the log-frequencies of musical tones are linearly related to the number *k*.

ii) The Wavelet Transform

Fourier analysis dominates the field of stationary signal processing, but as it is a technique inherently requiring a wide time span, it is less effective for unstable transient signals. The Short Time Fourier Transform (STFT) can analyze signals in both time and frequency using suitable fixed-length windows, but onset transients with their rapidly changing frequencies and amplitudes require more flexible and specific time segments.

In contrast to STFT which uses fixed-length windows, the Constant-Q method varies the length of windows according to the frequency being analyzed. That means it uses broad time windows (narrow frequency intervals) to analyze low frequencies and narrow time windows on high frequencies. The schemes for implementing the Constant-Q spectral analysis have been reviewed in the reference [39]. However, the Constant-Q spectral analysis is not computationally efficient. To overcome this problem, Brown and Puckette [40] proposed an efficient method of transforming a discrete Fourier transform into a Constant-Q transform, taking advantage of the speed of the FFT calculation. Besides the heavy computational load, another issue is that the existence of a Constant-Q filter bank does not necessarily imply a method for resynthesis [2]. The wavelet transform, which can be considered as a special case of the Constant-Q method in a general sense, does not have the above two potential

problems (computationally inefficient and no resynthesis in some cases). In 1988, Mallat produced a fast wavelet decomposition and reconstruction algorithm [41], following which the wavelet transform can always be carried out in two reversible directions: analysis and resynthesis. Various wavelet-based techniques have been applied in the field of music sound processing. These investigations include wavelet representations of musical signals (the time-frequency grid [42] and pitch-synchronous representation [43]), removing noise from music [44], compression [45], and analysis / resynthesis of musical sounds [42, 46].

1.2.3 Sound Synthesis Techniques

This section will explain the basic principles of contemporary synthesis methods. A typical digital sound synthesis technique uses a time varying mathematical equation with a few adjustable parameters to compute the time varying output waveform, which if then sent to a loudspeaker, will produce a physical sound waveform. The parameters contained in such a mathematical equation can be looked upon as the control functions of the equation or algorithm. Over the last few decades, many synthesis techniques have been proposed. Among them are physical modeling synthesis, additive synthesis, subtractive synthesis, multiple wavetable synthesis and modulation synthesis, each of which will be briefly introduced respectively.

Physical modeling synthesis produces a particular musical instrument's sounds by designing a mathematical model to simulate the physical sound production

mechanism of this musical instrument, involving all the essential physical and acoustical behavior of the real instrument. It can be imagined that this synthesis technique is extremely complicated. The important work in this field includes Hiller's finite difference approximations of the wave equation [47], the Karplus-Strong algorithm [48] and Julius O. Smith III's digital waveguide model [49].

Additive synthesis [2] emulates tones by Fourier series analysis, a powerful mathematical tool that can express any periodic function as the sum of trigonometric functions such as sine or cosine functions.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nt) + b_n \sin(nt)] \qquad (1\text{-}5)$$

where

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt) dt \qquad (1\text{-}6)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt) dt \qquad (1\text{-}7)$$

To obtain a time varying sound spectrum, the assigned amplitudes of each harmonic, i.e., the coefficients a_n and b_n can be further allowed to vary with time. Additive synthesis has a well-understood methodology and is able to reproduce very closely any sound that can be expressed as the spectral components of the waveform's harmonics in the frequency domain. Through varying the control functions, new musical sounds can be easily generated. However the drawback of additive synthesis is also obvious. Localized sound events or non-periodic sounds, e.g. onset transients or inharmonicity of string instruments like the piano, are difficult to generate. Therefore, Wavelet analysis, which is applied in the work described in this dissertation, attempts to overcome such difficulties.

Multiple wavetable synthesis [5] is another popular synthesis method. In multiple wavetable synthesis, the lookup table contains several general waveforms shapes. A mechanism exists for dynamically changing the wave shape as the musical tone evolves. More sophisticated methods have been proposed by a few authors, for example Horner [50]. As an enhanced extension of additive synthesis, the multiple wavetable synthesis is well suited for synthesizing quasi-periodic sounds.

Modulation synthesis can be divided into 2 categories, amplitude modulation (AM) synthesis and frequency modulation (FM) synthesis. In AM synthesis, an amplitude envelope is applied to an oscillating waveform in the time domain, thereby producing the modulated signal. The formula is shown below

$$f(t) = [1 + k_a m(t)] A_c \cos(2\pi f_c t) \qquad (1\text{-}8)$$

where A_c and f_c are namely the carrier amplitude and carrier frequency. k_a is the modulation index and $m(t)$ is the modulating function.

In FM synthesis, a modulator oscillator modulates the frequency of the carrier oscillator. With relatively few control parameters, frequency modulation synthesis can create a very complex waveform. The formula is

$$f(t) = A(t)\sin[2\pi f_c t + I\sin(2\pi f_m t)] \qquad (1\text{-}9)$$

where $A(t)$ is the amplitude, f_c is the carrier frequency, f_m is the modulation frequency and I is the modulation index. For FM synthesis, A is normally time constant.

John Chowning [3] first discovered the frequency modulation synthesis at Stanford university in 1973. Since then, many new variants of FM synthesis have

been proposed. These include the Asymmetrical Frequency Modulation (AFM) synthesis technique [51] [52] and Double Frequency Modulation (DFM) synthesis technique [53] [54] [55].

1.3 Piano Tones and Their Analysis

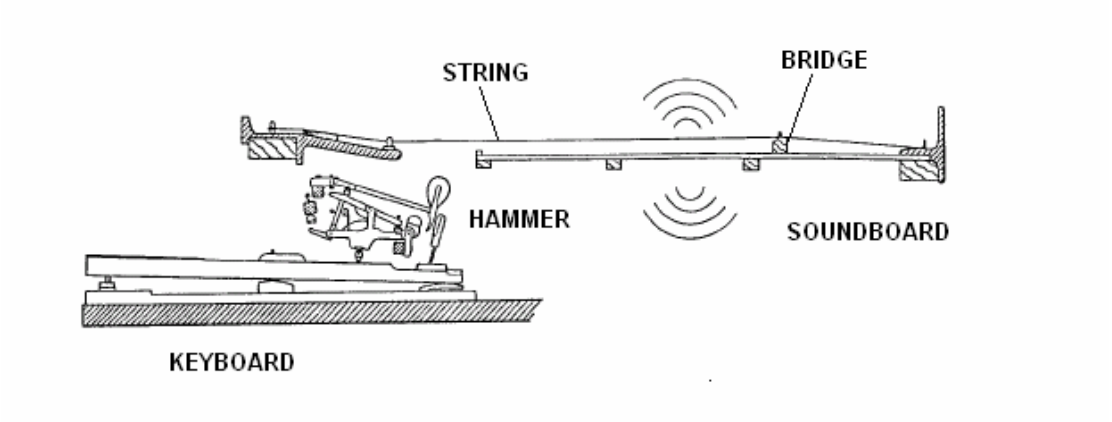


Fig 1.2 Production of piano sounds

The piano, or pianoforte, is among the most important instruments used in classical music. The piano's sound production and the underlying physical phenomena are very complicated. But we still can attempt to explain its sound production using a diagram like Fig 1.2. A piano produces sound by striking metal strings with felt covered hammers. The hammer rebounds, which allows the string to vibrate on its own frequency. These vibrations are transmitted through a bridge to a soundboard that amplifies them. Therefore according to the above-mentioned descriptions, the sound-production mechanism of the piano can be divided into 3 stages.

(1) When the hammer strikes the string, vibrations are excited on the string.

After the hammer bounces off the string, the kinetic energy of the hammer is transformed to the string's vibration energy.

(2) The string's vibration energy is stored by the string in the form of normal vibration modes. Although internal losses may dissipate some of this energy, most of the string's vibration energy is transmitted to the soundboard through the bridge, causing the soundboard to vibrate.

(3) Finally, the soundboard's energy is converted to the vibrational energy of a sound wave, and the sound wave's energy travels through the air to arrive at the listener's ears.

In the following, how each component contributes to the production of piano sounds and thus their acoustical properties will be discussed.

(1) The hammer

Many researchers have tried to examine the interaction between the hammer and the string [56-58].

In summary, the hammer and string interaction proceeds like this: the hammer, accelerated by the pianist, is thrown onto the string. But since its mass is not negligible, the hammer does not bounce clear off the string immediately. It is slowed down a little on impact on the string. The impact excites pulses on the string which will travel to both ends of the string. The reflected pulses from the ends will decelerate the hammer and give rise to secondary pulses. With more reflections, the hammer is thrown off the string.

Some particular properties or behaviors of the hammer can characterize piano

sounds. We know that the hammers are usually covered by wool felt. If the felt is harder, the piano will produce stronger partials and thus a brighter tone. On the contrary, softer hammers will result in less partials and a more mellow tone. The impact velocity of the hammer is also important. With increasing velocity, more high-frequency components of the tone are produced. Furthermore, the spectra of piano sounds also depend on the hammer-string contact point as well. For example, those modes of the string having a node near the contact point may not be excited effectively.

(2) The string

The strings of the piano are made of steel wire. In order to achieve high efficiency, the string is required to be at a high tension. The hammer motion mainly gives rise to two transverse polarizations (the vertical polarization and the horizontal polarization) in the string. Compared to horizontal polarization, greater vertical polarization is excited by the hammer and the energy transmission to the soundboard is more effective in the vertical direction as well. As a result, vertical polarization dominates at the beginning of piano tones. With more vertical polarization energy transmitted to the soundboard, the vibrations of the vertical polarization decay faster in the string. Therefore the horizontal polarization of the string determines the tail part of the tone. However, since these two polarizations are coupled to each other, the real situation is more complicated than the description given here. For listeners, the perceptual effect is such that the note appears to be not only loud but also sustained.

To obtain higher acoustic energy output usually requires two or more strings for

the same pitch. For each pitch, these strings are not tuned in perfect unison. So the use of multiple strings for the same note may also help to give the compound decay mentioned previously. As the piano tone decays, the strings are out of phase and they no longer move the bridge synchronously together. As the result, the bridge impedance increases, and the rate of energy transfer to the soundboard is much lower , resulting in a slowing down of the decay in the energy of string.

Beside the decay rate, another interesting issue about the string is its stiffness, which results in a slightly inharmonic tone. In music, inharmonicity is the degree to which the frequencies of partials depart from integer multiples of the fundamental frequency. The stiffness of the strings, and particularly of the lower strings which are thicker, may lead to inharmonicity. Generally, the wavelength of the transverse wave on a stretched string is much greater than the diameter of the string, which makes wave velocity on the string constant and thus partials show a harmonic structure. However, for higher partials with very short wavelengths, the diameter of the string cannot be considered negligible any more, particulalrly for thicker strings. The mechanical resistance of the string to bending becomes an additional force, resulting in increased wave speeds and hence of higher pitch than the expected harmonics, leading to inharmonicity of these partials.

Whether the inharmonicity is a desired factor is also an interesting question. Conlin [57] suggested that inharmonicity is an important factor of piano sound, but it should be as little as possible. For bass tones the amplitude of the fundamental frequency is weak and the pitch is mainly determined by higher partials. Due to

inharmonicity, the frequency difference between the partials increases with partial number. Accordingly, the definition of pitch becomes uncertain for such bass notes.

However, experiments [59] have found that synthesized tones with no inharmonicity i.e. with partials which are exact integer multiples of the fundamental frequency, are usually perceived as sounding dull. On the other hand, tones with too high inharmonicity are judged as sounding metallic. Thus the ears seem to expect a certain amount of inharmonicity.

The Fletcher equation [60] shows that

$$f_n = nf_0\sqrt{1+Bn^2} \qquad (1\text{-}10)$$

where f_0 is the fundamental frequency of the ideal string, n is the number of the partial and B is the inharmonicity coefficient. This relationship has been confirmed by several experimental studies [61, 62]. It can be seen from the equation that the degree of inharmonicity should increase with partial number. The inharmonicity coefficient, B also increases with the fundamental frequency of the string. That means treble tones should have more inharmonicity i.e. their partials are more inharmonic than bass tones. However when listening to piano sounds, people perceive more inharmonicity in the bass tones than in the middle or treble tones. This may be explained by the fact that the number of partials which can be heard is much higher for the bass tones. Another possible reason is a psychoacoustic phenomenon: there is higher threshold of perception for inharmonicity for tones with higher fundamental frequencies [63].

In Chapter 6, the application of wavelet packets is used for the measurement of inharmonicity coefficients of piano tones, B. The inharmonicity coefficient in the

above formula is solely determined by each string's material characteristics such as its length, diameter and Young's modulus, etc.

$$B = \frac{\pi^3 E d^4}{64 l^2 \tau} \qquad (1\text{-}11)$$

Here, E :Young's modulus for the string

d : the diameter of the string

l : the string length

τ : the tension

Once B has been determined, formula (1-10) enables us to predict any partial's frequency. Conversely, if we can measure the frequencies of the partials of real piano sounds, the inharmonicity coefficient B may be determined by using formula (1-10) to calculate B, as has been done by previous researchers [64, 65]. Some may argue that B could be obtained directly from formula (1-11). In practice, the tension τ may not be convenient to measure if the piano is not accessible or if the piano tones were obtained from a recording. In addition, the measurement of E, d and l in a real piano may be laborious. That is why most researchers try their best to estimate B indirectly from formula (1-10).

Galembo and Askenfelt designed an inharmonic comb filter to estimate the inharmonicity coefficient in the frequency domain [64]. Furthermore they also tried pitch extraction techniques such as cepstral analysis and the harmonic product spectrum [66]. Klapuri [67] tackled the inharmonicity measurement by estimating the fundamental frequency in subbands. Rauhala [65] used an iterative process designed to minimize the deviation of the expected partial frequencies compared to the

frequencies of the high amplitude peaks in the spectrum.

However, most previous work has been based on Fourier analysis and very few have used wavelet analysis. In Chapter 6, a method based on wavelet impulse synthesis is used to estimate the inharmonicity coefficient *B* from real piano sound samples. Compared to Fourier-based approaches, whose success largely depends on applying additional optimizing algorithms or signal processing techniques to the Fourier spectrum to 'extract' partials from among frequency peaks clustered together, our wavelet-based method does not require such sophisticated techniques. Moreover, compared to Fourier-based approaches, our wavelet-based method also considers the temporal aspect of each partial's frequency variation.

(3) The soundboard and the bridge

We have known that the vibration of the strings is transmitted to the soundboard through the bridge. The bridge functions as an impedance transformer, providing higher impedance to the string. If the strings were to be directly connected to the soundboard without the bridge, the decay times of the string enegy would be too short as the energy transfer would be too rapid, and no standing wave could be set up in the string for the tone to be sustained. However, the impedance must allow some energy to be transferred from the string to the soundboard. By carefully designing the soundboard and the bridge, the loudness and the decay times of the partials can be optimized.

From the preceding brief introduction, we can see that several factors together determine the timbre of the piano tones. Firstly, the piano is a struck string instrument,

which implies that the string is affected by impulse excitations and results in decaying amplitudes. The string determines the fundamental frequency of the note as well.

The decay and the transmitted energy also depend on the impedance provided by the bridge and the soundboard. With higher impedance, the partials deliver less energy to the soundboard. Therefore, the energy in the string is better conserved and decay times are longer.

The stiffness of the string gives rise to a high dispersion. No other western string instrument has inharmonicity as high as the piano. The characteristic attack noise of the piano sound comes mainly from the impulse response of the soundboard, but also from the noise of the piano action.

1.4 The Structure of This Dissertation

In the last section, we have mentioned some applications of the wavelet analysis in the field of computer music. Although these applications seem diverse, the mechanism behind them can be very similar. Generally speaking, most applications can be categorized into two different groups depending on which one of two important features of the wavelet analysis is adopted. These two important features of the wavelet analysis are namely, localizing 'unusual' events and resolving time-frequency information with flexible analysis windows.

In this dissertation, our motivation is to apply wavelet analysis to piano tones. As stated in Section 1.2.2, onset transients and inharmonicity are two essential factors

that play an important role in discriminating a piano sound from that of other musical instruments. Therefore in Chapter 3, we will localize onset transients of piano tones and measure their durations by a variant approach based on wavelet multiresolution analysis. Furthermore in Chapters 4, 5 and 6, we will use time-frequency analysis to resolve onset transients' spectral content, to reconstruct the original signal and to estimate the inharmonicity coefficients of piano tones respectively.

However, to understand how wavelet analysis works, knowledge of basic wavelet transforms is a prerequisite. Hence, starting from basic concepts like vector space and inner product, Chapter 2 provides a review on mathematical fundamentals of wavelet theory, where multiresolution analysis and the wavelet transform filter banks implementation will be introduced.

Chapter 2 Wavelet Fundamentals

In this Chapter, we provide a brief review of the fundamentals of wavelets. At the same time, some terminology like the wavelet transform *(WT)* and *multiresolution analysis (MRA)* etc will be clarified. From these techniques, a fast filter bank implementation of the wavelet transform will also be introduced. We are describing these wavelet techniques since our work is largely based on them.

Simply put, serving as a bridge, this Chapter's main purposes are:

1) Introducing the wavelet transform from the most basic mathematical concepts such as vector space, inner product and orthogonality, etc.

2) Linking the implementation of the wavelet transform to filter banks in practical signal processing applications.

2.1 General scheme for analyzing a signal

2.1.1 Vector space and inner product

In mathematics, a vector space (or simply *space*) is a set including a collection of vectors, on which addition and multiplication-by-scalar operations are defined and satisfy some natural axioms. Among all kinds of spaces, the most familiar one may be the 3-dimensional Euclidean space R^3 where a Cartesian coordinate $X = (x_1, x_2, x_3)$ identifies any of its member vectors (as shown in Fig 2.1).

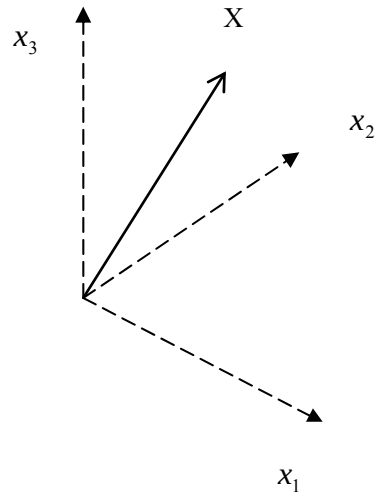


Fig 2.1A member vector *X* in R^3 space

In order to measure the length of its member vectors, a notion called the inner product may also be defined on the space. For instance, the inner product on R^3 has

been defined as below.

Definition 2.1 For two vectors $X = (x_1, x_2, x_3)$ *and* $Y = (y_1, y_2, y_3)$ *in* R^3, *the inner product is*

$$\langle X, Y \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 \qquad (2\text{-}1)$$

Thus the length of any vector X is given by the square root of the inner product of X with itself,

$$\textit{Length of } X = \sqrt{\langle X, X \rangle} = \sqrt{x_1^2 + x_2^2 + x_3^2} \qquad (2\text{-}2)$$

where the mathematical term for $\sqrt{\langle X, X \rangle}$ is "*the norm of X* " denoted as $\|X\| = \sqrt{\langle X, X \rangle}$. The notion of norm can also give meaning to the distance between two different member vectors. The distance between vector X and Y is

$$\|X - Y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \qquad (2\text{-}3)$$

In practical applications, given a signal which has been sampled at three distinct time points (e.g. t_1 , t_2 , t_3), we may represent the resultant sample as $f = (f_{t1}, f_{t2}, f_{t3})$. Therefore, all such 3-point sampled signals may constitute a real example of R^3 space. However in many cases, we have to sample far more samples than just three in order to represent the signal more accurately. This leads to a natural generalization from R^3 to R^N when the dimension is increased to a large number, *N*.

Similar to the definition of inner product on R^3, the corresponding inner product on R^N real space is defined as

$$\langle X, Y \rangle = \sum_{j=1}^{N} x_j y_j \qquad (2\text{-}4)$$

If *X* and *Y* are vectors in complex space C^N, then

$$\langle X, Y \rangle = \sum_{j=1}^{N} x_j \overline{y_j} \qquad (2\text{-}5)$$

2.1.2 Orthogonality and orthogonal projections

When we say that two vectors are orthogonal this is generally understood to mean that they are geometrically perpendicular to each other. Mathematically, a formal definition for orthogonality is derived from the idea of the inner product introduced in the last section.

Definition 2.2 Suppose V is an N-dimensional inner product space. Then

1) The vectors X and Y are said to be orthogonal if $\langle X, Y \rangle = 0$.

2) The collection of vectors e_i *, i=1,…,N, is said to be an orthonormal basis if each* e_i *has unit length,* $\|e_i\| = 1$*, and* e_i *and* e_j *are orthogonal for* $i \neq j$.

3) The subspaces V_1 *and* V_2 *of V are said to be orthogonal if each vector in* V_1 *is orthogonal to every vector in* V_2.

Once the orthonormal basis $\{e_{1,\cdots,}e_N\}$ have been found for a N-dimensional inner product space such as R^N, then any vectors $X \in R^N$ can be uniquely expanded as a linear combination of such basis vectors:

$$X = \sum_{j=1}^{N} \alpha_j e_j \qquad (2\text{-}6)$$

where α_j is the coefficient or the length of the projection of *X* on e_j.

To evaluate any coefficient α_k, taking the inner product with e_k on both sides of the equation (2-6), we have

$$\langle X, e_k \rangle = \sum_{j=1}^{N} \langle \alpha_j e_j, e_k \rangle$$

$$= \alpha_k \langle e_k, e_k \rangle \qquad (2\text{-}7)$$

$$= \alpha_k$$

From equations (2-6) and (2-7), we can summarize the general scheme for analyzing a signal:

1) Try to find a set of orthonormal bases for the vector space that describes the signal.
2) Use these orthonormal bases to decompose the signal using Equation (2-6).
3) The desired coefficients defining the signal are obtained by computing the inner product of the signal data with each basis vector according to the Equation (2-7).

Until now, we have summarized the general scheme for analyzing a signal. From this scheme, we can see that the most essential step is to find (or construct) the right set of orthonormal basis vectors. For the Fourier transform, these orthonormal bases are sine or cosine functions. Similarly, the wavelet transform also falls under this scheme, that is, by constructing orthonormal bases called *wavelets* and then decomposing a signal by calculating their inner products with the signal.

2.2 Wavelets and multiresolution analysis

2.2.1 About Wavelet

Wavelet (short for “wavelet basis”) is a set of basic functions associated with building a model for a signal. Just as their name’s literal meaning, these basic functions in wavelet theory are little waves. They must be oscillatory (waves) and have amplitudes which quickly decay to zero along both the positive and negative directions (little). Similar to sinusoids, the required oscillatory condition makes them good as building blocks. The quick decay condition is actually a windowing operation. Fig 2.2 shows the pattern of a typical wavelet.

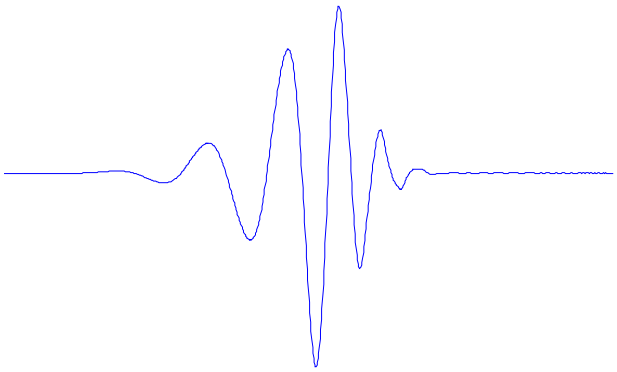


Fig 2.2 The example of a wavelet

Wavelet theory represents signals by breaking them into many interrelated component pieces. When these pieces are “scaled” and “translated” wavelets, this breaking down process is called *wavelet decomposition* or a *wavelet transform*. In contrast, wavelet reconstruction or an inverse wavelet transform puts the wavelet

pieces back together to retrieve the original object. Thus we can see that wavelet theory essentially consists of the study of these pieces (wavelets), their properties and interrelationships, and how to put them back together.

In the above statements, we have mentioned two operations involved in wavelet theory: the scaling operation and the translation operation. The scaling operation changes the independent variable (e.g., time in many cases) to create new functions, but have the same structure except they are either compressed or expanded. The translation operation moves the entire function along the independent variable's axis.

If the scaling operation is combined with the translation operation, mathematically, we have the following mapping relation between the original independent variable *x* and the new independent variable *x*'

$$x' = \frac{x - b}{a}$$

where *a* is the scale parameter and *b* is the translation parameter.

A wavelet coefficient at a particular scale and translation represents how well the original signal and the scaled and translated wavelet match. The more similarity there is the larger magnitude the coefficient has.

2.2.2 Multiresolution analysis

In the wavelet theory, two operations (*scaling* and *translation*) and two classes of functions (the scaling function $\phi(t)$ and the wavelet function $\psi(t)$) also play a vital role in constructing orthonormal wavelet bases. Both $\phi(t)$ and $\psi(t)$ must satisfy an

orthogonal relation with their respective *translation* functions. That is, for any $k \in Z, k \neq 0$,

$$\langle \phi(t), \phi(t-k) \rangle = 0 \qquad (2\text{-}8)$$

$$\langle \psi(t), \psi(t-k) \rangle = 0 \qquad (2\text{-}9)$$

The same orthogonality must also hold between any pair of these *translation* functions,

$$\langle \phi(t-j), \phi(t-k) \rangle = 0 \qquad (2\text{-}10)$$

$$\langle \psi(t-j), \psi(t-k) \rangle = 0 \qquad (2\text{-}11)$$

where $j \neq k$.

The scaling function $\phi(t)$ is constructed such that the set $\phi(t-k)$, for all $k \in Z$ constitutes an orthonormal basis, and let V_0 be the subspace spanned by the basis. The scaled basis comprising $\phi(2t-k)$, for all $k \in Z$, will then span the subspace V_1, with $V_1 \supset V_0$. The wavelet function $\psi(t)$ is constructed such that the set $\psi(t-k)$, for all $k \in Z$ will span the subspace W_0, with $W_0 \oplus V_0 = V_1$, and $W_0 \perp V_0$.

Similarly, the *scaling* operation should also impose constraints on $\phi(t)$ and $\psi(t)$. Taking $\phi(t) \in V_0$ as an example, the function set $\phi_{j,k} = 2^{j/2}\phi(2^j t - k); j,k \in Z$ should constitute a set of orthonormal bases for another space V_j, where a nested relation $V_j \subset V_{j+1}$ should also hold for any $j \in \{\cdots,-2,-1,0,1,2,\cdots\}$. Such a nested relation is very important in signal processing since it means that no information is lost as the resolution gets finer. The same occurs for $\psi(2^j t - k)$ and its corresponding space W_j as well. Furthermore W_j should

also be the orthogonal complement space of V_j in V_{j+1}. That means $V_{j+1} = V_j \oplus W_j$ and $W_j \perp V_j$ must be satisfied simultaneously. By successive orthogonal decompositions,

$$\begin{aligned} V_j &= W_{j-1} \oplus V_{j-1} \\ &= W_{j-1} \oplus W_{j-2} \oplus V_{j-2} \\ &\cdots \\ &= W_{j-1} \oplus W_{j-2} \oplus \cdots \oplus V_0 \end{aligned} \qquad (2\text{-}12)$$

Equation (2-12) shows that a signal in vector space V_j can be decomposed into many related spaces with distinct resolutions. This process is usually called *multiresolution analysis*.

2.2.3 Linking wavelets to filters

Although the wavelet theory can be very complicated and abstruse, the central equation in the multiresolution analysis, *the scaling relation* [68], can be concisely expressed as below:

Assuming $\phi(t-k) \in V_0$ *and* $\phi(2^j t-k) \in V_j$ *exist for any* $j,k \in Z$, *where* $\phi(t-k)$ *and* $2^{j/2}\phi(2^j t-k)$ *are respectively orthonormal bases of* V_0 *and* V_j, *then it is necessary and sufficient that the following expression holds*

$$\phi(t) = \sum_{k \in Z} p_k \phi(2t-k) \qquad (2\text{-}13)$$

with $p_m = 2\int_{-\infty}^{\infty} \phi(t)\overline{\phi(2t-m)}dt, m \in Z$.

Substitute *t* in (2-13) by $2^{j-1}t-l$, and we can get a more general expression:

$$\phi(2^{j-1}t-l) = \sum_{k \in Z} p_{k-2l}\phi(2^j t-k) \qquad (2\text{-}14)$$

Similar to (2-13), there is also a formula which builds a bridge between scaling functions and wavelet functions

$$\psi(t) = \sum_{k \in Z} (-1)^k \overline{p_{1-k}} \phi(2t - k) \qquad (2\text{-}15)$$

Starting from the scaling relation, the mathematical conditions and rules p_k must satisfy are:

$$\begin{aligned} &\sum_{k \in Z} p_{k-2l} \overline{p_k} = 2\delta_{l,0} \\ &\sum_{k \in Z} |p_k|^2 = 2 \\ &\sum_{k \in Z} p_k = 2 \\ &\sum_{k \in Z} p_{2k} = 1 \text{ and } \sum_{k \in Z} p_{2k+1} = 1 \end{aligned} \qquad (2\text{-}16)$$

A direct construction of the sequence p_k that satisfies Equation (2-16) is difficult. Usually, the sequence is constructed with the help of the Fourier transform. According to the orthonormality condition,

$$\int_{-\infty}^{\infty} \phi(t) \overline{\phi(t-k)} dt = \delta_{0k} \qquad (2\text{-}17)$$

The Plancherel's identity for the Fourier transform states that

$$\int f(x) \overline{g(x)} dx = \int \hat{f}(\xi) \hat{g}(\xi) d\xi \qquad (2\text{-}18)$$

where $\hat{f}(\xi)$ is the result of $f(x)$'s Fourier transform

In addition, the Fourier transform also has the following property:

$$Fourier[f(x-l)](\xi) = e^{-il\xi} Fourier[f(x)](\xi) \qquad (2\text{-}19)$$

where $Fourier[f(x)](\xi)$ represents the operation of Fourier transform on the function $f(x)$ whose outcome is equal to $\hat{f}(\xi)$.

Thus the (2-17) becomes

$$\int_{-\infty}^{\infty} \hat{\phi}(\xi)\overline{\hat{\phi}(\xi)e^{-ik\xi}}d\xi = \delta_{0k} \quad \text{or} \quad \int_{-\infty}^{\infty} \left|\hat{\phi}(\xi)\right|^2 e^{ik\xi}d\xi = \delta_{0k} \qquad (2\text{-}20)$$

Dividing the whole integral range into small intervals [$2\pi j$, 2π (j+1)] for $j \in Z$, the above equation can be written as

$$\sum_{j\in Z}\int_{2\pi j}^{2\pi(j+1)} \left|\hat{\phi}(\xi)\right|^2 e^{ik\xi}d\xi = \delta_{0k} \qquad (2\text{-}21)$$

If we replace ξ by $\xi + 2\pi j$, the limits of integration will change to 0 and 2π :

$$\int_0^{2\pi} \sum_{j\in Z} \left|\hat{\phi}(\xi + 2\pi j)\right|^2 e^{ik(\xi+2\pi j)}d\xi = \delta_{0k} \qquad (2\text{-}22)$$

Since $e^{i2\pi j} = 1$, the above equation becomes

$$\int_0^{2\pi} \sum_{j\in Z} \left|\hat{\phi}(\xi + 2\pi j)\right|^2 e^{ik\xi}d\xi = \delta_{0k} \qquad (2\text{-}23)$$

Let $\quad F(\xi) = 2\pi\sum_{j\in Z} \left|\hat{\phi}(\xi + 2\pi j)\right|^2$

The Equation (2-23) becomes $\frac{1}{2\pi}\int_0^{2\pi} F(\xi)e^{ik\xi}d\xi = \delta_{0k}$ (2-24)

$F(\xi)$ is a 2π -periodic function because

$$F(\xi + 2\pi) = 2\pi\sum_{j\in Z} \left|\hat{\phi}(\xi + 2\pi(j+1)\right|^2$$

$$= 2\pi\sum_{q\in Z} \left|\hat{\phi}(\xi + 2\pi q)\right|^2 \qquad (q = j+1)$$

$$= F(\xi)$$

Since $F(\xi)$ is periodic, it has a Fourier series like $\sum \alpha_k e^{ikx}$, where $\alpha_k = \frac{1}{2\pi}\int_0^{2\pi} F(\xi)e^{-ik\xi}d\xi$ according to the definition of Fourier series. Comparing α_k with Equation (2-24), we have $\alpha_k = \delta_{0k}$, which in turn means that $F(\xi) = 1$.

We have thus obtained an important conclusion that can be summarized as the following theorem

A function ϕ *satisfies the orthonormal condition only if*

$$2\pi\sum_{j\in Z}\left|\hat{\phi}(\xi+2\pi j)\right|^2=1 \quad \text{for all} \quad \xi\in R \qquad (2\text{-}25)$$

We have introduced the scaling relation at the beginning of this section. That scaling relation is expressed in the time domain. In fact, the scaling relation has also the equivalent expression in the frequency domain. That is

The scaling relation $\phi(t)=\sum_{k\in Z}p_k\phi(2t-k)$ is equivalent to

$$\hat{\phi}(\xi)=\hat{\phi}(\xi/2)P(e^{-i\xi/2}) \qquad (2\text{-}26)$$

where the polynomial P is given by $P(z)=\frac{1}{2}\sum_{k\in Z}p_k z^k$

It can also be proved that if the function ϕ satisfies both the orthonormal condition (2-25) and the scaling relation (2-26), the polynomial $P(z)=\frac{1}{2}\sum_{k\in Z}p_k z^k$ satisfies the following equation

$$\left|P(z)\right|^2+\left|P(-z)\right|^2=1 \text{ for } z\in C \text{ with } |z|=1 \qquad (2\text{-}27)$$

or equivalently

$$\left|P(e^{-i\xi})\right|^2+\left|P(e^{-i(\xi+\pi)})\right|^2=1 \text{ for } 0\le\xi\le 2\pi \qquad (2\text{-}28)$$

Demonstrating as an example, we attempt to derive the p_k s for Daubechies wavelets with two vanishing moments.

Let $P(e^{-i\xi})=f(\xi)$, so (2-28) can be rewritten as

$$\left|f(\xi)\right|^2+\left|f(\xi+\pi)\right|^2=1 \qquad (2\text{-}29)$$

Raising both sides of the identity $\cos^2(\xi/2)+\sin^2(\xi/2)=1$ to the 3rd power,

$(\cos^2(\xi/2)+\sin^2(\xi/2))^3=1$

or $1=\cos^6(\xi/2)+3\cos^4(\xi/2)\sin^2(\xi/2)+3\cos^2(\xi/2)\sin^4(\xi/2)+\sin^6(\xi/2)$

Since $\cos(u)=\sin(u+\pi/2)$ and $\sin(u)=-\cos(u+\pi/2)$, we have

$1=\cos^6(\xi/2)+3\cos^4(\xi/2)\sin^2(\xi/2)+3\sin^2((\xi+\pi)/2)\cos^4((\xi+\pi)/2)+\cos^6((\xi+\pi)/2)$

Let $|f(\xi)|^2=\cos^6(\xi/2)+3\cos^4(\xi/2)\sin^2(\xi/2)$, then the above equation becomes $|f(\xi)|^2+|f(\xi+\pi)|^2=1$

So (2-29) is satisfied. Furthermore

$|f(\xi)|^2=\cos^4(\xi/2)(\cos^2(\xi/2)+3\sin^2(\xi/2))$

$=\cos^4(\xi/2)\left|\cos(\xi/2)+\sqrt{3}i\sin(\xi/2)\right|^2$

Any function of the form $\cos^2(\xi/2)[\cos(\xi/2)+\sqrt{3}i\sin(\xi/2)]\alpha(\xi)$, where $|\alpha(\xi)|=1$ can be used as the function $f(\xi)$.

Since $\cos(\xi/2)=\dfrac{e^{i\xi/2}+e^{-i\xi/2}}{2}$ and $\sin(\xi/2)=\dfrac{e^{i\xi/2}-e^{-i\xi/2}}{2i}$, we have

$f(\xi)=\dfrac{1}{8}(e^{i\xi}+2+e^{-i\xi})(e^{i\xi/2}+e^{-i\xi/2}+\sqrt{3}e^{i\xi/2}-\sqrt{3}e^{-i\xi/2})\alpha(\xi)$

Selecting $\alpha(\xi)=e^{-3i\xi/2}$,

$f(\xi)=(\dfrac{1+\sqrt{3}}{8})+(\dfrac{3+\sqrt{3}}{8})e^{-i\xi}+(\dfrac{3-\sqrt{3}}{8})e^{-2i\xi}+(\dfrac{1-\sqrt{3}}{8})e^{-3i\xi}$

Remembering that $f(\xi)=P(e^{-i\xi})=\dfrac{1}{2}\sum_{k\in Z}p_k e^{-ik\xi}$, we have

$$p_0=\frac{1+\sqrt{3}}{4}\quad p_1=\frac{3+\sqrt{3}}{4}\quad p_2=\frac{3-\sqrt{3}}{4}\quad p_3=\frac{1-\sqrt{3}}{4}$$

Given that a sequence p_k is found, scaling functions $\phi(x)$ can be constructed by the cascade algorithm [69] and wavelet functions $\psi(x)$ can be constructed from

the formula $\psi(t)=\sum_{k\in Z}(-1)^k\overline{p_{1-k}}\phi(2t-k)$. That is why in cases given in many textbook expositions, a certain type of wavelet is declared by a sequence p_k instead of the curve of $\phi(x)$ or $\psi(x)$.

2.2.4 Fast filter bank implementations of wavelet transform

In the preceding sections, we have mentioned that $\phi_{j,k}=2^{j/2}\phi(2^j t-k);\ j,k\in Z$ and $\psi_{j,k}=2^{j/2}\psi(2^j t-k)$ are orthonormal bases of vector spaces V_j and W_j respectively. Suppose that we are dealing with a signal f that is already in V_j. Then in terms of orthonormal bases, we have

$$f=\sum_{k\in Z}\left\langle f,\phi_{j,k}\right\rangle\phi_{j,k} \tag{2-30}$$

Since $V_j=V_{j-1}\oplus W_{j-1}$, then

$$f=\sum_{k\in Z}\left\langle f,\phi_{j-1,k}\right\rangle\phi_{j-1,k}+\sum_{k\in Z}\left\langle f,\psi_{j-1,k}\right\rangle\psi_{j-1,k} \tag{2-31}$$

In this way, projection of the signal f in V_{j-1} can be further decomposed into V_{j-2} and W_{j-2} and so on until V_0 and W_0 as shown in Equation (2-12). Usually, it is more convenient to use bases $\phi(2^j t-k);\ j,k\in Z$, although they are not orthonormal bases. Hence, rewriting the equation (2-30), we have

$$\begin{aligned} f&=\sum_{k\in Z}\left\langle f,\phi_{j,k}\right\rangle\phi_{j,k}\\ &=\sum_{k\in Z}\underbrace{2^{j/2}\left\langle f,\phi_{j,k}\right\rangle}_{a_k^j}\times\underbrace{2^{-j/2}\phi_{j,k}}_{\phi(2^j t-k)}\\ &=\sum_{k\in Z}a_k^j\phi(2^j t-k) \end{aligned} \tag{2-32}$$

Therefore a_k^j is looked on as the wavelet transform's coefficients in V_j and $b_k^j=2^{j/2}\left\langle f,\psi_{j,k}\right\rangle$ as the coefficients in W_j.

Likewise,

$$a_l^{j-1} = 2^{(j-1)/2}\left\langle f, \phi_{j-1,l}\right\rangle \tag{2-33}$$

Since $\phi_{j,k}$ is defined as $\phi_{j,k} = 2^{j/2}\phi(2^j t - k);\ j,k \in Z$, we have

$$\phi_{j-1,l} = 2^{(j-1)/2}\phi(2^{j-1} t - l);\ j,l \in Z \tag{2-34}$$

Insert (2-14) into (2-34)

$$\begin{aligned}\phi_{j-1,l} &= 2^{(j-1)/2}\sum_{k\in Z} p_{k-2l}\phi(2^j t - k) \\ &= \sum_k 2^{-1/2}\, p_{k-2l}\phi_{j,k}\end{aligned} \tag{2-35}$$

From (2-32), there is

$$f = \sum_{k\in Z} 2^{-j/2} a_k^j \phi_{j,k} \tag{2-36}$$

By the Parseval's equation*

$$\left\langle f, \phi_{j-1,l}\right\rangle = \sum_k 2^{-(j+1)/2}\,\overline{p_{k-2l}}\, a_k^j \tag{2-37}$$

Compare (2-33) with (2-37), we have

$$\begin{aligned}a_l^{j-1} &= 2^{(j-1)/2}\left\langle f, \phi_{j-1,l}\right\rangle \\ &= 2^{(j-1)/2}\sum_k 2^{-(j+1)/2}\,\overline{p_{k-2l}}\, a_k^j \\ &= 2^{-1}\sum_{k\in Z}\overline{p_{k-2l}}\, a_k^j\end{aligned} \tag{2-38}$$

In the same way, since $b_k^j = 2^{j/2}\left\langle f, \psi_{j,k}\right\rangle$ there is also

$$b_l^{j-1} = 2^{-1}\sum_{k\in Z}(-1)^k\, p_{1-k+2l}\, a_k^j \tag{2-39}$$

Equations (2-38) and (2-39) are together called decomposition formulas.

---

* Let V be a complex inner product space with an orthonormal basis $\{u_k\}_{k=1}^{\infty}$. If $f \in V$ and $g \in V$, that is

$$f = \sum_{k=1}^{\infty} a_k u_k \quad \text{and} \quad g = \sum_{k=1}^{\infty} b_k u_k$$

then

$$\left\langle f, g\right\rangle = \sum_{k=1}^{\infty} a_k \overline{b_k}$$

This is called Parseval's equation.

Repeating the decomposition formulas, we can obtain all wavelet coefficients associated with all the scale from *j* to 0, if the wavelet coefficients a_k^j are known.

Furthermore, let *h* and *g* be the sequences

$$h_k = \frac{1}{2}\overline{p_{-k}} \tag{2-40}$$

and

$$g_k = \frac{1}{2}(-1)^k p_{k+1} \tag{2-41}$$

Define two discrete filters (convolution operators*) *H* and G via

$$H(x) = h * x \tag{2-42}$$

and

$$G(x) = g * x \tag{2-43}$$

Embodying the *x* in the preceding *H(x)* and *G(x)* by a specific expression, e.g. $x = a^j$, we have

$$H(a^j)_l = \frac{1}{2}\sum_{k\in Z}\overline{p_{k-l}}a_k^j \tag{2-44}$$

Comparing (2-44) with (2-38), we get $a_l^{j-1} = H(a^j)_{2l}$. Similarly, there is also $b_l^{j-1} = G(a^j)_{2l}$. $H(a^j)_{2l}$ or $G(a^j)_{2l}$ here means that we have discarded all odd coefficients. This operation is so called down-sampling. If we use $\downarrow 2$ to represent down-sampling, then the 1-level wavelet transform can be illustrated by Fig 2.3

---

* Convolution of two sequences $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ and $y = (\ldots, y_{-1}, y_0, y_1, \ldots)$ is defined by

$$(x * y)_l = \sum_{k\in Z} x_k y_{l-k}$$

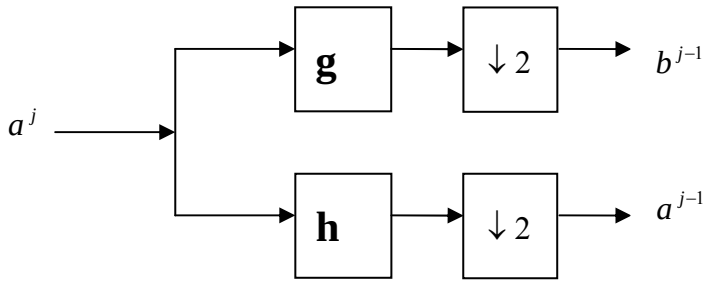


Fig 2.3 One level wavelet transform

For the multi-level wavelet transform, the 1-level wavelet transform in Fig 2.3 will be iteratively applied to a^{j-1}, a^{j-2} and so on. This will be discussed in detail in the next Chapter where we will use multiresolution analysis and a variant multi-level wavelets transform to quickly estimate the durations of onset transients of piano tones.

To end this Chapter, we will summarize the general procedures on how to process a signal by basic wavelet analysis.

1. *Sample*. This step is actually a preprocessing step. For a continuous signal, it must be sampled at a sufficient frequency rate so that we can capture its essential details. The specific sampling rate could depend on a variety of factors. For example, a good rule of thumb frequently used in sampling a sound signal is to use a rate that is at least above 40 kHz, double the audible frequency limit of human hearing, roughly 20 kHz, due to the Nyquist theorem. But because of aliasing effect, before entering the sampling process, all frequency components higher than the desired frequency should not be presented in the signal. This is often done by sending the signal through a low-pass filter with a cut-off frequency 20 kHz.
2. *Pre-processing*. In Fig 2.3, the input parameters are the wavelet coefficients

a^j, instead of the signal $x[n]$. If we directly input $x[n]$ into the wavelet transform as shown in Fig 2.3, we commit the 'wavelet crime' (Strang and Nguyen) One solution Strang and Nguyen recommend is that the samples $x(n)$ are converted to coefficients *a(k)* by

$$a(k) = \sum_n x(n)\overline{\phi(n-k)} \qquad (2\text{-}45)$$

3. *Decompose.* Once the signal has been sampled, we iterate equation (2-38) and (2-39) until reaching an appropriate level (for instance $j = 0$). The output of this step is coefficients of all levels in the wavelet transform.
4. *Process.* At this step, the decomposed (or transformed) signal can be processed depending on what we wish to accomplish. For instance, if our purpose is to compress the signal, we can discard those insignificant coefficients. If we want to detect a singularity event involved in a signal, we should pay attention to those abnormally large coefficients. After various modifications, the output may be stored or immediately reconstructed to recover the signal in the next step.
5. *Reconstruct.* To recover the signal, the reconstruction algorithm is invoked here. The specific reconstruction algorithm is summarized in the following formulas and figures.

   Define,

$$rh = p_k \qquad (2\text{-}46)$$

$$rg = \overline{p_{1-k}}(-1)^k \qquad (2\text{-}47)$$

$$RH(x) = rh * x \qquad (2\text{-}48)$$

$$RG(x) = rg * x \qquad (2\text{-}49)$$

thus, the 1-level inverse wavelet transform is shown in Fig 2.4

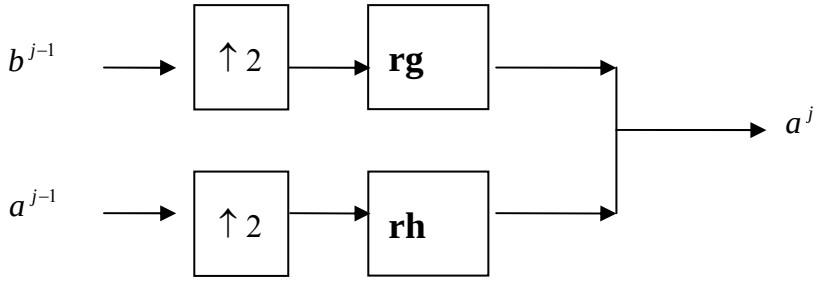


Fig 2.4 One level inverse wavelet transform

where $\uparrow 2$ denotes up-sampling which is done by inserting zero samples between each incoming sample)

Chapter 3 Waveform Analysis of Piano tones' Onset Transients

Consider a modern piano keyboard (as shown in Fig 3.1) with 88 keys. The leftmost key (top in Fig 3.1) is an A0 tone with a fundamental frequency 27.5 Hz, the number besides the piano key in units of cycles per seconds, or Hertz. This distribution of frequencies on the piano is well-known as equal temperament tuning. That is each successive pitch is obtained by multiplying the pitch of the preceding note by the twelfth root of two. For example, A4 is normally tuned to 440 Hz as its fundamental frequency. The pitch of the next semitone A4# can be derived by $440 \times \sqrt[12]{2} = 466.16$ Hz. Likewise, to get B4's fundamental frequency, multiply 440 by $\left(\sqrt[12]{2}\right)^2$. But we should note that such a list of frequencies is only for a theoretical ideal piano. On an actual piano, the ratio between semitones may be tuned differently, especially at the high and low ends, as is done by piano tuners for greater sonority and brilliance.

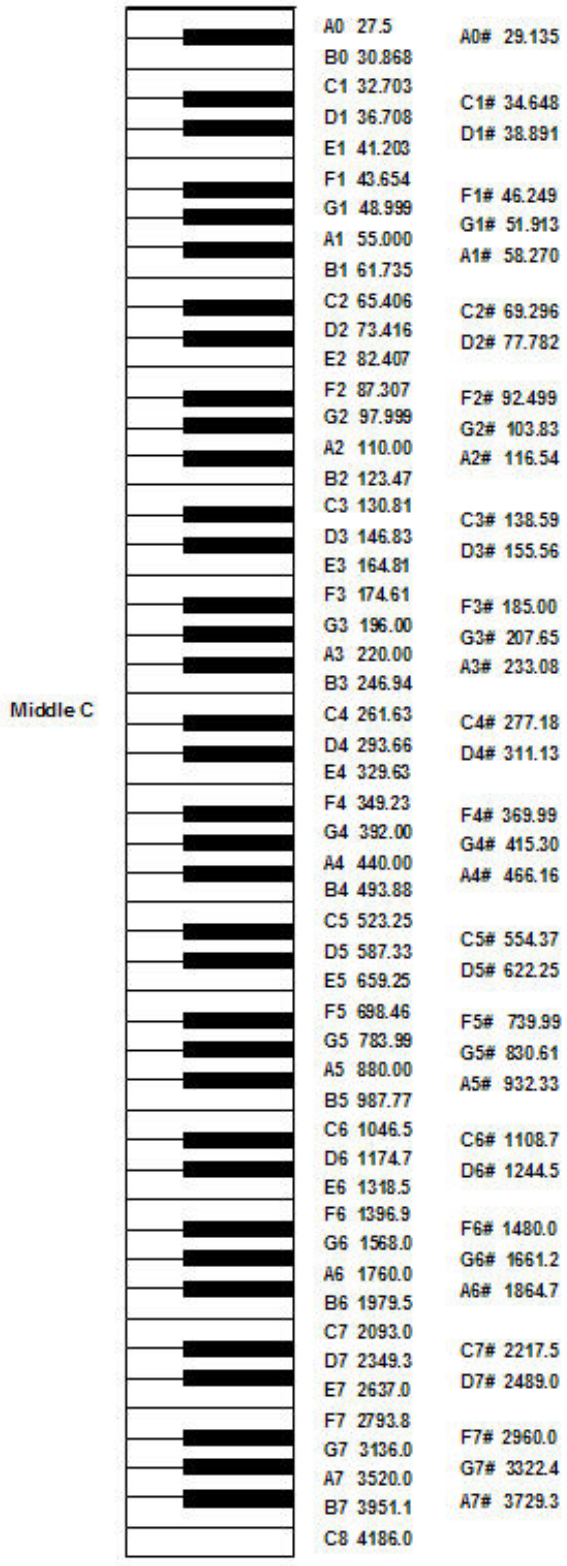


Fig 3.1 A modern standard piano keyboard with the distribution of fundamental frequencies

As shown in Fig 3.2, the waveform of a typical piano tone increase rapidly to its peak amplitude within a very short transient duration, followed by a steady and much slower exponential-like decay. This short transient duration is usually called the *onset transient* (or *onset attack).* As the stage which gives birth to all the tone's partials, the onset transient of a piano tone contains the most rapid changes in spectral energy. Accordingly, this stage might also give rise to many important properties of piano tones that have already been introduced in Chapter 1.

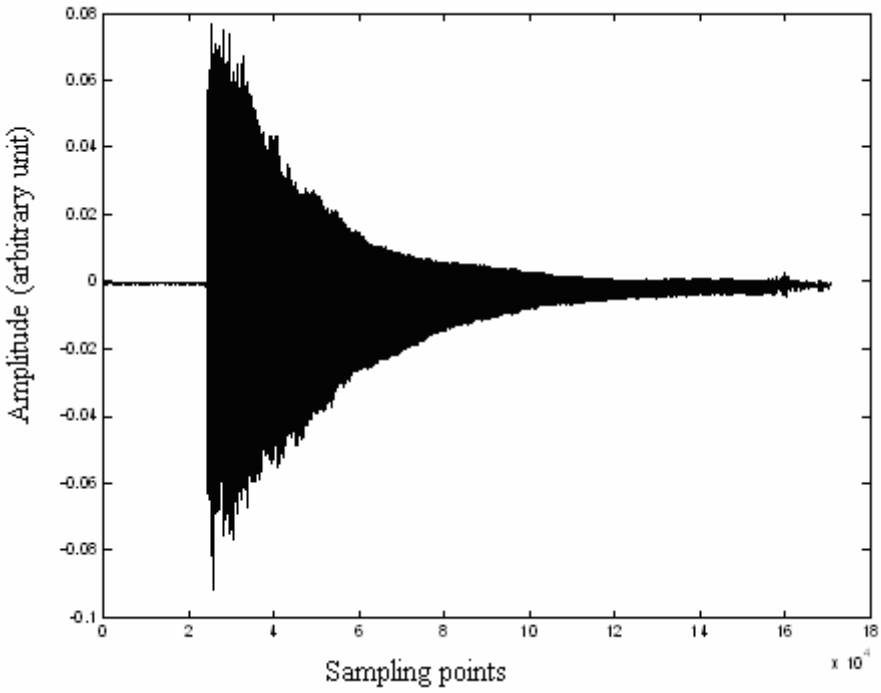


Fig 3.2 The waveform of a piano tone C4 whose corresponding key is located in the middle of the piano keyboard

In this dissertation, piano onset transients will be analyzed from two perspectives. Firstly in this Chapter, attention will be paid on their most straightforward feature, namely their waveforms. Although the '*very-short*' feature of onset transients has been repeatedly highlighted in preceding Chapters, we avoided giving a direct answer

to the question ‘how short are these transients?’ Hence this Chapter will attempt to find a way to rapidly estimate their durations with the help of the discrete wavelet transform (DWT) and its multiresolution techniques. As for the second perspective, the spectral analysis of onset transients, it will be dealt with in the next Chapter.

3.1 Definitions for Onset transients

As stated in the introductory chapter, the onset transient is generally taken to be the initial part of a musical tone. However, such a rough description only gives its relative temporal position in each tone. This general definition is not sufficient to differentiate the onset transient from the following more stable portion of a tone. A more specific definition, derived from the properties of the onset transients of piano tones, is necessary if it is to be of practical usefulness.

1. Definition by *Peak Amplitude*: The onset transient may be defined as starting from the first measurable instant of the piano tone to its peak amplitude point. Also as seen in Fig 3.2, a very rapid rise to its peak amplitude is one apparent but important common property for all piano tones. The waveforms of some other piano tones are also listed in Fig 3.3 and Fig 3.4.

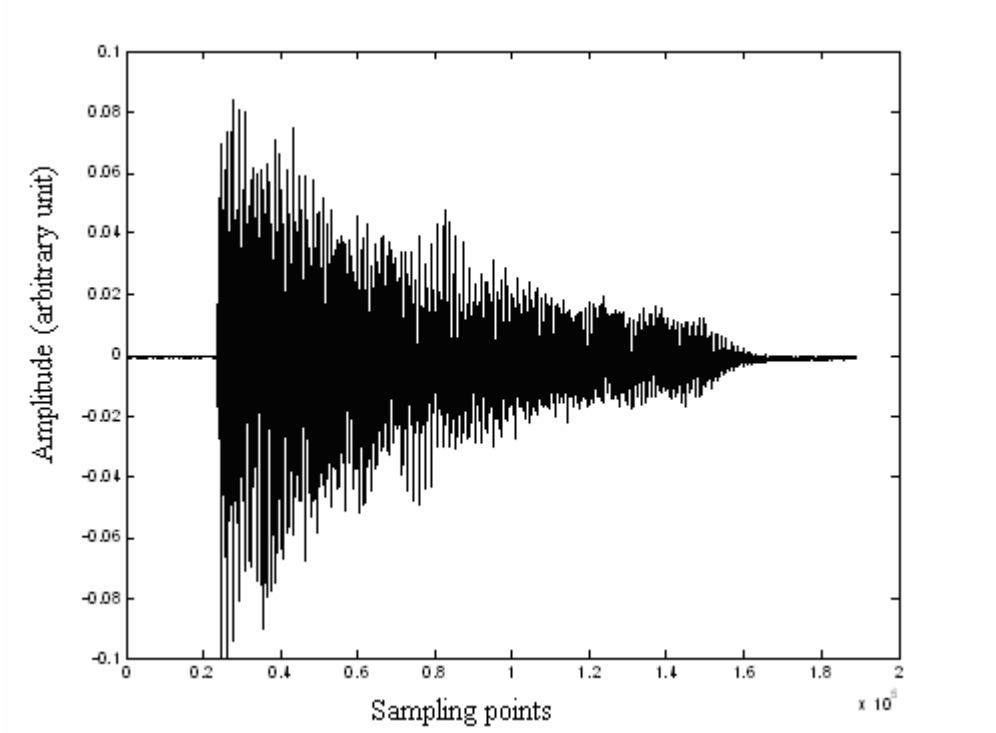


Fig 3.3 The waveform of piano tone A0 whose corresponding key is located on the extreme left of the piano keyboard

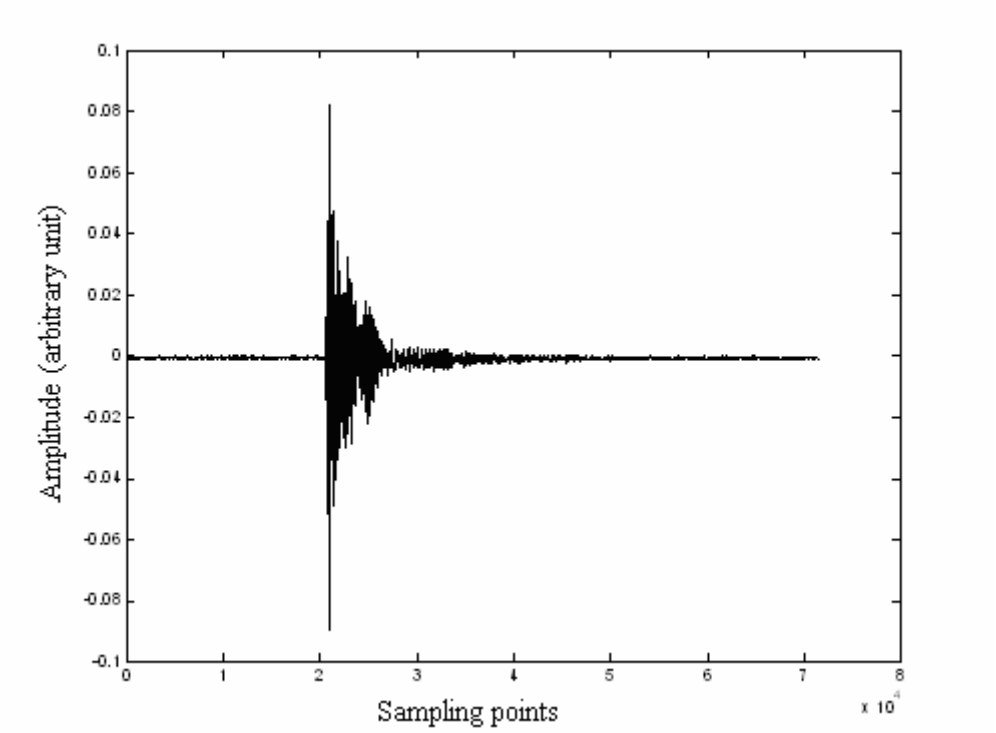


Fig 3.4 The waveform of piano tone C8 whose corresponding key is located on the extreme right of the piano keyboard

2. Definition by *Evolving Process*: An alternative way is to define the onset transient as starting from the first instant which is acoustically measurable, to the point after which the period of the tone's waveform is completely established. Similar to Florian Keiler's paper [70] where he plotted the evolving process for the harpsichord's onset transient waveform, Fig 3.5 illustrates how a piano tone, C4, evolves. In the figure, **A** is the initial noise; **B** is the time during which the periodic motion is not established yet; **C** is the quasi-periodic period; **D** marks the start of the stable period. The complete onset transient of a piano tone thus includes **A+B+C+D.** But we should note that parts **B** or **C** as defined in Fig 3.5 for different piano tones may not occur or may occur more than once. Some bass tones have two or three quasi-periods while for some treble tones B and C are not always obvious.

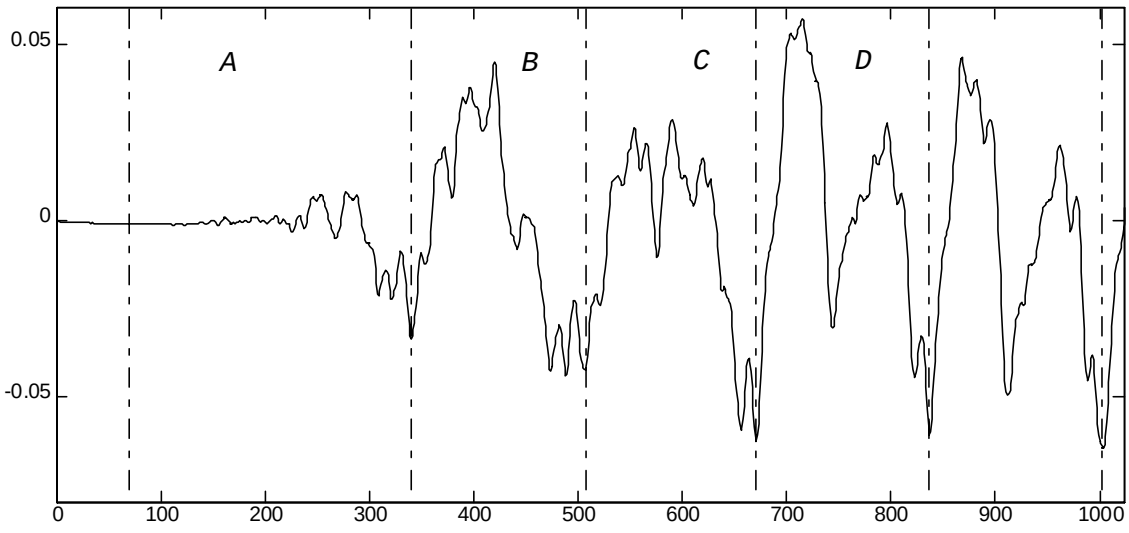


Fig 3.5 The evolving process of the piano tone C4, roughly the initial 1,024 sampled points as the x-axis shows.

Both definitions are derived from the direct observation and measurement of the

piano tone waveform. The difference between the two definitions is that the first one looks at the overall shape of the waveform while the latter looks more closely at the details of the waveform.

Using the second definition of onset transient, if we assume that the duration of each part A, B, C, or D in Fig 3.5 is approximately equal to the fundamental period of the tone, then the whole onset transient duration is equal to about four times the fundamental period. Assuming this relationship, the relationship of the onset transient duration versus the piano tone fundamental frequency is presented in Fig 3.6.

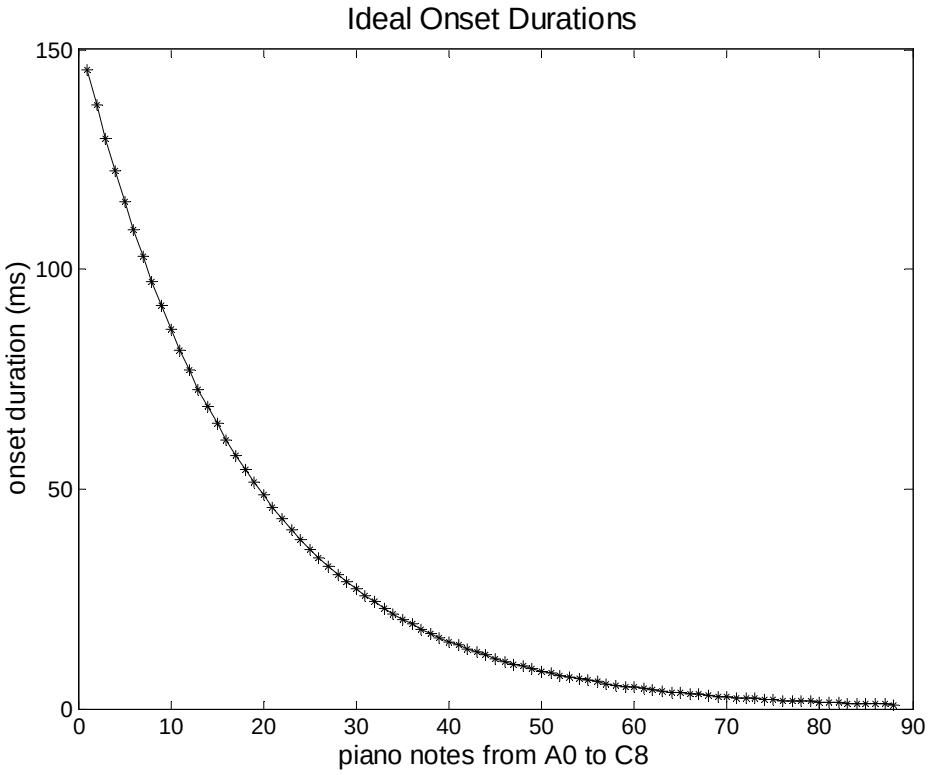


Fig 3.6 Onset durations of all piano tones in the ideal theoretical situation

However in practice, A, the initial noise part, may not be equal to the fundamental period, and A is also negligible for some piano tones. In the B part, the

periodic behavior has not yet been established. More significantly, as stated before, the parts B or C may not occur or may occur more than once. For these reasons, the actual durations of the onset transient may be either higher or lower for each frequency than that depicted in the ideal curve, and the actual curve may depart from the ideal curve. Nevertheless, the above ideal curve gives us a starting basis for understanding the behavior of the duration of the onset transient with respect to fundamental frequency.

3.2 Measuring Durations of piano onset transients

In the last section, we have attempted to define onset transients of piano tones in two possible ways, for the purpose of facilitating the measurement of the onset duration of each piano tone. Based on the latter definition, onset durations of all piano tones in the ideal theoretical situation have been plotted in Fig 3.6 to help us to anticipate the variation trends of onset durations with different piano tones with distinctly different fundamental frequencies. Later in this section, we will investigate the measurement of onset duration in practical situations by a wavelet-based method.

3.2.1 The challenges

The estimation and determination of the durations of parts A, B, C and D of each onset transient can prove to be a very time consuming task. It may be thought that the first definition would give us a simpler and fast method of computing the onset

transient duration by measuring the time distance between the beginning instant of each onset transient and the time point when the peak amplitude occurs. If the piano tone is directly measured from a real piano, it might indeed be simple to determine the so-called first measurable acoustical instant in the definition. Our piano tones were obtained from the McGill University Master Samples (MUMS) compact discs as is the normal practice, so that the recordings are known to have been consistently recorded. All these tones are arranged in sound tracks with silent intervals between each other as shown in Fig 3.7.

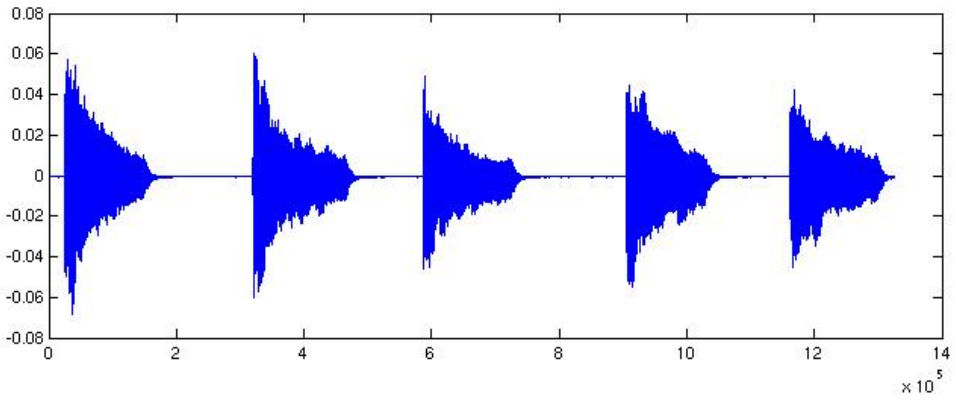


Fig 3.7 The arrangement of piano tones in a segment of MUMS CD sound tracks

Therefore it is difficult to locate the so-called first measurable acoustical instant because even if the acoustical amplitudes in the intervals are very minute, they are not actually zero. The conventional solution to this problem is to set a threshold where any amplitude value below it will be set to zero. However, the threshold that works well for a certain tone may be quite inappropriate for others. It would be very laborious and inconsistent if one has to repeatedly zoom in on each waveform and set

the threshold for each tone. In this section, we will use *wavelet multiresolution decomposition* (WMD) to automatically locate the beginning of each tone and the peak of a piano tone computationally. Once the beginning point has been determined, the duration can be easily obtained by calculating the time difference between the beginning point and the peak point according to the first definition.

3.2.2 Wavelet Multiresolution Decomposition by filter banks and 'wavelet crime'

The discrete wavelet transform (DWT) decomposes a digitized waveform by passing the discrete time sequence of the digital samples of the waveform through pairs of digital filters (one high-pass and one low-pass) iteratively. The basic decomposition unit consisting of a low-pass filter *h* and a high-pass filter *g* is shown in Fig 3.8 (a).

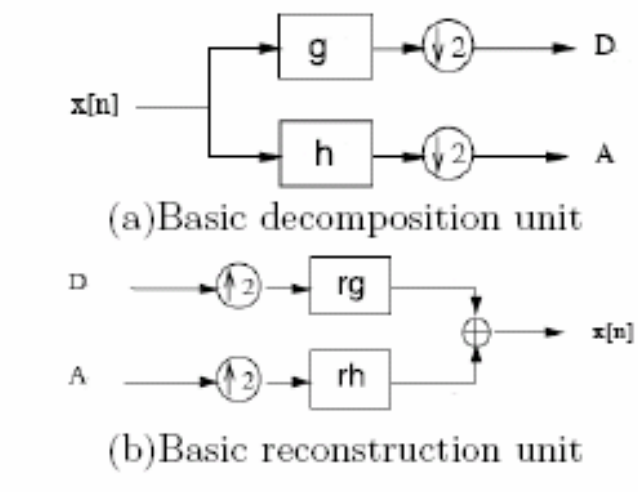


Fig 3.8 One stage 1-D wavelet transform

Let $x[n]$ be the original discrete time sequence of the digitized audio waveform and *h*, *g* be the discrete wavelet transform's low-pass and high-pass filters. In Fig 3.8, ↓ 2 and ↑ 2 denote down-sampling and up-sampling respectively. The discrete wavelet transform can be briefly summarized as in Fig 3.8(a) and Fig 3.9. The input sequence $x[n]$ passes through filters *g* and *h* respectively, whose outputs are the high-pass *detail* subbands *D* and the low-pass *approximation* subbands *A* respectively. As *D* and *A* would each have the same number of samples as the original signal, both *D* and *A* then undergo a down-sampling operation to reduce the number of samples by half, so that the total number of original samples is unchanged and divided equally between the outputs of *g* and *h*. Multi-level discrete wavelet decomposition is achieved by applying the *g* and *h* filters followed by down-sampling on each *approximation* subband *A* iteratively.

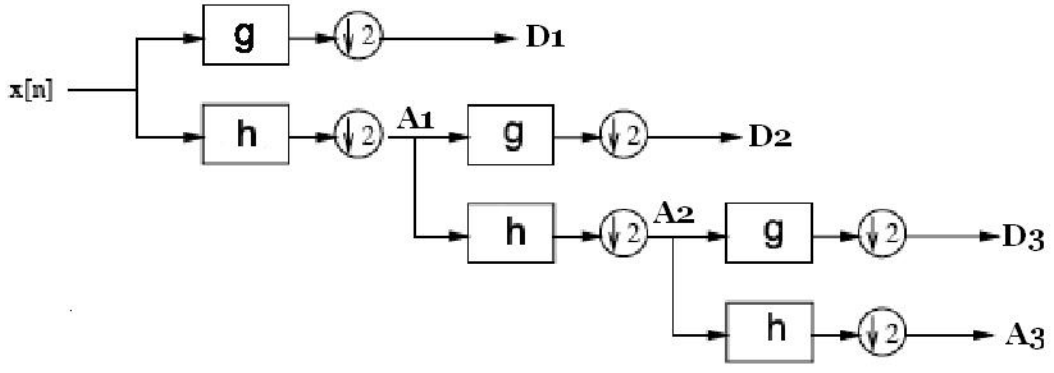


Fig 3.9 Multi-level decomposition

The basic reconstruction unit consists of a pair of low- and high-pass filters as shown in Fig 3.8 (b). Pairs of the *approximation* subbands *A* and the *detail* subbands

D are first upsampled to double the number of samples. The upsampled *A* and *D* are then passed through the corresponding reconstruction filters, *rh* and *rg*, to reconstruct the *approximation* subband of the preceding level and ultimately to obtain $x[n]$ as shown in Fig 3.10. The reconstruction process constitutes the inverse Discrete Wavelet Transform or the IDWT.

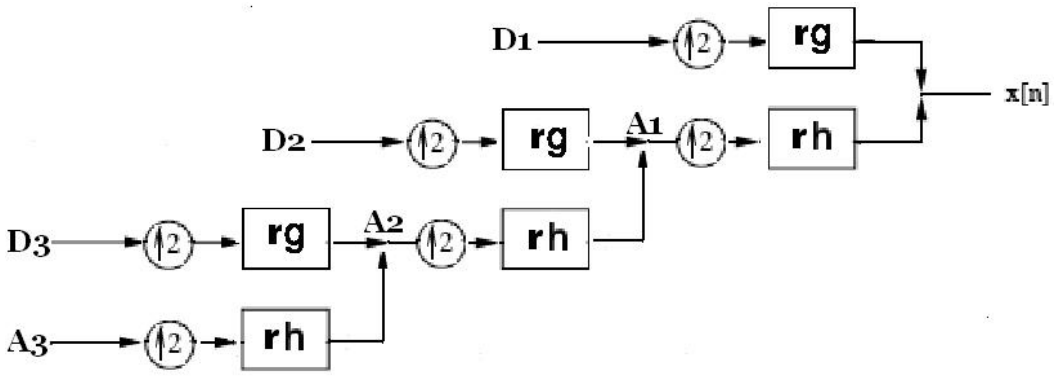


Fig 3.10 Multi-level inverse Discrete Wavelet Transform

However, in our multiresolution analysis in this Chapter, we do not use such an inverse discrete wavelet transform to restore the signal. Instead, each subband is reconstructed separately (or independently) from the other particular subbands. That means we are able to obtain the signal waveform associated with each subband. The whole process is described in Fig 3.11 and a_x^j and d_x^j are used to respectively denote the j^{th} level's approximation signal and detail signal.

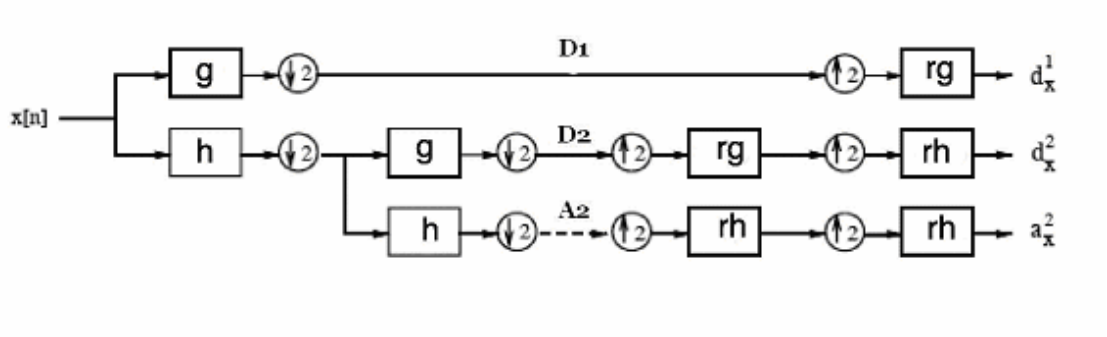


Fig 3.11 Diagram of multiresolution decomposition

It can be theoretically proved that the summation of all components a_x^j, d_x^j, d_x^{j-1}…and d_x^1 is exactly equal to the original signal $x[n]$ itself. To demonstrate this property of multiresolution analysis, we will use Matlab to do the verification.

In Matlab, we usually use the following group of commands to simulate a sampling procedure for a sine wave of f Hz (e.g 200 Hz) by a sampling frequency Fs (e.g 1024 Hz)

```
% Sampling frequency
   Fs = 1024;
% Time vector of 1 second
   t = 0 : 1/Fs : 1;
% Create a sine wave of 200 Hz.
   s = sin(2 * pi * t * 200);
```

In this way, we can fabricate an imaginary signal composed of 4 sine waves with different frequencies (e.g. respectively 10 Hz, 40 Hz, 190 Hz and 450 Hz). We also arrange these four components in different time locations. These fabricated four components are shown in Fig 3.12.

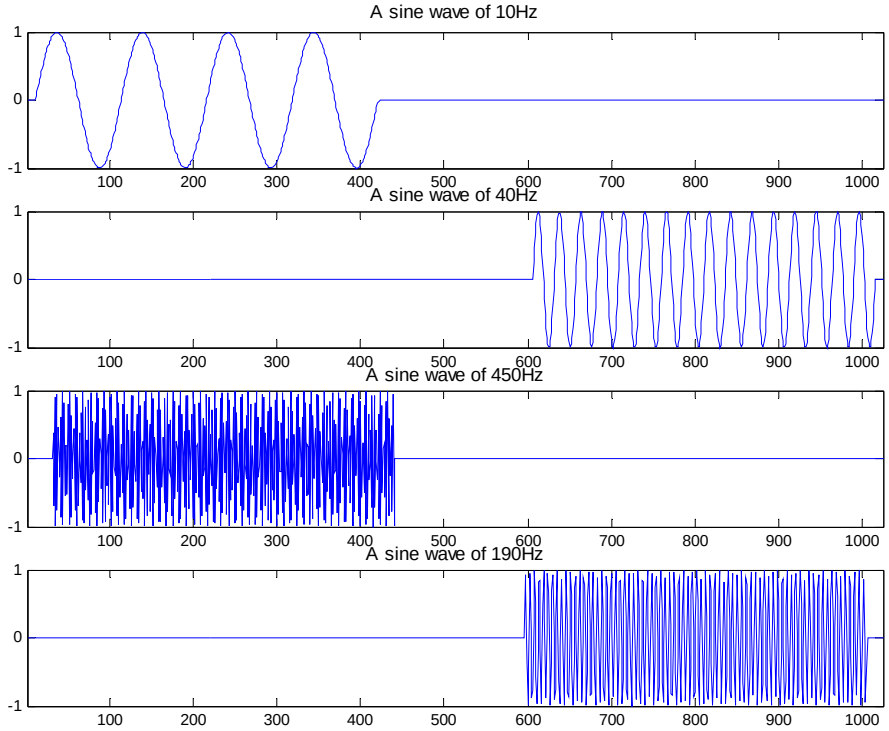


Fig 3.12 Four sine functions with different frequencies at different time

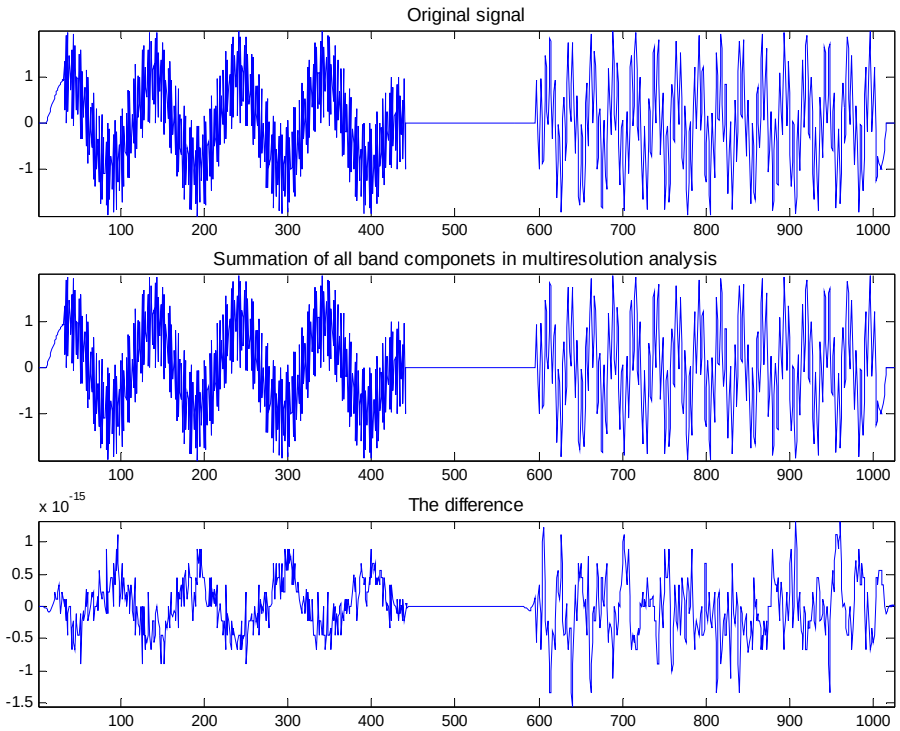


Fig 3.13 The comparison between the original signal and the summation of all subbands in using multiresolution analysis

After applying the multiresolution analysis displayed in Fig 3.11 on the fabricated signal to three levels and then summing all a_x^j and d_x^j, we find that the difference between the original signal and the summation of all a_x^j and d_x^j is on the order of 10^{-15}, small enough to be ignored. The result is shown in Fig 3.13.

The specific contents of all a_x^j and d_x^j are shown in Fig 3.14.

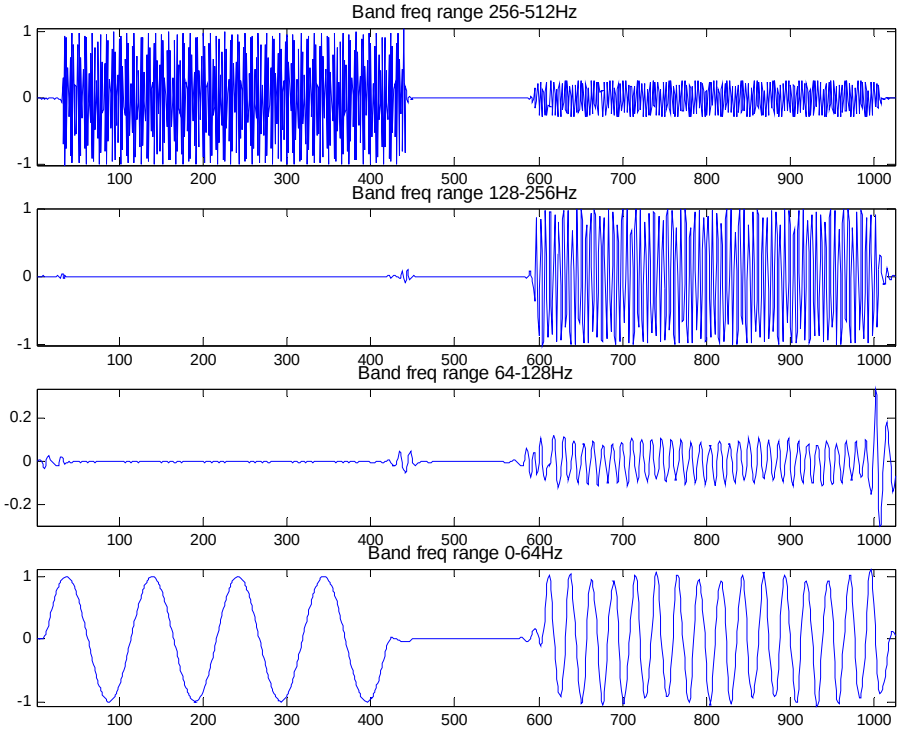


Fig 3.14 The contents of every subband in the three level multiresolution analysis. From top to bottom, each subband respectively corresponds to d_x^1, d_x^2, d_x^3 and a_x^3

From Fig 3.14, we see that four components in Fig 3.12 can be easily identified and be put both in the correct frequency bands and the correct time locations. As shown in Fig 3.14, components with different frequencies (e.g., the sine functions with 450 Hz and 190 Hz) are filtered into d_x^1 and d_x^2. Although the frequencies of

components at 10 Hz and 40 Hz are too close to be separated by only three level multiresolution analysis, we can still separate them in different time locations in subband a_x^3.

Through this demonstration, we can conclude that there are at least two advantages in our multiresolution analysis. Firstly, a full reconstruction of the original signal may be performed by simply summing the waveforms of each subband. Secondly, listing the waveforms of each subband, we break the original signal into different frequency and time slots and thus can reveal the signal's time-frequency features to some degree.

3.2.3 Measurement and Analysis

Our target here is to acquire the pattern of a signal's energy envelope. This process can be done by squaring sampling values and then filtering out high frequencies. For example, a simple signal has the form $f(t) = g(t)\sin(t)$ where $f(t)'s$ energy envelope is $g^2(t)$. Squaring $f(t)$, we have $f^2(t) = g^2(t)\sin^2(t) = g^2(t)(\frac{1-\cos(2t)}{2})$. After filtering out the high frequency content, only the item $\frac{1}{2}g^2(t)$ is left. Although there is a constant multiplying-factor (i.e., 1/2) before $g^2(t)$, $\frac{1}{2}g^2(t)$ has the same pattern as the signal's energy envelope. For the same reason, after applying the multiresolution analysis, we can ignore all high frequency components (d_x^j in Fig 3.11) to get the pattern of energy envelope (i.e. a_x^j)

In our measurement, each piano tone in the MUMS CD was sampled at 44,100 Hz and stored as a .wav file. After importing this .wav file into Matlab, we obtained its 1-Dimensional amplitude array. Each amplitude value was then squared to acquire the energy array. Finally the multiresolution analysis shown in Fig 3.11 was applied to this energy array. Here we used the Coiflet 1 wavelet. Since the sampling rate is 44,100 Hz, the bandwidth of the sampled signals is from 0 Hz up to half of the sampling rate according to the Nyquist theorem, (i.e., 22,050 Hz), which is roughly the upper limit of human hearing. For a wavelet transform with *b* levels, the signal is decomposed into *b+1* subbands as shown in Fig 3.11, which segment the frequency axis into [0, 22,050/ 2^b], [22,050/ 2^b , 22,050/ 2^{b-1}], ... , [22,050/2, 22,050] respectively. The value of *b* is chosen so that we obtain a good envelope curve in a_x^b. Following this, all oscillations within the energy envelope were filtered out, leaving only the energy envelope as the *b* level's approximation signal (i.e. a_x^b) which is shown in Fig 3.15.

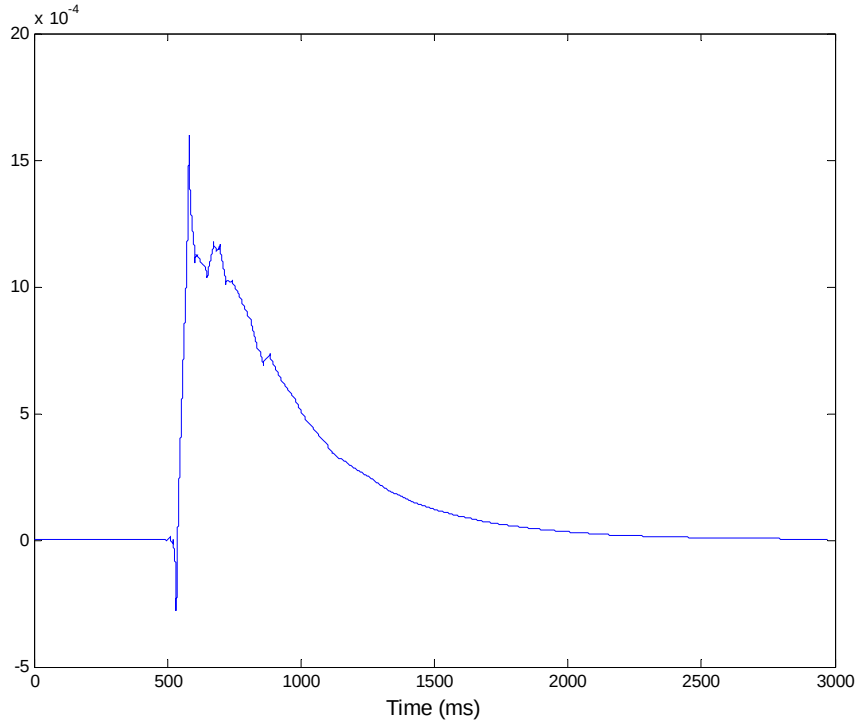


Fig 3.15 The energy envelope of C4 piano tone

At the beginning of the resulting envelope, there is a negative dip followed by a huge upward peak. The dip is surprising because energy cannot take a negative value. However, we may explain the appearance of this negative dip as follows. Firstly, just as sine or cosine functions are used as bases in the Fourier transform, a wavelet basis set is employed to analyze the signal in the wavelet transform. This wavelet basis set includes scale functions $\phi_{b,k}$ and wavelet functions $\psi_{b,k}$ where *k* is the translation value and b represents the scale or the level value. Secondly, the multiresolution decomposition here is a linear operation, which means that by summing a_x^b and all d_x (as shown in Fig 3.11) from level 1 to level *b* together, we obtain the same signal as the original. Combining these two facts leads to the following formula for a given known *b*:

$$f = \underbrace{\sum_k \langle f, \phi_{b,k} \rangle \phi_{b,k}}_{a_x^b} + \underbrace{\sum_k \langle f, \psi_{b,k} \rangle \psi_{b,k}}_{d_x^b} + \cdots + \underbrace{\sum_k \langle f, \psi_{1,k} \rangle \psi_{1,k}}_{d_x^1} \qquad (3\text{-}1)$$

Suppose that the envelope peak just falls into the range of $\phi_{b,m}$. To match the envelope peak, the whole waveform of $\phi_{b,m}$, has to be enlarged in both the positive and negative y-axis directions. If the targeted signal rapidly changes its amplitude from 0 to the maximum in a short time, as the piano tone does, the enlarged positive peak of $\phi_{b,m}$ (B in Fig 3.16) may possibly match the positive peak of the signal envelope, and the negative peak of $\phi_{b,m}$ (A in Fig 3.16) leaves a negative dip at the start of the tone.

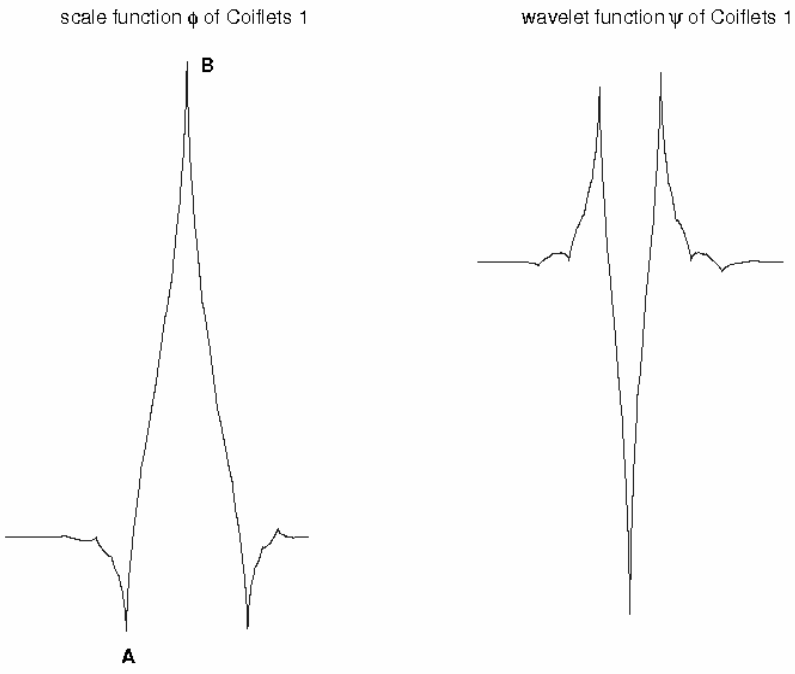


Fig 3.16 Scaling and wavelet function of wavelet bases Coiflet 1

Corresponding to this negative dip, we should be able to locate a positive hump in the d_x^b waveform which adds linearly to the negative dip to give us the zero of the actual piano tone envelope in question. Our wavelet analysis has verified this hypothesis (Fig 3.17).

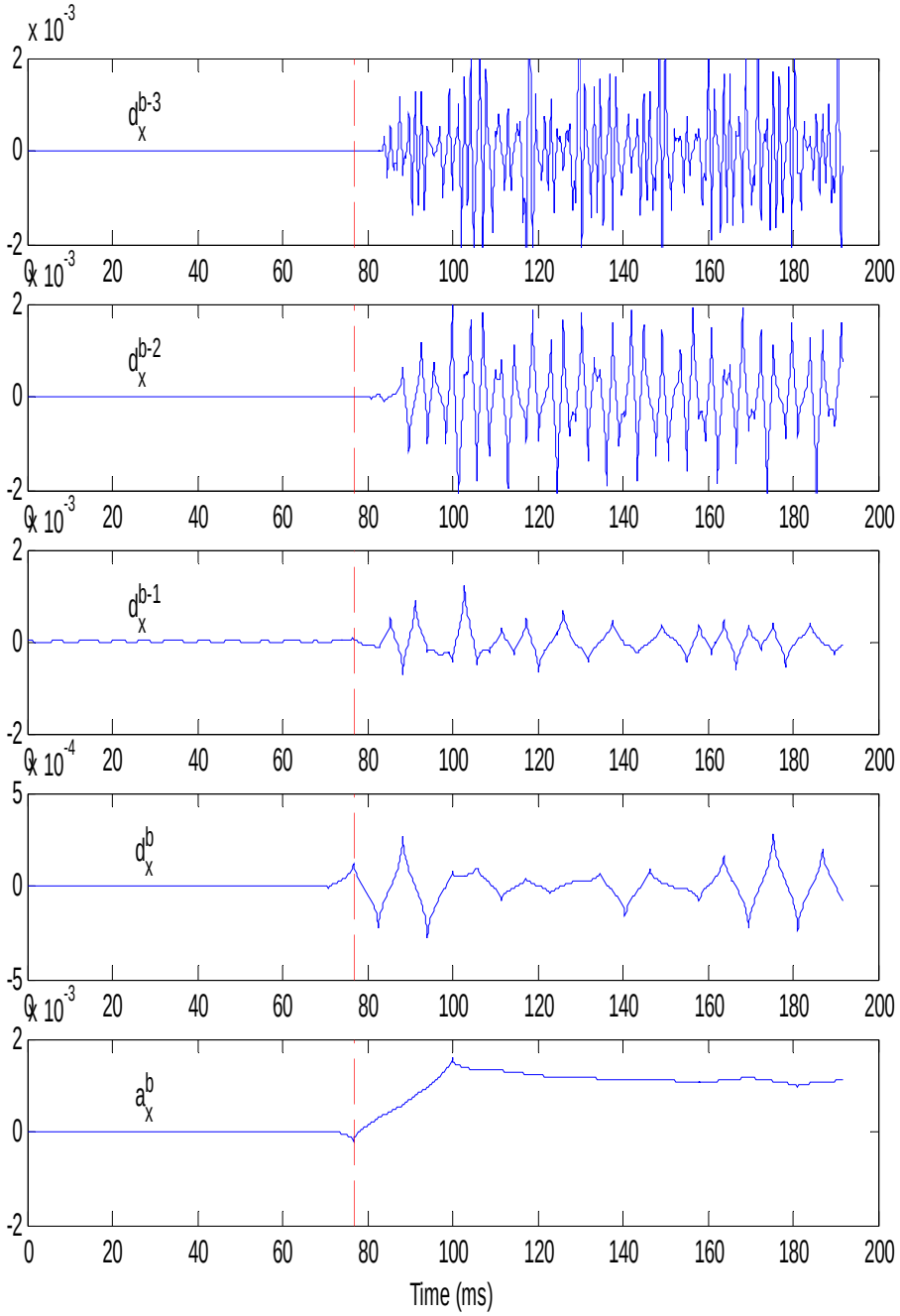


Fig 3.17 The waveforms of some subbands in the multiresolution analysis of C4 piano tone

Our results also indicate that this negative dip points to the neighborhood of beginning of the piano tone and thus can be approximately used as the starting point of that tone. This is also shown in Fig 3.18.

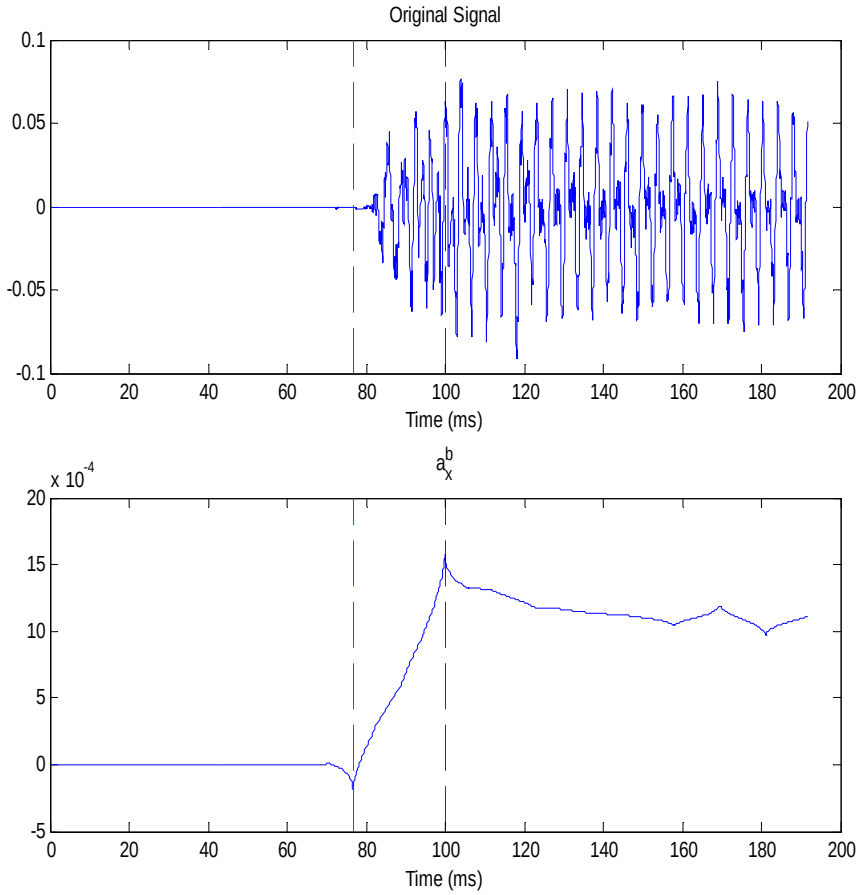
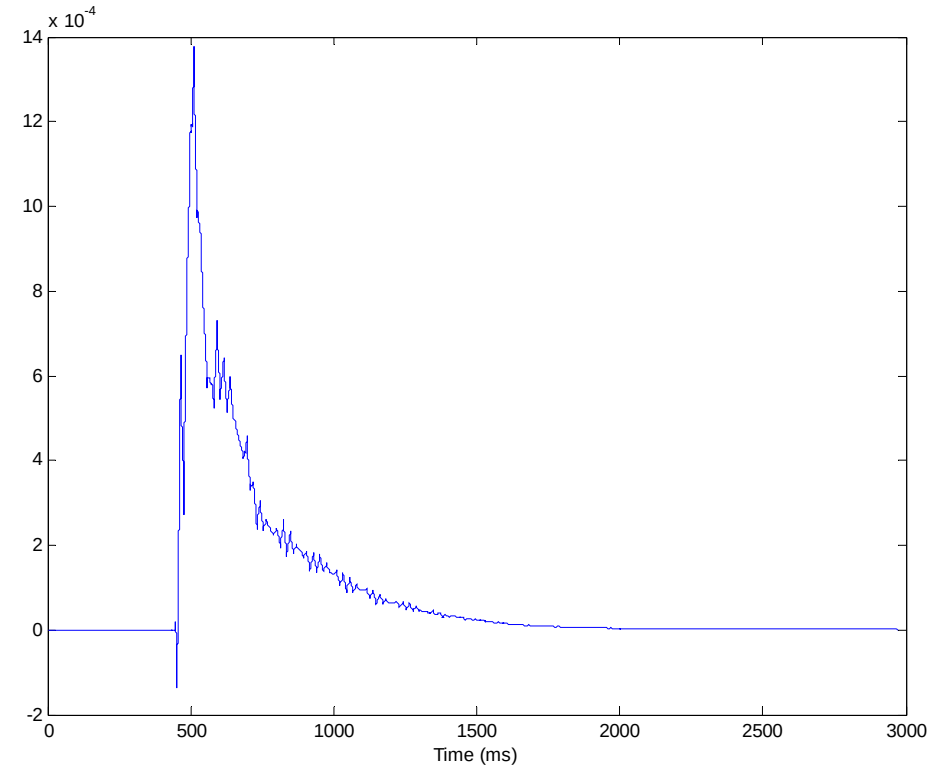


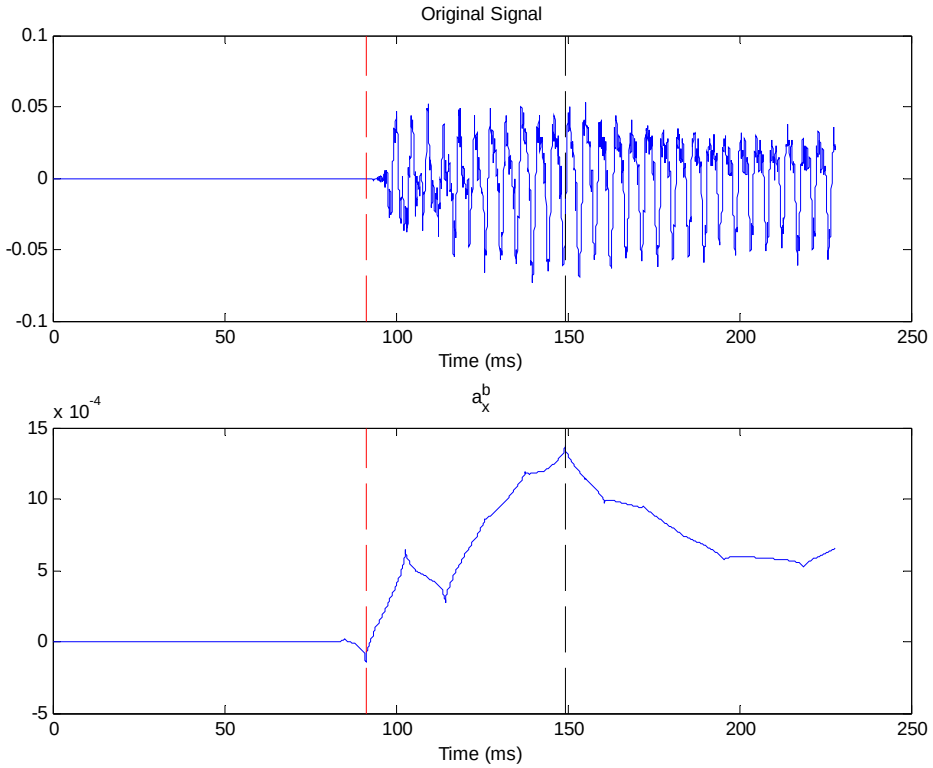
Fig 3.18 Results of the multiresolution analysis for C4 piano tone

With the wavelet transform going to more levels, we actually do some kind of 'averaging' operations on the signal in the approximation band. For example in the extreme case, the approximation band of the Haar wavelet transform is from the mean values of paired coefficients in the preceding level. So the peak of the energy envelope we have computed may not be identical to the exact peak point but we can ensure that the 'averaged' magnitudes are approximately around the largest magnitudes. Similarly, the dip may not exactly fall on the first oscillation of the signal but it will very be close to it.

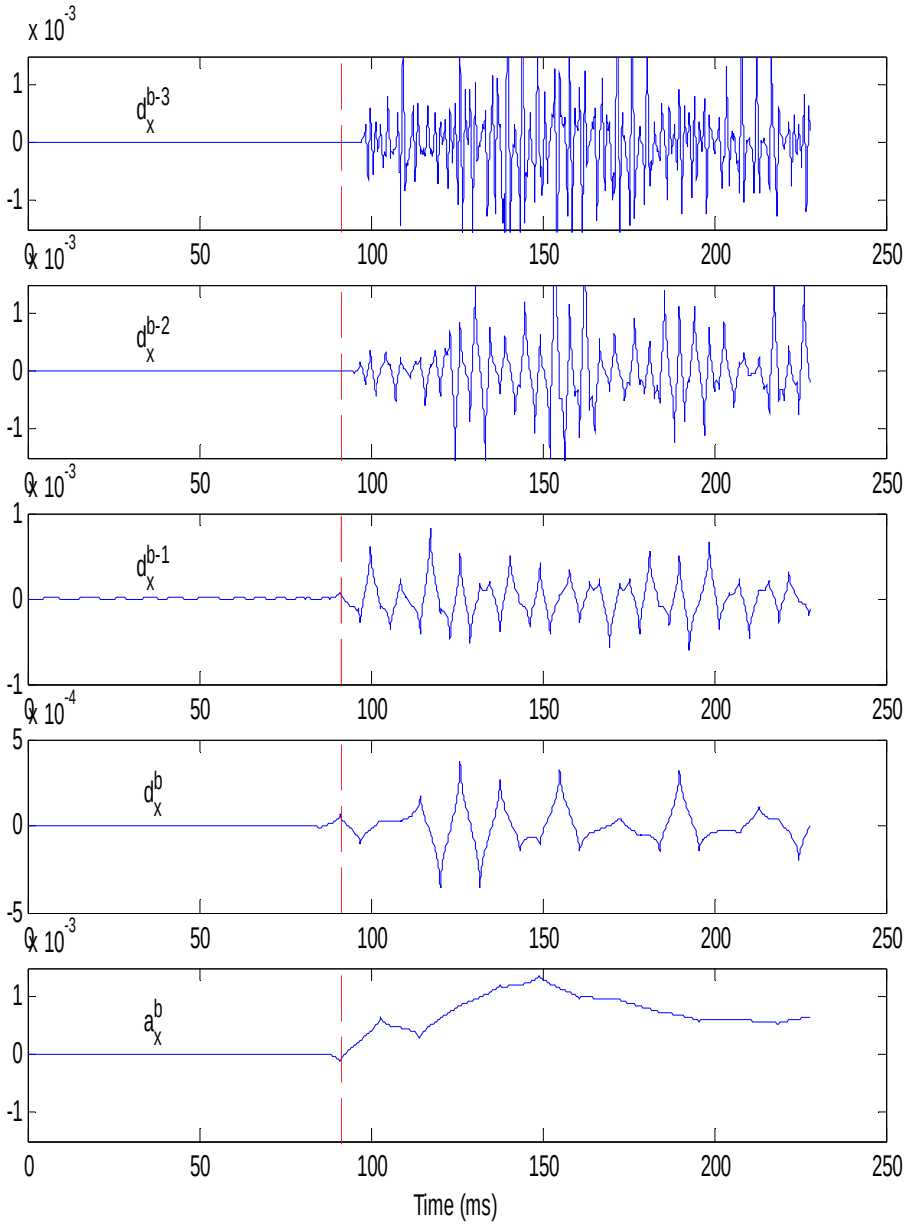
In order to test the applicability of multiresolution analysis to onset duration measurements, more experiments were conducted on different pitches of the piano. The results further confirmed our theory, which can be seen from Fig 3.19 to Fig 3.23 for A3, D1, F5, B0 and G7 piano tones.



(a) Energy envelope of A3 piano tone

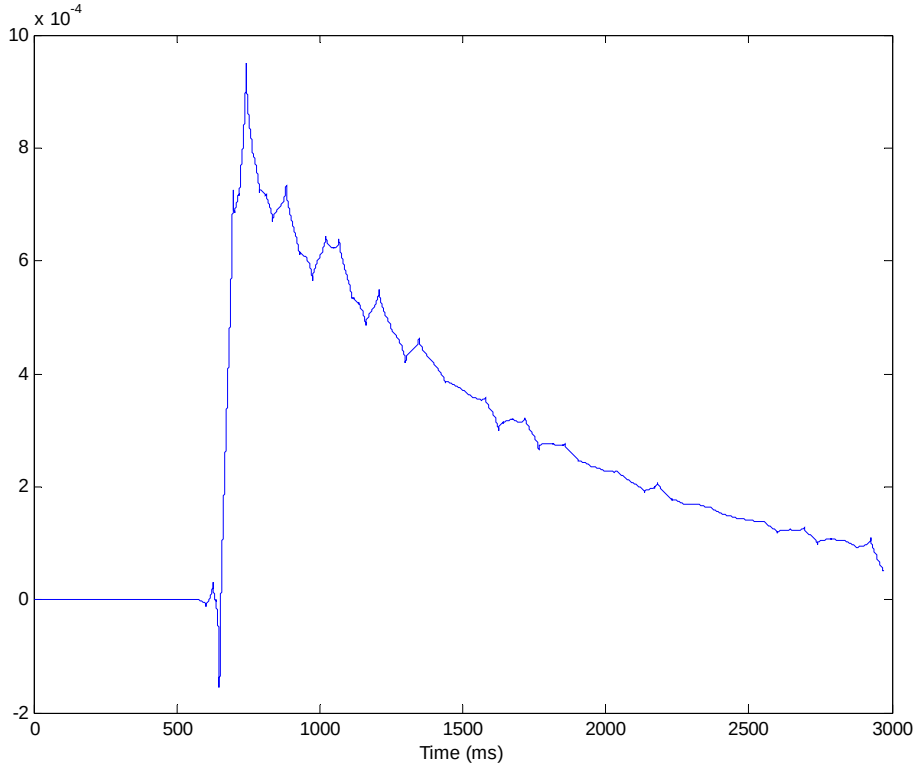


(b) Final result of multiresolution analysis for A3 piano tone. Only the zoomed-in small segment around the negative dip is shown.

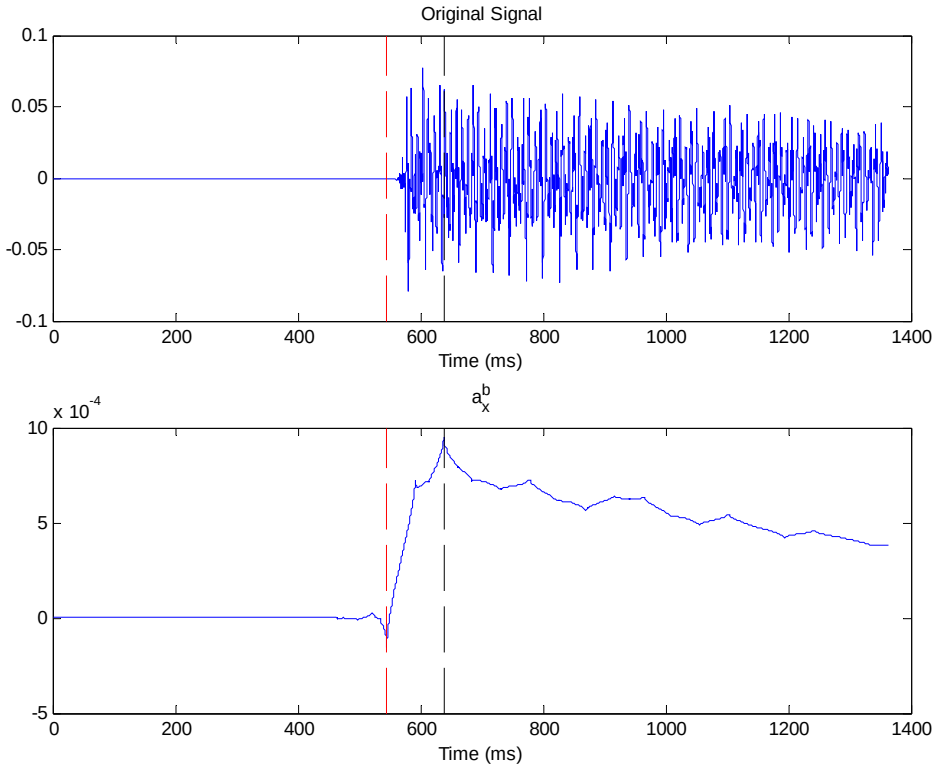


(c) Waveforms of some subbands in multiresolution analysis of A3 piano tone

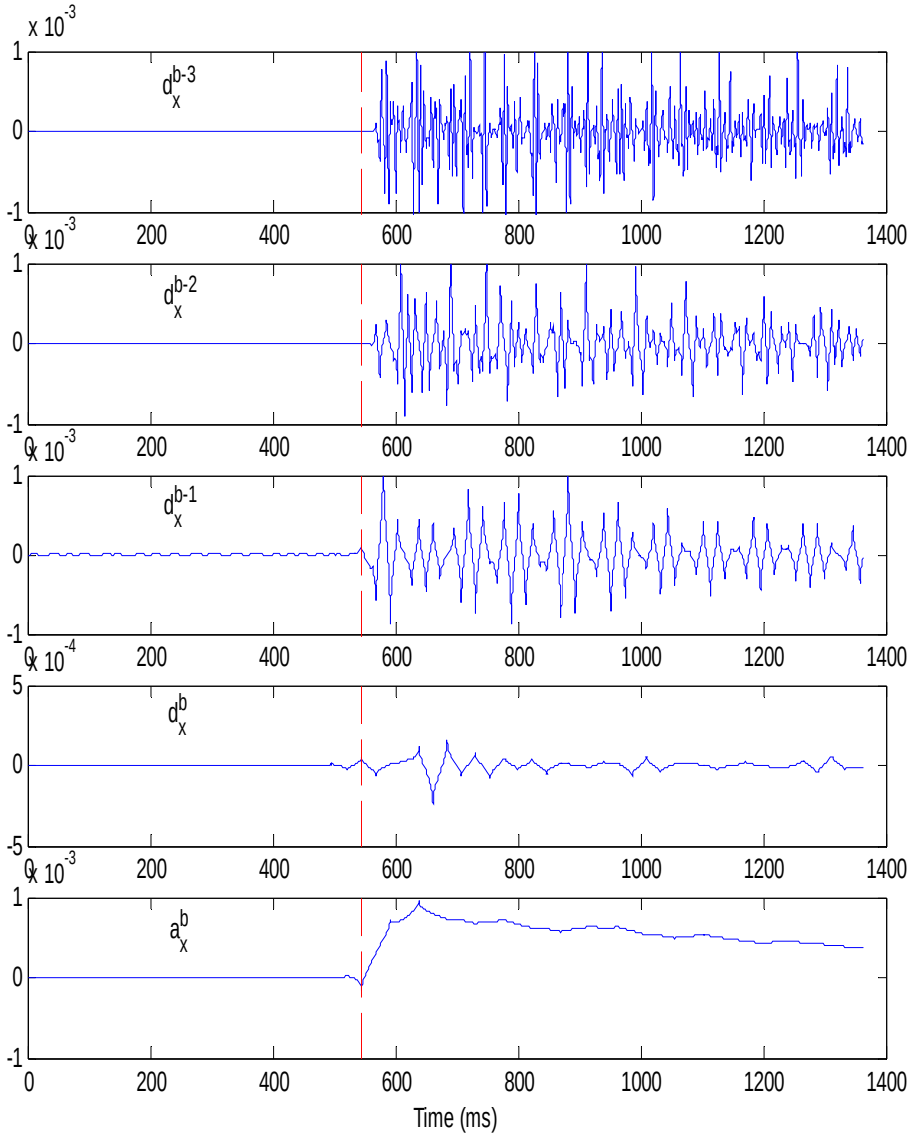
Fig 3.19 The measurement of A3 piano tone



(a) Energy envelope of D1 piano tone

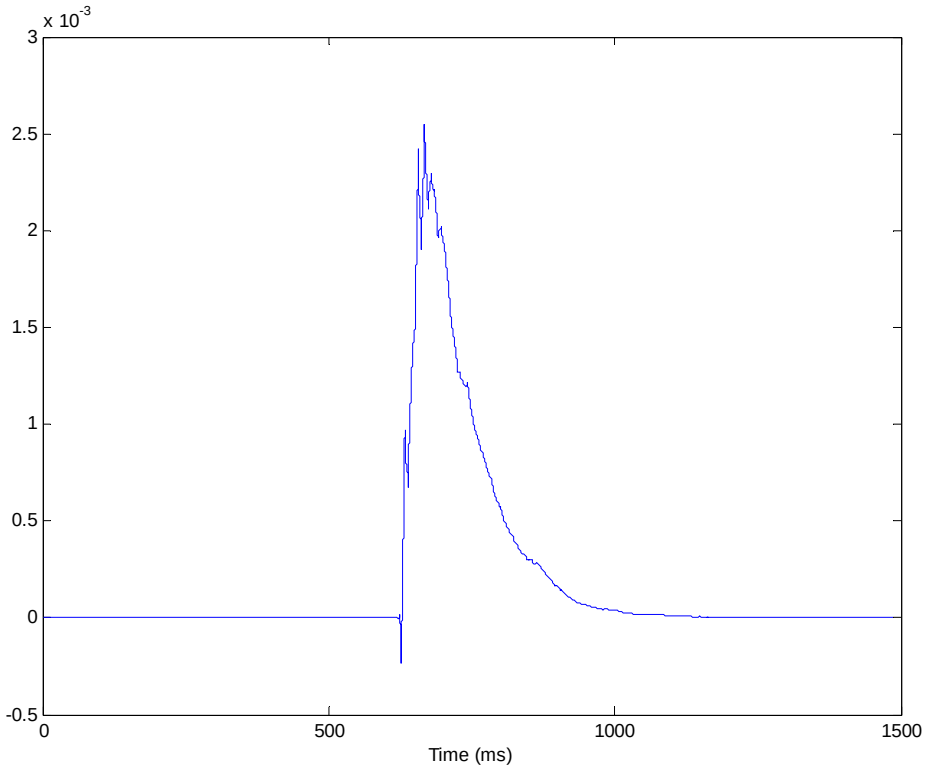


(b) Final result of multiresolution analysis for D1 piano tone. Only the zoomed-in small segment around the negative dip is shown.

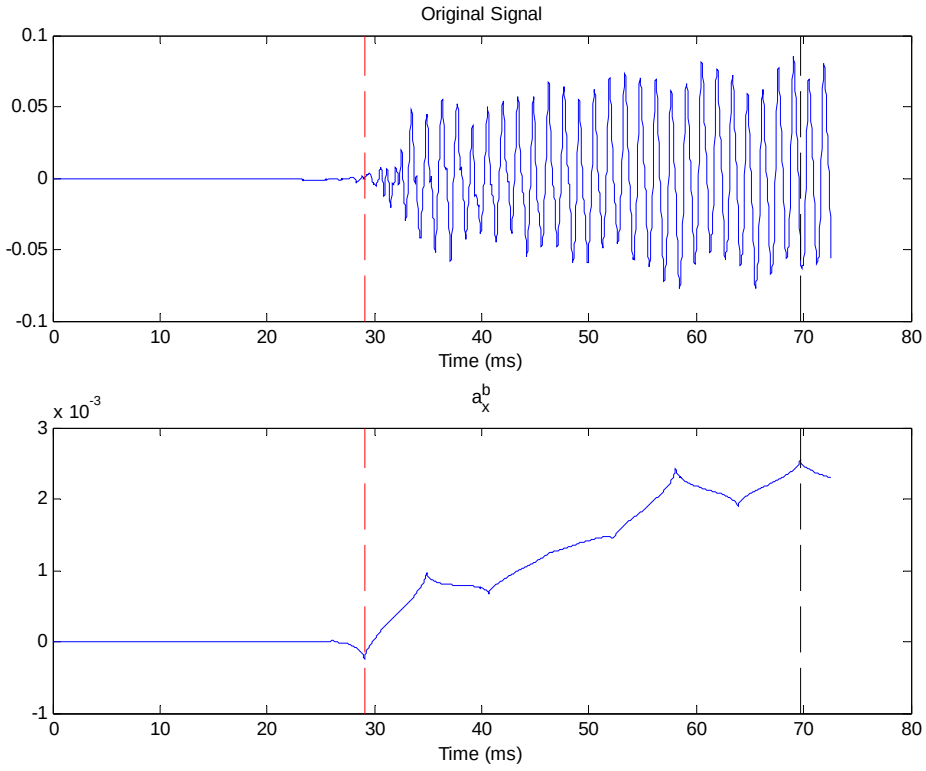


(c) Waveforms of some subbands in multiresolution analysis of D1 piano tone

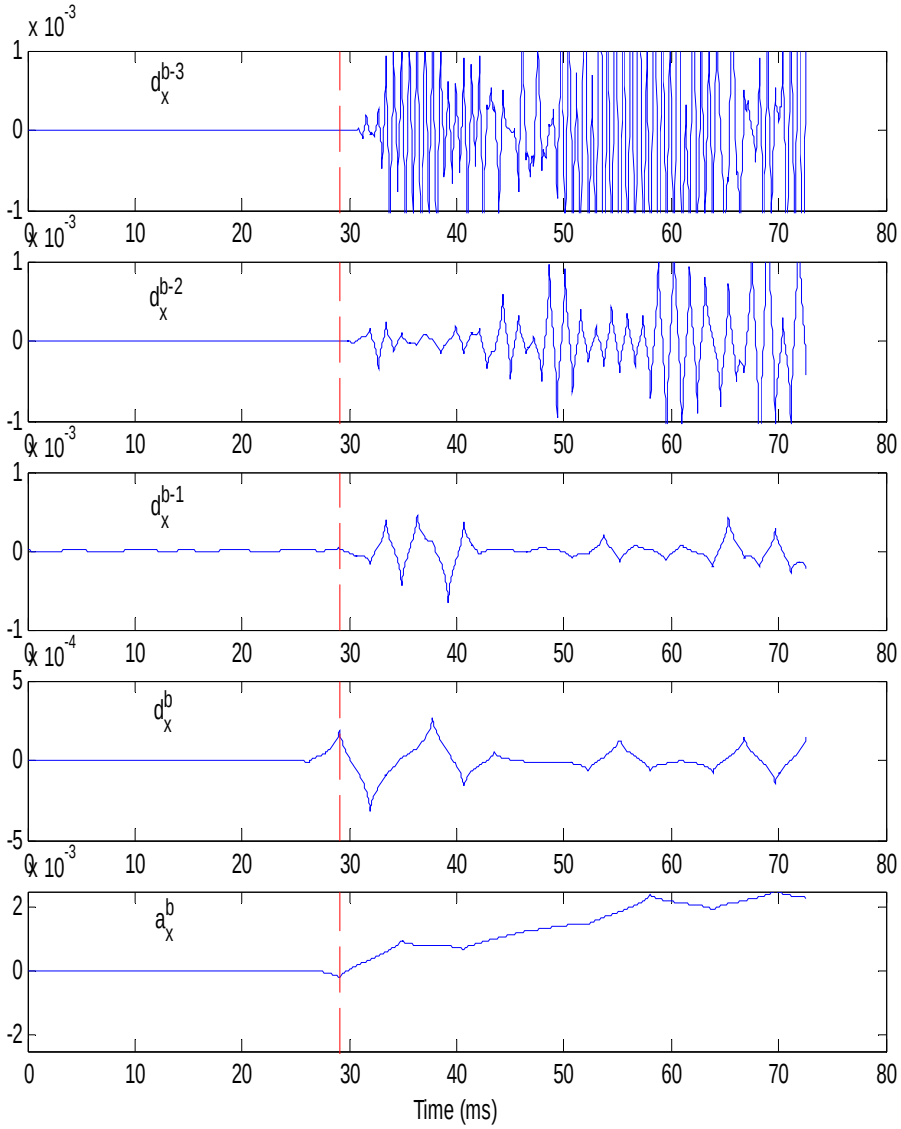
Fig 3.20 The measurement of D1 piano tone



(a) Energy envelope of F5 piano tone

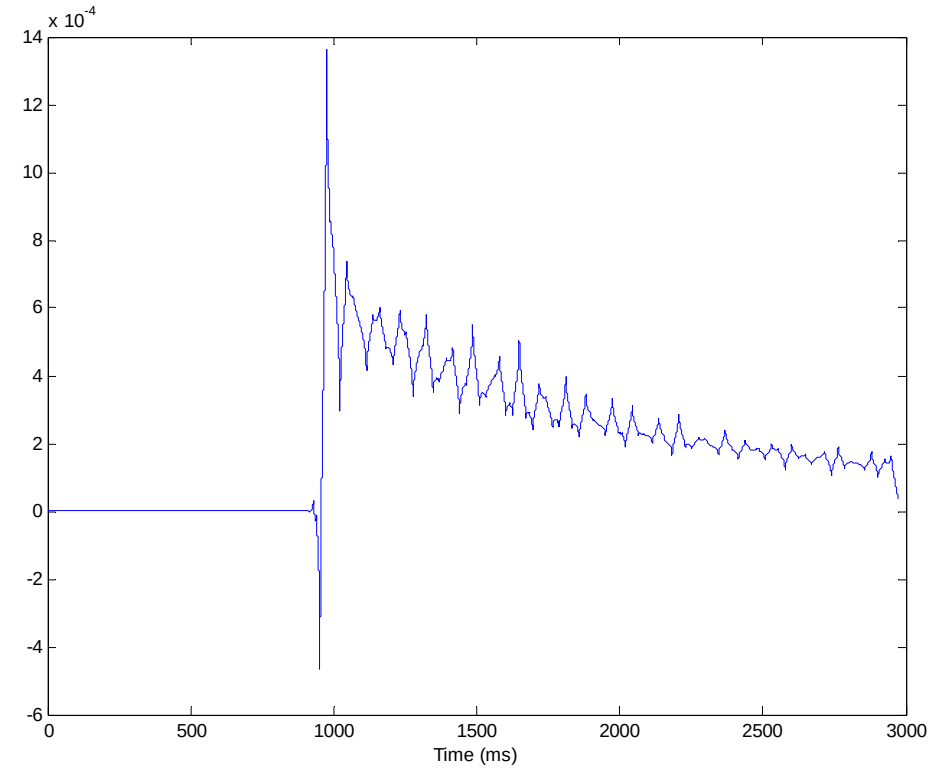


(b) Final result of multiresolution analysis for F5 piano tone. Only the zoomed-in small segment around the negative dip is shown.

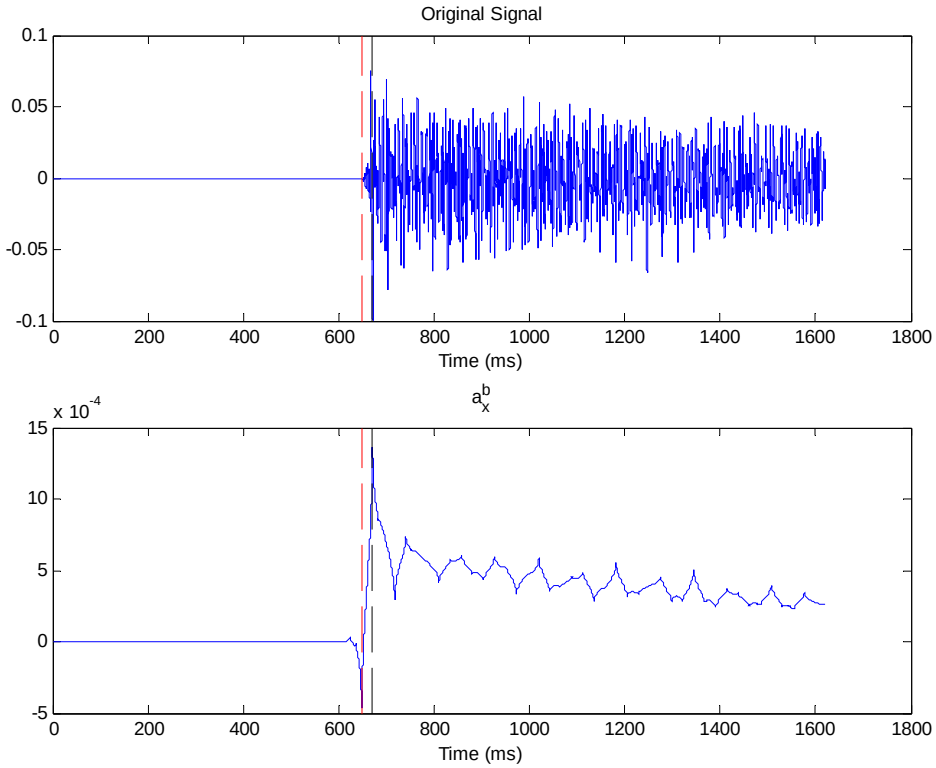


(c) Waveforms of some subbands in multiresolution analysis of F5 piano tone

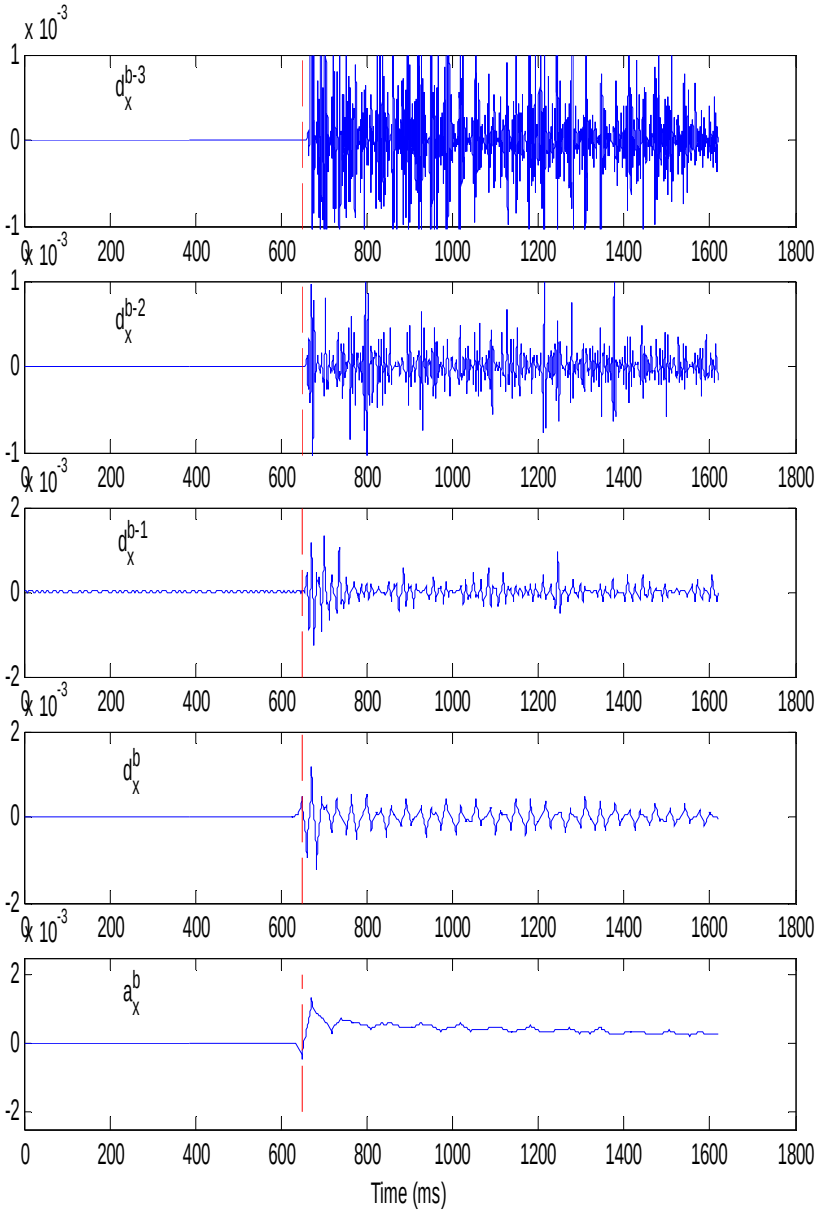
Fig 3.21 The measurement of F5 piano tone



(a) Energy envelope of B0 piano tone

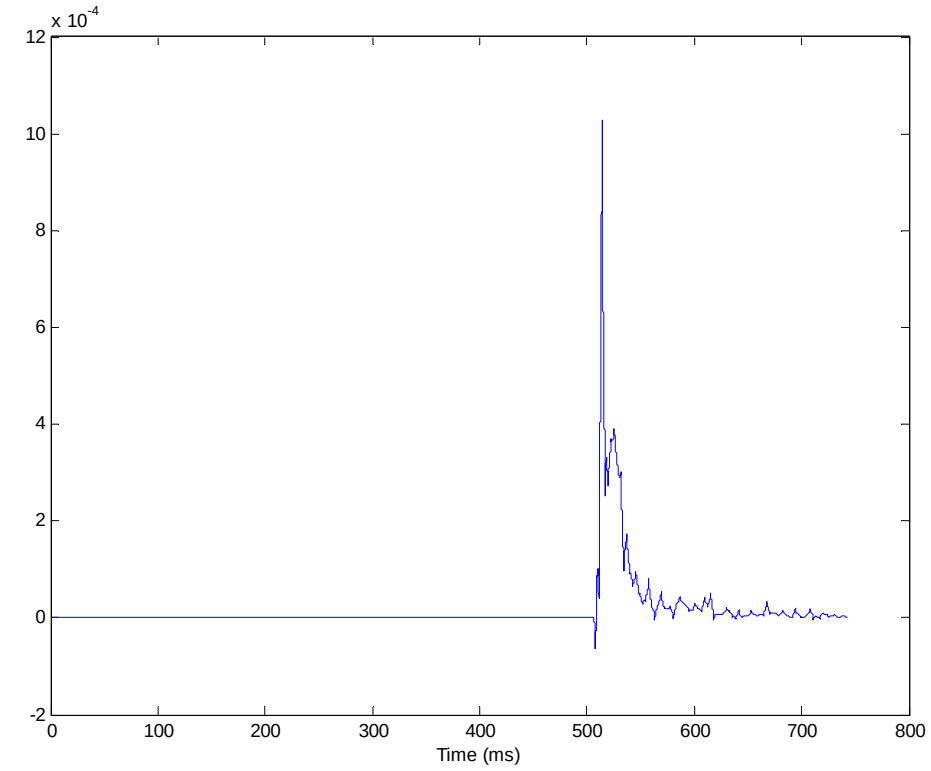


(b) Final result of the multiresolution analysis for B0 piano tone. Only the zoomed-in small segment around the negative dip is shown.

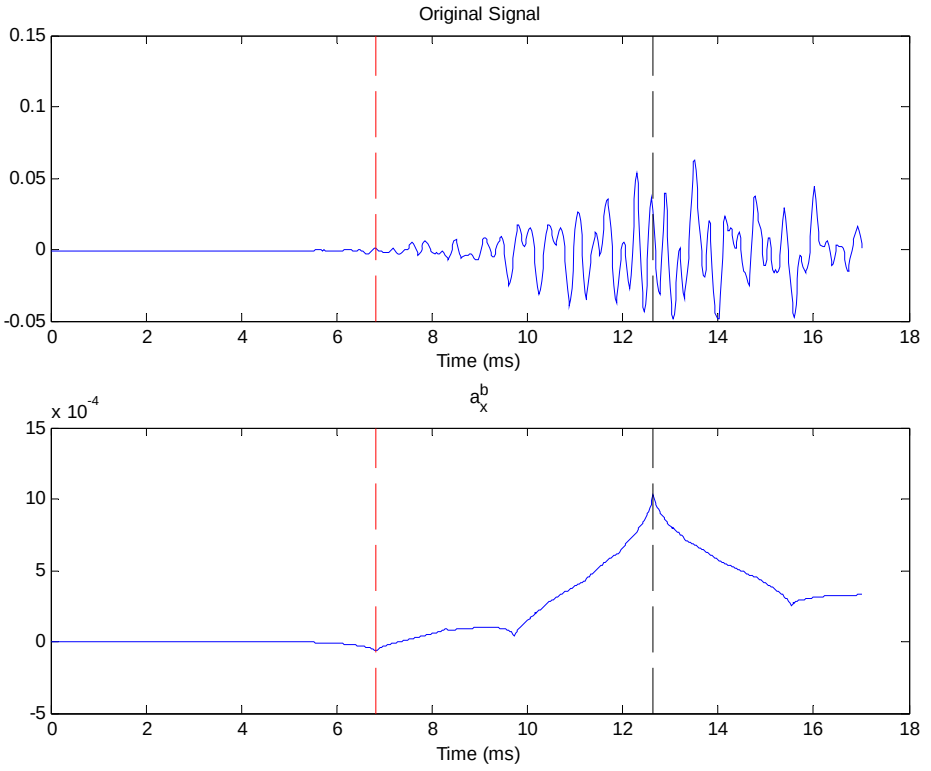


(c) Waveforms of some subbands in multiresolution analysis of B0 piano tone

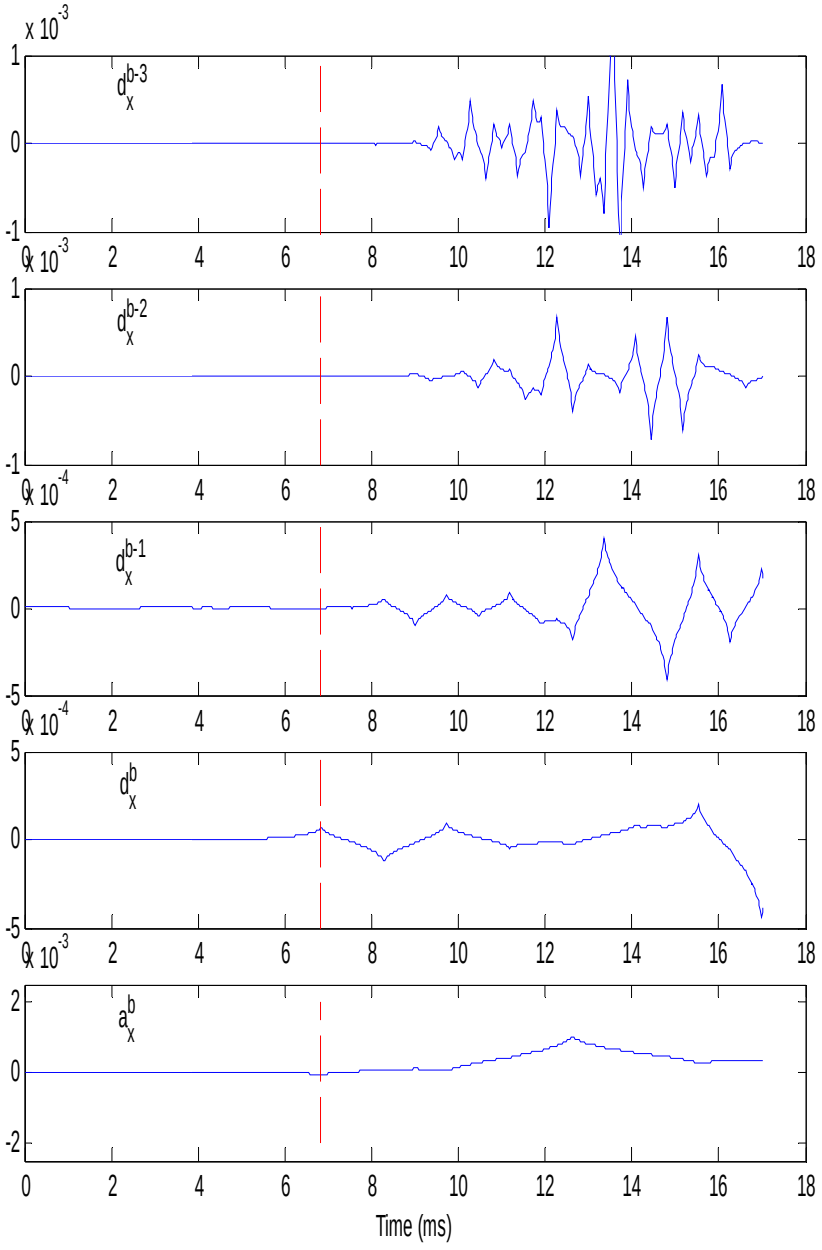
Fig 3.22 The measurement of B0 piano tone



(a) Energy envelope of G7 piano tone



(b) Final result of multiresolution analysis for G7 piano tone. Only the zoomed-in small segment around the negative dip is shown.



(c) Waveforms of some subbands in multiresolution analysis of G7 piano tone

Fig 3.23 The measurement of G7 piano tone

To validate our estimation, the results are compared with direct visual inspection on the waveform.

Piano tone	Our method	Visual inspection	Piano tone	Our method	Visual inspection
C4	23.22ms	25ms	A3	58.05ms	61ms
D1	92ms	54ms	F5	40.63ms	42ms
B0	23.22ms	24ms	G7	5.8ms	6.9ms

Table 3.1 Comparison with visual inspection

From the table, we can see that our results are a good match with values from visual inspection except for D1. As explained previously, the peak of the energy envelope we have computed may not be identical to the exact peak point but we can ensure that the 'averaged' magnitudes are approximately around the largest magnitudes. In Fig.3.20 (b) where D1's energy envelope and waveform are presented, it can be seen that the waveform peak (top diagram) seems to coincide with the first lower peak in the energy envelope (bottom diagram). However, our estimation uses the peak of energy envelope as the end of the onset transient. Although the waveform corresponding to this peak does not have the largest amplitude, the 'averaged' waveform amplitude over this neighborhood is at a maximum. From this perspective, our estimation through the energy envelope is more reasonable.

Therefore, by measuring the duration between the negative dip and the positive peak of the envelope in a_x^b, we can obtain the onset transient duration by the first definition. Fig 3.24 gives the computed duration of the onset transient for all 88 piano

tones which we have computed by this method. The solid line is the exponential fitting curve which reflects how the onset transient durations change with increased pitch. From A0 to C8, the durations gradually decrease from over 100ms for the lowest bass tones to approximately 10ms for very high treble tones. That is consistent with common sense about the durations of piano onset transients.

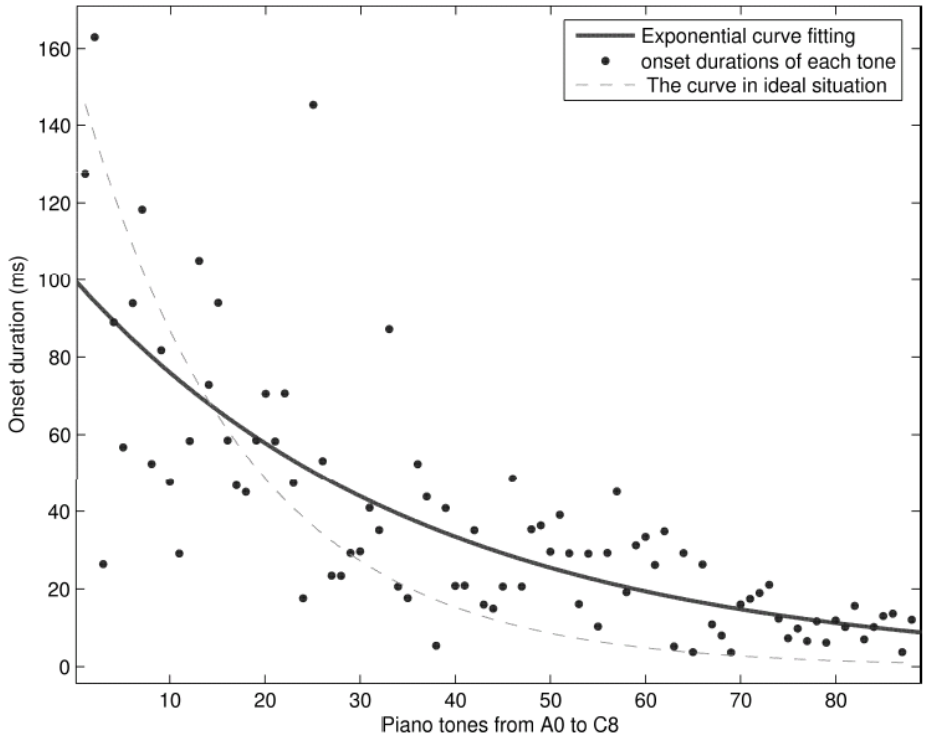


Fig 3.24 Onset durations of all piano tones (from A0 to C8) as computed by multiresolution analysis

We have not found much data about the measured onset transient duration of real pianos. This might be due to the fact that no standard definition exists for the onset transient. It is also impossible to tell which exact point is the boundary between the onset transient and the following stationary part. Furthermore, the waveform of each piano tone's onset transient is also not unique and greatly influenced by how hard we press the key. But even an approximate knowledge of each onset transient's duration

can still help us to determine the time resolution needed for onset detection. It may be observed that our result are scattered a little widely. Firstly, this may due to the limitation of our method. The wavelet transform is dyadic, which means that the time resolution of the next level is two times of that in current level and so on. With more levels, the time resolution exponentially increases and thus the time interval between neighborhood windows could be large. Secondly, the transient evolves in an unpredictable way. Many incidental factors can influence the peaks of the waveform and thus make the measurement under the first definition widely scattered.

Chapter 4 Time-Frequency Analysis of Piano tones

The time and frequency representations together form the whole of the information about a signal. In most cases, the time representation may appear in a way (e.g., in terms of a waveform). By contrast, the frequency information is usually hidden behind the waveform and has to be presented more indirectly. Therefore, disclosing a signal's detailed frequency content is a main goal of many signal processing techniques, including Fourier analysis and Wavelet analysis. However, separately individual time and frequency representations may miss how the energy in

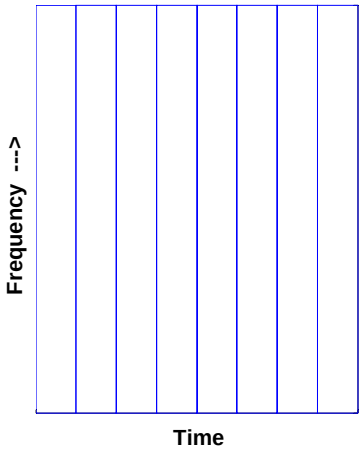
a signal is distributed with respect to both the time and frequency variables. Short-time Fourier transform (STFT) analysis uses spectrograms to indicate a time-frequency representation. Similar to spectrograms, wavelet analysis integrates both time and frequency by a concept called the time-frequency plane. The main difference between the STFT and Wavelet time-frequency representation lies in the fixed or flexible window size which has been explained in Chapter 1. Therefore, in this Chapter, we will try to examine more closely some features of onset transients in terms of the wavelet time-frequency plane.

4.1 Wavelets Packet Transform and Time-Frequency Plane

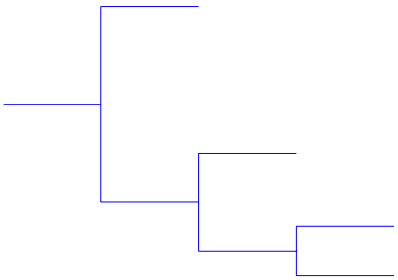
As summarized in Chapter 3, the DWT passes a sampled signal through a low-pass filter *h* and a high-pass filter *g* respectively, which is then followed by a down-sampling operation. After this decomposition, two ”branches” or nodes result: the high-frequency output (the *detail* subband) and the low-frequency output (the *approximation* subband). Due to the down-sampling operation, the number of coefficients in each offspring node is reduced to one half of that in the parent node, so that the total number of samples does not change. That means we are obtaining finer frequency information (splitting the frequency range into half for each node) at the expense of the time resolution (with only half of the original number of samples in each node).

Initially the sampling process determines the time resolution Δt (the time interval between successive samples) and the frequency resolution Δf (the frequency range which is half of the sampling frequency). Following decomposition by a pair of filters, Δf is halved due to the filtering process while Δt is doubled due to down-sampling as the number of samples is halved for the same overall time period, thus keeping $\Delta t \Delta f$ constant in accordance with the Heisenberg Uncertainty Principle. In this manner, the wavelet transform converts the original signal from a pure time domain to a mixed time-frequency domain. Each iterative process that passes an *approximation* subband through a pair of filters further doubles Δt, and halves Δf. Decomposition and reconstruction filters are always selected to complement each other, so that aliasing effects do not occur in down-sampling. The complementary set of decomposition and reconstruction filters form a quadrature mirror filter, and the choice of the particular quadrature mirror filter determines the type of wavelet basis (e.g., a Haar, Debauchies or Battle-Lemarie wavelet), used for the wavelet analysis.

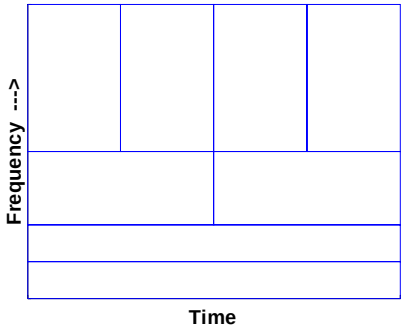
Taking an eight-point signal as an example, we will use the following diagrams (Fig 4.1) to show the relationship between tree structures in the wavelet transform and time-frequency planes.



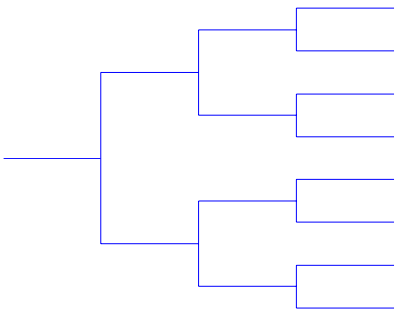
(a) An eight point signal $x[1, \ldots, 8]$'s T-F plane



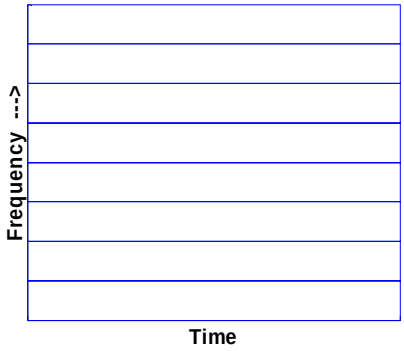
(b) 3 stage wavelet transform



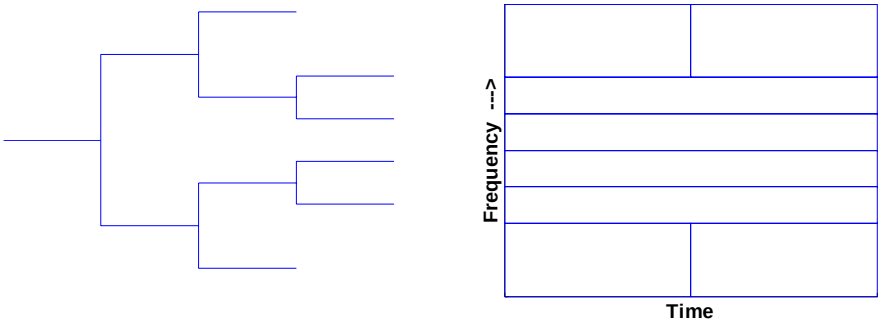
T-F plane of the left tree



(c) Full wavelet packets tree



T-F plane of the left tree



(d) Another wavelet packets tree T-F plane of the left tree

Fig 4.1 Some T-F planes for an 8 points signal

In Fig 4.1, we start with a basic discrete series of 8 samples $x[1,...,8]$, which corresponds to a pure time domain series with 8 vertical divisions, each corresponding to one sample and spanning the whole frequency range. After decomposition through one filter pair (and down-sampling), each node D_1 and A_1 takes up half the frequency range but has coarser time resolution with only four divisions, each of which is twice as wide as the original time divisions of the original signal. If the *approximation* node A_1 is split up again, the corresponding lower half of the time-frequency diagram is split into two nodes D_2 and A_2 corresponding to the halved frequency ranges, each with only two time divisions. A further split of the *approximation* subband A_2 results in two nodes D_3 and A_3 with frequency ranges halved again, each spanning the whole time interval (as the corresponding nodes now consist of only one sample).

In this DWT example (for Fig 4.1 (b)), the original signal can be reconstructed by 'adding' A_1 and D_1, *or* A_2 and D_2 and D_1 *or* A_3 and D_3 and D_2 and D_1.

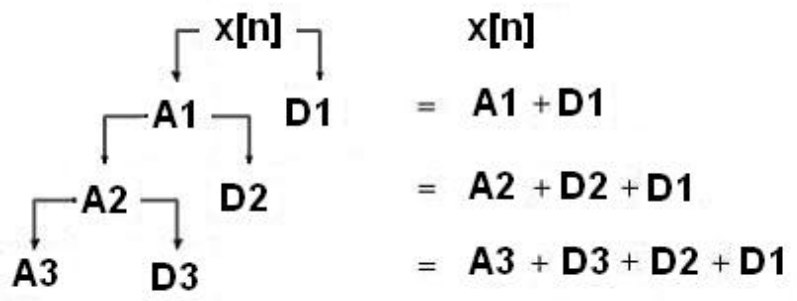


Fig 4.2 The hierarchy diagram of DWT for 8 points, corresponding to the Fig 4.1(b) (Note: the '+' here does not mean the ordinary plus operation in mathematics. It only means that A3, D3, D2 and D1 together may make up one possible result among the DWT decomposition.)

In comparison with the DWT, the wavelets packet transform (WPT) also iteratively subjects the detail coefficients to the basic decomposition unit, which could result in a full tree structure as indicated in Fig 4.1(c) (or Fig 4.3) for the 3 level case.

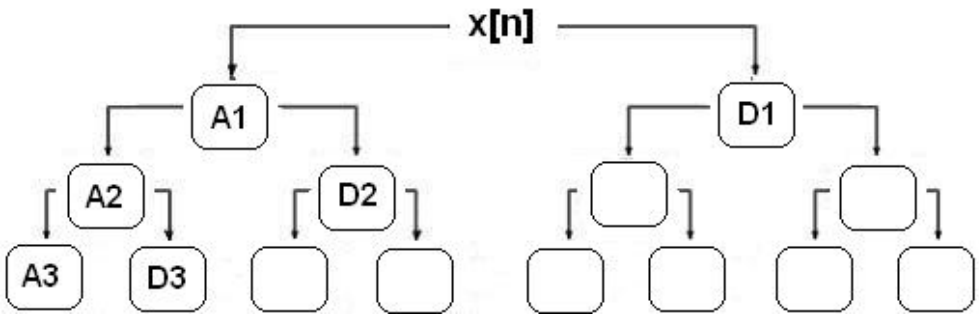


Fig 4.3 The full tree hierarchy diagram of WPT for 8 points, corresponding to the Fig 4.1 (c)

Previously, we have discussed the wavelet theory. In Equation (2-12) $V_j = W_{j-1} \oplus V_{j-1}$, we know that the pair of conjugate mirror filters can split space V_j into two subspaces. Coifman, Meyer and Wickerhauser have generalized this result to any space U_j which has an orthonormal basis $\{\theta_j(2^j t - k)\}$. Let h and g be a pair of conjugate mirror filters and define $\theta_{j-1}^0(t) = \sum_{k \in Z} h(k)\theta_j(2^j t - k)$ and

$\theta^1_{j-1}(t) = \sum_{k \in Z} g(k)\theta_j(2^j t - k)$ whose corresponding spaces are U^0_{j-1} and U^1_{j-1} respectively. They have shown that $U_j = U^0_{j-1} \oplus U^1_{j-1}$. If we set $U_j = W_j$, we have $W_j = W^0_{j-1} \oplus W^1_{j-1}$. This means that the same conjugate mirror filters *h* and *g* which used to split $V_j = W_{j-1} \oplus V_{j-1}$ can also decompose $W_j = W^0_{j-1} \oplus W^1_{j-1}$.

Besides such a full tree structure (Fig 4.1 (c)), there are many other possible 'underdeveloped' tree structures (e.g. Fig 4.1 (d)), each of which is a valid representation of the original signal. In other words, the signal can be represented by the full tree or by partial trees, with the terminal nodes of each tree adding up to form the original signal. For more levels or stages, the number of different possible WPT trees can become very large. For example, a 9 level WPT generates about 10^{77} trees [71], each of which is a possible representation of the signal. For any particular signal, there will be one tree structure which 'best' represents the signal. Using the Shannon entropy function ($-\sum_n x[n]^2 \log(x[n]^2)$ for vector *x*[*n*] of finite length), Coifman [72] has designed an algorithm to efficiently select the 'best' tree to represent any particular signal. Coifman's algorithm chooses the best tree for which the Shannon entropy function has the lowest value, which means that the signal's energy is most efficiently represented. Firstly, calculate the cost value of every node according to the Shannon entropy function $-\sum_n x[n]^2 \log(x[n]^2)$. Let v_1 represent the cost value of the parent node and v_2 represent the sum of two offspring nodes' cost value. If $v_1 \le v_2$, the two offspring nodes are deleted. If $v_1 > v_2$, the two offspring nodes are marked as the terminal nodes and the cost value of the parent node (originally is v_1) is replaced by v_2. The above process continues until the top node is involved and

therefore all terminal nodes are determined.

To visualize the results, each tree can uniquely correspond to a particular division of the time-frequency (T-F) plane [4, 5] into $\Delta t \Delta f$ "boxes". The T-F plane shows how the energy in a signal is distributed with respect to the time and frequency variables. Each $\Delta t \Delta f$ box corresponds to one of the terminal nodes at the right of the corresponding tree structure as shown in Fig 4.1

Furthermore in practical computations of the wavelet packets transform natural/frequency ordering and boundary effect should be paid some attentions. In Fig 4.3 where a full tree is shown, the frequencies of the nodes (or subbands) should be increased from left to right. We use the series 0 (lowest frequency), 1 , 2, 3, 4, 5, 6, 7 (highest frequency) to represent the relative ordering. In theory, the frequency order in Fig 4.3 from left to right should be 0, 1, 2, 3, 4, 5, 6, 7. But due to the aliasing of subbands, the frequency distribution after the wavelet packet transform is 0, 1, 3, 2, 6, 7, 5, 4. From left to right, the nodes' frequencies do not monotonically rise. We usually call the frequency distribution of the immediate output of the wavelet packet transform the *natural ordering* or *filter bank ordering*. In order to convert the natural ordering (i.e., 0, 1, 3, 2, 6, 7, 5, 4) to expected *frequency ordering* (i.e. 0, 1, 2, 3, 4, 5, 6, 7), the Grey code permutation is used. Given an integer x, write it in binary notation as $n_{N+1}n_N \ldots n_2 n_1$. For example, x=6, we have n_4=0 n_3=1, n_2=1 and n_1=0. The Grey code permutation is $GC(n)_i = n_{i+1} + n_i \mod 2$. For example, after the wavelet transform, from left to right, the 6th node is not the second largest frequency subband (i.e., the order index =6), but is the third largest frequency subband (i.e., the order index =5)

after the Grey code permutation.

Another difficulty encountered when we implement the wavelet packets transform is the boundary effect. We know that the wavelet transform passes the signal through a pair of filters followed by the down-sampling. If the signal has infinite length, there should be no problem. But in practice, the signal is finite and boundary problem will arise, i.e., the low-pass or high-pass filter will spill over samples which are not available. For example, we have n=8 samples, s0, s1….s7. When we compute some coefficients we need s8 or s9. But we don't actually have s8 and s9. Therefore techniques like zero padding, period extension, smooth padding etc are used. In zero padding, we extend the signal by adding zeros at the two ends. The periodic extension technique makes the data periodic by letting s(k+n)=s(k). In smooth padding, we extend the signal by linearly extrapolating the data near the two ends.

4.2 The Time-Frequency Planes of Onset Transients by WPT bases

From Section 4.1 we know that the time-frequency plane is partitioned into rectangles which represent the decomposition of the signal according to the wavelet basis. The position of the rectangle indicates the time span and frequency range which it covers, and its dimensions indicate the relative uncertainties of the two quantities which are defined by Heisenberg's inequality principle. Hence these rectangles are known as Heisenberg boxes. The values of coefficients corresponding to each $\Delta t \Delta f$

box are assigned to a grey scale that fills the box. A WPT analysis of a signal can be visually represented by a time-frequency plane, which consists of $\Delta t \Delta f$ boxes, whose particular layout optimally represents the signal and the darkness of each box represents its energy content.

This T-F plane is similar to the spectrum of the short-time Fourier transform (STFT). However, instead of constant width windows in the STFT, the WPT (including Coifman's algorithm) can achieve much greater flexibility, compared to the DWT, with variable width windows and time-frequency blocks which are best suited to the nature of the signal to reflect its transient or steady-state condition, so that the window size is varied to determine more accurately either time or frequency as is appropriate for any particular part of the signal.

We modified some routines in the Wavelab toolbox [73] for our own purpose. The modified routines performed the wavelet packets transform in this section. The local cosine bases transform and matching pursuit from the wavelet packets transform were then implemented in the following sections.

The piano tones examined in our work were also obtained from the McGill University Master Samples (MUMS) compact discs. Fig 4.4 shows on the top the WPT time-frequency plane for the onset transient portion, and on the bottom for the stationary portion of the C4 tone. In this WPT process we have applied Daubechies wavelets and the corresponding quadrature mirror filters, using the Coifman entropy algorithm to select the optimum tree structure. The time-frequency blocks starting from around 14ms are darker and more regular than those before, indicating that the

quasi-periodic behavior of the tone is beginning to be established. A few square time-frequency blocks are subsequently observed sandwiched by two thin lines. The two lines are the 1st and 3rd partials respectively, which are verified by comparing their frequency positions with the time-frequency plane of the stationary part of the same tone.

An interpretation of the behavior of the square blocks is that they are adapting themselves to the rapid changes of frequency, and because of these rapid changes, the time resolution is enhanced at the expense of the frequency resolution. On the other hand, the T-F plane for the stationary part of the tone has a clearly discrete frequency level structure, showing that the fundamental frequency and its partials have settled to a stable set of values. However, for the onset transient, such a clear discrete structure is absent, and certain frequency levels are missing and may merge into wider frequency bands. The square boxes may thus be thought of as frequency bands.

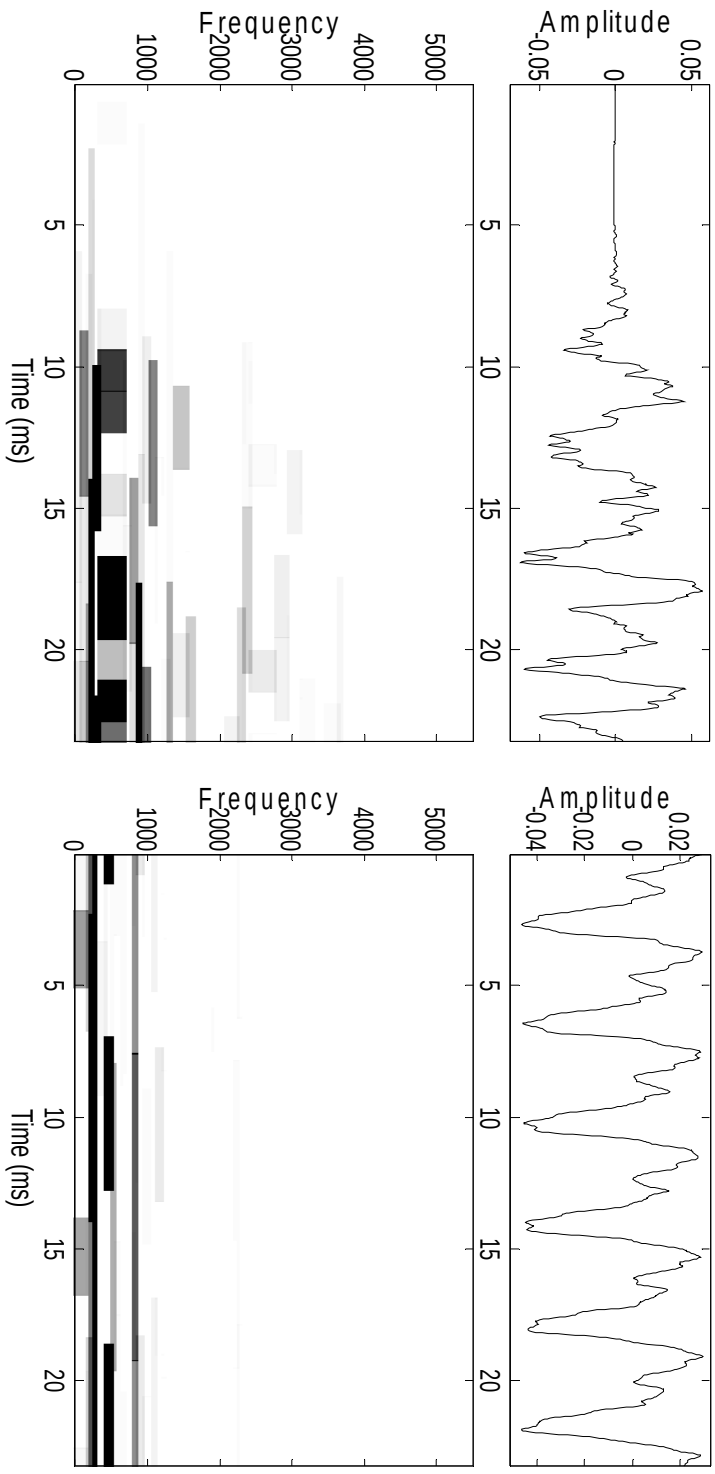


Fig 4.4 The time-frequency plane for tone C4 by the wavelet packets transform: onset transients (top) and stationary part (bottom)

This frequency level and frequency band structure in the onset transient appears to hold true for all the piano tones we investigated (see Figs 4.5 to 4.9). The existence of frequency bands is therefore the main difference between the onset transient and

the stationary part of piano tones.

In Fig 4.5, tone D7 is a treble note (whose key is located on the right side of the piano keyboard and whose fundamental frequency is thus relatively high). From 9.5 ms in the tone, the harmonic structure begins to come into being. Through comparison of the time-frequency plane between D7's onset transient and its stationary part, we can see that there is nearly no difference in the number of partials shown in both diagrams. But in Fig 4.6 where the E2 piano tone is presented, the number of partials in the onset transient is significantly larger than in its stationary part. These results seem to suggest that higher partials would decay away more quickly than lower partials.

In the A3 piano tone's scenario (Fig 4.7), there is a messy layout of T-F blocks around the frequency band from 500 Hz to 1000 Hz in the onset transient. Thus, in this band, we would expect irregular and rapid changes in its frequency evolution.

In Fig 4.8, which shows the F5 tone's time-frequency plane, the second partial is strong during the onset transient. At the same time, the third partial has not been fully established. However, during the stationary part, the situation reverses. The third partial becomes more important and the second partial seems to have lost much of its energy.

For the B0 piano tone, a low bass tone, the frequency gap between two neighboring partials is sometimes so narrow that the frequency bands could easily obscure them. All these results point to the fact that the time-frequency characteristics of the onset transient are significantly different from that of the stationary part, and

that such differences could be diverse in nature.

But we also should note that we cannot use the beginning of the time-frequency blocks as the starting point of the onset transient. This is due to some practical wavelet computation issues. For example, only a few time-frequency blocks are visible and others are too white to be seen in the grey scale graph. When solving boundary problems in the wavelet computation, many wavelet toolbox including 'Wavelab' use periodic extension, which let s(k+n)=s(k). Thus, some time-frequency blocks calculated at the end boundary will use the data at the beginning of the signal and then be wrapped back at the beginning to show the periodic effect.

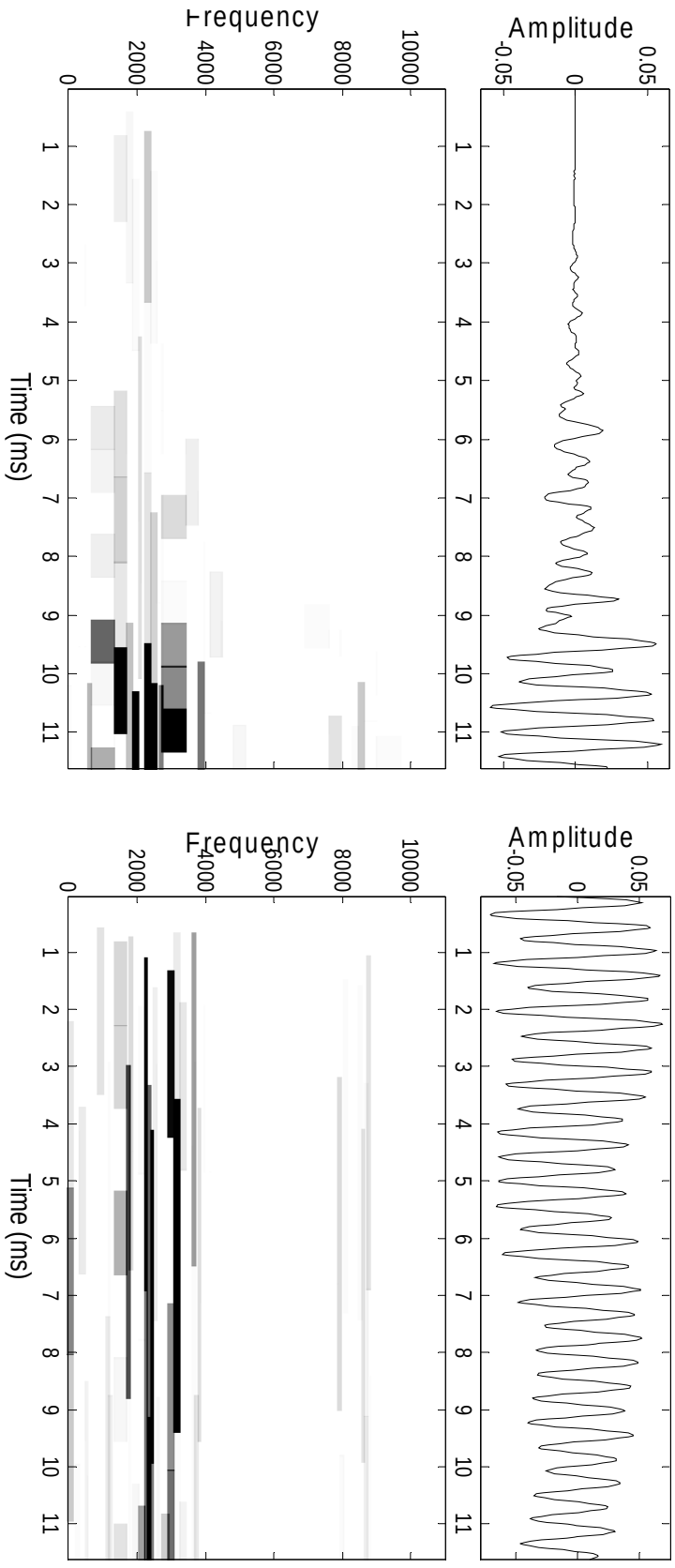


Fig 4.5 Onset transient (top) and stationary part (bottom) of D7 piano tone

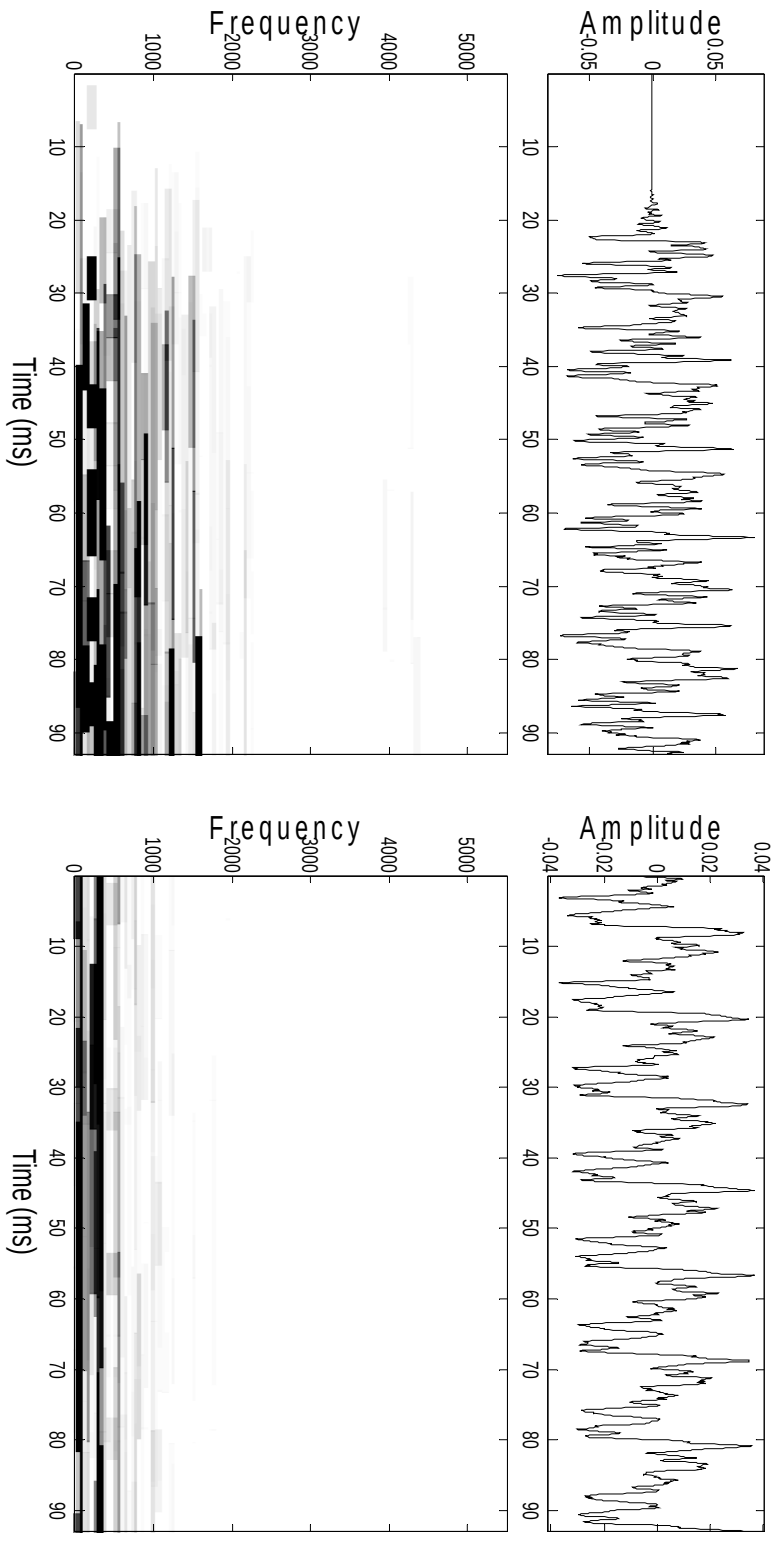


Fig 4.6 Onset transient (top) and stationary part (bottom) of E2 piano tone

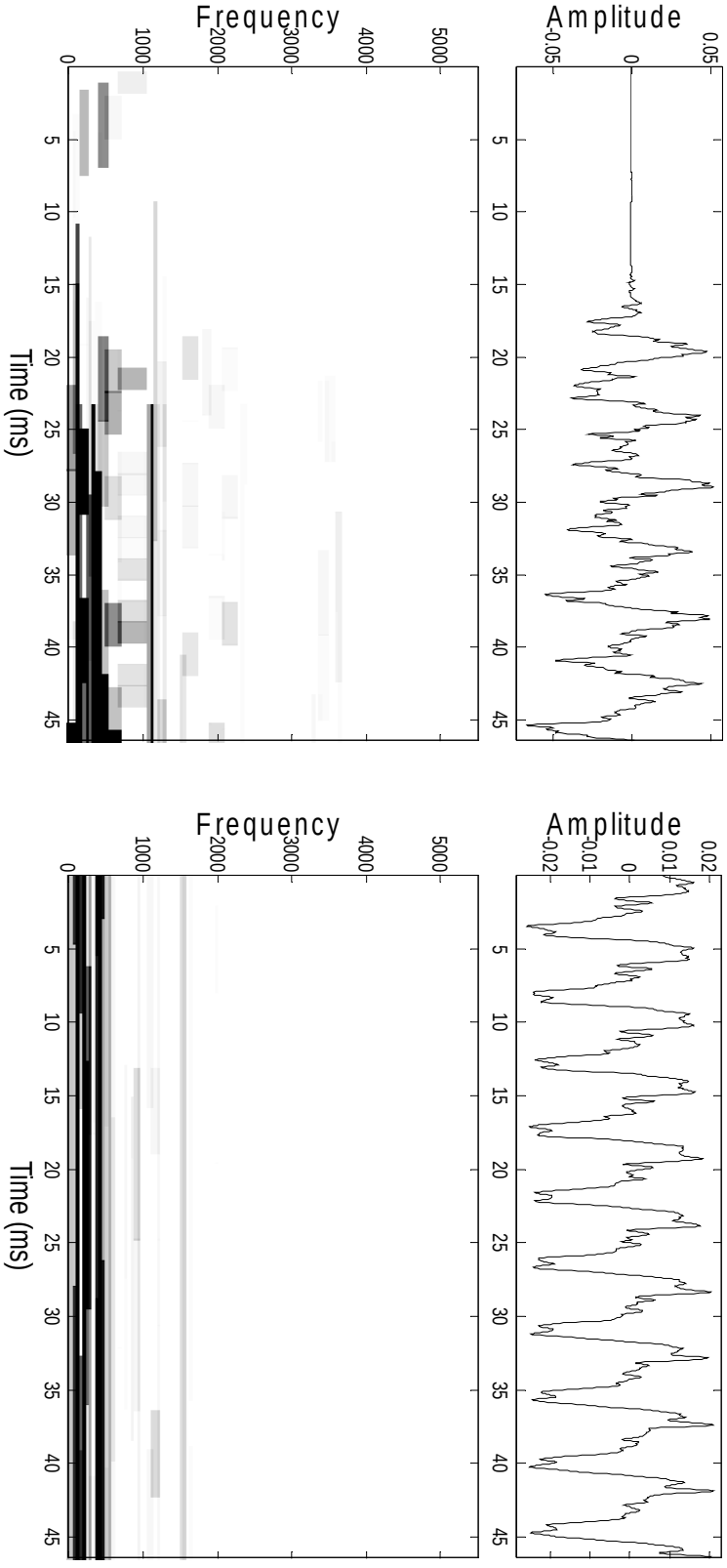


Fig 4.7 Onset transient (top) and stationary part (bottom) of A3 piano tone

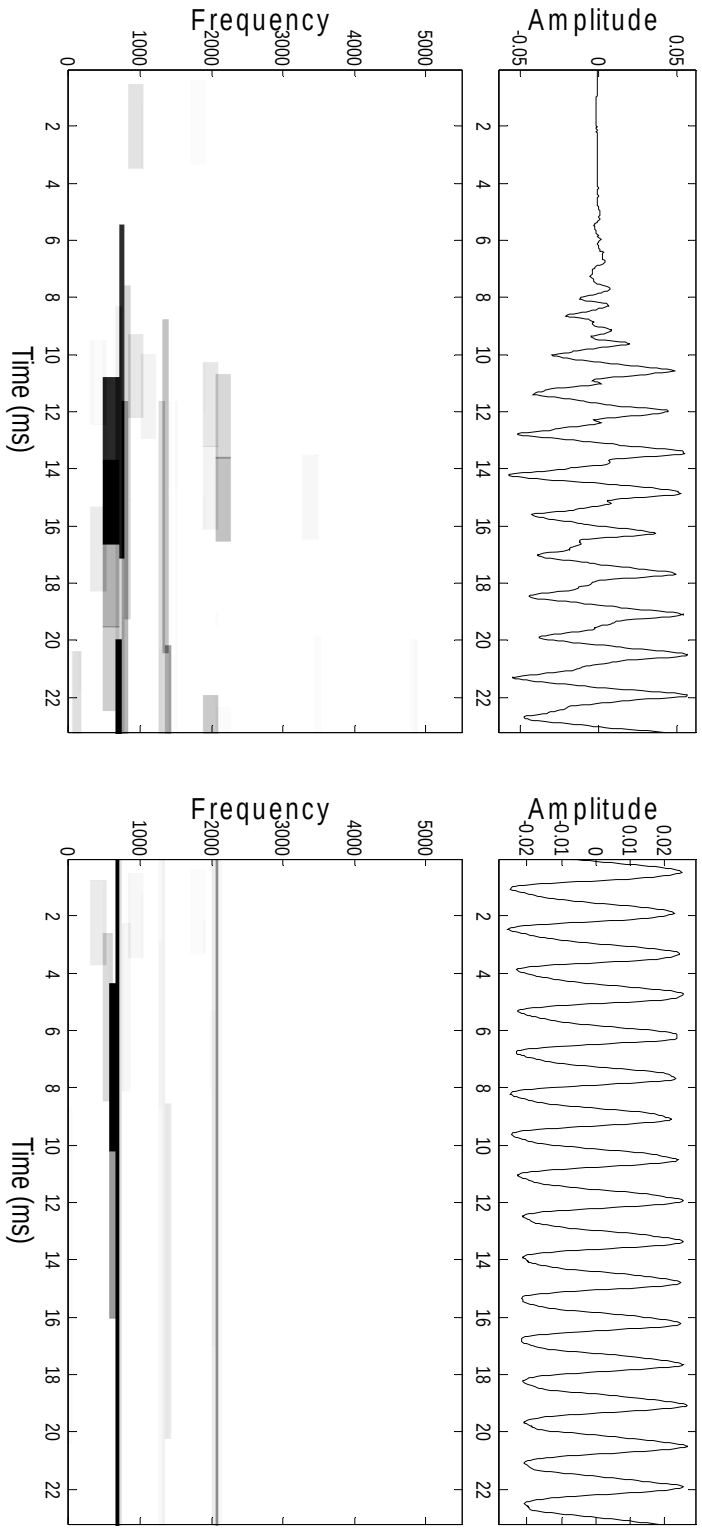


Fig 4.8 Onset transient (top) and stationary part (bottom) of F5 piano tone

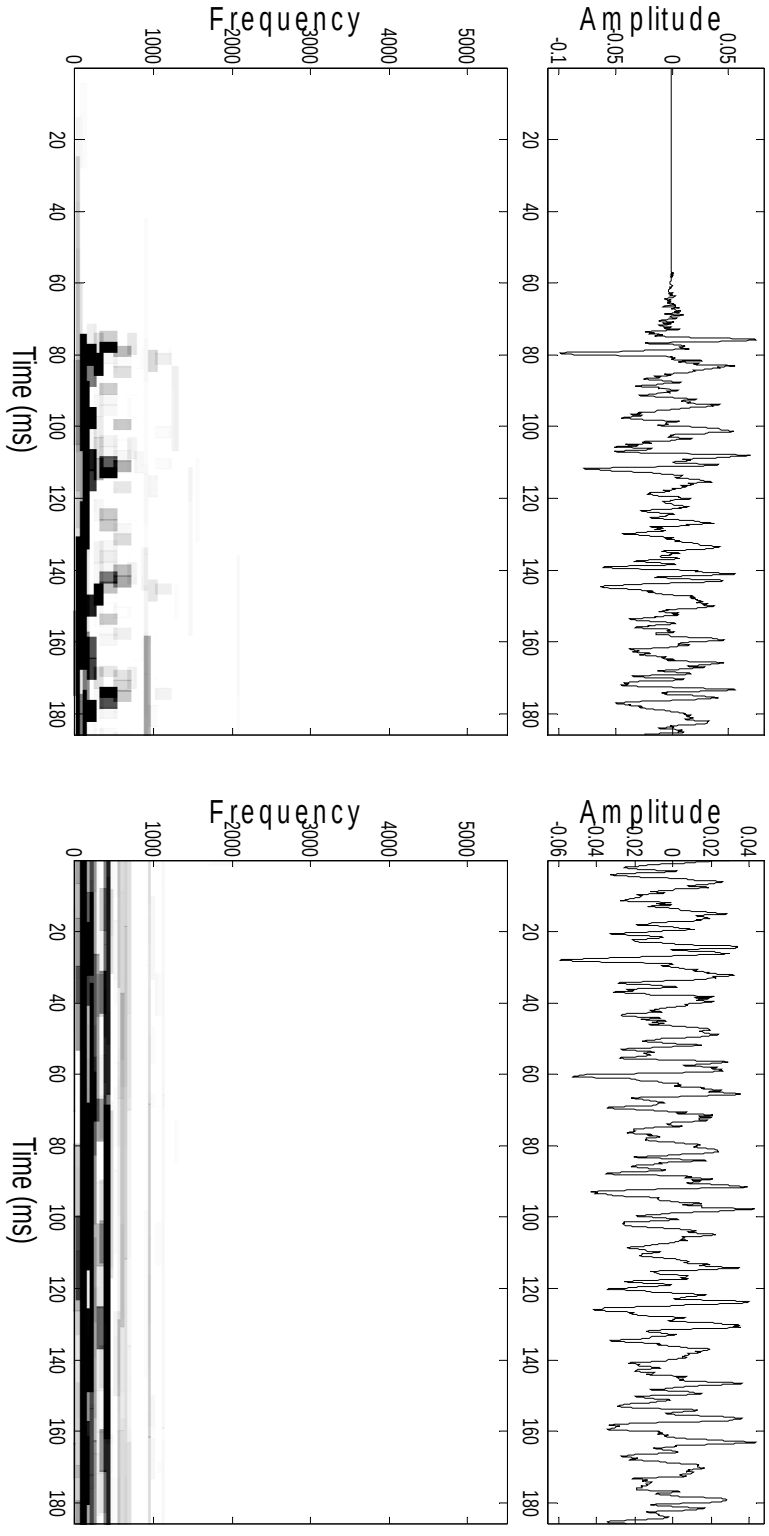
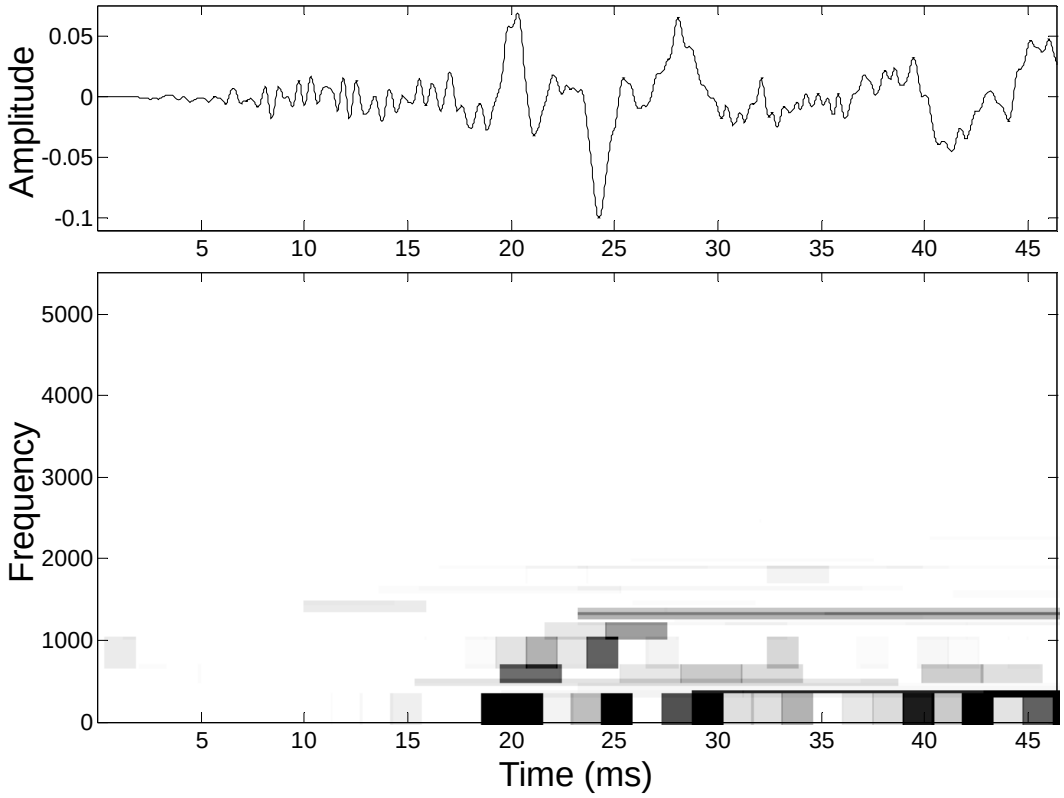


Fig 4.9 Onset transient (top) and stationary part (bottom) of B0 piano tone

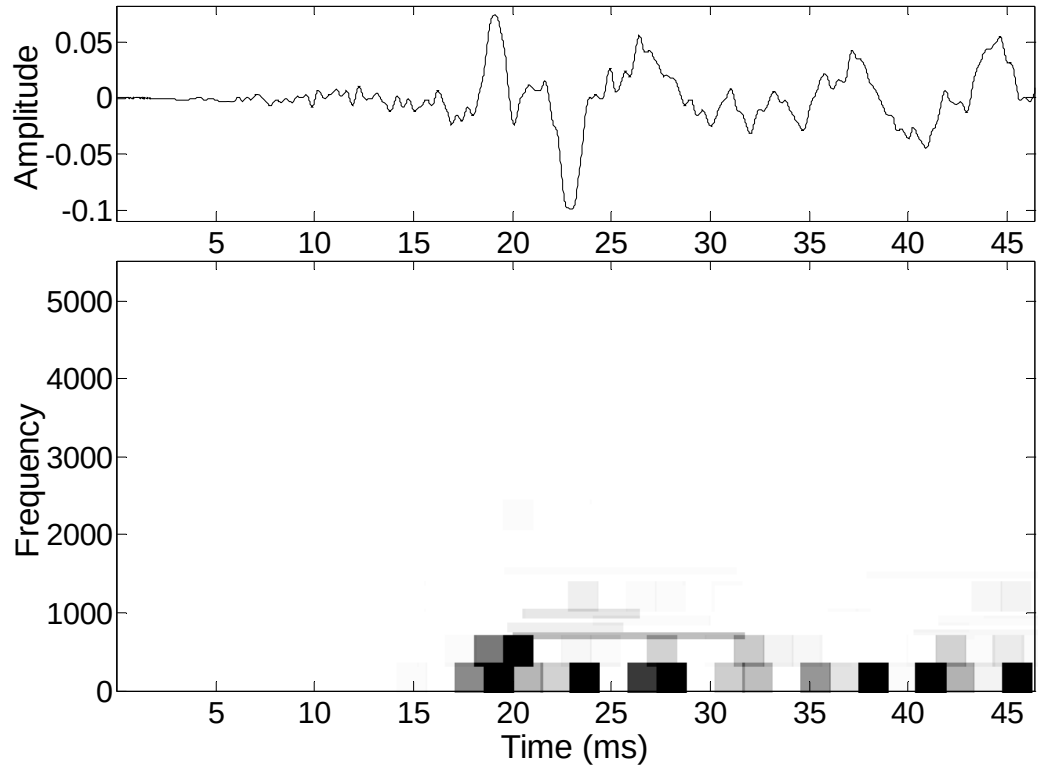
In Chapter 3 (section 3.1), we have presented two 'reasonable' definitions for onset transients (since there is no exact definition). They are respectively the

definition by *Peak Amplitude* and the definition by *Evolving Process.* Both definitions are derived from the direct observation and measurement of the piano tone waveform. The difference between the two definitions is that the first one looks at the overall shape of the waveform while the latter looks more closely at the details of the waveform. Though these are different definitions, the estimated durations of the onset transient are not very divergent. In Chapter 3 or particularly Fig3.22(b), the 'macro' definition, i.e., the peak amplitude is used. While in Chapter 4, we want to explore the inner time-frequency relationship and thus the 'micro' definition is used. That is why in Fig 3.22(b) and Fig4.9 the ending of the onset transient is different. Since the ending of the onset transient cannot be exactly marked, in Fig 4.9, we use a little longer duration which can fully contain the entire onset transient, so that we can further compare the spectral difference between the onset transients and the stationary part.

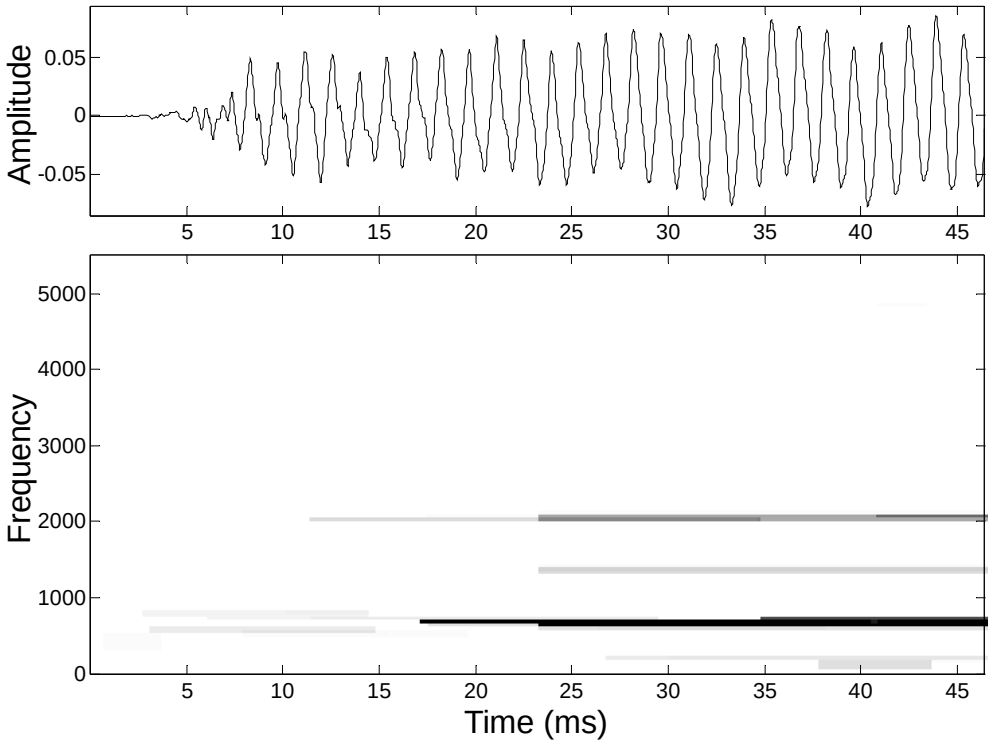
Moreover, the time constant of the ear is approximately 50ms [74], which is longer than the duration of the medium-pitch onset transients, but much less than that of most low tones' onset transients, which is of the order of 100ms. From Fig 4.10, we can see that most frequency levels are replaced or obscured by frequency bands during the first 50ms of the bass tones' onset transient. Such a messy time-frequency structure may give the listener a greater sense of inharmonicity for the bass tones as compared to the medium pitch piano tones.



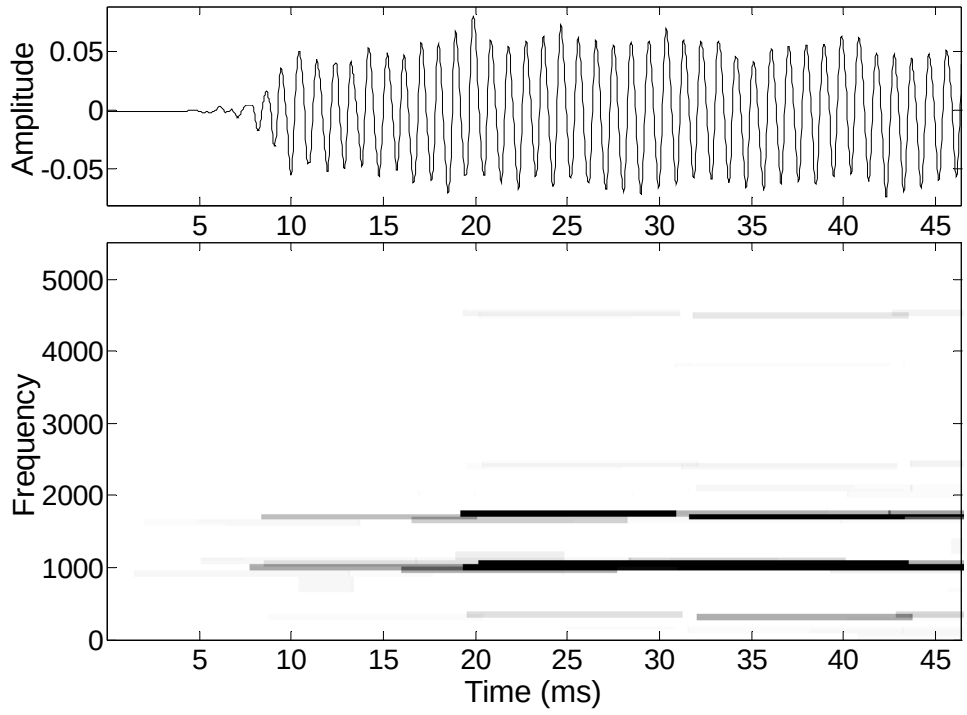
(a) First 50 ms for A0 piano tone



(b) first 50 ms for B0 piano tone



(c) First 50 ms for F5 piano tone



(d) first 50 ms for C6 piano tone

Fig 4.10 Time-frequency plane of approximately first 50 ms for (a) A0, (b) B0, (c) F5 and (d) C6 piano tone

Although the wavelet packet technique has effectively revealed new aspects of the spectra of piano tones, it also has weaknesses. The wavelet packet bases divide the frequency axis into intervals of varying sizes. In each interval, wavelet packet bases are uniformly translated in time. In his book, Mallat [75] has given a pithy evaluation of the wavelet packets: "*These bases are particularly well adapted to decomposing signals that have different behavior in different frequency intervals. If a signal has properties that vary in time, it's then more appropriate to decompose the signal in a block basis that segments the time axis in intervals whose sizes are adapted to the signal structures.*" The next section will therefore attempt to segment the time axis using local cosine bases.

4.3 Local cosine bases

The local cosine basis is one kind of block basis, which allows two adjacent time intervals to overlap while maintaining the orthogonality of the basis function for the two intervals. The signal is decomposed into a local cosine basis set that is composed of functions

$$g_{p,k}(t) = g_p(t)\sqrt{\frac{2}{l_p}}\cos[\pi(k+\frac{1}{2})\frac{t-a_p}{l_p}] \qquad (4\text{-}1)$$

with $t \in [a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ and η represents a small value. a_p and a_{p+1} are two points in the time axis and the size of the interval between them is l_p: $l_p = a_{p+1} - a_p$. Fig 4.11 illustrates the time-frequency partition by local cosine bases.

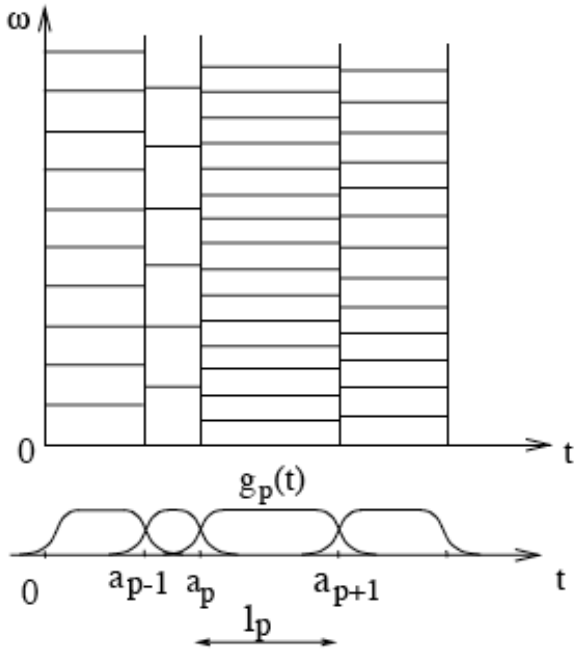


Fig 4.11 Time-Frequency Partition by local cosine bases (Source: from Mallat [75])

g_p depends on a profile function β which is defined as

$$\beta^2(t) + \beta^2(-t) = 1 \quad \text{for} \quad t \in [-1,1] \tag{4-2}$$

Outside $t \in [-1,1]$, $\beta(t)$ satisfies

$$\beta(t) = 0 \quad \text{if} \quad t < -1 \quad \text{and} \quad \beta(t) = 1 \quad \text{if} \quad t > 1 \tag{4-3}$$

We divide the time range of g_p that is $I_p = [a_p - \eta_p, a_{p+1} + \eta_{p+1}]$ into two overlapping interval O_p, O_{p+1} and a central interval C_p:

$$I_p = [a_p - \eta_p, a_{p+1} + \eta_{p+1}] = O_p \cup C_p \cup O_{p+1} \tag{4-4}$$

with $O_p = [a_p - \eta_p, a_p + \eta_p]$ and $C_p = [a_p + \eta_p, a_{p+1} - \eta_{p+1}]$

With the function β, g_p can be expressed as

$$g_p = \begin{cases} 0 & \text{if } t \notin I_p \\ \beta(\dfrac{t - a_p}{\eta_p}) & \text{if } t \in O_p \\ 1 & \text{if } t \in C_p \\ \beta(\dfrac{a_{p+1} - t}{\eta_{p+1}}) & \text{if } t \in O_{p+1} \end{cases} \tag{4-5}$$

Finding a specific form of profile function β is called 'window design'. One popular 'window' is $\beta = \sin(\frac{\pi}{4}(1+t))$ for $t \in [-1,1]$

From Equation (4-1), we can see that the local cosine basis $g_{p,k}(t)$ is equal to multiplying a cosine function $\sqrt{\frac{2}{l_p}}\cos[\pi(k+\frac{1}{2})\frac{t-a_p}{l_p}]$ by a smooth window function $g_p(t)$. That means we should first apply the smooth window operation on the signal followed by a fourier transform. Skipping all intermediate steps, the final result of local cosine basis transform is

$$\left\langle f, g_{p,k} \right\rangle = \sum_{a_p \le t < a_{p+1}} h_p(t)\sqrt{\frac{2}{l_p}}\cos[\pi(k+\frac{1}{2})\frac{t-a_p}{l_p}] \qquad (4\text{-}6)$$

where

$$h_p(t) = \begin{cases} g_p(t)f(t) + g_p(2a_p - t)f(2a_p - t) & \text{if } t \in O_p \\ f(t) & \text{if } t \in C_p \\ g_p(t)f(t) - g_p(2a_{p+1} - t)f(2a_{p+1} - t) & \text{if } t \in O_{p+1} \end{cases} \qquad (4\text{-}7)$$

For any $j \ge 0$, the interval [0, T] is divided in 2^j intervals of length $2^{-j}T$ by replacing a_p by $a_{p,j} = p2^{-j}T$

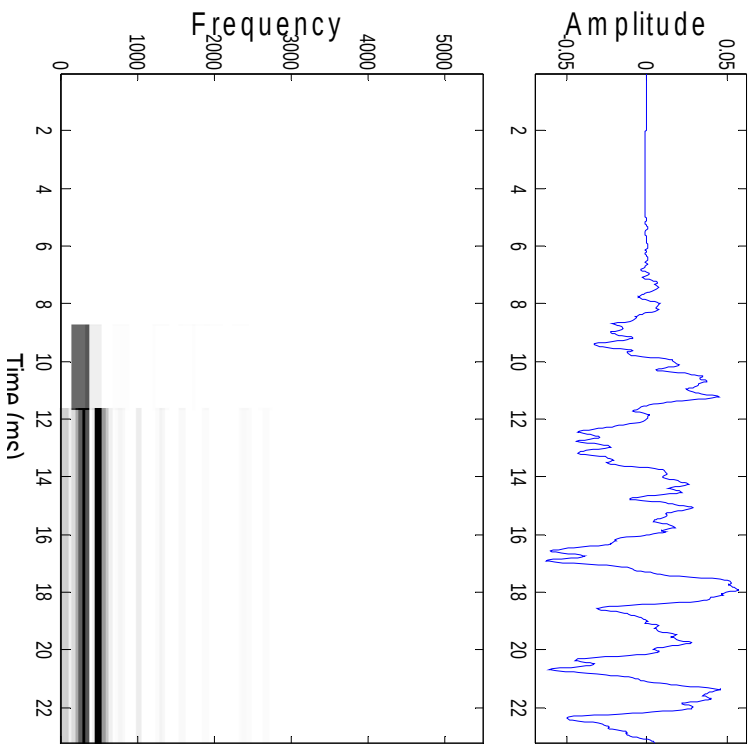
So the local cosine basis transform can also establish a binary tree structure similar to that of the wavelet packet transform. The difference is that the local cosine basis transform repeatedly segments the time axis while the wavelet packet transform segments the frequency axis. Because of the binary tree structure, the best basis selection can also use Coifman's algorithm which has been introduced in preceding section.

These local cosine bases segment the time axis and then translate themselves along the frequency axis. Thus we can expect to see how the targeted signal varies

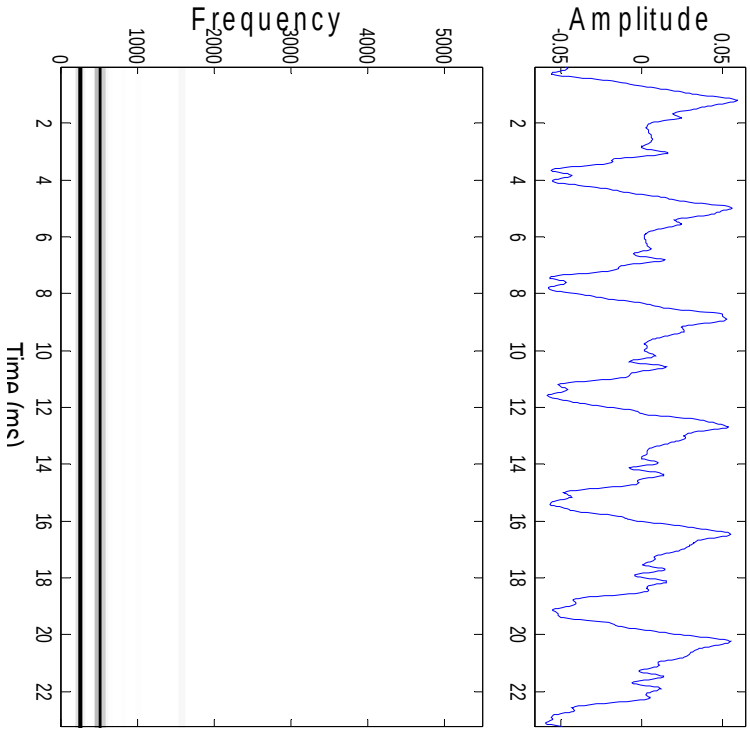
through its onset transient and as the harmonic structure of the stationary part is formed while time progresses. After all, it is the onset part of the tone which gives rise to all harmonics of the stationary part.

The experimental results are shown in Fig 4.12 for a C4 tone. As we can see, the time-frequency plane reflects the corresponding variations of the C4 waveform. There are two distinct parts of the C4 onset transient. The frequency levels of the later portion can be deduced to be the steady harmonics of the stationary part of the signal. This can be observed by comparison with the pattern obtained from the stationary part of the signal, where the fundamental and the first harmonic in Fig 4.12 (b) are distinctly laid out. The energy of the signal is concentrated in the first two harmonics since these two lower harmonics in Fig 4.12 (a) are very much darker than the other harmonics. In the earlier portion of the onset transient figure, we observe that the frequency resolution is broader for the short initial portion, as seen in Fig 4.12 (a). This implies that the fundamental partial undergoes a wide variation of frequency during that short time period, before its frequency finally fixes on the fundamental frequency established during the stationary portion of the tone.

Since the local cosine basis transform half partitions the time axis repeatedly, in Fig 4.12(a), the longer basis covers T/2, and the short basis covers T/9. In Fig 4.13 to Fig 4.17, there are many time-frequency blocks. They all have the length $2^{-j}T$ with varying j.



(a) Onset transient



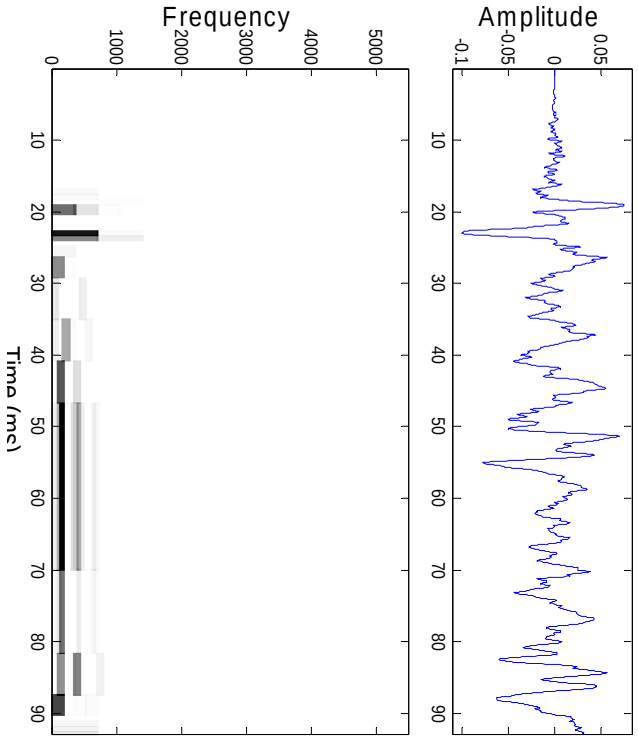
(b) Stationary part

Fig 4.12 The time-frequency plane for C4 by local cosine bases

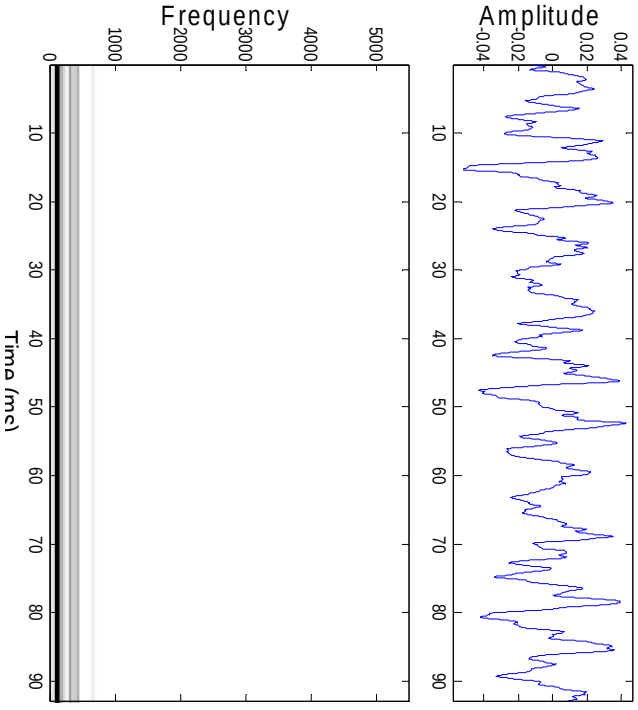
Some bass piano tones, however, present a slightly more complex scenario. The onset of the B0 piano tone is shown in Fig 4.13. The two distinct peaks seen at the beginning of the waveform (one positive and one negative) are clearly indicated by the narrow initial dark blocks in the time-frequency plane. After this very short initial phase, we can see from the many short blocks which follow that the signal undergoes a rather complicated series of changes before stabilizing. Finally, the harmonic structure of the signal then starts to stabilize and take shape. This is also evident if we compare this later portion of the onset transient with the harmonic structure obtained from the stationary part of the tone as seen in Fig 4.13.(b).

Although for the same tone B0, the result by the local cosine transform (shown in Fig 4.13) is a little different from that by the wavelet packet transform. The reason is that wavelet packet transform partitions the time-frequency plane from the frequency perspective while the perspective of the local cosine transform is with respect to time. The wavelet packet transform optimally partitions the frequency axis into some subbands and then the time width of time-frequency blocks within the subbands are determined by the uncertainty relation ($\Delta t \Delta f$ =constant). On the other hand, the local cosine transform optimally partitions the time axis and then determines the frequency width of time-frequency blocks by the uncertainty relation. Therefore, the T-F plane by the wavelet packet transform (Fig 4.9) is sensitive to frequency variations while that by the local cosine transform (Fig.4.13) is sensitive to events with time variations. So the Fig 4.13 can detect the a few peaks at the beginning of the signal and Fig 4.9 indicates that the frequencies of lower partials vary significantly.

More examples of time-frequency plane by local cosine basis are displayed in Figs 4.14 to 4.16.

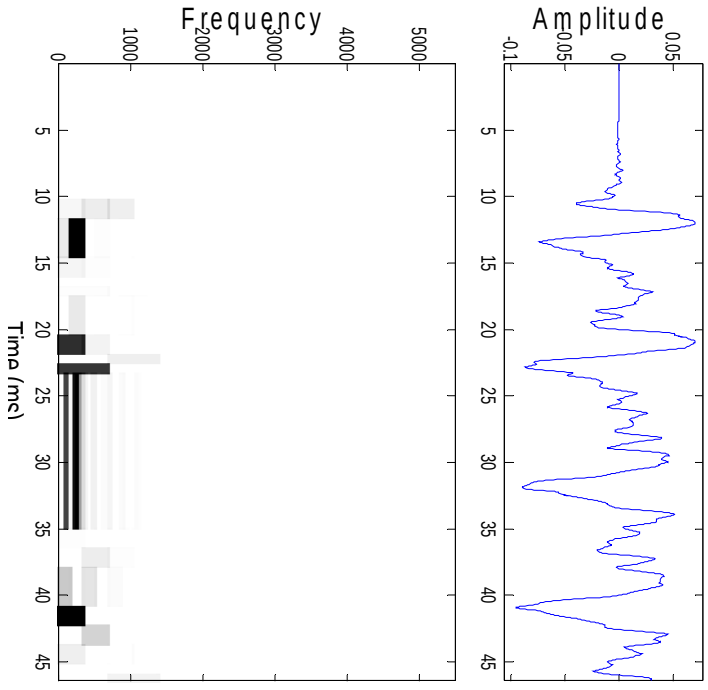


(a) Onset transient

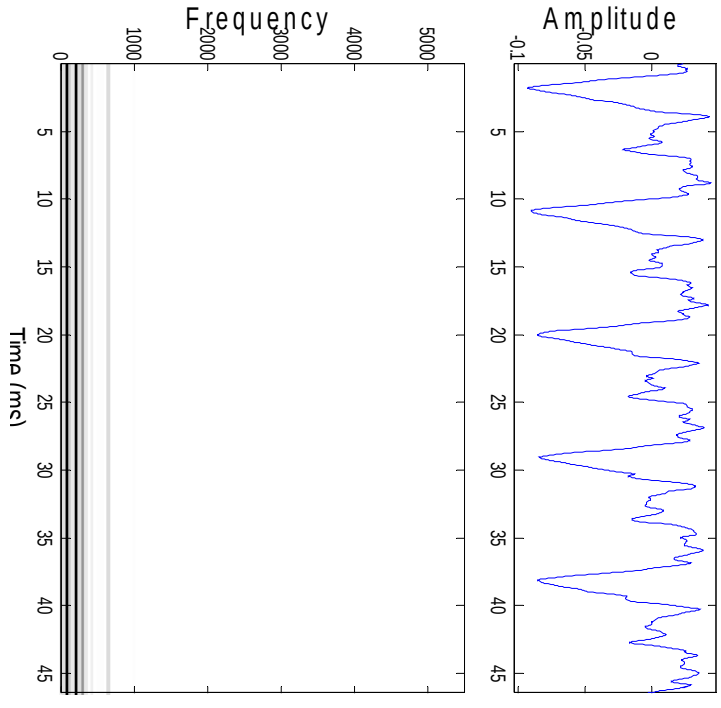


(b) Stationary part

Fig 4.13 The time-frequency plane for B0 by local cosine bases

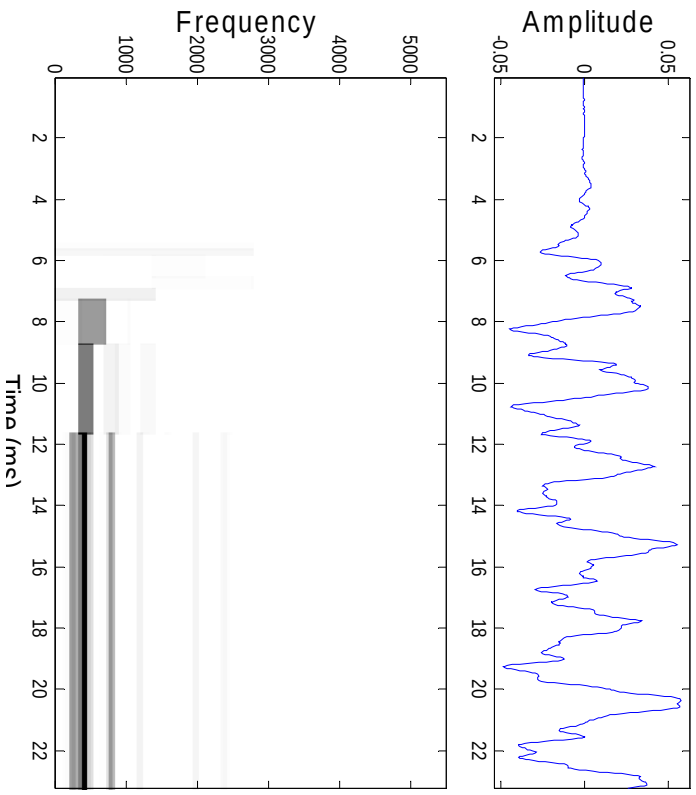


(a) Onset transient

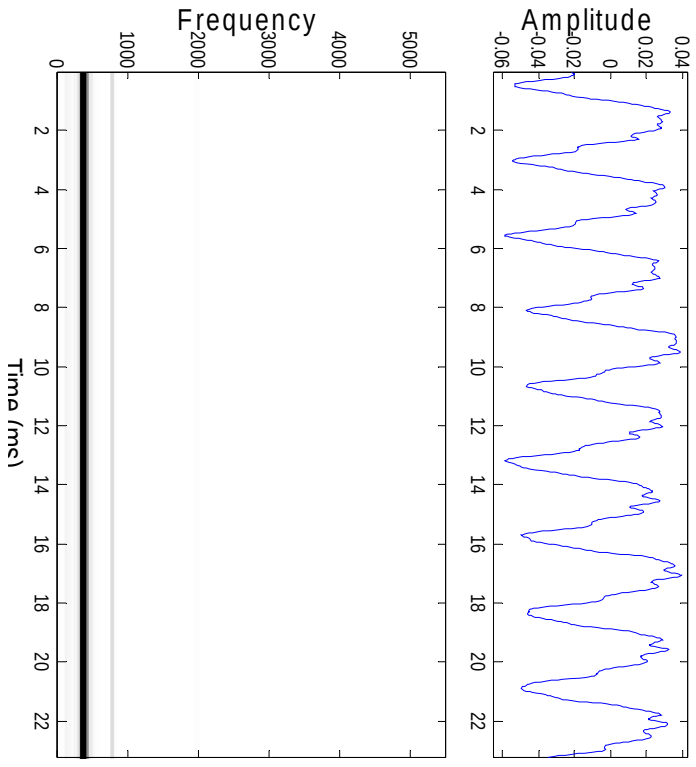


(b) Stationary part

Fig 4.14 The time-frequency plane for A2 by local cosine bases

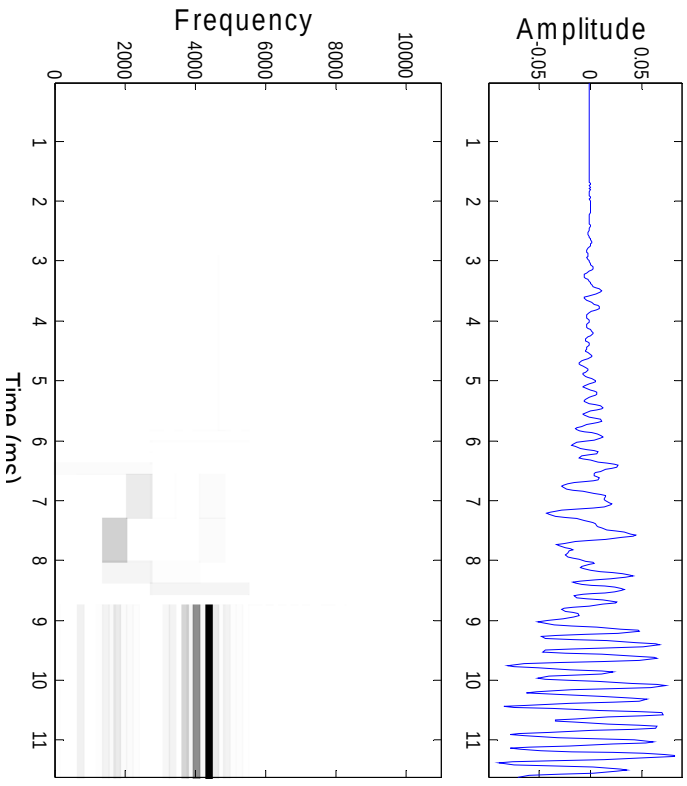


(a) Onset transient

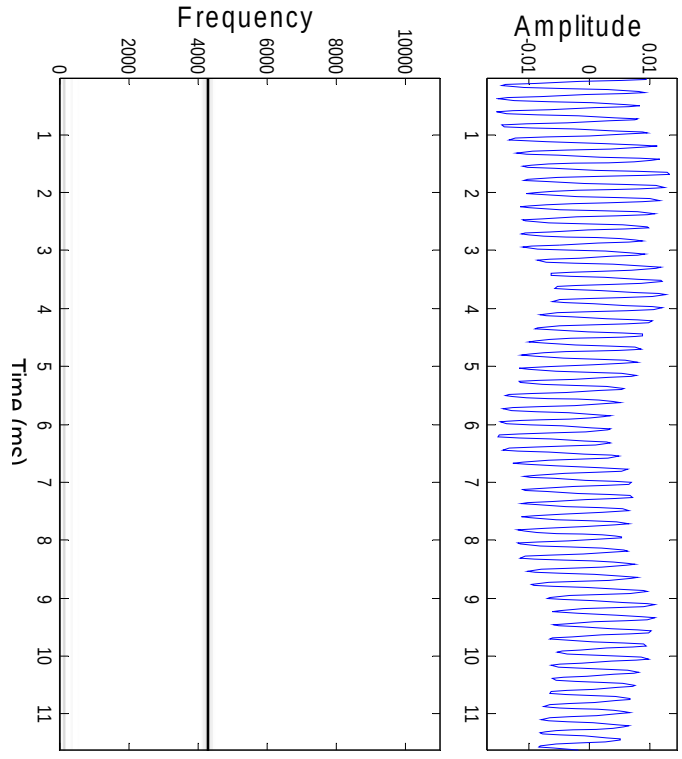


(b) Stationary part

Fig 4.15 The time-frequency plane for G4 by local cosine bases



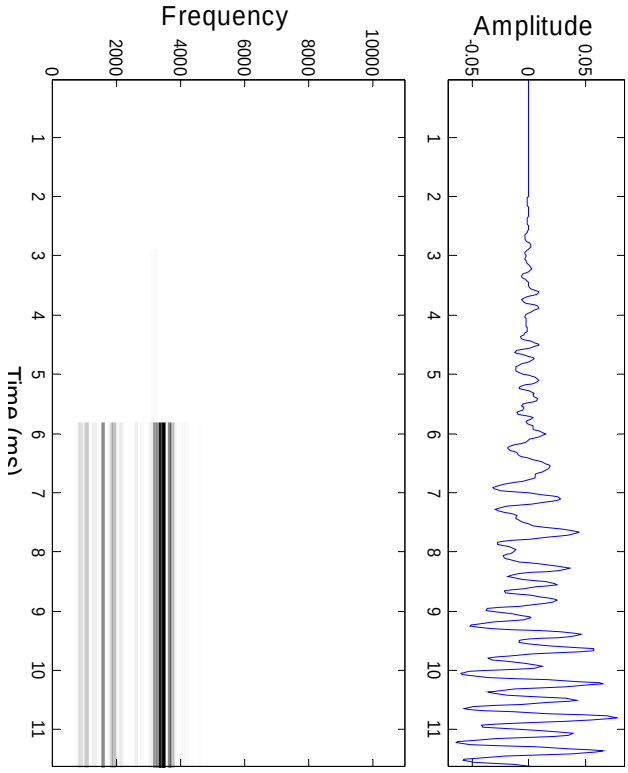
(a) Onset transient



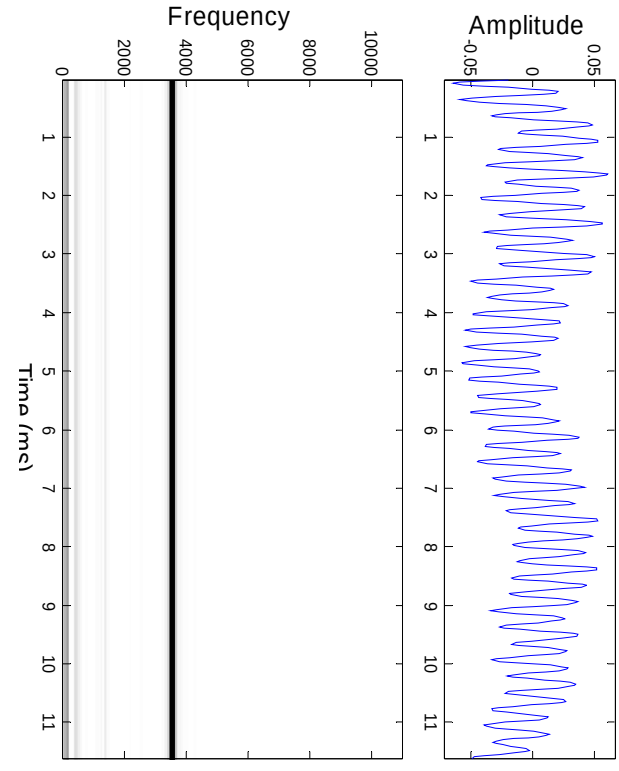
(b) Stationary part

Fig 4.16 The time-frequency plane for C8 by local cosine bases

In Fig 4.16, there is a narrow line at a frequency slightly above 4000 Hz. Checking the Appendix A, we know it is the fundamental frequency of C8 (4186 Hz). However, in Fig 4.16 (a), there are some lines below the C8's fundamental, though the fundamental frequency is still most significant. A similar pattern happens in other treble notes, for example the A7 piano tone in Fig 4.17. We recall the second definition of onset transient in Chapter 3 and Fig 3.5 where the onset transients comprise of the noise, the time during which the periodic motion is not established, the quasi-period period and the stable period. These observations may suggest the noise (part A in Fig 3.5) and the duration in which the periodic motion is not established yet (part B in Fig 3.5) play an important role in the onset transient of treble notes and thus the frequency distribution is much more varied.



(a) Onset transient



(b) Stationary part

Fig 4.17 The time-frequency plane for A7 by local cosine bases

4.4 Matching pursuit

Until now, both wavelet packet analysis and local cosine bases have disclosed much useful information about the piano onset transients. But they may have missed other information due to their limitations. They may, for example, only select the best Heisenberg box in any frequency or time interval and then tile the entire interval with such a box.

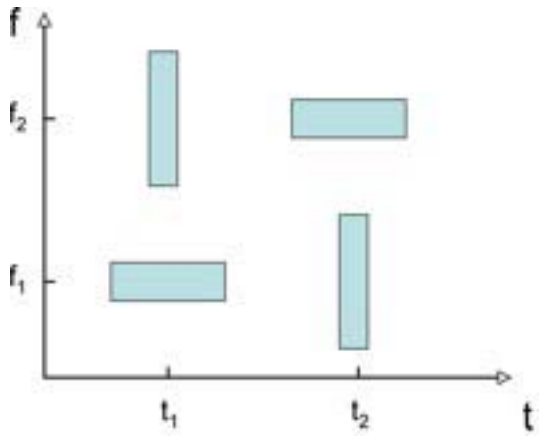


Fig 4.18 box1(t_1, f_1); box2(t_2, f_1); box3(t_1, f_2); box4(t_2, f_2)

Referring to Fig 4.18, assume that a signal is composed of such four Heisenberg boxes in the time frequency plane. Each Heisenberg box corresponds to a certain single wavelet packet basis or local cosine basis. Considering the case for the wavelet packets transform, the transformation algorithm needs to decide between choosing either box1 or box2 for analysis since they are in the same neighborhood of frequency f_1. If box1 is of larger weighting or importance than that of box2, the latter will be ignored. This however, presents a potential error as box2 has an equal

probability of describing essential features of the signal which box1 cannot do alone. Once box1 is decided upon, it will be used to analyze the frequency and those at its frequency neighborhood. Positions such as at (t_2, f_1) will have a "copy" of box1 replacing box2. Therefore, the contents of box2 which are not included in this "copy" of box1 will be spread across many other new boxes around (t_2, f_1). Hence the actual time-frequency content of the signal at (t_2, f_1) may not be described in an optimal fashion as might have been if box2 had been used. Similarly, for the local cosine bases case which segment the time axis, if box1 carries more weight than box3, a "copy" of box1 will be moved to (t_1, f_2) and box3 will not be utilized.

In comparison, the matching pursuit algorithm is not restricted by such tiling considerations, and can freely choose any basis from wavelet packet basis or local cosine basis dictionaries at any desired time-frequency position.

Consider a signal space where there are at least *N* linearly independent vectors. Let $D = \{g_p\}_{0 \le p \le P}$ be a redundant dictionary of *P*>*N* vectors. Given any $M \ge 1$, an approximation f_M of the original signal f may be obtained from a linear combination of any *M* dictionary vectors:

$$f_M = \sum_{m=0}^{M-1} a(p_m) g_{p_m} \qquad (4\text{-}8)$$

Since g_p can be freely chosen from the dictionary, the number of possible combinations will increase rapidly when *M* increase. However, the best approximation f_M is only one that minimizes the $\|f - f_M\|$. Such a kind of optimization is an NP hard problem which means that there is no known polynomial time algorithm which can solve it. Any algorithm that is especially created to solve

this optimization is called a pursuit algorithm. A matching pursuit uses a greedy algorithm [76] that chooses at each iteration a waveform from the given dictionary that is best adapted to approximate part of the signal. Firstly, we can find the basis g_{p0} from the dictionary D which makes $\left|\left\langle f, g_{p0}\right\rangle\right|$ maximum. Then project signal f on g_{p0} and there is $f = \left\langle f, g_{p0}\right\rangle g_{p0} + Rf$ where Rf is the residue. Then pick the basis g_{p1} that makes the $\left|\left\langle Rf, g_{p1}\right\rangle\right|$ maximum. Projecting Rf on g_{p1}, we have $Rf = \left\langle f, g_{p1}\right\rangle g_{p1} + R^1 f$. If this process continues, we have $f = \sum \left\langle f, g_{pm}\right\rangle g_{pm} + R^m f$. Mallat has proved that $\left\|R^m f\right\|$ converges exponentially to 0 when m tends to infinity.

So when $\left\|R^m f\right\|$ is less than a predefined threshold at M iterations, f can be approximated by $f_M = \sum \left\langle f, g_{pm}\right\rangle g_{pm}$

In our cases here, the dictionary is composed of wavelet packet bases or in other words, the set of all possible time-frequency blocks. Although the wavelet packet transform previously mentioned and matching pursuit here are based on the same set of bases, one distinction is that to find out the best representation of the signal, the wavelet packet transform uses Coifman's 'best basis algorithm' while the matching pursuit uses Mallat's greedy algorithm.

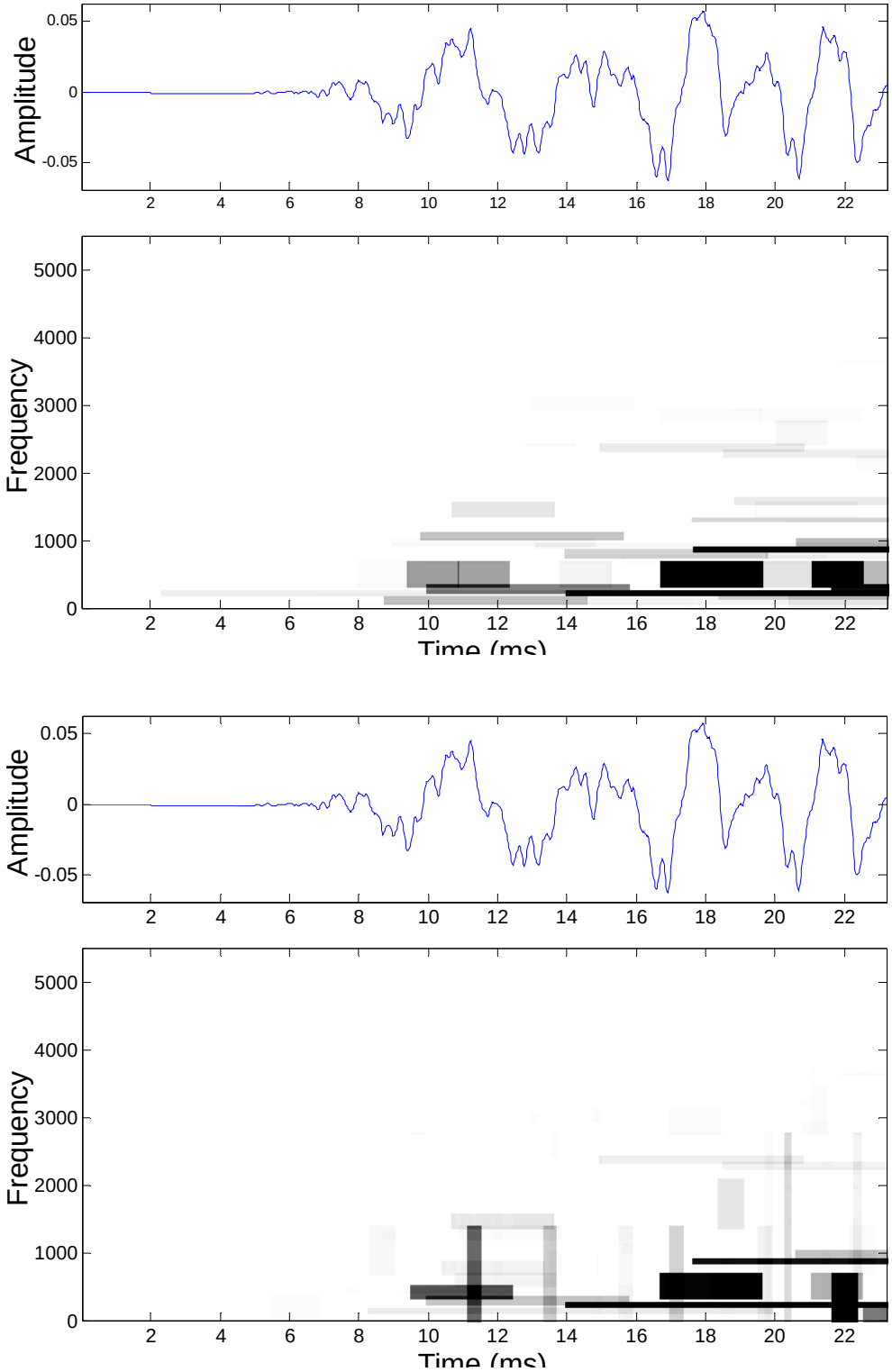


Fig 4.19 Comparison: the time-frequency plane for tone C4 by wavelet packet (top) and matching pursuit (bottom)

Fig 4.19 compares the results using the wavelet packet analysis and matching

pursuit using the wavelet packet dictionary. In comparison with wavelet packets, the matching pursuit technique retains the main frequency structures and on top of that, includes a few additional vertically standing time-narrow bars crossing the horizontal frequency lines. These vertical bars reflect the abrupt change in the relative amplitudes of the waveform of the signal within their respective short time resolutions Δt. Such abrupt changes are not part of the periodic behavior which forms the steady harmonics of the tone, but they are rather just isolated changes on top of the usual periodic behavior. The periodic behavior, as we have seen, is best described by the horizontal frequency lines, whose assigned darkness depicts the magnitudes of the harmonic frequencies of the periodic behavior. The observed overlapping of vertical bars and horizontal lines means that the signal is undergoing two different behaviors at the same time. Further observations of other piano tones reveal that the vertical bars tend to cross the first few horizontal frequency lines at the lowest frequencies. This indicates that the violent changes mainly occur in the fundamental or lower harmonics. Combining this with the fact that the fundamental and the lower harmonics contain most of the energy, which has already been stated in previous sections, this agrees with Palmer's [77] statement that little energy was generated in the upper harmonics for each piano sound event, and peak amplitude was primarily determined by the first few (i.e., the lowest frequency) spectral components.

Taking the same examples which we have used for wavelet packets, we will also obtain their time-frequency planes by matching pursuit as below for piano tones D7, E2, A3, F5 and B0 in Figs 4.20, 4.21, 4.22, 4.23 and 4.24 respectively. Besides the

similar band structure with some vertical bars as mentioned before, matching pursuit also provides a more concise and clear time-frequency plane compared to t-f planes by wavelet packets, especially for bass tones (e.g. piano tone B0). But we can not conclude too hastily to which one, the matching pursuit or wavelet packets, is better. Only when used in specific applications, can these two wavelet-based techniques be compared for their relative advantages or disadvantages.

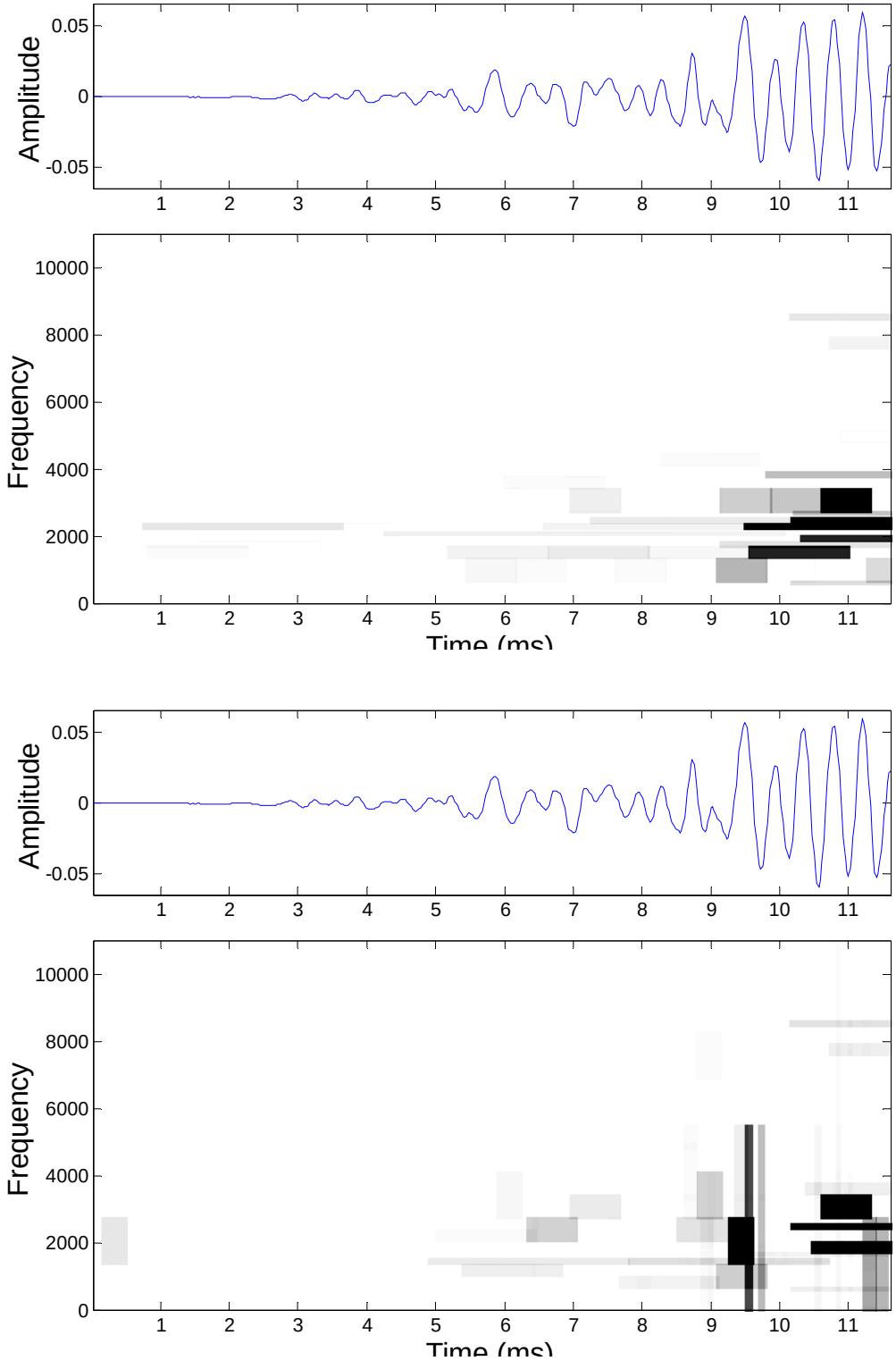


Fig 4.20 Comparison: the time-frequency plane for tone D7 by wavelet packets (top) and matching pursuit (bottom)

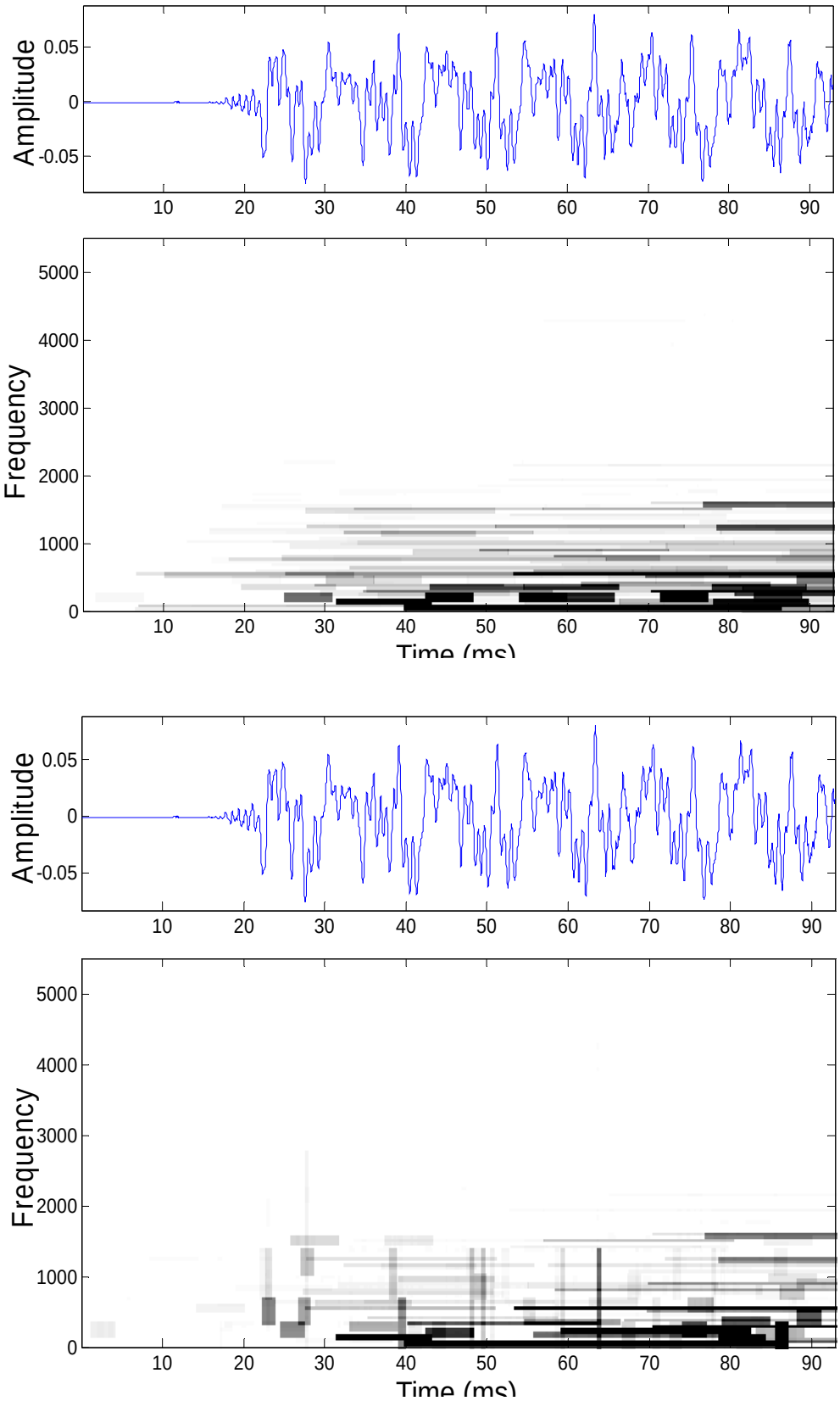


Fig 4.21 Comparison: the time-frequency plane for tone E2 by wavelet packets (top) and matching pursuit (bottom)

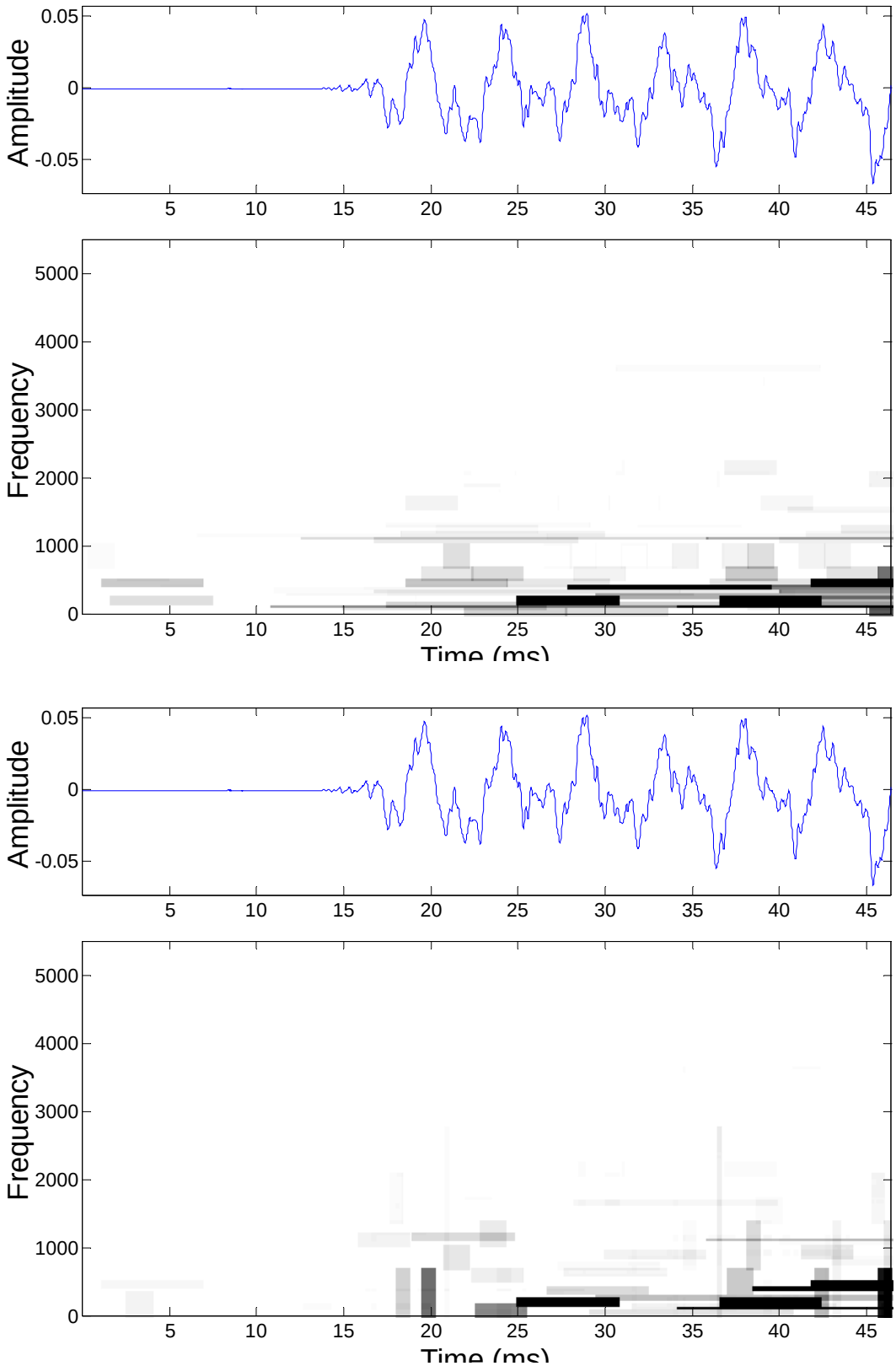


Fig 4.22 Comparison: the time-frequency plane for tone A3 by wavelet packets (top) and matching pursuit (bottom)

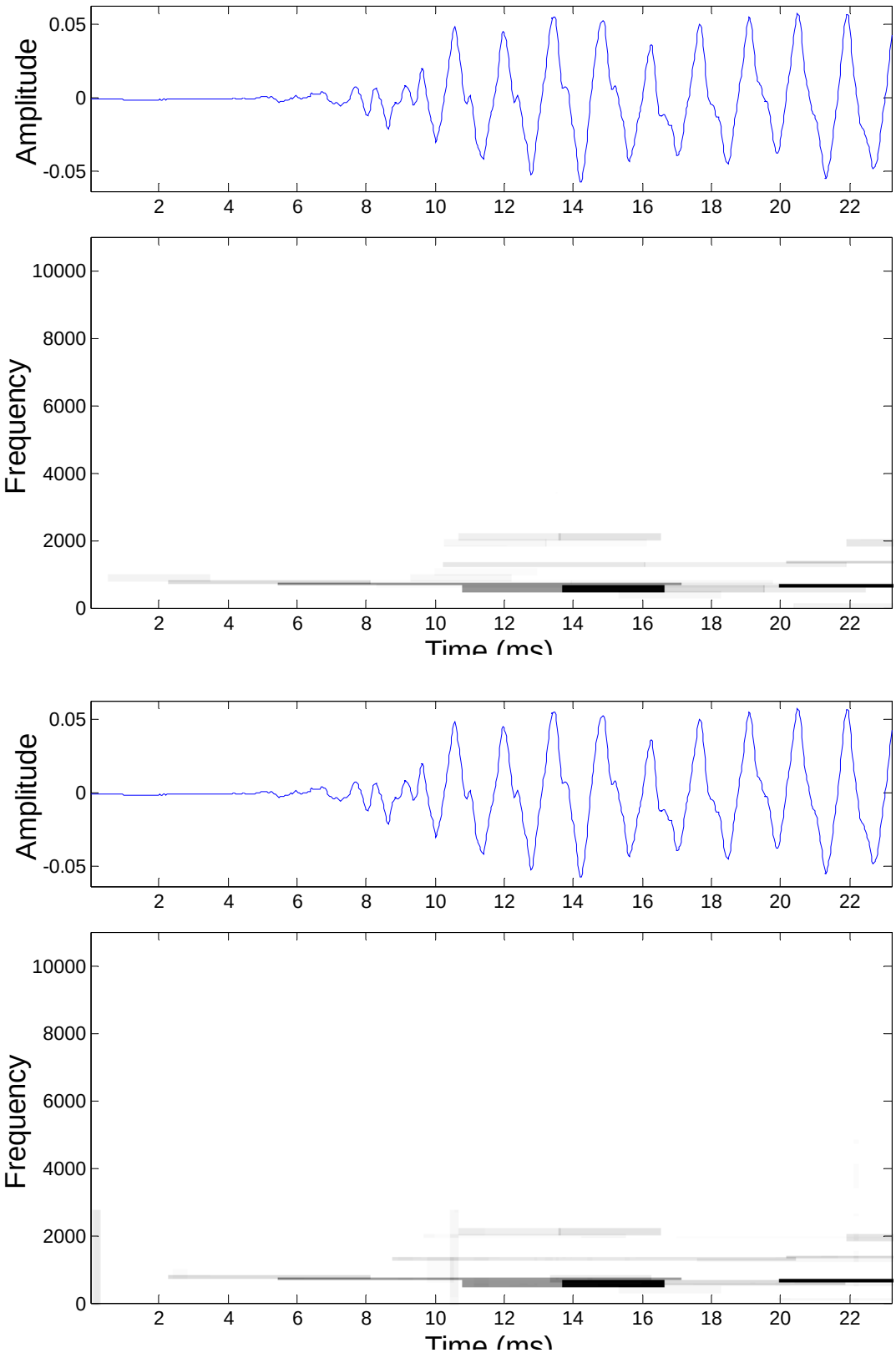


Fig 4.23 Comparison: the time-frequency plane for tone F5 by wavelet packets (top) and matching pursuit (bottom)

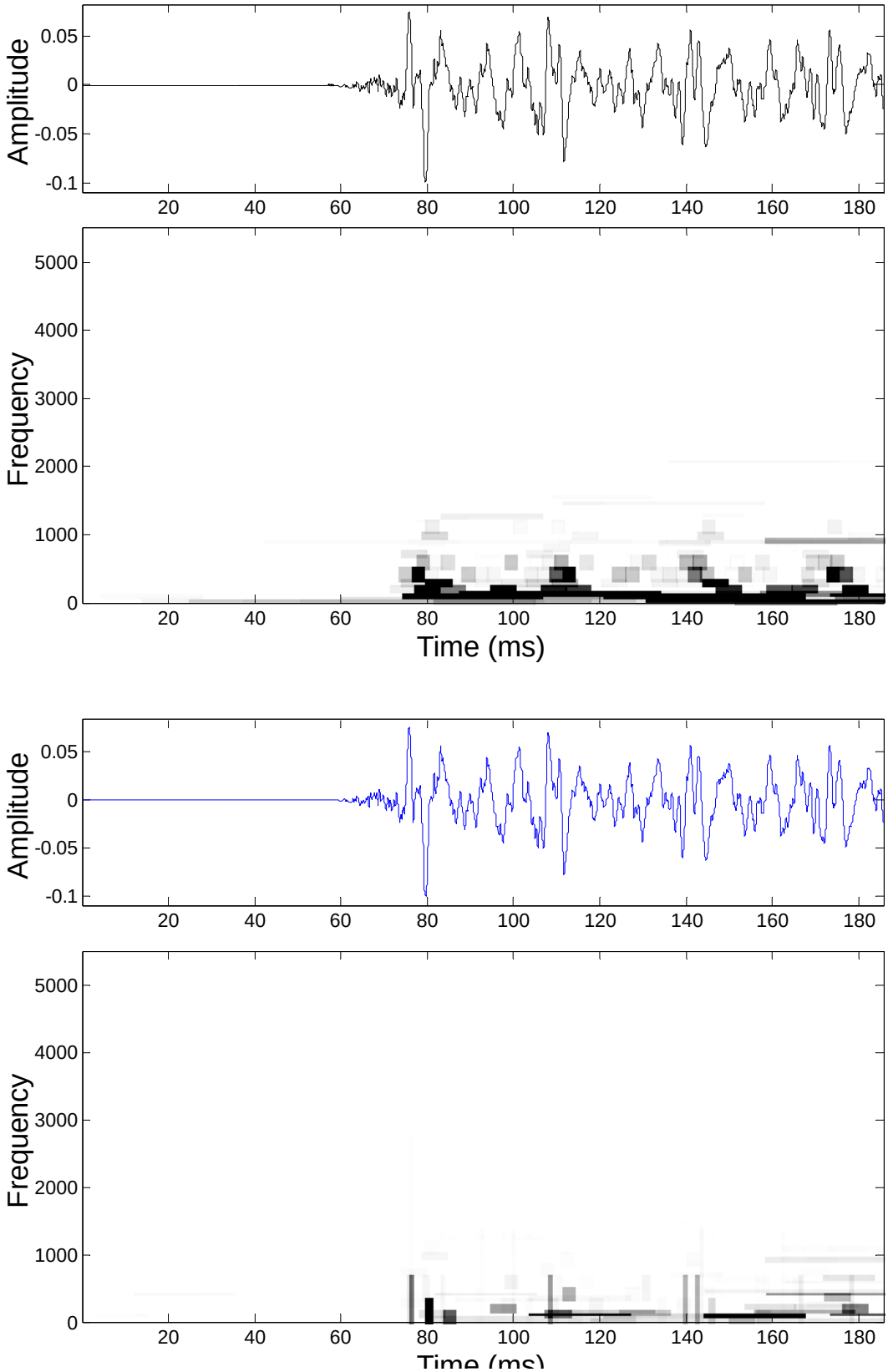


Fig 4.24 Comparison: the time-frequency plane for tone B0 by wavelet packets (top) and matching pursuit (bottom)

Chapter 5 Reconstructing Waveforms By Wavelet Impulse Synthesis

In the last Chapter, we have used wavelet packet analysis (WPT) and its tree structure, and established the relationship between the tree structure and the time-frequency plane. In this Chapter, we will further elaborate on this relationship and use it to derive a new approach to waveform synthesis - the wavelet impulse synthesis. Just as AM and FM techniques from the radio engineering field have been applied to computer music, we have linked the wavelet impulse algorithm to the measurement of inharmonicity coefficients of piano sounds. The wavelet impulse synthesis is very important in the context of this dissertation and will build a foundation for measuring the inharmonicity coefficients of piano tones in the next Chapter.

5.1 Wavelet Impulse Synthesis

As explained in Chapter 4, in *wavelet packet transform* (WPT) analysis the signal is analyzed into a nested set of coefficients (reflecting the tree structure, or equivalently the time-frequency plane) by WPT filters. In order to quickly identify a particular coefficient in a tree structure, we assign *d* and *b* to respectively label the tree depth (or level) and the node to which the coefficient belongs, and use k as the index of this coefficient within the node. The position of each coefficient will thus be uniquely identified by the combination (*d, b, k*), where *d, b*, and *k* all start from 0.

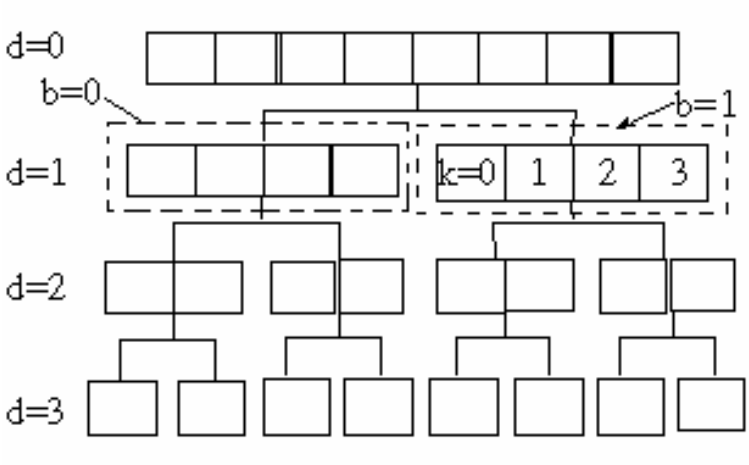


Fig 5.1 An 8-point 3 level full tree WPT: any coefficient can be uniquely identified by (*d,b,k*), where $d \equiv depth$, $b \equiv node$, $k \equiv index\ within\ node$

The d, b and k do not exist independently of each other and some relationships can be easily observed from Fig.5.1. At the d^{th} level, there are 2^d nodes and the node (d, b) comprises $n/2^d$ coefficients where n is the length of the original signal. Moreover, the node (d, b) represents the frequency band in the time-frequency plane

ranging from $b\frac{f}{2^d}$ to $(b+1)\frac{f}{2^d}$ Hz where f is equal to half of the sampling frequency. Likewise, the coefficient (d, b, k) corresponds to the time-frequency block which spans the time axis from $k \cdot 2^d \cdot \Delta t$ to $(k+1) \cdot 2^d \cdot \Delta t$ where Δt is defined as the sampling time interval.

Similar to the Fourier transform where trigonometric functions are used as bases to decompose the signal being analyzed, and whose coefficients are actually the amplitudes of the bases, the WPT also breaks the signal into selected WPT bases and their amplitudes are also called coefficients, which have already been identified by (*d, b, k*) previously. Let us denote $\psi_{d,b,k}(t)$ as such a WPT basis, where to keep things simple, we do not distinguish it as a scale function or a wavelet function. The tree structure for a certain signal, *x*, has a valid representation that can be described as follows:

$$x = \sum_{\text{all\_termainal\_nodes}} \underbrace{\sum_{\text{k}} \left\langle x(t), \psi_{d,b,k}(t) \right\rangle \psi_{d,b,k}(t)}_{\text{a certain terminal node (d,b)}} \qquad (5\text{-}1)$$

where $\left\langle x, \psi_{d,b,k} \right\rangle$ is the value of the coefficient (*d, b, k*) and a *best basis algorithm* (such as Coifman's) decides the selection of terminal nodes and the particular tree structure.

It is evident from the formula that similar to the Fourier transform, the wavelet packet transform is also a linear transform. We can see that a WPT basis $\psi_{d,b,k}$ can represent the position of coefficient (d, b, k) in the tree or in the time-frequency block (d, b, k) in the time-frequency plane. The two summing symbols, $\sum_{\text{all\_terminal\_nodes}}$ and $\sum_{k}$ indicate that if we know all coefficients and their corresponding $\psi_{d,b,k}$, we can reconstruct the original sound.

To reconstruct the waveform of a certain WPT basis $\psi_{d,b,k}$, the toolbox *Wavelab* uses a so-called *impuls*e method. It keeps the resultant tree structure from the analysis process intact and resets coefficient $\langle x, \psi_{d,b,k} \rangle$ to 1 but all other coefficients to 0, thus generating an impulse in the position (*d, b, k*) of the tree. The inverse wavelet packet transform (IWPT) is then applied to this modified "impulse tree" to realize the waveform of $\psi_{d,b,k}$ spanning the full time duration of the original sound.

But in the actual calculation, the original signal's tree structure can be ignored because the waveform of a basis has nothing to do with the structure of the targeted signal. If we call a node whose coefficients are all zero a *zero-node*, extending that node downwards to two offspring *zero-nodes* does not change the function represented by the tree. Repeating the extending operations, no matter what the original tree structure was, we always can replace it with a d-level full tree which has only one non-zero coefficient (i.e., the coefficient (*d, b, k*) at level *d* with the value 1).

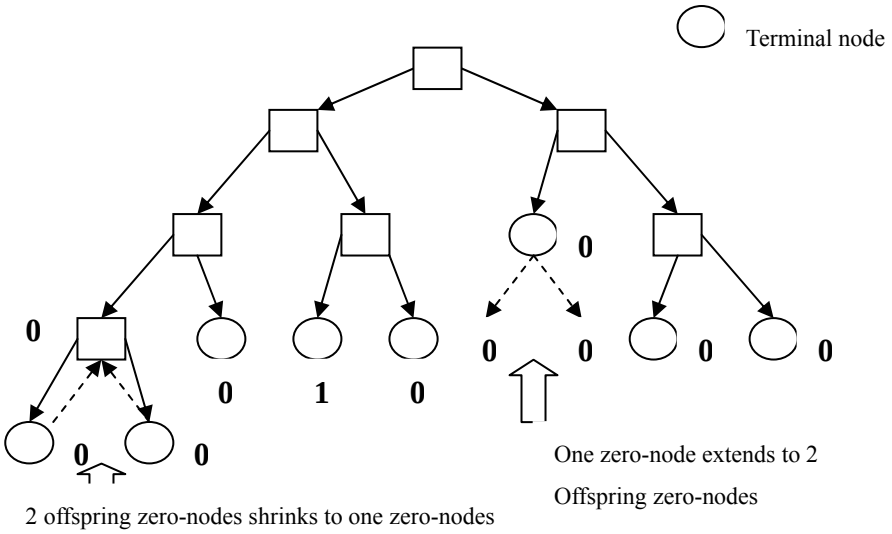


Fig 5.2 The demonstration for zero-nodes' extending or shrinking

By multiplying each basis's waveform $\psi_{d,b,k}$ with its weighting $\langle x,\psi_{d,b,k}\rangle$, we obtain the weighted decomposed component $W_{d,b,k}=\langle x,\psi_{d,b,k}\rangle\psi_{d,b,k}$. The summation of all these weighted basis components gives the original signal.

Mathematically, we can represent the analysis process according to the following expressions

$$WPT\{x\}=WPT\{\sum_{\text{all\_termainal\_node}}\underbrace{\sum_{\text{k}}\langle x(t),\psi_{d,b,k}(t)\rangle\psi_{d,b,k}(t)\}}_{\text{a certain terminal node (d,b)}} \tag{5-2}$$

Due to the linearity of WPT, the above expression can be expressed as

$$\begin{aligned} WPT\{x\} &= \sum_{all\_ter\min al\_nodes}\underbrace{\sum_{k}\langle x,\psi_{d,b,k}\rangle WPT\{\psi_{d,b,k}\}}_{a\_certain\_ter\min al\_node\_(d,b)} \\ &= \sum_{all\_time-frequency\_blocks}\underbrace{\langle x,\psi_{d,b,k}\rangle WPT\{\psi_{d,b,k}\}}_{any\_time-frequency\_block(d,b,k)} \end{aligned} \tag{5-3}$$

where $WPT\{\psi_{d,b,k}\}$ is a d-level full tree for the $\psi_{d,b,k}$ basis, which would be the *impulse* tree described above which has only one non-zero coefficient at position (d, b, k)

For the reconstruction process, the inverse wavelet packet transform or IWPT is applied on both sides of the Equation (5-3),

$$\begin{aligned} IWPT\bullet\{WPT\{x\}\} &= IWPT\{\sum_{all\_time-frequency\_blocks}\underbrace{\langle x,\psi_{d,b,k}\rangle WPT\{\psi_{d,b,k}\}\}}_{any\_time-frequency\_block(d,b,k)} \\ &= \sum_{all\_time-frequency\_blocks}\underbrace{\langle x,\psi_{d,b,k}\rangle IWPT\bullet\{WPT\{\psi_{d,b,k}\}\}}_{any\_time-frequency\_block(d,b,k)} \end{aligned} \tag{5-4}$$

where we have $IWPT\bullet WPT=\vec{1}$.

The left side of the Equation (5-4) shows the reconstruction by the usual inverse wavelet packet transform (IWPT), i.e.

Fig 5.3 Traditional Wavelet Packet Analysis and Synthesis

while the right side shows the reconstruction by the wavelet impulse synthesis (WIS) approach. Although mathematically identical, the WIS approach is very different from the IWPT approach in physical meaning and also in terms of practical computational demands. Starting from the terminal nodes of the tree structure, the IWPT retraces the path of nodes upwards and shrinks the tree to its top level. In other words, the IWPT combines frequency bands along the frequency axis in the time-frequency plane.

However, in the wavelet impulse method, we are able to focus on each coefficient separately and compute its related basis $\psi_{d,b,k}$ by a d-level full tree to obtain the component waveform for each time-frequency block.

To illustrate how the wavelet impulse synthesis works, we also take the onset portion of a C4 piano tone in Chapter 4 as the example. After wavelet packets transform and the best basis selection algorithm, we sorted all coefficients in descending order. Picking the largest coefficients and finding out their associated (*d, b, k*) positions, we can obtain the following waveforms corresponding time-frequency planes as shown in Fig 5.5 to 5.9. Fig 5.4 shows the T-F plane and waveform for the original C4 piano tone. Fig 5.10 shows a reconstructed waveform and T-F plane of this piano tone using only these five largest coefficients.

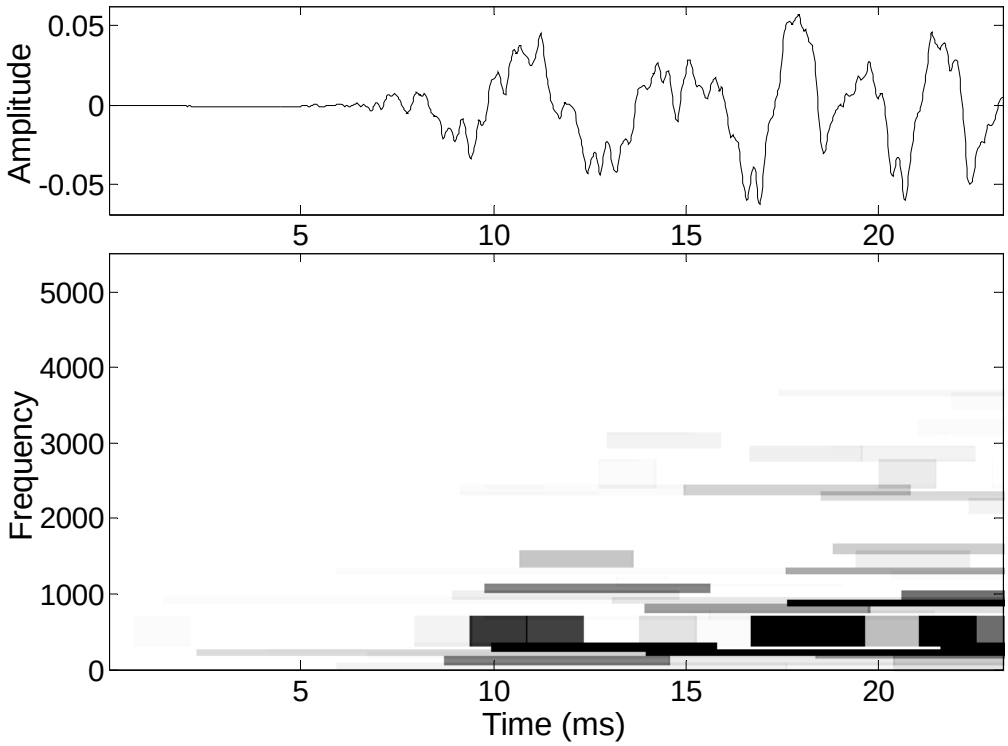


Fig 5.4 The T-F plane of the onset transient of C4 piano tone

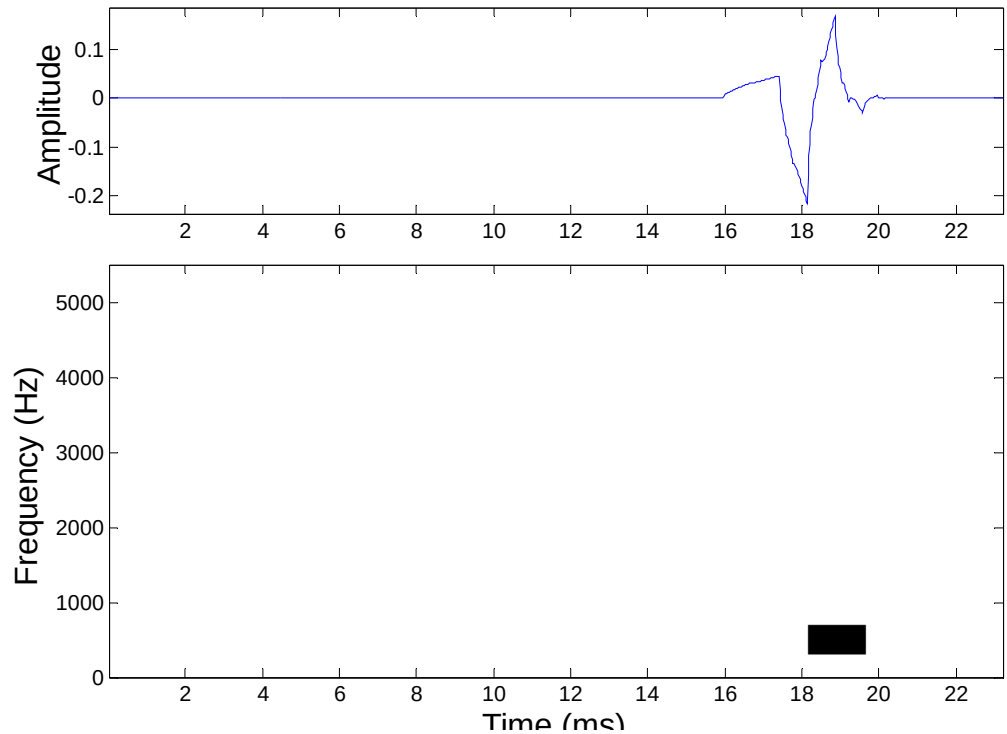


Fig 5.5 The T-F block whose coefficient is largest (bottom) and the waveform of the basis this T-F block corresponds (top)

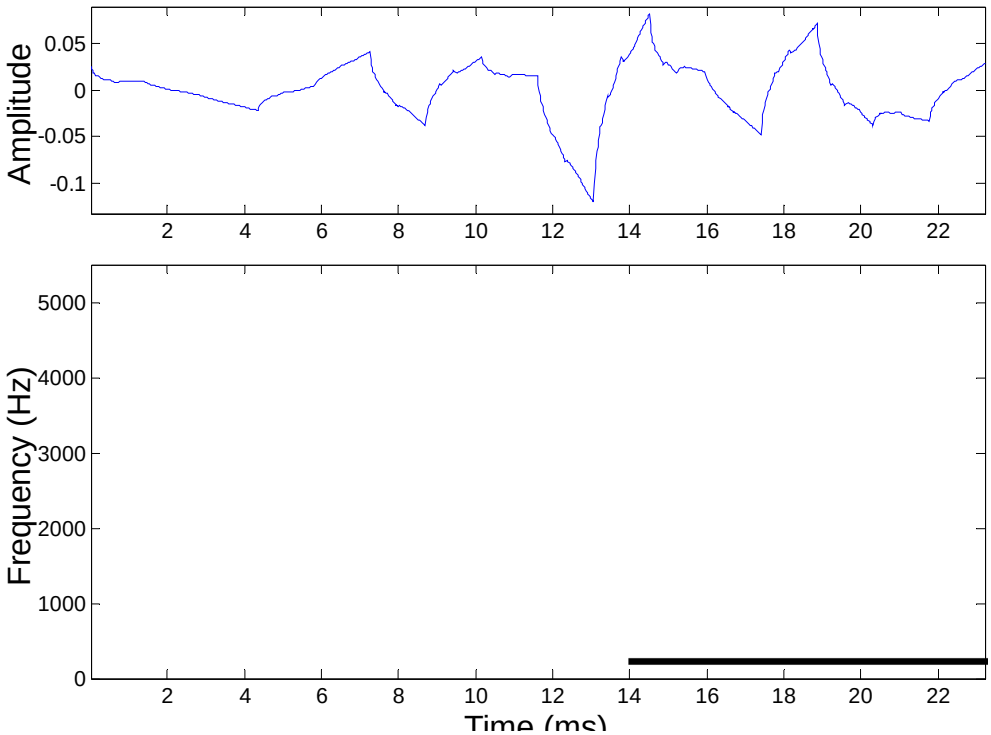


Fig 5.6 The T-F block whose coefficient is 2nd largest (bottom) and the waveform of the basis this T-F block corresponds (top)

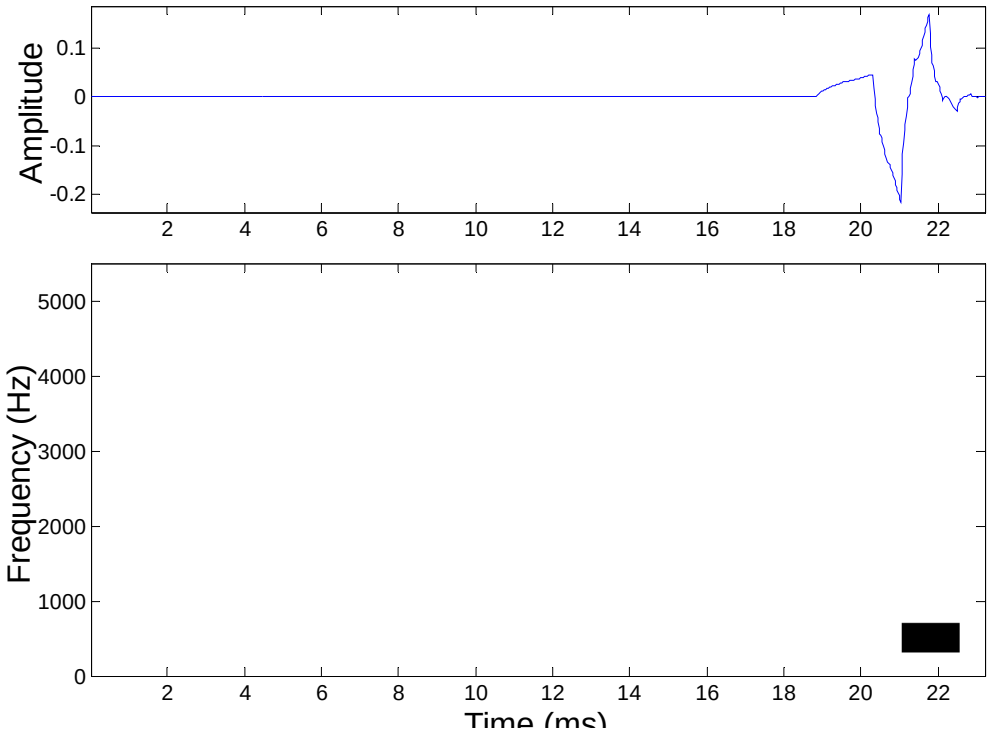


Fig 5.7 The T-F block whose coefficient is 3rd largest (bottom) and the waveform of the basis this T-F block corresponds (top)

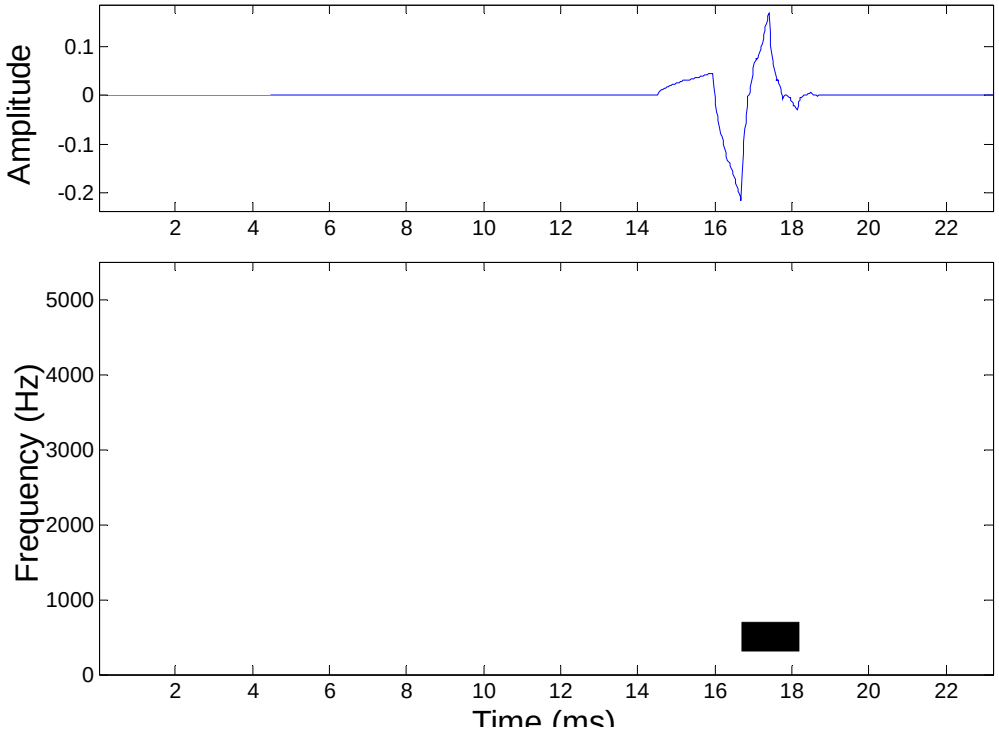


Fig 5.8 The T-F block whose coefficient is 4th largest (bottom) and the waveform of the basis this T-F block corresponds (top)

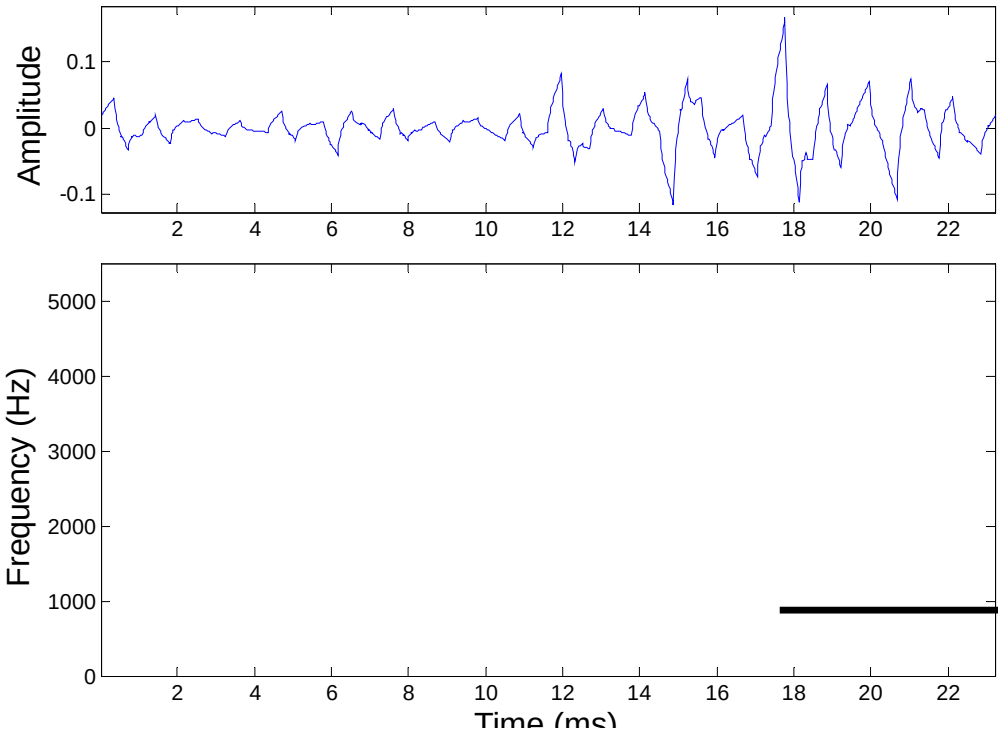


Fig 5.9 The T-F block whose coefficient is 5th largest (bottom) and the waveform of the basis this T-F block corresponds (top)

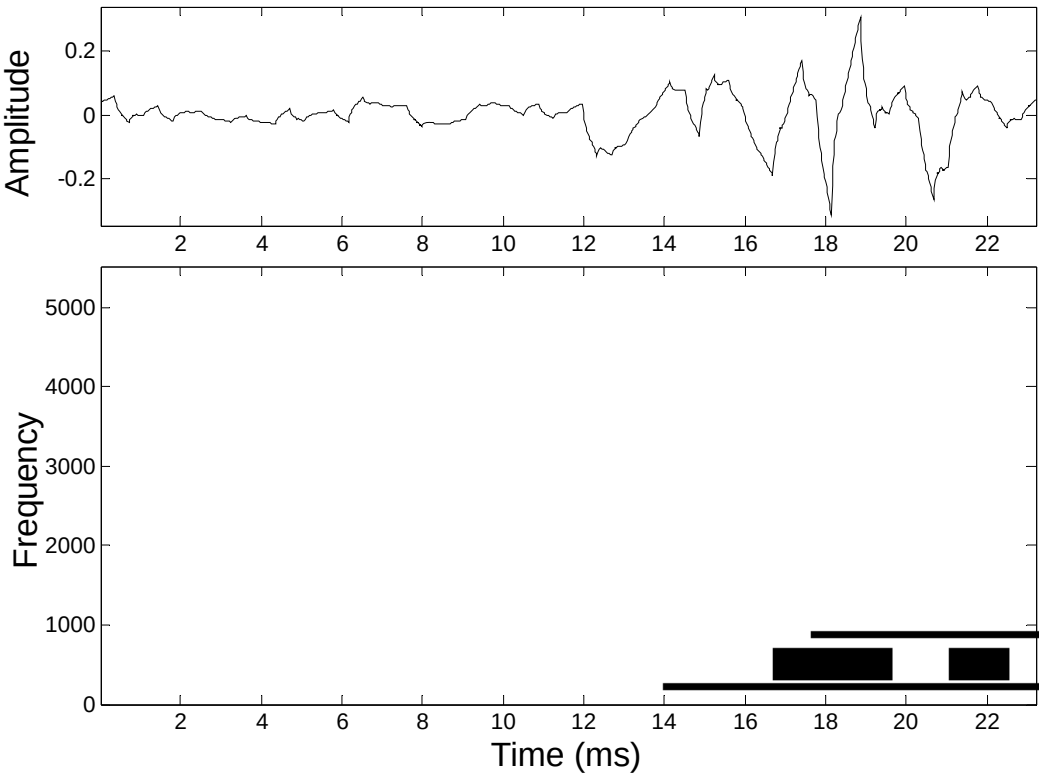


Fig 5.10 The synthesis by five largest T-F blocks

WPT partitions the time-frequency plane from the perspective of frequency. That means it firstly measures each basis's frequency value. Then the time range is determined according to the uncertainty principle (i.e. $\Delta t \Delta f$ =constant). Therefore it is impossible that the time-frequency block does not coincide with the exact time span of the waveform because time span is ruled by the uncertainty principle and is not directly gauged from the signal. But the 'mass' center of the waveform will fall within the time span. That is why we also include the local cosine basis transform in Chapter 4. The local cosine basis directly gauges the time span from the signal and determines the frequency interval by the uncertainty principle.

5.2 Effective Approximation And Waveform Reconstruction

From equation (5-1), if some coefficients' weighted values are significantly larger than the others, it may be possible to ignore the smaller components during the reconstruction of the original signal in order to reduce the computational load in the reconstruction. This suggests that the wavelet transform when used for storage or transmission of waveforms, may lead to significant compression of the digital waveform data as compared to storing it in normal uncompressed digital formats. This was indeed the case for the piano tones, for which only a small portion of all the time-frequency blocks has been 'visually' presented (e.g., Fig 4.3 and Fig 4.4). The others are too white in the grey scale to be seen in the figures (i.e., their coefficients are very small relative to those of the 'visualized' blocks).

Fig 5.11 to Fig 5.16 shows a number of different reconstructions of the digitized B0 piano tone which has 32,768 sampled points (for which there are 32,768 coefficients in the tree structure and therefore the same number of wave components). These reconstructions only use a small number of the most significant components from these 32,768 components. The number of time-frequency blocks used for each reconstruction increases from 100 to 2000 blocks, each case respectively shown in Figs 5.11 to 5.16. The results indicate that if more than 1000 most significant T-F blocks are used for the reconstructing, wavelet impulse synthesis can give reasonably accurate approximations of the original waveform. Since 1000 is very much less than 32768, the compression capability of the wavelet impulse synthesis for piano tones is

potentially very powerful.

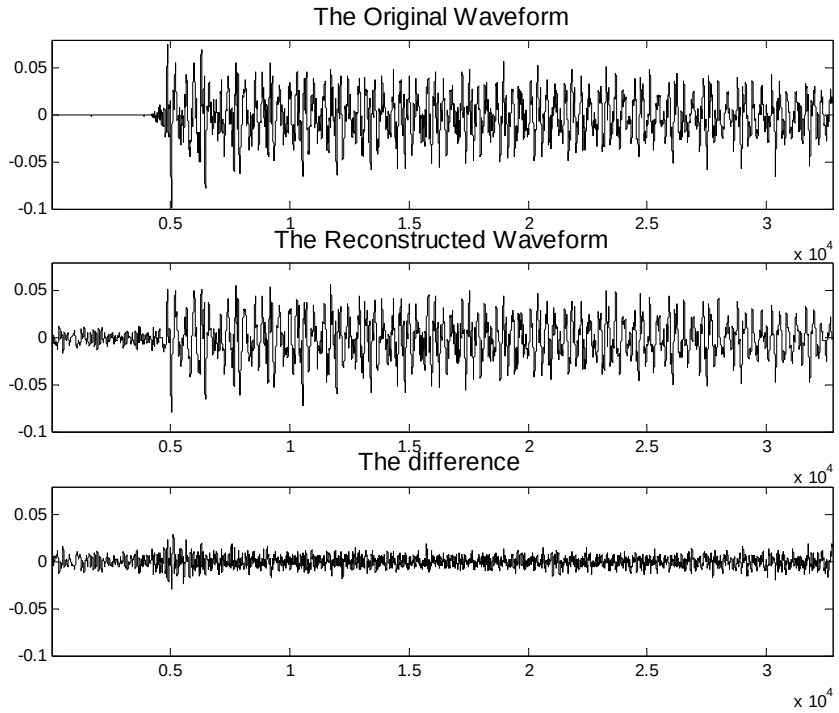


Fig 5.11 Reconstruction of B0 piano tone by 100 most significant T-F blocks

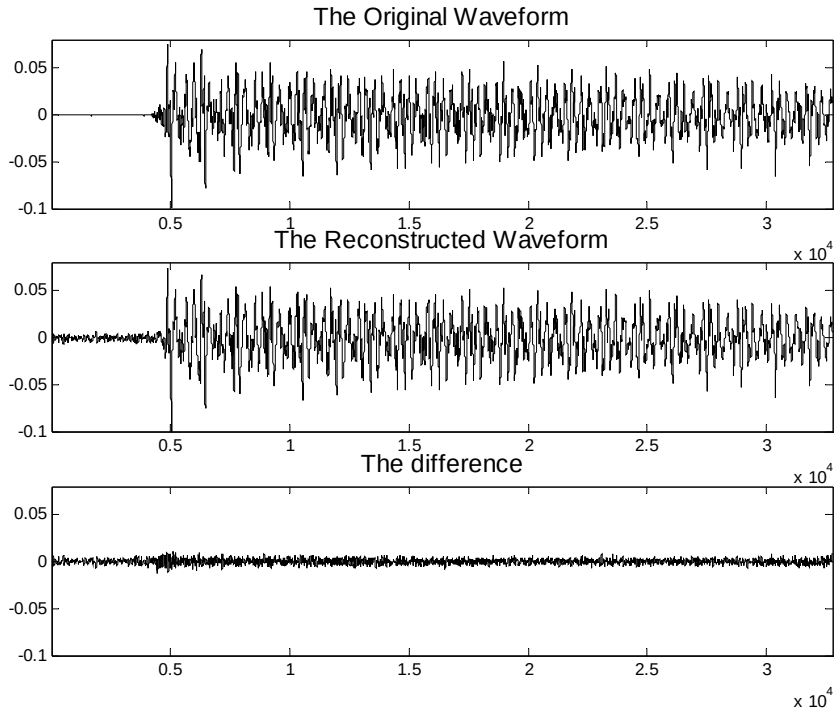


Fig 5.12 Reconstruction of B0 piano tone by 300 most significant T-F blocks

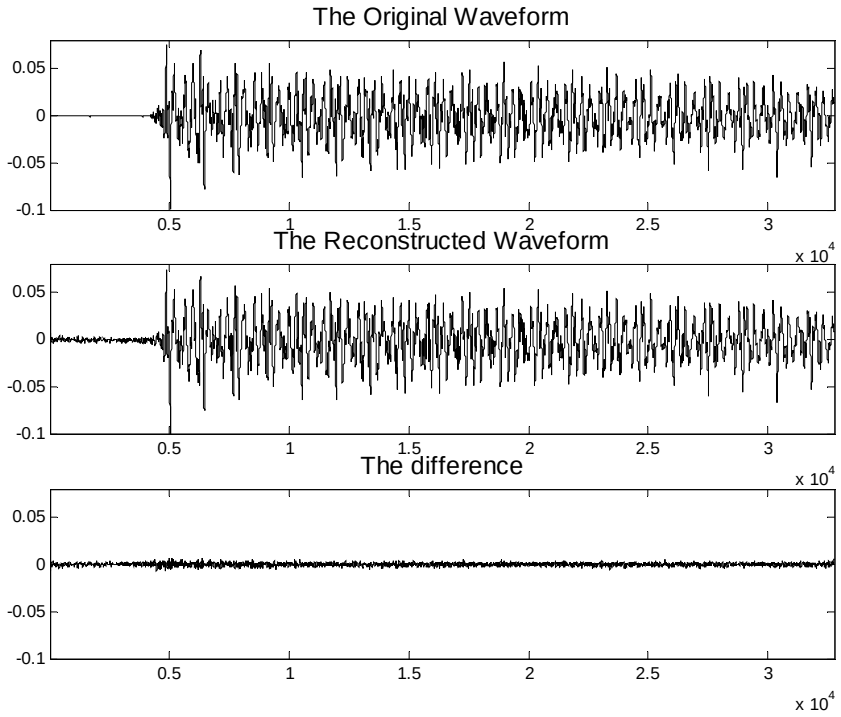


Fig 5.13 Reconstruction of B0 piano tone by 500 most significant T-F blocks

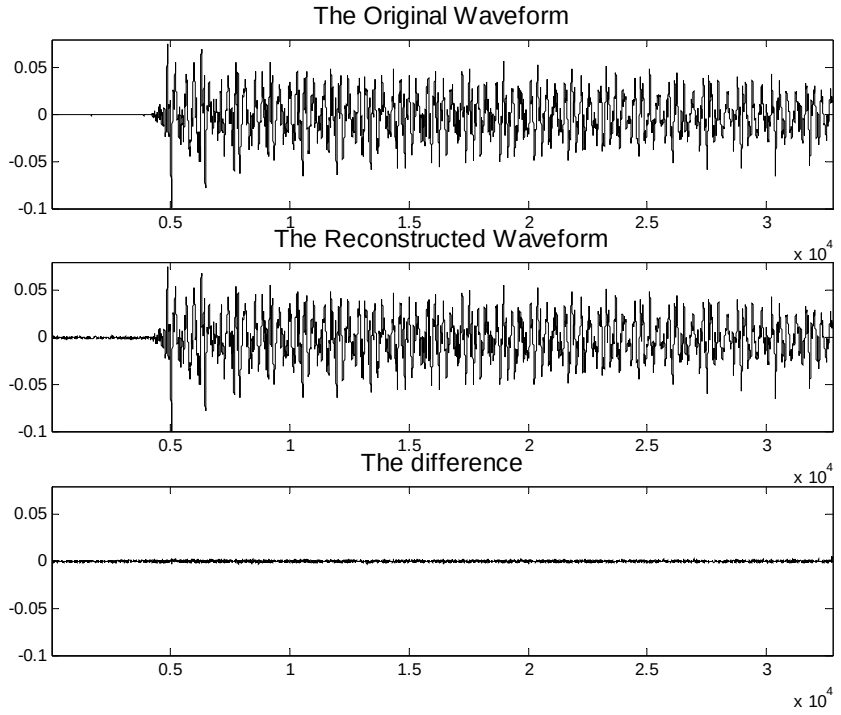


Fig 5.14 Reconstruction of B0 piano tone by 1000 most significant T-F blocks

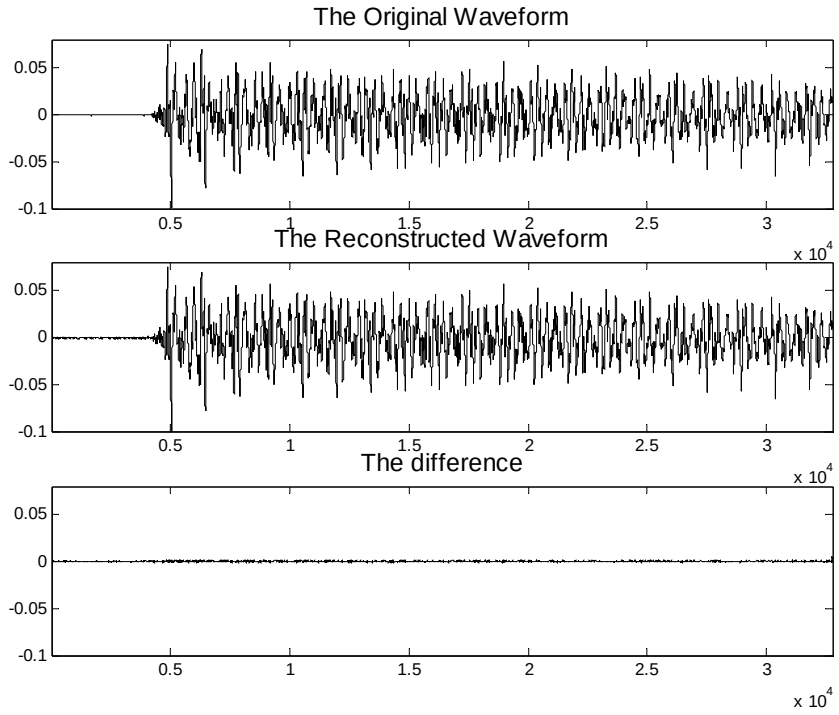


Fig 5.15 Reconstruction of B0 piano tone by 1500 most significant T-F blocks

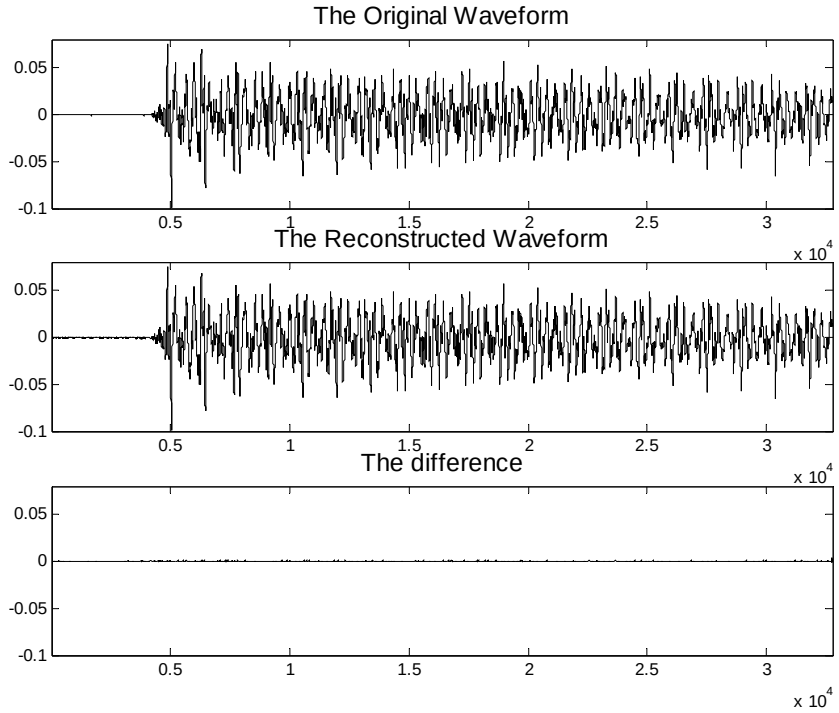


Fig 5.16 Reconstruction of B0 piano tone by 2000 most significant T-F blocks

The high quality of such a reconstruction with relatively few blocks is firstly due to the compression ability of the wavelet transform (using the *best basis algorithm*). Secondly from the perspective of the piano tones' characteristics, we may also get some idea of why the compression can be effective. During the onset transient of a piano tone, little energy is generated in the upper harmonics for each piano sound event, and the peak amplitude is primarily determined by the first few (i.e., the lowest frequency) spectral components [77]. During the stationary part of a piano tone, the high frequency partials decay very quickly, and the energy is still mainly concentrated in the lower harmonics. Hence it is not unreasonable that the inclusion of only the more important coefficients of the wavelet transform can give a good-quality reconstruction.

For further verification of the reconstruction, more experiments have been conducted on a different F1 piano tone and their results are shown in following figures.

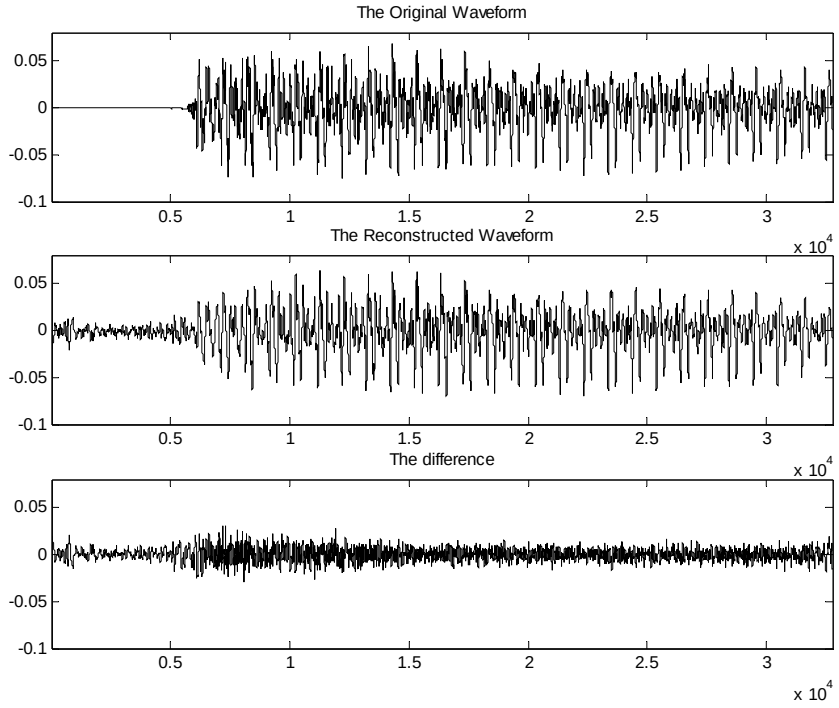


Fig 5.17 Reconstruction of F1 piano tone by 100 most significant T-F blocks

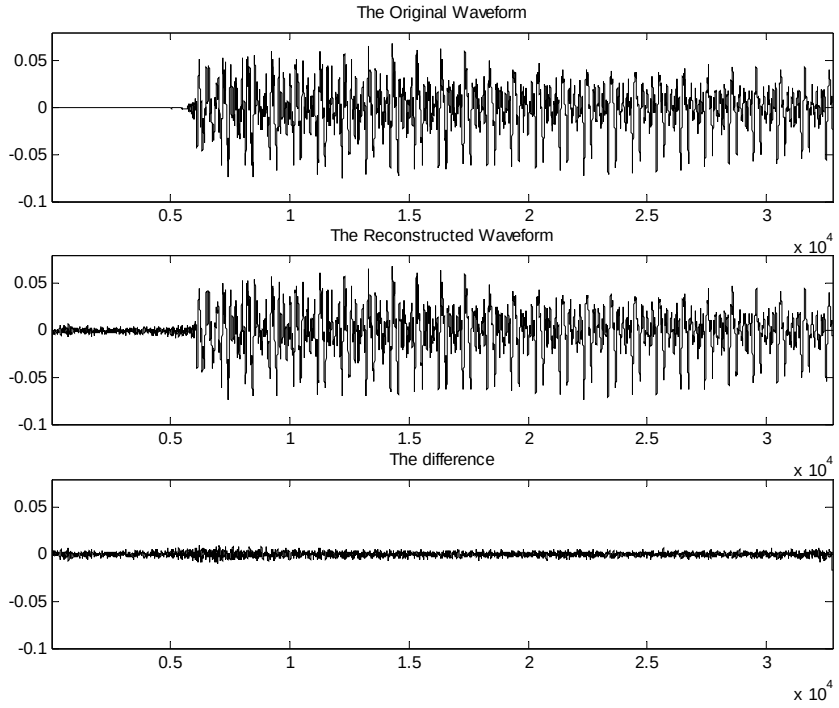


Fig 5.18 Reconstruction of F1 piano tone by 500 most significant T-F blocks

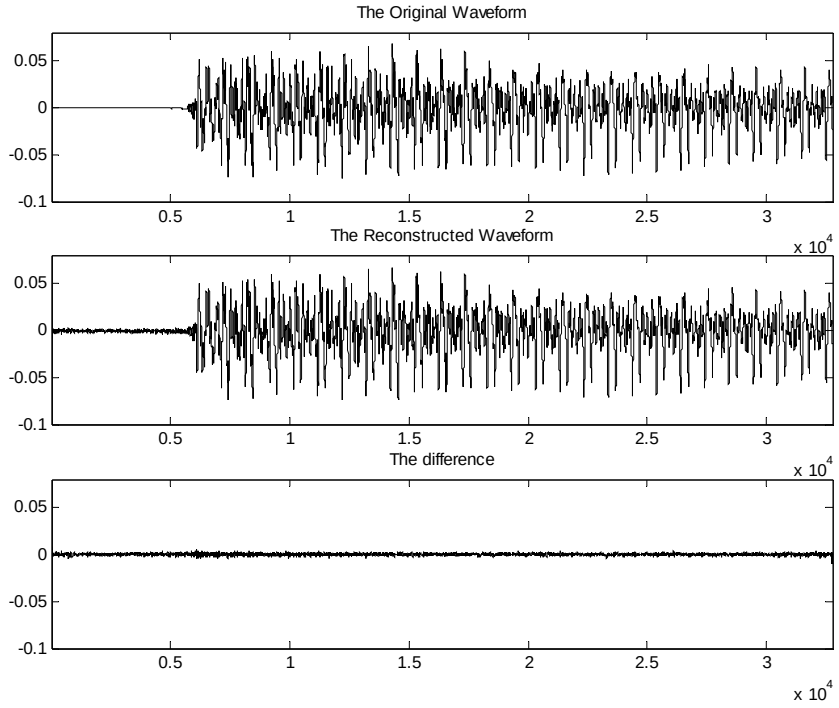


Fig 5.19 Reconstruction of F1 piano tone by 1000 most significant T-F blocks

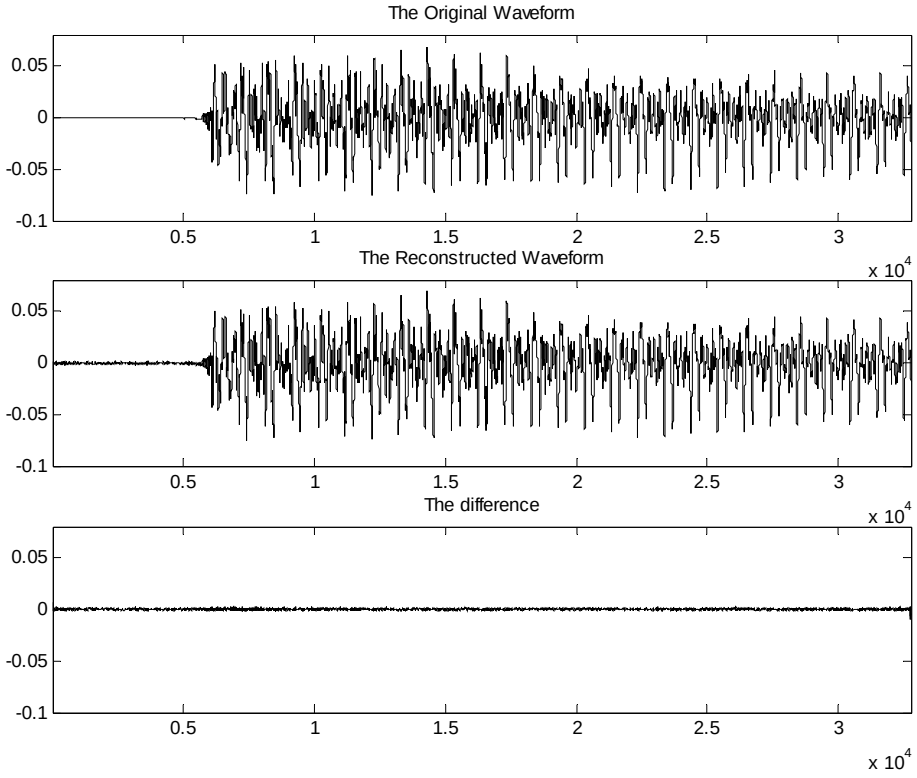


Fig 5.20 Reconstruction of F1 piano tone by 1500 most significant T-F blocks

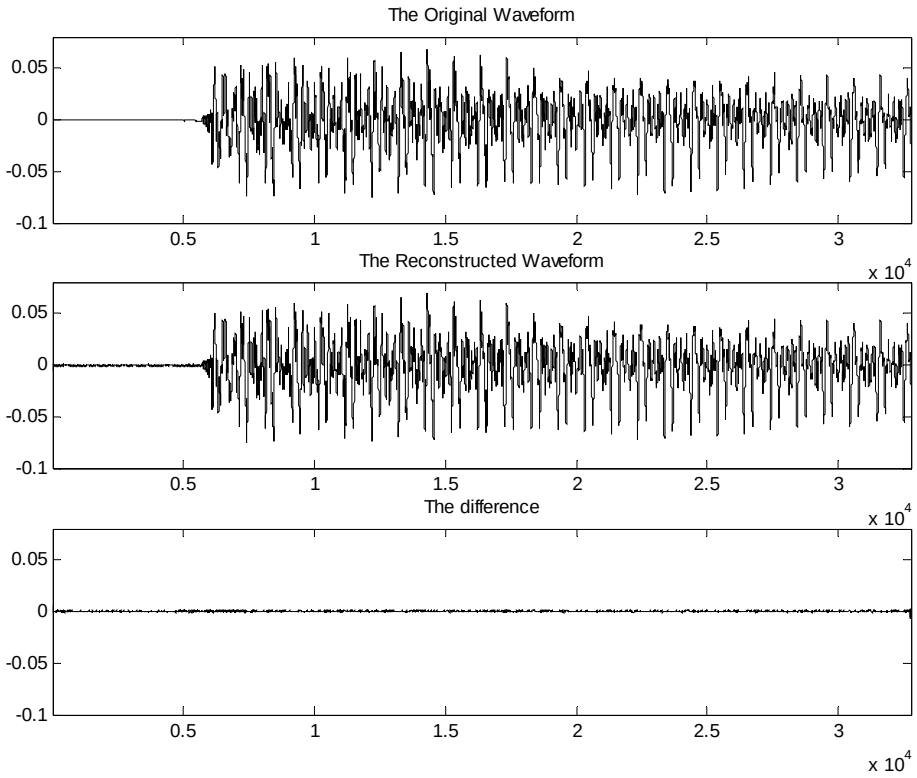


Fig 5.21 Reconstruction of F1 piano tone by 2000 most significant T-F blocks

5.3 A listening test

A listening test was conducted to verify the quality of the reconstruction. The silences at the beginning of both the original and reconstructed sounds were removed because they did not belong to the real parts of these sounds, and to remove any extraneous cues which might help to identify the sounds. The listener was presented with the B0 original tone and the reconstructed one (using only the 1,000 most significant components for reconstruction). Then the two sounds were randomly selected with equal probability for the listening tests.

For the listening test, we used an independent the listener who has been trained

in music. The listener was presented with either the original tone or the reconstructed tone. The listener heard sounds through a sennheiser headphone. The listener could then choose whether he/she thought that the tone presented was the original or reconstructed tones, but an additional choice the listener could make was to indicate whether the choice was one he/she was sure of, or whether it was a guess. Each run consisted to 20 presentations of either the original or reconstructed tones. The results of the listening tests for one musically-trained listener are shown in Table 5.1. There were 10 runs, and the listener's choices were always by guessing, which shows that the close similarity of the original to the reconstructed tone did not allow the listener to be confident about any of the choices. The 10 runs of 20 presentations each meant that there were a total of 200 choices made by the listener. Of these 200 choices, only 108 or 54% were correct, which convincingly shows that the reconstructed tone cannot be easily distinguished from the original tone.

Run No.	No. of choices for Original (Sure)	No. of choices for Original (Guess)	No. of choices for Reconstructed (Sure)	No. of choices for Reconstructed (Guess)
1	0	9 (7:2)	0	11 (7:4)
2	0	14 (6:8)	0	6 (3:3)
3	0	9 (4:5)	0	11 (5:6)
4	0	8 (5:3)	0	12 (6:6)
5	0	15 (7:8)	0	5 (3:2)

6	0	8 (4:4)	0	12 (6:6)
7	0	14 (10:4)	0	6 (4:2)
8	0	9 (7:2)	0	11 (6:5)
9	0	8 (5:3)	0	12 (4:8)
10	0	16 (7:9)	0	4 (2:2)
Total	0	110 (62:48)	0	90 (46:44)

Table 5.1 The results of the listening test for tone B0, where the numbers outside the brackets are the No. of choices by the listener and the number pairs within the bracket show the correction rate expressed by (correct : wrong). For a total of 200 choices (110 original and 90 reconstructed), 108 (62 plus 46) were correct and 92 (48 plus 44) were wrong.

We have therefore shown that by using only the most significant 1,000 components out of the total 32,768 components, we can obtain a reconstruction of the original waveform which is good enough to satisfy a musically trained listener.

Chapter 6 Determining the inharmonicity coefficients for piano tones

In this Chapter, the application of wavelet packet is extended to the measurement of inharmonicity coefficients of piano tones, another essential feature of piano sounds besides their onset transients. Assuming that a piano tone's fundamental frequency is F_1, ideally, the frequencies of all its partials are expected to be harmonic, (i.e., for the *n*th partial, its frequency $F_n = nF_1$). However in the case of real piano strings, due to the inherent stiffness a piano tone's partial frequency is raised by a small amount [78],

$$F_n = nF_1\sqrt{1+Bn^2} \qquad (6\text{-}1)$$

B in the above formula is called the inharmonicity coefficient which is solely determined by each string's material characteristics such as its length, diameter and Young's modulus, etc.

$$B = \frac{\pi^3 E d^4}{64 l^2 \tau} \qquad (6\text{-}2)$$

Here, E :Young's modulus for the string

d : the diameter of the string

l : the string length

τ : the tension

As mentioned in Chapter 1, in many situations, direct measurement of B according to formula (6-2) is either impractical or inconvenient. That is why most researchers try their best to estimate B indirectly from formula (6-1).

Recall from Chapter 1, Galembo and Askenfelt designed an inharmonic comb filter to estimate the inharmonicity coefficient in the frequency domain [64]. Furthermore they also tried pitch extraction techniques such as cepstral analysis and the harmonic product spectrum [66]. Klapuri [67] tackled the inharmonicity measurement by estimating the fundamental frequency in subbands. Rauhala [65] used an iterative process designed to minimize the deviation of the expected partial frequencies compared to the frequencies of the high amplitude peaks in the spectrum.

However, most previous work has been based on Fourier analysis and very few have used wavelet analysis. In this Chapter, a method based on wavelet impulse synthesis is used to estimate the inharmonicity coefficient B from real piano sound samples. Compared to Fourier-based approaches, whose success largely depends on

applying additional optimizing algorithms or signal processing techniques to the Fourier spectrum to 'extract' partials from among frequency peaks clustered together, our wavelet-based method does not require such sophisticated techniques. Moreover, compared to Fourier-based approaches, our wavelet-based method also considers the temporal aspect of each partial's frequency variation.

6.1 Theoretical Preparation

From this section, we will describe the determination of inharmonicity coefficients for piano tones. The basic idea has been stated before -- estimating frequencies of respective partials and subsequently obtaining the inharmonicity coefficients from Fletcher's formula $F_n = nF_1\sqrt{1+Bn^2}$. Here our dataset contains not only the onset transient but also the stationary part of the tone (i.e., we will use the entire waveform of the piano tone). This is for the following three reasons.

Firstly, for good frequency resolution, the number of data points cannot be too small.

Secondly, it may not be possible to determine the partial frequencies of the piano tones only from their onset transients (i.e. the non-steady state of the tone). We also cannot ignore the onset transient because onset transients have rich frequency components and many of them are inharmonic. If ignore this part, we may not guarantee that the dataset is complete especially when our target is to measure all inharmonicity coefficients.

Finally, the inharmonicity coefficients do not change with time although the perceptual effect of inharmonicity is most effective in the initial portion of the tone [79].

Our method for estimating inharmonicity coefficients is mainly built on the concepts of wavelet impulse synthesis and the time-frequency plane which have already been elaborated on in the preceding Chapters. Thus we will not repeat them again. Here we will discuss another important factor that we will not be able to ignore the choice of wavelet bases.

6.1.1 Choice of Wavelet Bases

Choosing the type of wavelet basis (i.e. the quadrature mirror filters h and g) is also an important issue because not all types of wavelet bases are suitable for determining the frequencies of partials. Fig 6.1 shows the Daubechies bases to be quite dispersed in terms of frequency. The Daubechies basis for the coefficient (6, 1, 6) that belong to the 2nd partial of the C4 tone (refer to the time-frequency planes in previous Chapters), has its main frequency peak around 500 Hz, but also has other minor peaks. In contrast, the Battle-Lemarie bases are more localized in both time and frequency (also shown in Fig 6.1). They have a single frequency peak and their width is narrower than that of the main frequency peak of the Daubechies basis, making them more suitable for frequency determination.

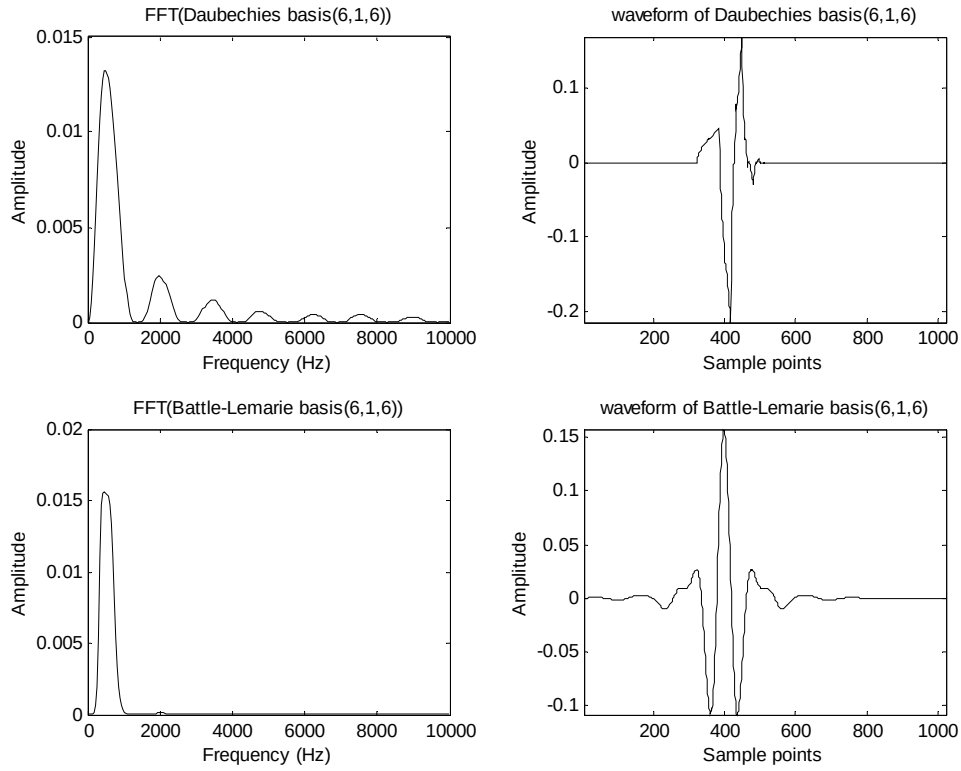


Fig 6.1 Comparison between Daubechies bases (6,1,6) and Battle-Lemarie bases (6,1,6)

If we vary the *k*'s (i.e. the index of within a node) value (recall that the basis is denoted as (*d*, *b*, *k*)) while keeping the *d* and *b* constant, we will get Fig 6.2. Likewise, by only adjusting *b* (the node), we have the basis (6, 0, 6) in Fig 6.3.

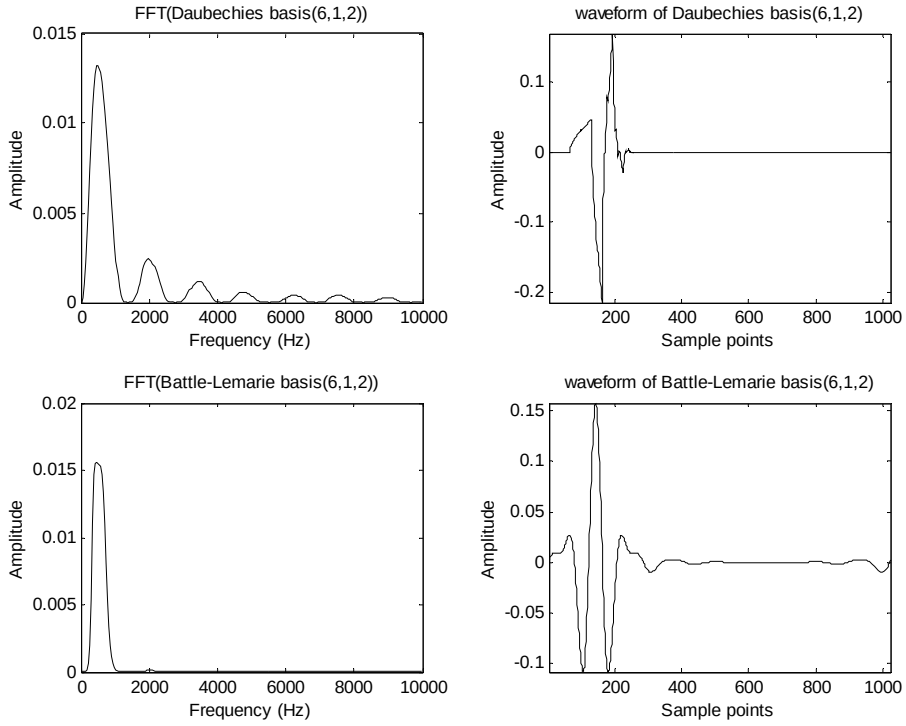


Fig 6.2 Comparison between Daubechies bases (6,1,2) and Battle-Lemarie bases (6,1,2)

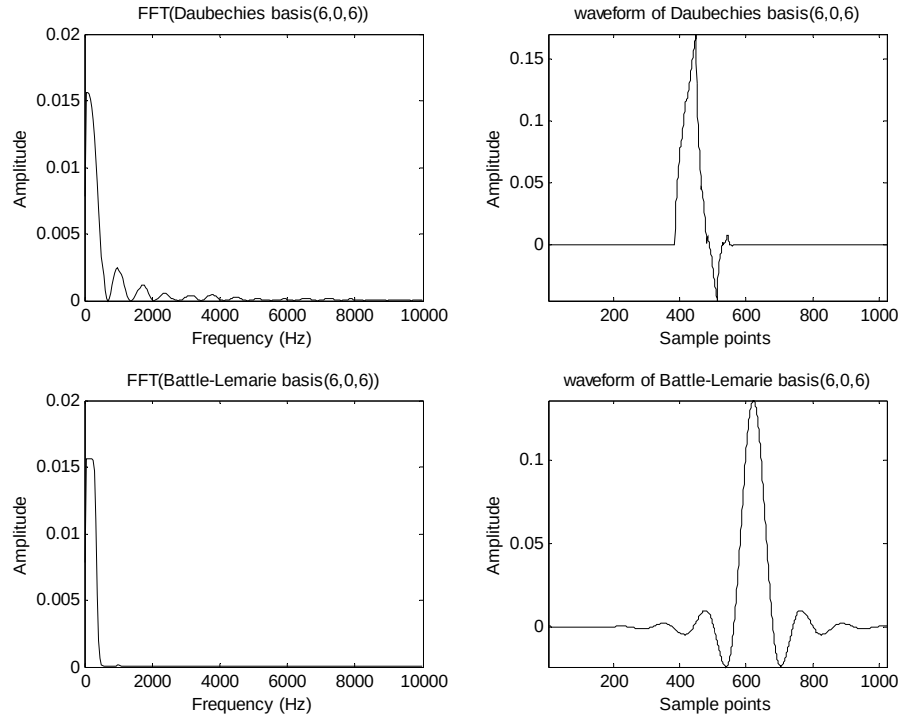


Fig 6.3 Comparison between Daubechies bases (6,0,6) and Battle-Lemarie bases (6,0,6)

From Fig 6.2, we see the frequency structure and value of basis (6, 1, 2). The frequency content of both the Daubechies wavelet and Battle-Lemarie wavelet do not change compared to the corresponding frequency content of basis (6, 1, 6) in Fig 6.1. The waveforms of both bases remain except that the waveform of basis (6, 1, 2) is moved ahead along the time axis. This is in accordance with the translation operation in wavelet theory stated in Chapter 2.

In Fig 6.3, where *b*'s value is changed while *d* and *k* are kept intact, the frequency peak of basis (6, 0, 6) is left-shifted along the frequency axis, which means that the frequency of basis (6, 0, 6) is less than that of basis (6, 1, 6). This is reasonable because the time-frequency plane subband (6,0) (or equivalently node (6,0) in the tree structure) is below subband (6,1). It should be noticed that the shape of basis (6,0,6) is a little different from that of basis (6,1,6). This is because (6,0,6) gives the approximation part represented by the scaling function, while (6,1,6) give the complementary detail part represented by the wavelet function.

Besides the translation effect, we also illustrate the scaling effects in Figs 6.4 and 6.5. That is, we will increase or decrease *d*, the depth, while keeping *b* and *k* unchanged. The scaling effects are clearly seen through the shrinking and expanding waveforms of basis (4, 1, 6) and (7, 1, 6). In Fig 6.4, the amplitude of the second frequency peak of the Daubechies wavelet for (4, 1, 6) is relatively large enough so that we cannot ignore it. In addition, the two frequency peak bands of Daubechies are too broad in Fig 6.4 to be useful for detecting the partial's frequency. Hence we can conclude that for lower values of *d,* the Daubechies wavelet is less suitable as the

analysis wavelet in inharmonicity estimation, as the wavelet frequency bands become too broad.

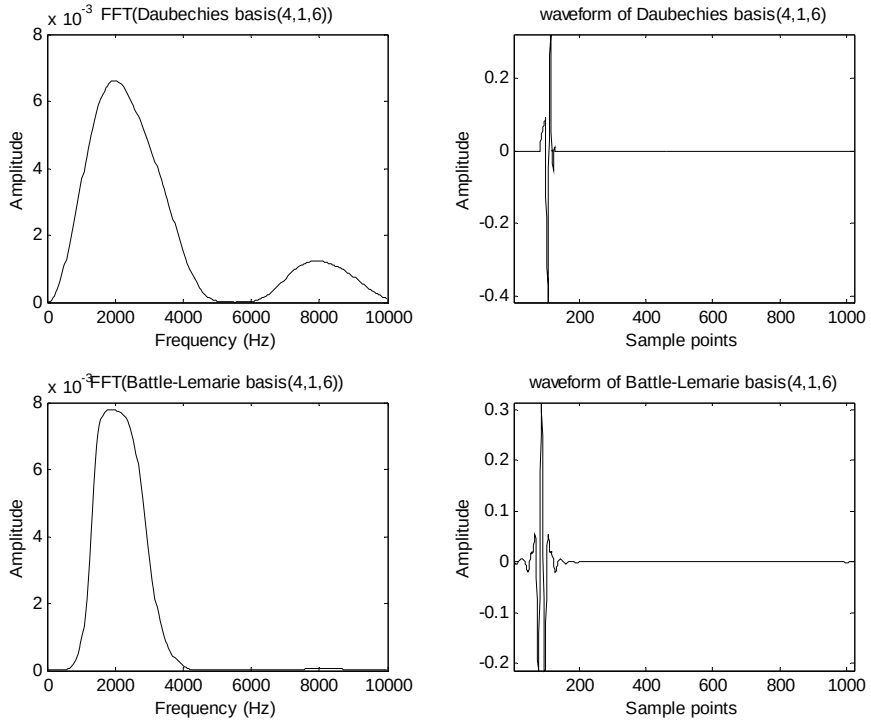


Fig 6.4 Comparison between Daubechies bases (4,1,6) and Battle-Lemarie bases (4,1,6)

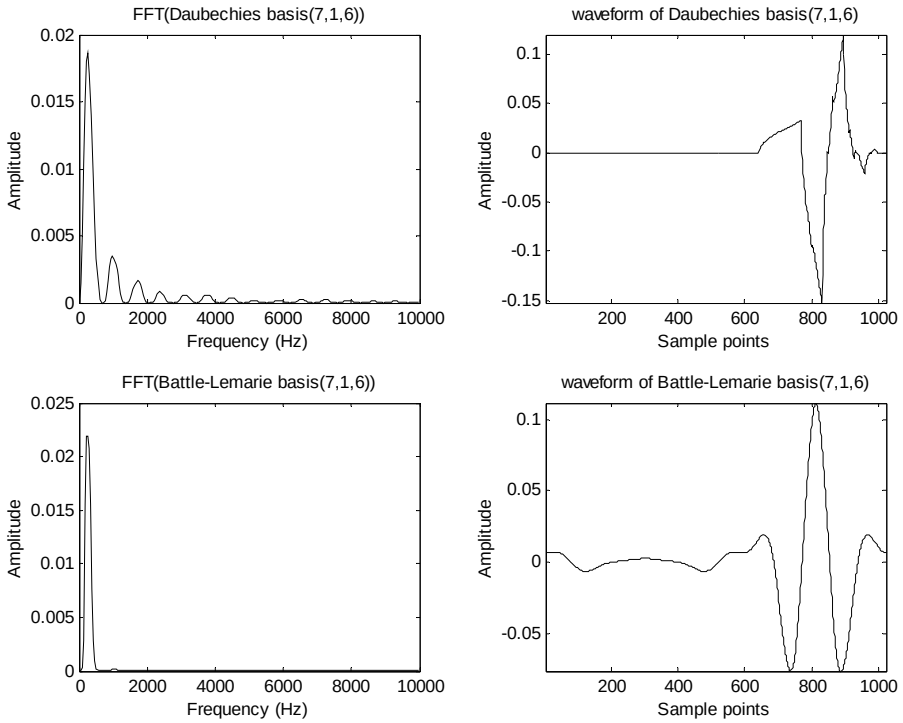


Fig 6.5 Comparison between Daubechies bases (7,1,6) and Battle-Lemarie bases (7,1,6)

6.2 Experiments and Results

Taking a F1 piano tone as an example, the sample length n was set to $2^{15} = 32768$. In our experiments, we selected the m = 1,500 most significant components produced by the WPT using the best basis procedure.

In general, our approach consisted of a rough estimation of the partial frequencies and the inharmonicity coefficients, followed by a correction process that is achieved by a number of iteration loops to optimize the combination of fundamental frequencies and inharmonicity coefficients. The detailed procedures are described as below:

Step 1: The sample is represented by m components (m=1,500 in this example),

with the waveform $W_i(t) = \langle x, \psi_{d,b,k}(t) \rangle \psi_{d,b,k}(t)$ $(1 \le t \le n \quad 1 \le i \le m)$.

Step 2: Each component $W_i(t)$ is then subjected to an FFT operation and the frequency of the peak of the spectrum is labeled for each $W_i(t)$ frequency as f_i. In this step, we simply use the peak frequency value to represent the basis. That is we 'label' this peak as f_i. We will do some optimization in Step 4 where we consider time factor and average many values for f_i according to time.

Step 3: Let jF_0 ($j = 1,2,3,\cdots$) be a series of *analysis frequencies*, where F_0 is the ideal fundamental frequency of F1 (43.65 Hz, using A4=440 Hz and the Equal-tempered scale) (note that the measured fundamental frequency is denoted as F_1). If f_i is within the frequency interval $jF_0 \pm F_0/2$, it is classified as belonging to the j^{th} partial. After such classification, each partial consists of a few (e.g., *u*) frequency components (f_i) distributed around its *analysis frequency* jF_0.

Step 4: For a piano tone with inharmonic partials, we assume that the j^{th} partial's frequency has the value f_i with a probability of $p_i(t) = \dfrac{W_i^2(t)}{\sum_i W_i^2(t)} (1 \le t \le n \quad 1 \le i \le u)$ and the expected frequency of the j^{th} partial at time *t* is thus given by $F_j(t) = \sum_i p_i(t) f_i$. The function $F_j(t)\,(1 \le t \le n)$ describes the frequency variation of the j^{th} partial with time. In $p_i(t) = \dfrac{W_i^2(t)}{\sum_i W_i^2(t)}$, the sum

symbol is in terms of *i*. That means we sum different time frequency blocks at the cross-section (or at a certain time point *t*).

However, another interpretation of $F_j(t)$ is that in order to determine the partial frequency, we have done *n* trials at different time points to measure its value and each trial's result is $F_j(t)\,(1 \le t \le n)$ respectively. Therefore, the average of $F_j(t)$ according to *t* can reasonably be considered the measured value of this partial.

Table 6.1 presents the measured frequency values of some partials for tone F1, where the dashed line '-' means that the corresponding partial is not observed. Only some partials are listed here to save space.

Partial No.	Freq (Hz)	Partial No.	Freq (Hz)	Partial No.	Freq (Hz)	Partial No.	Freq (Hz)
1	47.17	31	1348.6	61	2672.0	91	3951.3
2	86.23	32	1402.2	62	2718.1	92	-
3	128.88	33	1437.6	63	2738.8	93	-
4	172.00	34	1470.1	64	2788.6	94	-
5	215.52	35	1523.2	65	2839.6	95	-
6	258.81	36	1577.4	66	2897.8	96	-
7	302.97	37	1626.2	67	2919.2	97	-
8	344.57	38	1666.6	68	2958.9	98	-
9	390.27	39	1716.6	69	3007.1	99	-
10	432.81	40	1730.9	70	3064.7	100	-

11	478.28	41	1774.2	71	3108.3	101	-
12	518.82	42	1823.9	72	3129.0	102	-
13	566.16	43	1883.8	73	3180.7	103	-
14	609.11	44	1909.6	74	3238.1	104	-
15	653.45	45	1977.2	75	3280.5	105	-
16	698.63	46	1991.8	76	3305.9	106	4625.6
17	724.99	47	2056.8	77	-	107	-
18	777.55	48	2077.0	78	-	108	-
19	830.85	49	2132.3	79	-	109	-
20	879.21	50	2179.9	80	-	110	-
21	924.36	51	2236.4	81	3542.2	111	-
22	967.28	52	2281.7	82	-	112	-
23	1013.9	53	2309.4	83	-	113	-
24	1057.4	54	2351.0	84	3666.0	114	-
25	1099.9	55	-	85	-	115	-
26	1146.5	56	2444.8	86	-	116	-
27	1192.3	57	2498.6	87	-	117	-
28	1217.0	58	2552.7	88	-	118	-
29	1249.7	59	2561.4	89	-	119	-
30	1304.4	60	2610.1	90	3950.0	120	-

Table 6.1 Frequencies of some partials of F1 piano tone after rough estimation

Step 5: Once the frequencies of the partials are obtained, the estimation process

proceeds to calculate the inharmonicity coefficient *B*. Substituting any two partial frequencies (e.g., F_p and F_q) into Fletcher's equation $F_n = nF_1\sqrt{1+Bn^2}$, the two unknown variables, F_1 and *B* can be easily solved from those two equations. For further simplification, Fletcher also derived formula (6-3) based on the fact that Bn^2 is always very small [78].

$$F_n - nF_1 = BF_1n^3/2 \qquad (6\text{-}3)$$

Fletcher selected frequencies F_p and F_q (in Fletcher's paper, partial frequencies are obtained by manual measurement) such as $q = 2p$, reducing the solution of F_1 and *B* to:

$$F_1 = (8F_p - F_{2p})/6p \qquad (6\text{-}4)$$

$$B = (2/p^2)[(r-2)/(8-r)] \qquad (6\text{-}5)$$

where $r = F_{2p}/F_p$.

By repeatedly substituting the partial frequencies (Table 6.1) into Equations (6-4) and (6-5), a number of possible values for F_1 and *B* are obtained as in Table 6.2. The absolute value of *B* in each case is then obtained and the final average values for F_1 and $|B|$ are 43.55 Hz and 0.000028 respectively.

Partial No.	F_1 (Hz)	*B*	Partial No.	F_1 (Hz)	*B*
4 and 8	42.975	0.000070	23 and 46	44.342	-0.000022
5 and 10	43.045	0.000110	24 and 48	44.323	-0.000021
6 and 12	43.103	0.000043	25 and 50	44.131	-0.000010

7 and 14	43.206	0.000071	26 and 52	44.167	-0.000005
8 and 16	42.874	0.000144	27 and 54	44.365	-0.000013
9 and 18	43.419	-0.000032	28 and 56	43.401	0.000004
10 and 20	43.055	0.000105	29 and 58	42.787	0.000017
11 and 22	43.318	0.000062	30 and 60	43.472	0.000000
12 and 24	42.960	0.000089	31 and 62	43.391	0.000005
13 and 26	43.369	0.000050	32 and 64	43.900	-0.000004
14 and 28	43.522	-0.000003	33 and 66	43.449	0.000005
15 and 30	43.592	-0.000006	34 and 68	43.146	0.000004
16 and 32	43.613	0.000009	35 and 70	43.433	0.000003
17 and 34	42.449	0.000032	36 and 72	43.936	-0.000004
18 and 36	42.991	0.000030	37 and 74	44.015	-0.000002
19 and 38	43.686	0.000005	38 and 76	43.978	-0.000004
20 and 40	44.190	-0.000026	42 and 84	43.354	0.000002
21 and 42	44.214	-0.000020	45 and 90	43.955	-0.000000
22 and 44	44.157	-0.000018	53 and 106	43.553	0.000000
Average	F1	|B|			
	43.55	0.000028			

Table 6.2 F_1 and B for a F1 piano tone

It was observed that quite a few *B* values are negative in Table 6.2. In these cases, the observed partial frequencies are lower than their corresponding values in the

harmonic series, which contradicts the expectation that inharmonicity increases in each partial frequency compared to the harmonic series. The measured partial frequencies in Table 6.2 also do not seem to fit the curve of $F_n = nF_1\sqrt{1+Bn^2}$ with F_1=43.55 Hz and B= 0.000028 (see Fig 6.6).

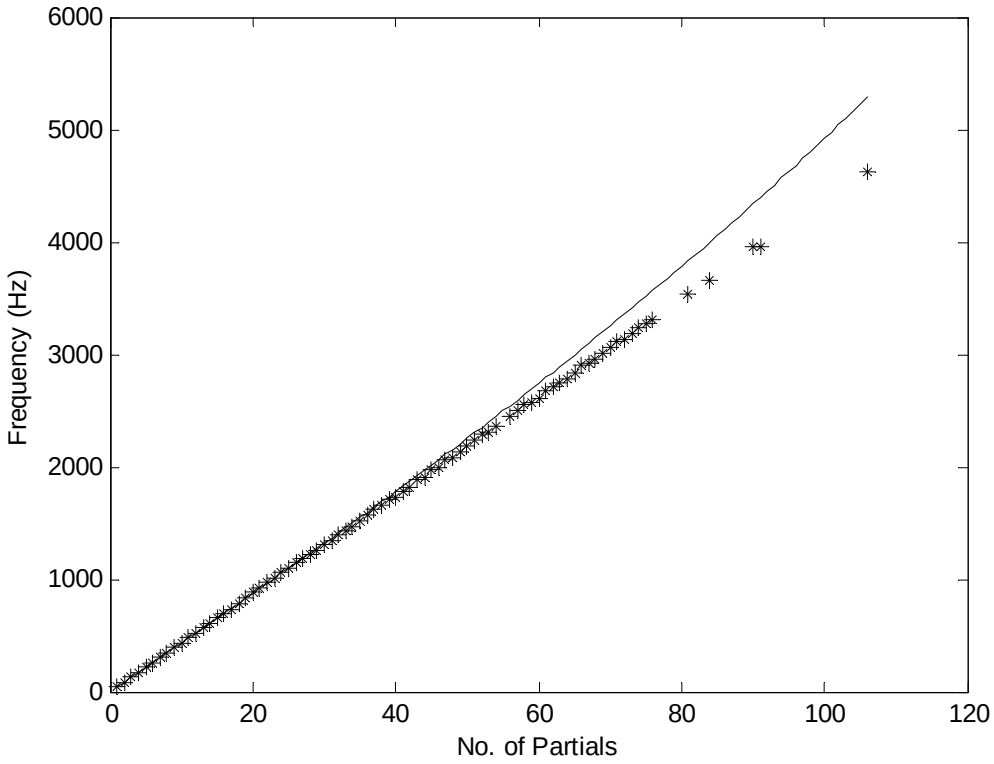


Fig 6.6 The result of rough estimation: the expected curve $nF_1\sqrt{1+Bn^2}$ *vs* measured partial frequencies

The shift in the partials can account for above problems. Suppose that there is a component $W_i(t)$ with frequency 1,831 Hz. Given that 43.65 Hz is F1's ideal fundamental frequency, if we assume no inharmonicity, this component is counted as belonging to the 42th partial in Step 3 above, and is used to estimate the 42th partial's frequency in Step 4. However, in the real situation where inharmonicity (e.g. for

B=0.000028) is considered, 1,831 Hz is actually the frequency of the 41st partial, calculated by the inharmonicity formula $F_n = nF_1\sqrt{1+Bn^2}$. Thus in the above rough estimation, many components have been erroneously classified, causing errors in all the succeeding computations.

To eliminate the impact of this 'partial shift' effect and refine the results based on the rough estimation, a correction is thus required. Starting from $F_1 = 43.55$ Hz and *B*=0.000028, the series of harmonic *analysis frequencies*, jF_0 in Step 3 is replaced by a new series of inharmonic *analysis frequencies* $jF_1\sqrt{1+Bj^2}$, while keeping the computations in Steps 4 and 5 unchanged. But the results in Table 6.1 and Table 6.2 will be upgraded to reflect the change due to this correction. More iteration loops are needed to fully correct the 'partial shift' effect. The values of F_1 and *B* obtained from the previous iteration are adopted to establish new *analysis frequencies* for the current iteration in Step 3. Following Steps 4 and 5, the new combination of F_1 and *B* is used in the next iteration and so on.

Additionally, in Step 3 of each iteration, the frequency range covering a partial should no longer be $jF_0 \pm F_0/2$ or $jF_1 \pm F_1/2$, because in an inharmonic tone, the interval between neighboring partials is broader and broader as we go up to higher partials. Thus, the frequency range of partial *n* should be $[F_n - (F_n - F_{n-1})/2$, $F_n + (F_{n+1} - F_n)/2]$ where $F_j = jF_1\sqrt{1+Bj^2}$. Otherwise, there is a frequency gap between neighboring partials and some blocks would be missed.

In Table 6.2, we observe that *B* values are getting smaller as higher partials are involved to calculate the *B* values. Moreover, nearly all *B* values calculated from the

lower partials, i.e. from the pair 4 and 5 to pair 13 and 26, are positive, while negative *B* values frequently appear for higher partial pairs. This may because in the rough estimation, we use the harmonic analysis frequencies to fit the distribution of time-frequency blocks. For the lower partials, the deviation from exact harmonic frequencies is still small, and therefore the assignment of the frequency components is largely to the correct partials. In other words, the interval $jF_0 \pm F_0/2$ is not too different from the range $[F_n - (F_n - F_{n-1})/2$, $F_n + (F_{n+1} - F_n)/2]$ where $F_j = jF_1\sqrt{1+Bj^2}$, since j^2 is small. As j increases, the previously mentioned 'shift effect' becomes larger. Many frequency components are thus wrongly assigned to higher partials than they should be assigned to, which results in negative *B* values as calculated by partial pairs with higher *j* values.

We have used the absolute values of these negative *B*s to obtain only positive inharmonicity coefficients for the first iteration. From the Table 6.3 where the output of the first iteration is shown, most *B* values from higher pairs have become positive now. Irrespective of the initial values, the B values from the first iteration have now become positive so that a few more iteration would fully eliminate the 'shift effect'.

Partial No.	F_1 (Hz)	*B*	Partial No.	F_1 (Hz)	*B*
4 and 8	42.975	0.000069	23 and 46	43.726	0.000031
5 and 10	43.045	0.00011	24 and 48	43.607	0.000036
6 and 12	43.106	0.000039	25 and 50	43.438	0.000041
7 and 14	43.204	0.000072	26 and 52	43.588	0.000036
8 and 16	42.872	0.000144	27 and 54	43.676	0.000041

9 and 18	43.405	-0.000024	28 and 56	43.554	0.000043
10 and 20	43.052	0.000106	29 and 58	43.887	0.000034
11 and 22	43.313	0.000064	30 and 60	43.431	0.000039
12 and 24	42.948	0.00009	31 and 62	43.698	0.000038
13 and 26	43.357	0.000052	32 and 64	43.694	0.000037
14 and 28	43.249	0.000061	33 and 66	43.011	0.000045
15 and 30	43.354	0.000043	34 and 68	43.648	0.000039
16 and 32	43.38	0.000051	35 and 70	44.169	0.000031
17 and 34	42.017	0.000108	36 and 72		
18 and 36	42.597	0.000093	37 and 74		
19 and 38	43.256	0.000061	38 and 76	44.143	0.000032
20 and 40	43.832	0.000016	39 and 78		
21 and 42	43.747	0.00003	40 and 80	42.71	0.000049
22 and 44	43.62	0.000034			

Table 6.3 The first iteration: absolute value operation applied

Taking absolute values is not the only possible approach. Another approach is to keep these negative B values intact. We note that most of the *B* values from lower partial pairs (Table 6.2) are not too far from the final iterative value of 0.000069 (Table 6.5). But on the contrary, it is noticeable that from the pair 14 and 28 upwards, the initial *B* values are generally very much, since they have been derived on the assumption that the partials are harmonic.(Table 6.2). If we include their *B* values in calculating the average values of *B*, these *B* values from the higher pairs of partials play a much less important role than lower pairs. Hence, the positive *B* values from

lower pairs will dominate the result. That means it doesn't matter whether we need to take the absolute value when we encounter negative *B* for the initial values. In Table 6.4 after the first iteration, most *B* values become positive even if we include the negative *B* values.

Partial No.	F_1 (Hz)	*B*	Partial No.	F_1 (Hz)	*B*
4 and 8	42.975	0.000069	23 and 46	43.894	0.000016
5 and 10	43.045	0.00011	24 and 48	43.939	0.00001
6 and 12	43.103	0.000043	25 and 50	43.767	0.000017
7 and 14	43.204	0.000072	26 and 52	43.856	0.000018
8 and 16	42.872	0.000144	27 and 54	43.912	0.000017
9 and 18	43.419	-0.000032	28 and 56	43.964	0.000009
10 and 20	43.055	0.000105	29 and 58	44.195	0.000009
11 and 22	43.316	0.000063	30 and 60	43.539	0.000014
12 and 24	42.96	0.000089	31 and 62	43.11	0.000023
13 and 26	43.36	0.000052	32 and 64	43.404	0.000022
14 and 28	43.301	0.000049	33 and 66	43.258	0.00002
15 and 30	43.479	0.000017	34 and 68	43.632	0.000015
16 and 32	43.59	0.000013	35 and 70	43.503	0.000018
17 and 34	42.195	0.000074	36 and 72	43.292	0.00002
18 and 36	42.976	0.000032	37 and 74		
19 and 38	43.611	0.000015	38 and 76		
20 and 40	43.836	0.000014	39 and 78		
21 and 42	43.842	0.000018	40 and 80	43.837	0.000014

22 and 44 43.818 0.000015

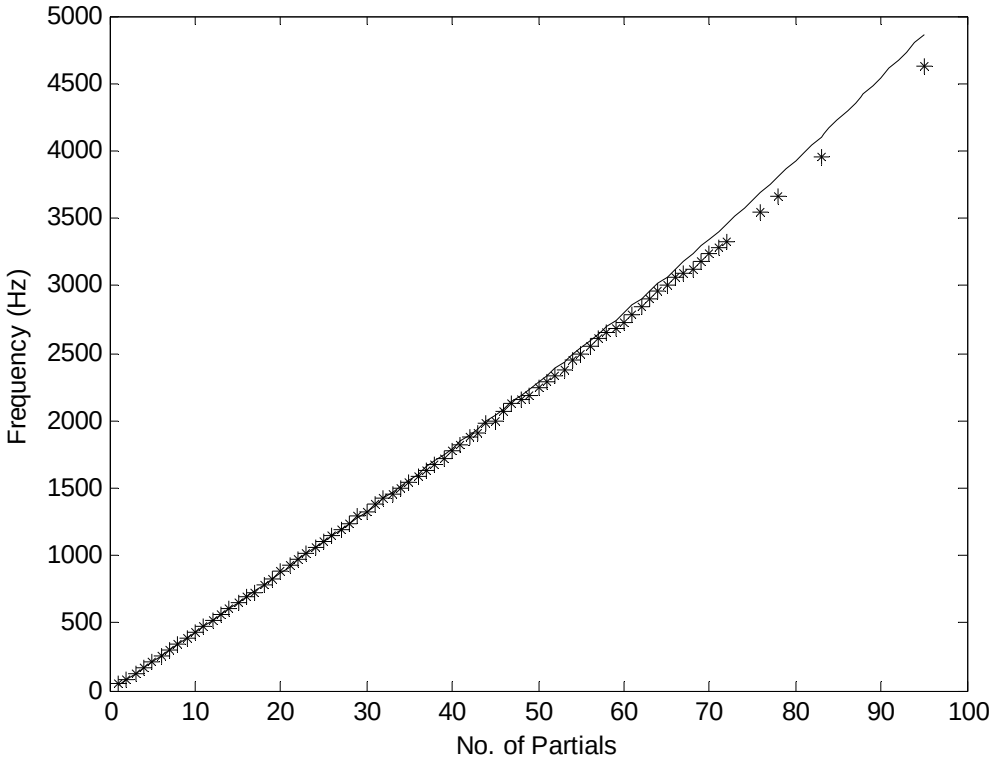
Table 6.4 The first iteration: nothing has been done on negative *B* values in the rough estimate

After 11 iterations (absolute value operation applied), the evolution of F_1 and *B* is summarized in Table 6.5.

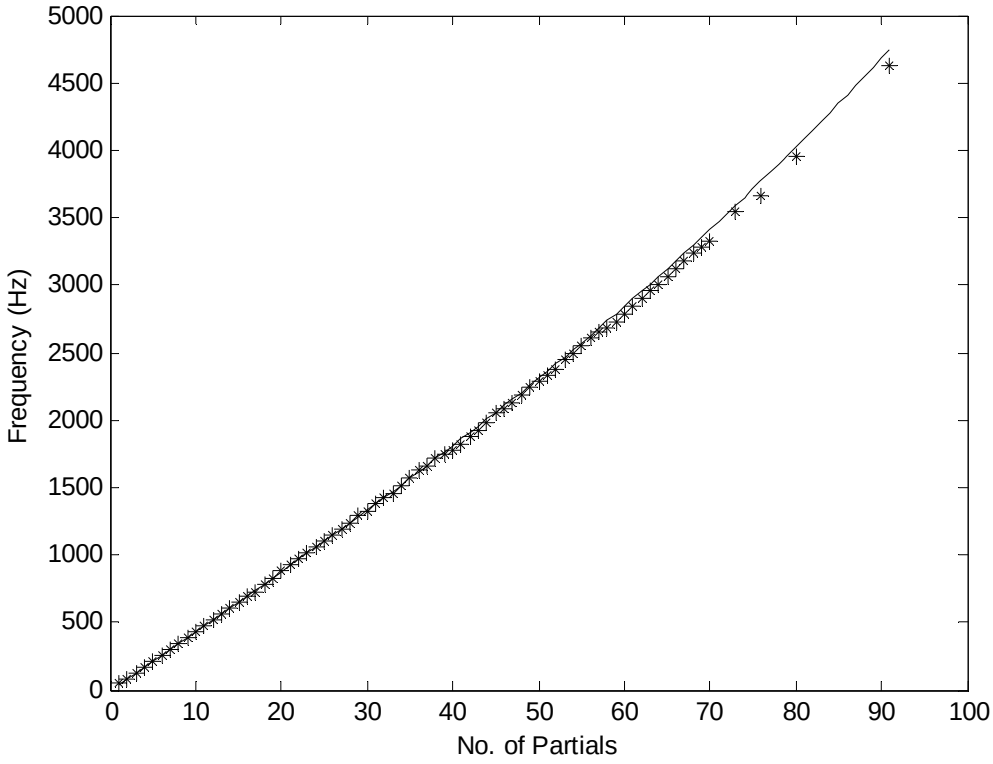
	F	B
Rough Estimate	43.55	0.000028
1st iteration	43.42	0.000043
2nd iteration	43.36	0.000054
3rd iteration	43.29	0.000062
4th iteration	43.22	0.000068
5th iteration	43.16	0.000069
6th iteration	43.16	0.000069

Table 6.5 F_1 and B calculated from rough estimation to the 6th iteration

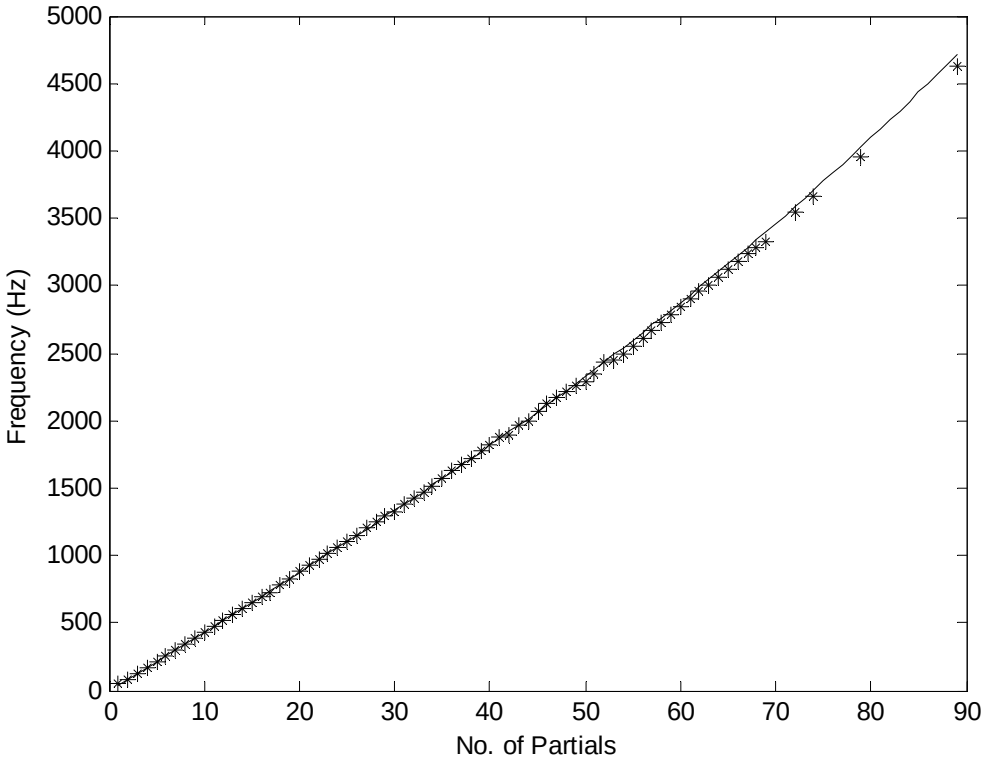
By using F_1 and B at the end of each iteration, the inharmonicity curve can be obtained by $F_n = nF_1\sqrt{1+Bn^2}$. Plotting the measured partial frequencies of each iteration(the upgraded Table 6.1 in each iteration) together with the inharmonicity curve in the same diagram, the correction process is illustrated in Fig 6.7, compared to the rough estimation in Fig 6.6.



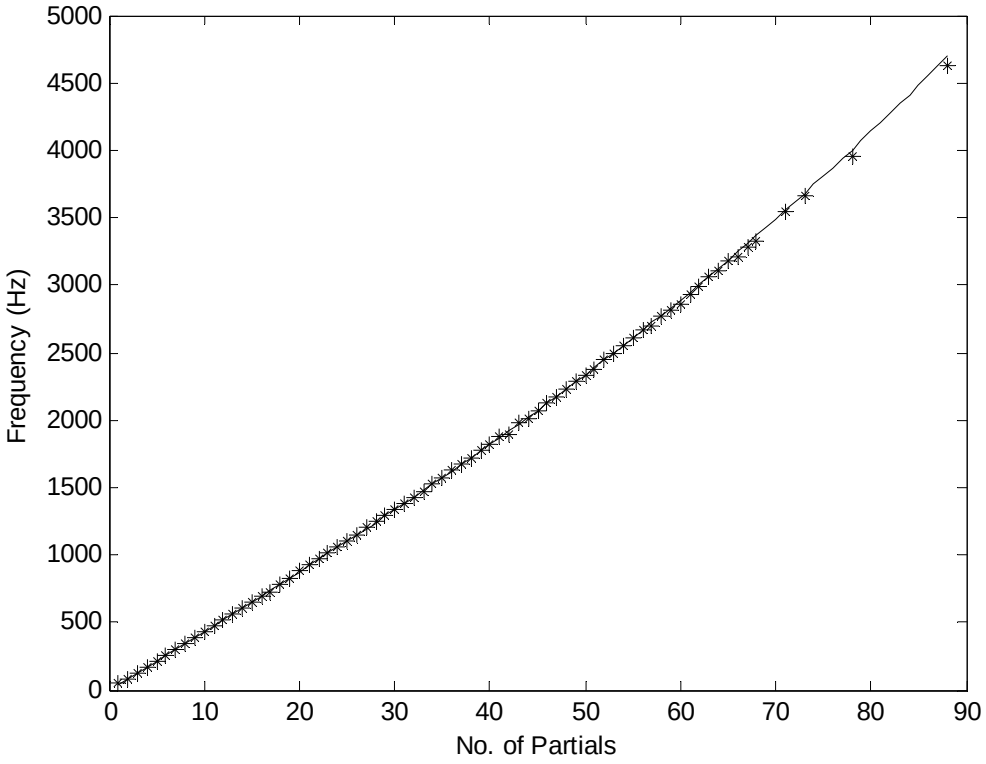
(a) The 1st iteration



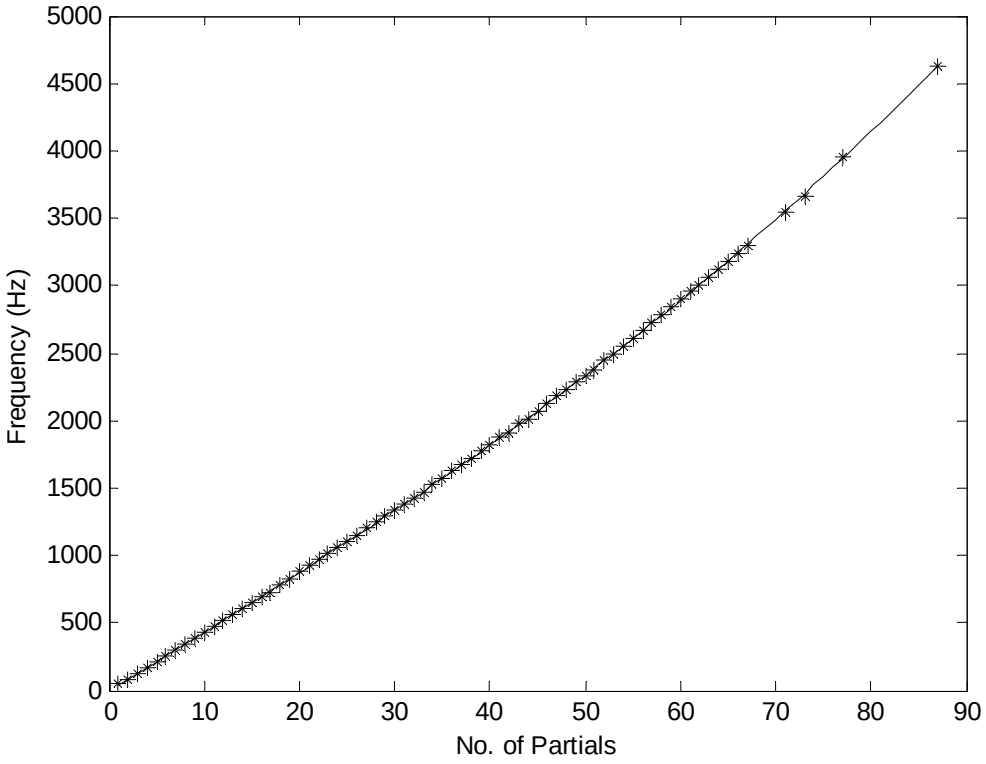
(b) The 2nd iteration



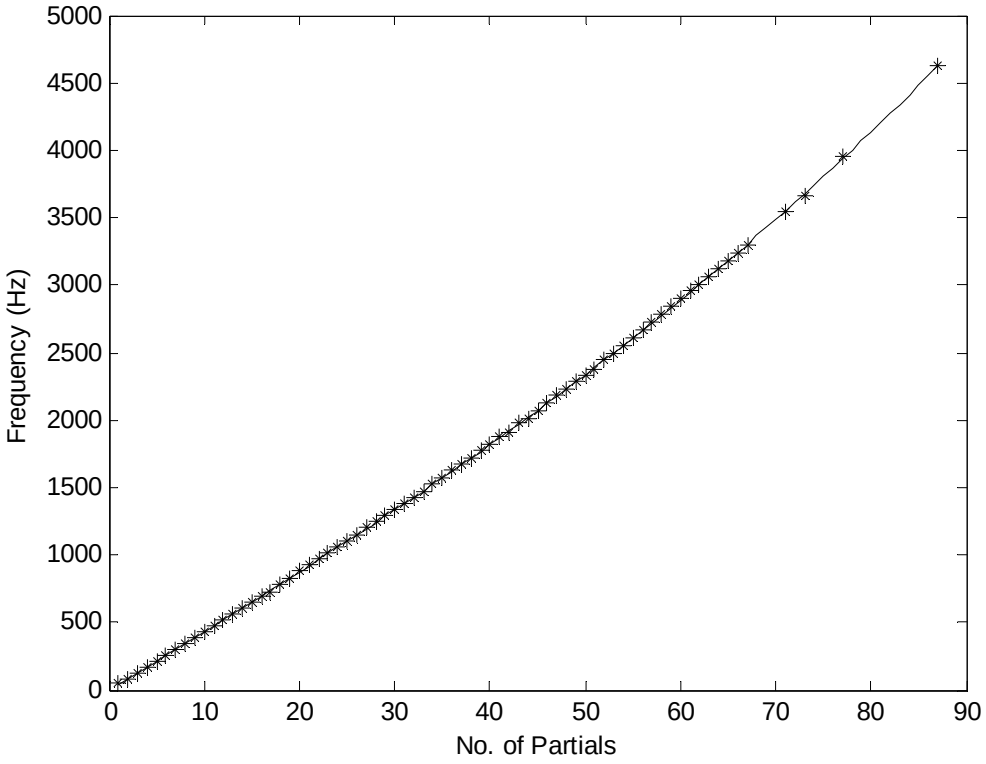
(c) The 3rd iteration



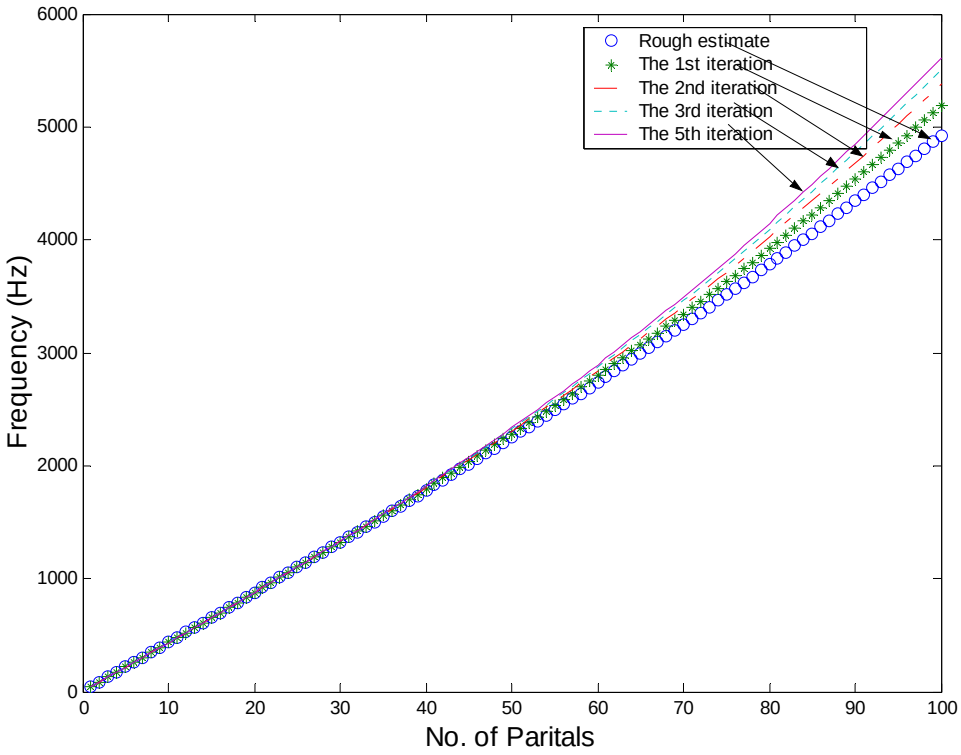
(d) The 4th iteration



(e) The 5th iteration



(f) The 6th iteration



(g) Fletcher curves

Fig 6.7 Results of the 6-iteration correction process for F1 piano tone

It can be seen from both Table 6.5 and Fig 6.7 that the results become stable from the 10^{th} iteration, implying that the 'partial shift' effect has been eliminated. Hence, it is appropriate to take $F_1 = 43.16$ and $B = 0.000069$ as our final estimate for the F1 piano tone. The technique introduced in this Chapter has been published in the Journal of the Audio Engineering Society [80].

In Table 6.1, the initially estimated fundamental frequency is 47.17, significant different from the ideal fundamental frequency of piano tone F1 (43.65 Hz). But the final calculated fundamental frequency is 43.16 in Table 6.2 which is more reasonable. The possible reason may be that usually lowest partials contain more number of broader time-frequency blocks that have wider frequency peaks. If $F_0 + F_0/2$ is used as the boundary of the partial one and partial two, the wide peaks may span

across the boundary line $F_0 + F_0/2$. For bass tones, the frequency interval between neighbor partials is very small. These wide peaks actually contribute to both partials. However, we classify them to either partial one or partial two. This makes the calculated fundamental frequency unreliable. Therefore in Step 5, we ignore lowest partials and start from the partial four as shown in Table 6.2.

To check whether our estimation of inharmonicity is correct, we applied the Fast Fourier Transform (FFT) on the sample of F1 piano tone. We also substituted $F_1 = 43.16$ and $B = 0.000069$ in Equation (6-1) and predicted some partial frequencies. Plotting these partial frequencies as a series of frequency lines in the FFT spectrum (Fig 6.8), we can find out how much our prediction matches the reality. In addition, to indicate that piano tones really have an inharmonicity effect, we also combine the FFT spectrum and the harmonic frequency lines for partials which are multiples of the fundamental frequency in Fig 6.9.

From Fig 6.8, we can see that our predicted inharmonic structure matches the FFT spectrum well. However in Fig 6.9, most upper partials can not fall on harmonic lines, although for those lower partials, there is a good match. This indicates how efficiently the formula $F_n = nF_1\sqrt{1+Bn^2}$ works. Since the inharmonicity coefficient B is generally very small (e.g. in our measurement of the F1 piano tone, $B = 0.000069$), when n is also very small, the item Bn^2 can be ignored. Hence, there is nearly no difference between an inharmonic pattern and a harmonic structure. But when n goes to a large number, the square operation makes Bn^2 much larger than 1. For such large n, the inharmonicity item Bn^2 will impose a significant

influence on partial frequencies with the help of n^2.

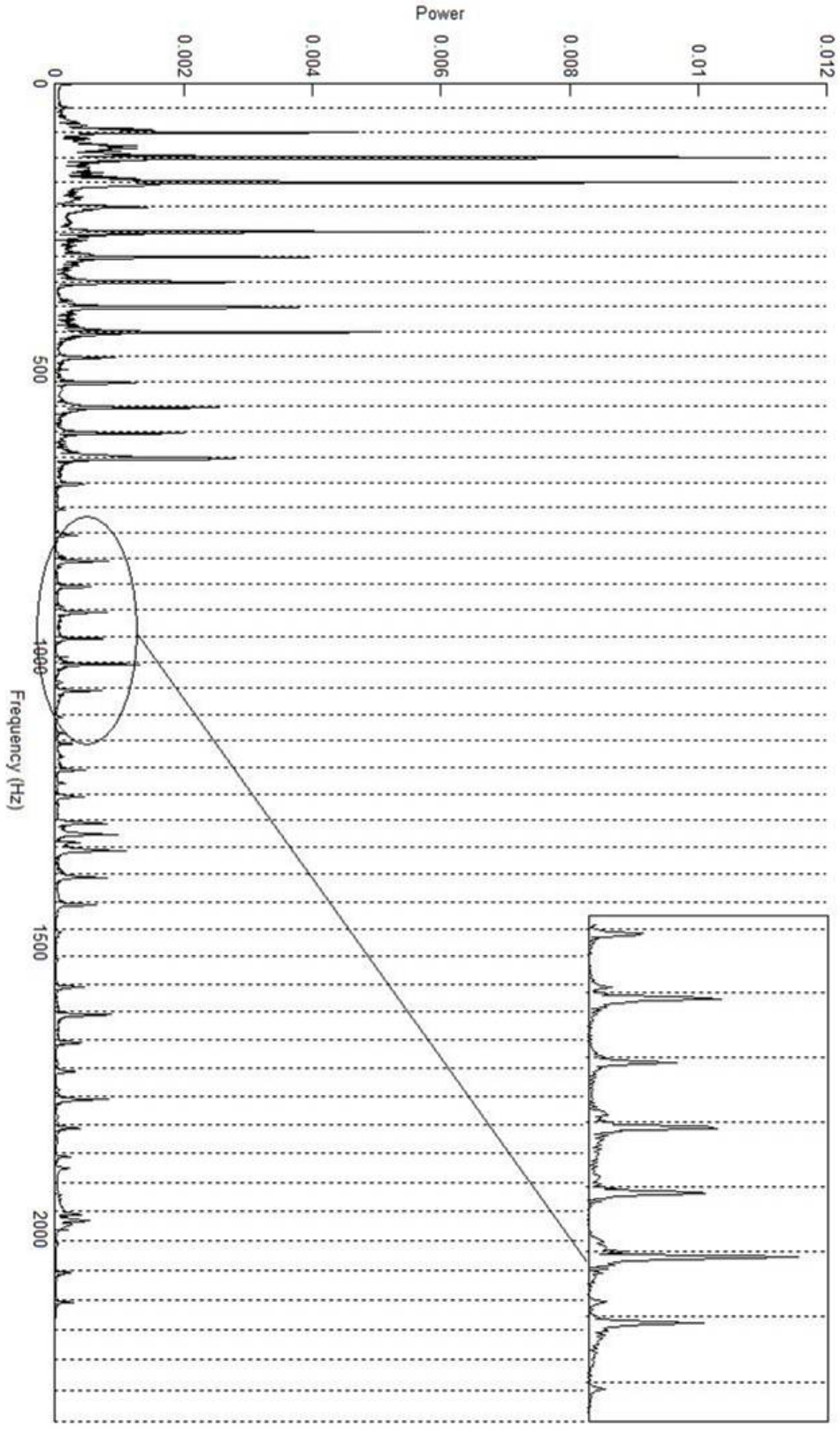


Fig 6.8 Our prediction on the F1 piano tone inharmonic frequency structure and its real FFT spectrum

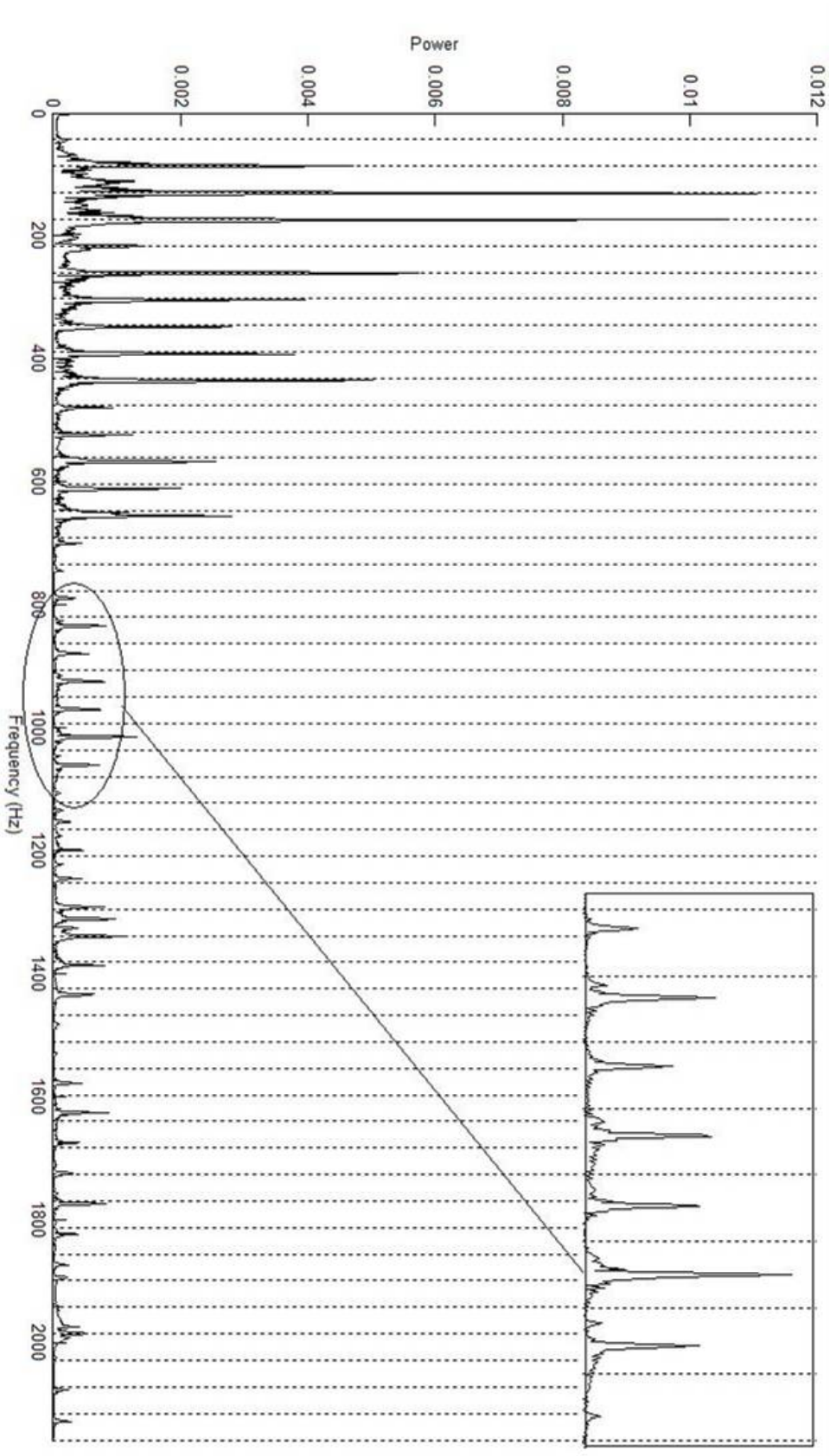


Fig 6.9 The assumed harmonic structure of the F1 piano tone and its real FFT spectrum Note the frequency range roughly from 800 Hz to 1200 Hz, and the frequencies around the 1600 Hz

To examine the applicability of our method to the measurement of inharmonicity coefficients, a number of other piano tones such as B0 (Figs 6.10-6.13 and Table 6.6), G2 (Fig 6.14-6.17 and Table 6.7), D3# (Fig 6.18-6.21 and Table 6.8), C4 (Fig 6.22-6.25 and Table 6.9) and A5 (Fig 6.26-6.29 and Table 6.10) were tested. The predicted inharmonic partial structure for each piano tone (Figs 6.12, 6.16, 6.20, 6.24 and 6.28) satisfactorily matches the real FFT spectrum, which confirms the feasibility and validity of our approach. However, From the Figs 6.13, 6.17, 6.21, 6.25 and 6.29 where the assumed harmonic structures are plotted with the real FFT spectrum, we can observe the mismatches for some partials which show that these partials are clearly inharmonic. The mismatched frequency locations have been noted in the captions of these figures. The mismatches are most noticeable in the higher partials.

Measurement for piano tone B0

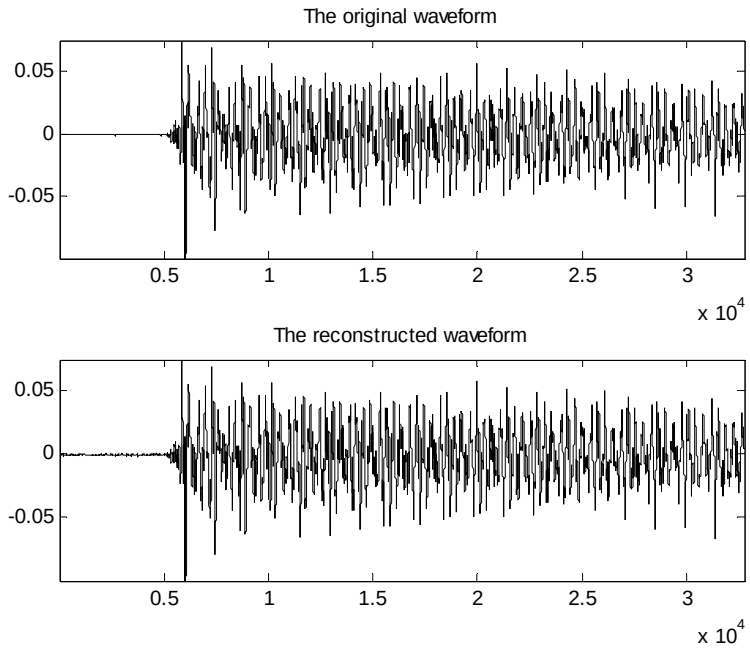
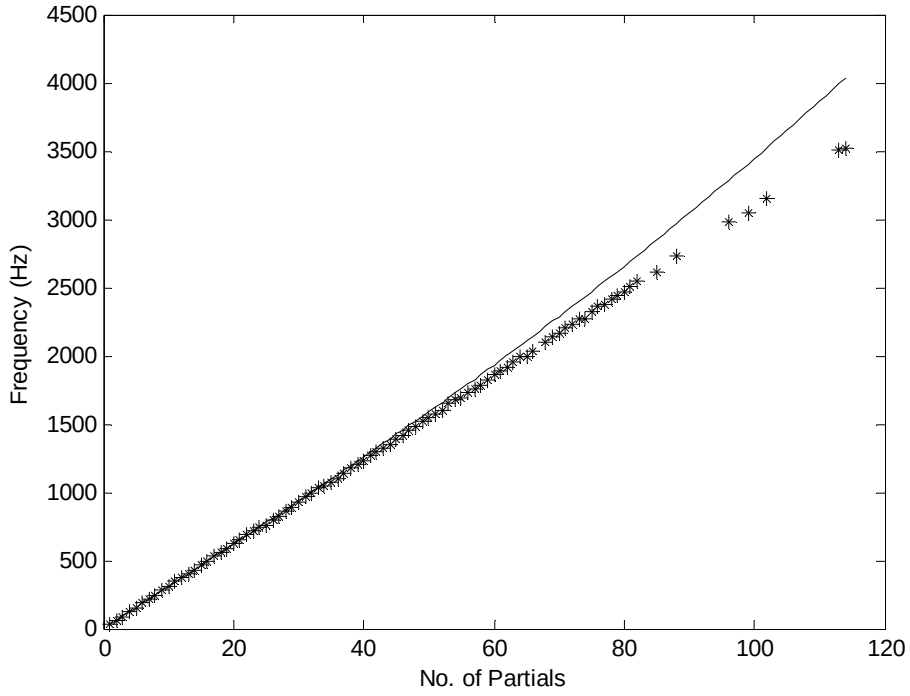


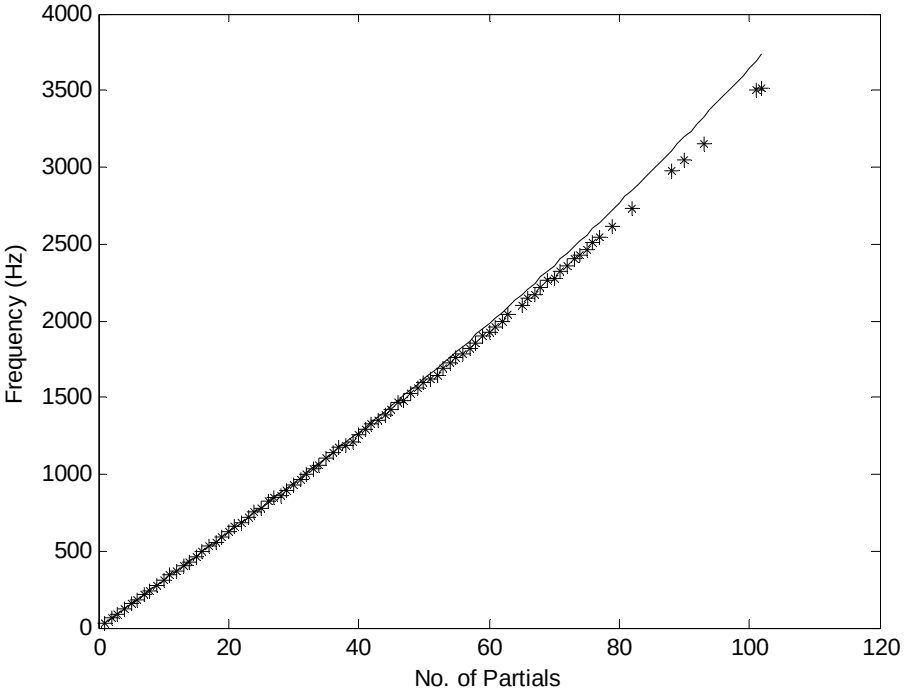
Fig 6.10 Reconstruction of a 32768-point tone B0 sample by *m*=1500 most significant time-frequency blocks

	F	B
Rough Estimate	30.87	0.000024
1st iteration	30.90	0.000039
2nd iteration	30.87	0.000054
3rd iteration	30.84	0.000063
4th iteration	30.87	0.000072
5th iteration	30.87	0.000081
6th iteration	30.84	0.000086
7th iteration	30.82	0.000089
8th iteration	30.78	0.000092

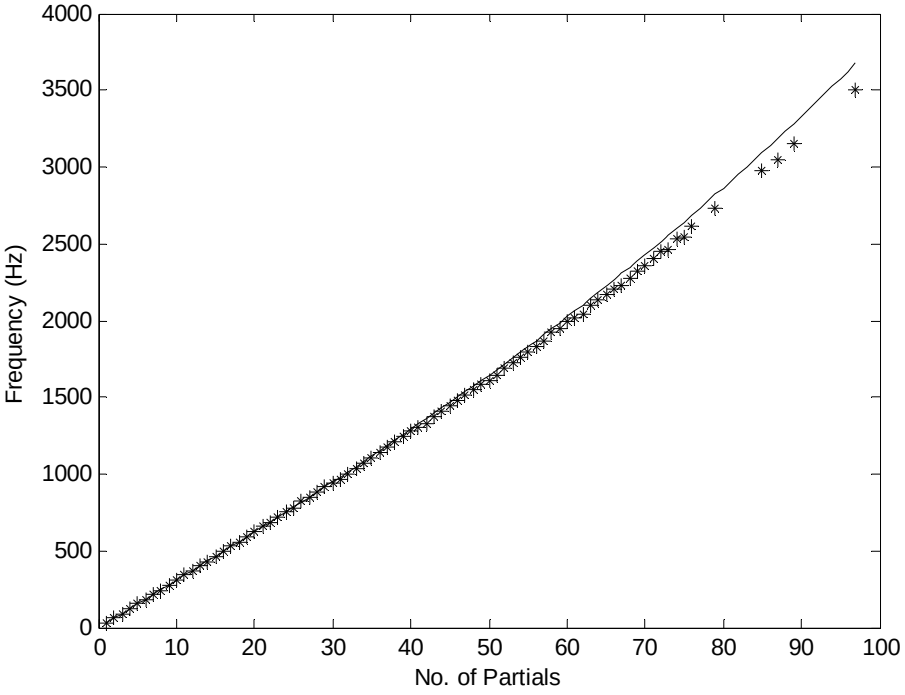
Table 6.6 F_1 and B calculated for the B0 piano tone



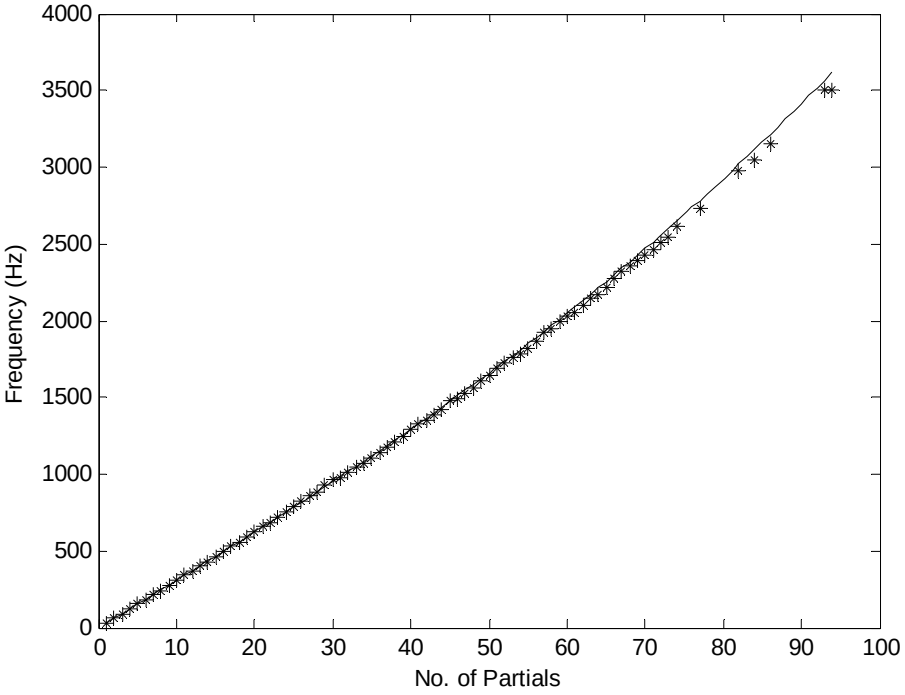
(a) Rough estimation



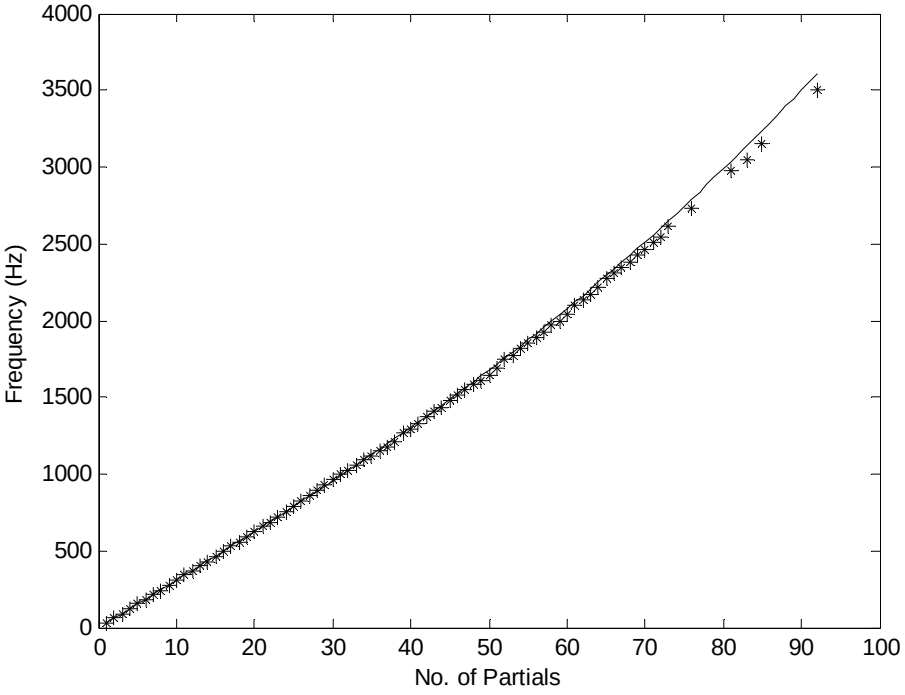
(b) The 1st iteration



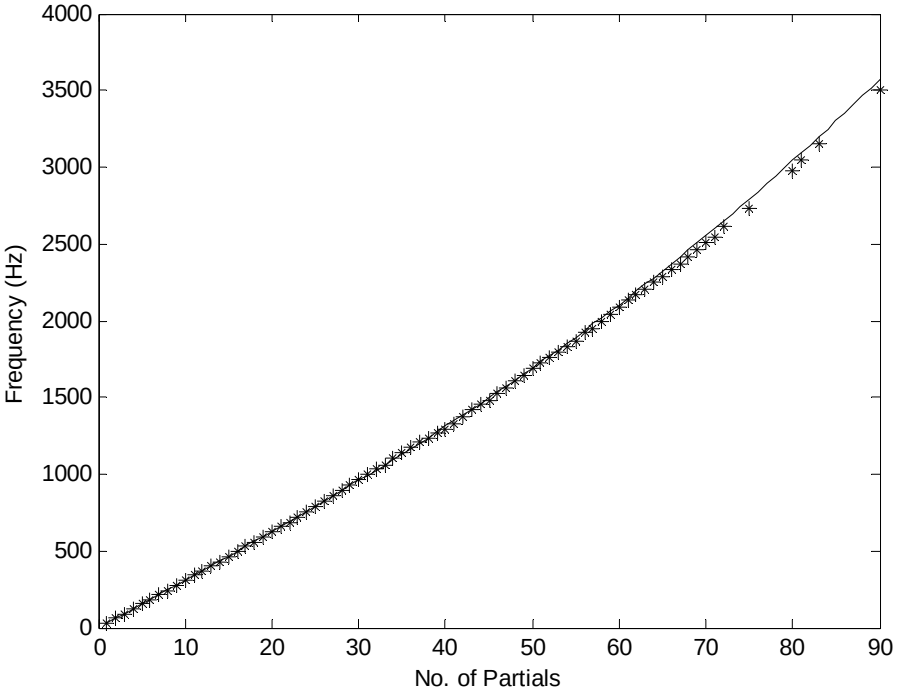
(c) The 2nd iteration



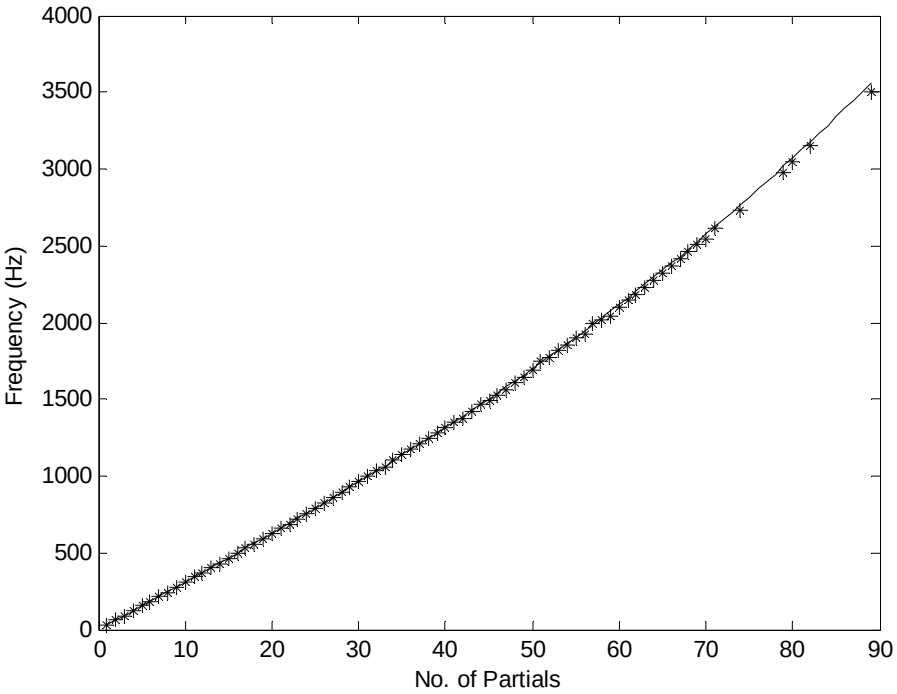
(d) The 3rd iteration



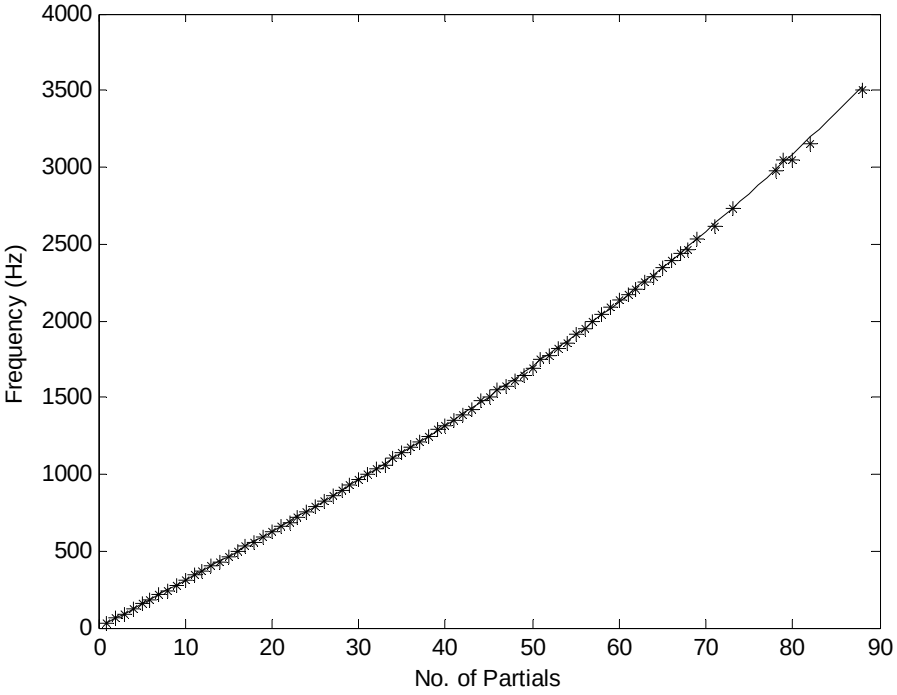
(e) The 4th iteration



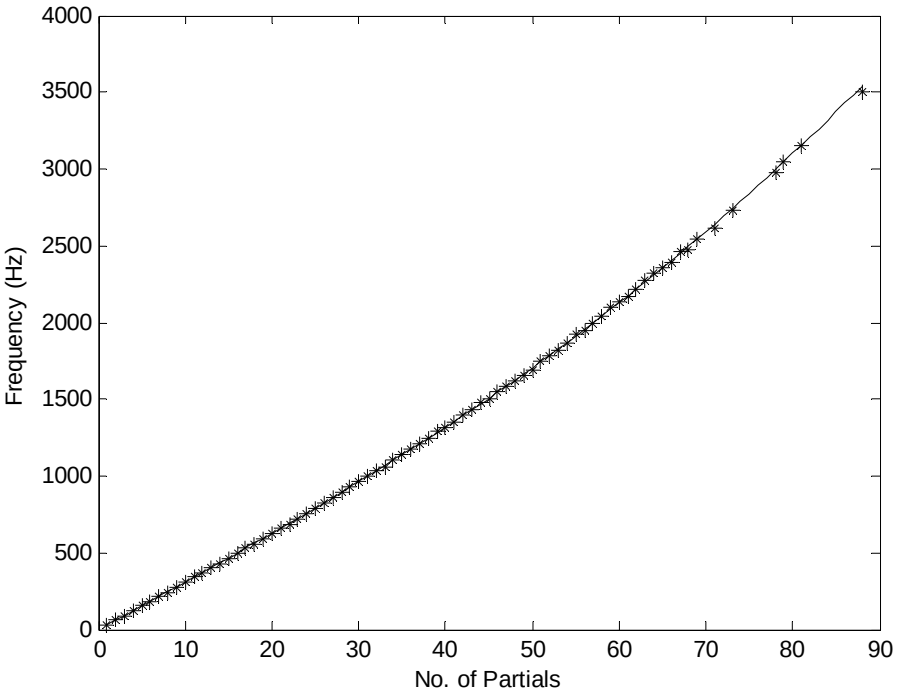
(f) The 5th iteration



(g) The 6th iteration



(h) The 7th iteration



(i) The 8th iteration

Fig 6.11 Results of the 8-iteration correction process for the B0 piano tone

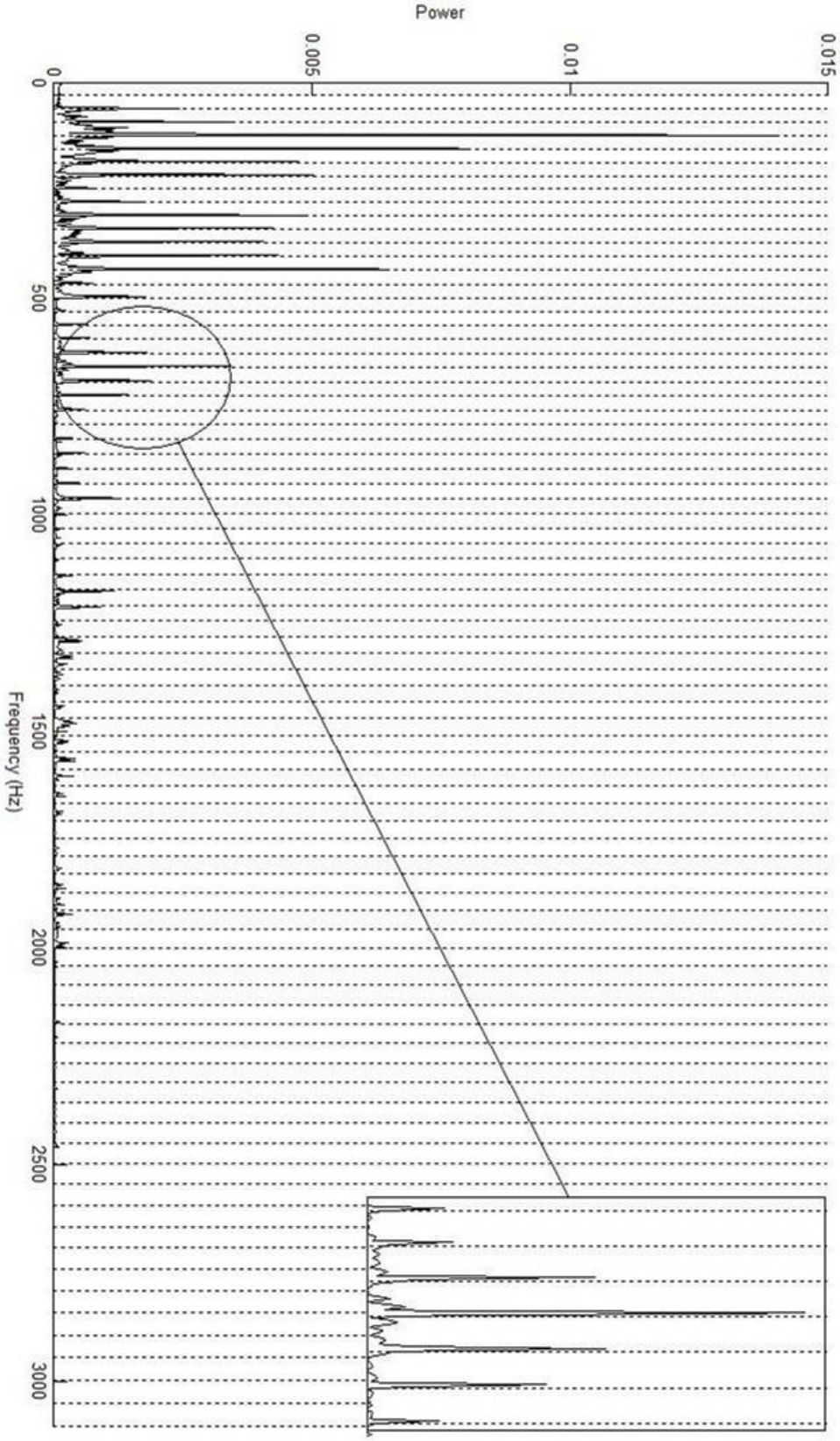


Fig 6.12 Our prediction on the B0 piano tone inharmonic frequency structure and its real FFT spectrum

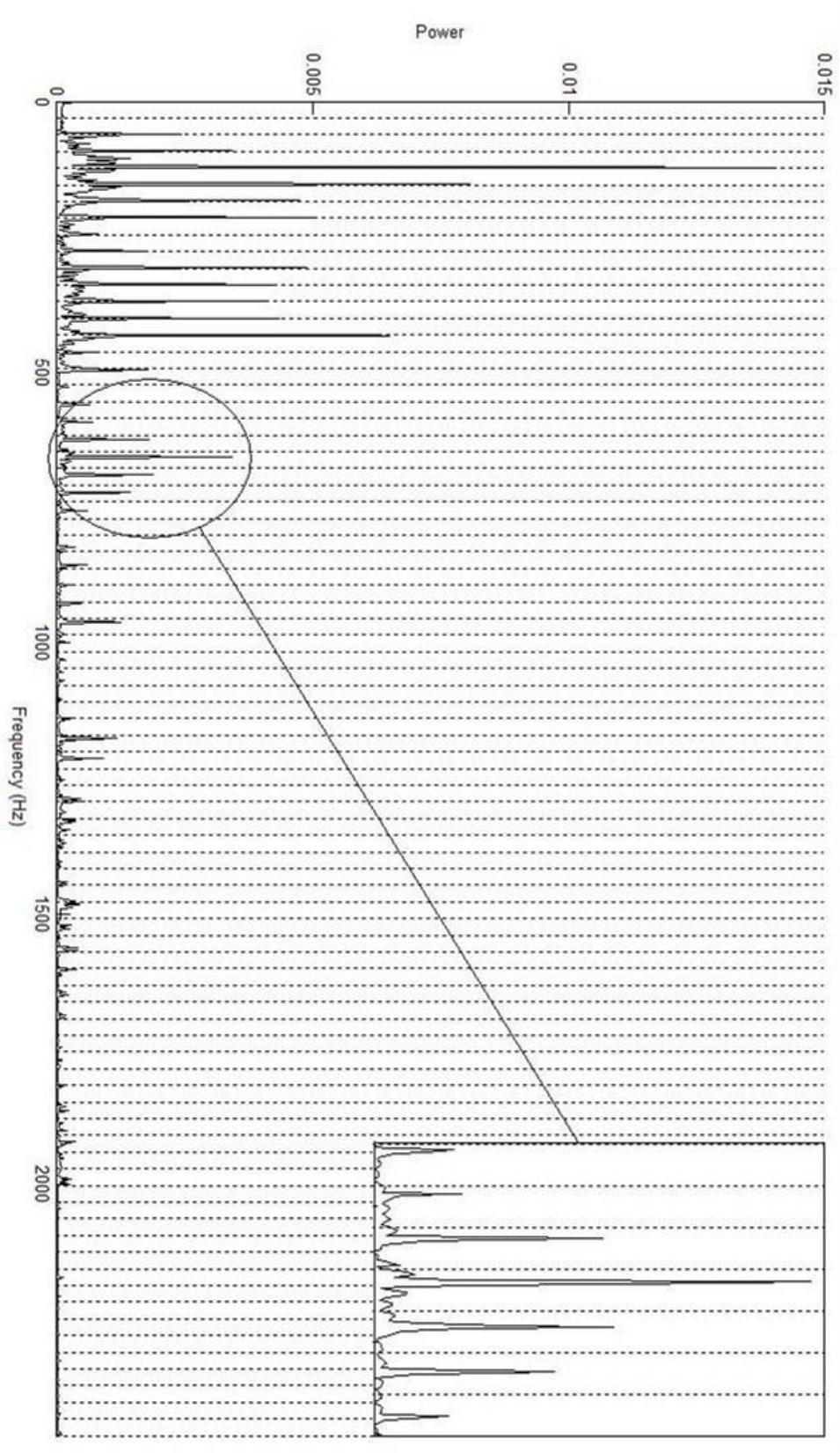


Fig 6.13 The assumed harmonic structure of the B0 piano tone and its real FFT spectrum Note the frequency range roughly from 600 Hz to 800 Hz

Measurement of piano tone G2

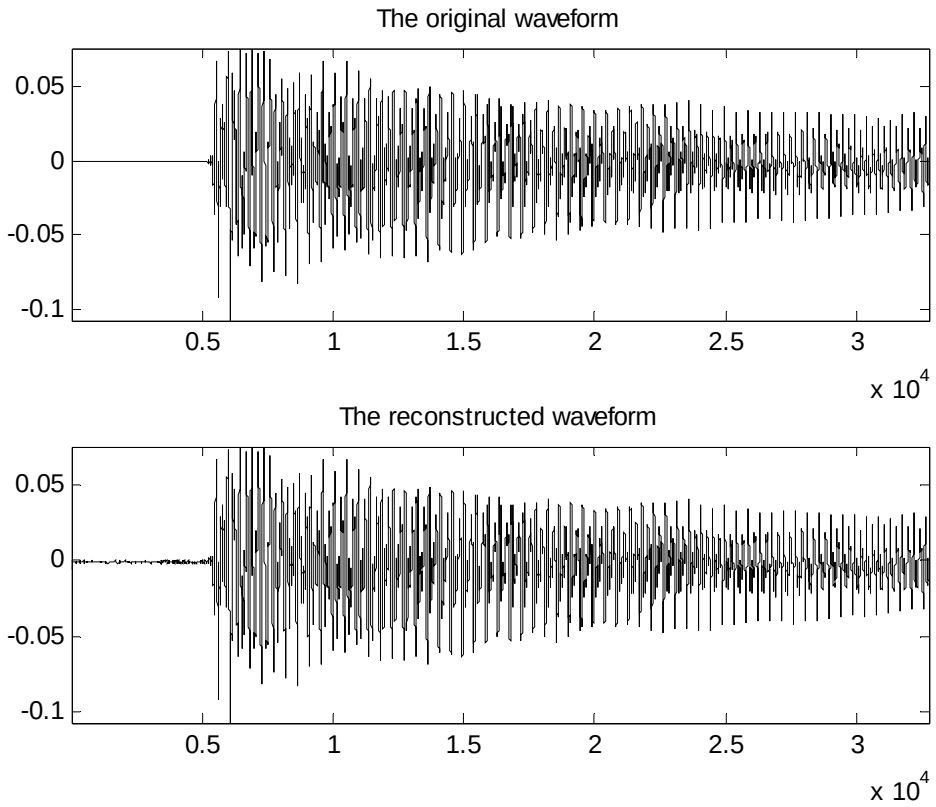
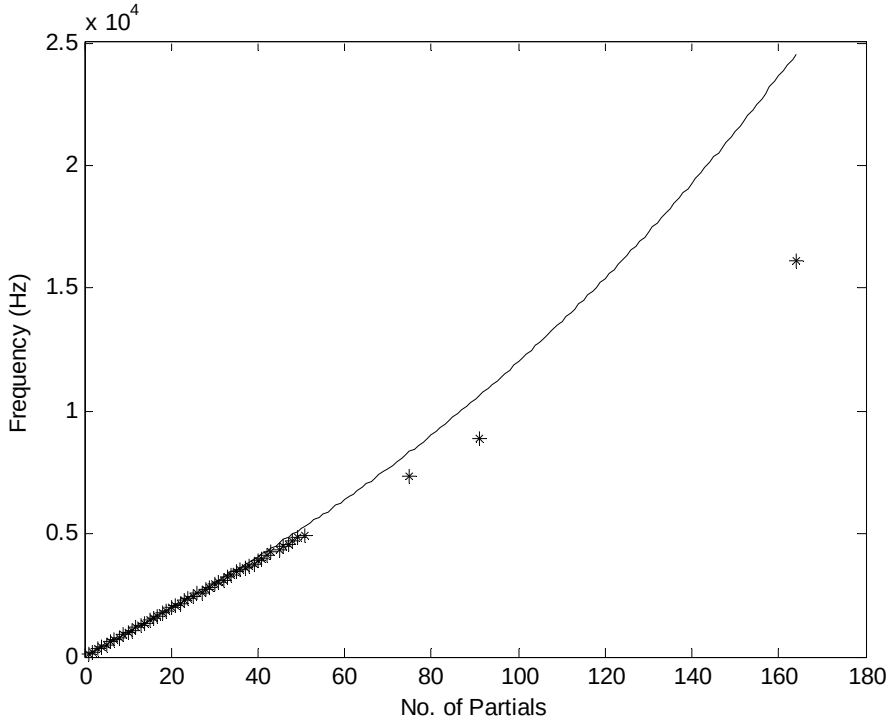


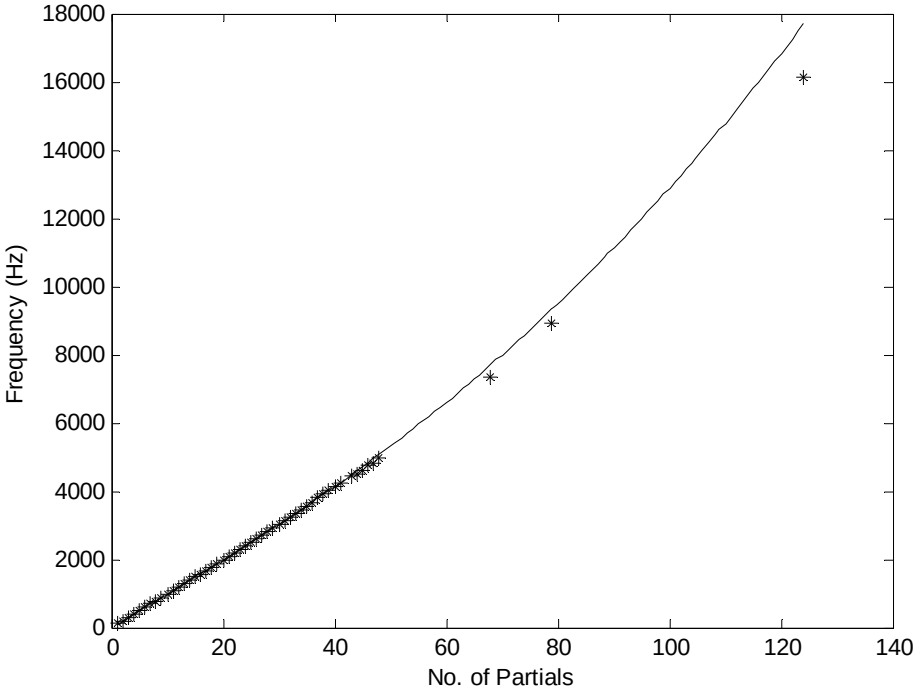
Fig 6.14 Reconstruction of a 32768-point tone G2 sample by *m*=1500 most significant time-frequency blocks

	F	B
Rough Estimate	98.4237	0.000049
1st iteration	97.8652	0.000073
2nd iteration	97.5967	0.000078
3rd iteration	97.6159	0.000079

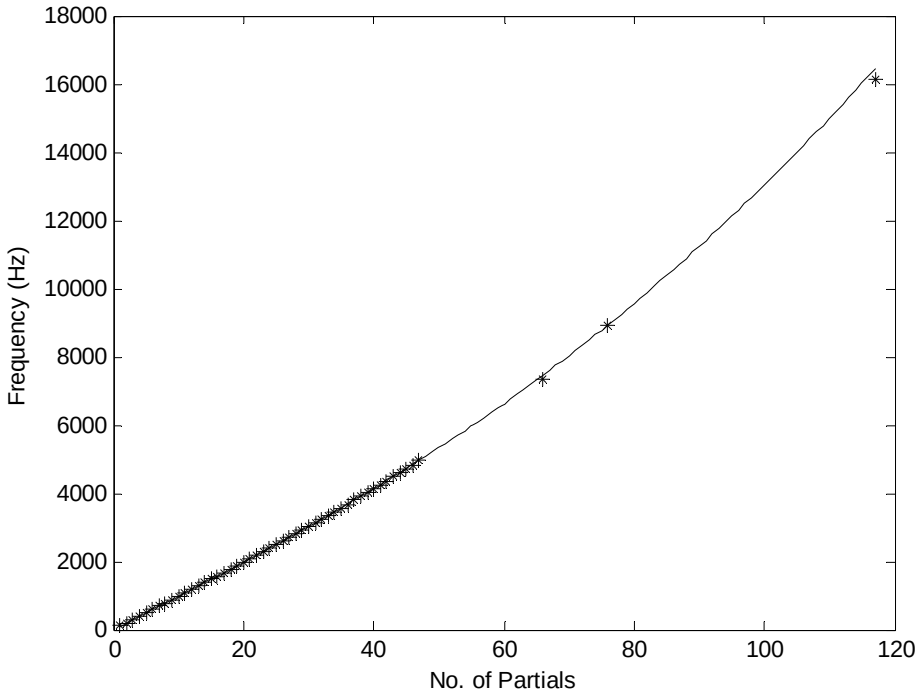
Table 6.7 F_1 and B calculated for the G2 piano tone



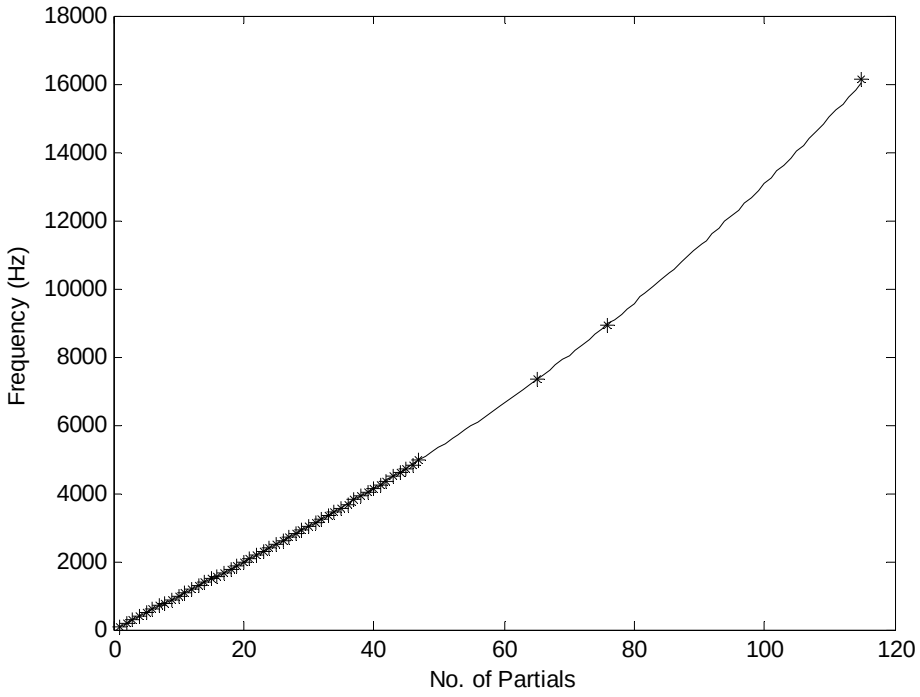
(a) Rough Estimation



(b) The 1st iteration



(c) The 2nd iteration



(d) The 3rd iteration

Fig 6.15 Results of the 3-iteration correction process for the G2 piano tone

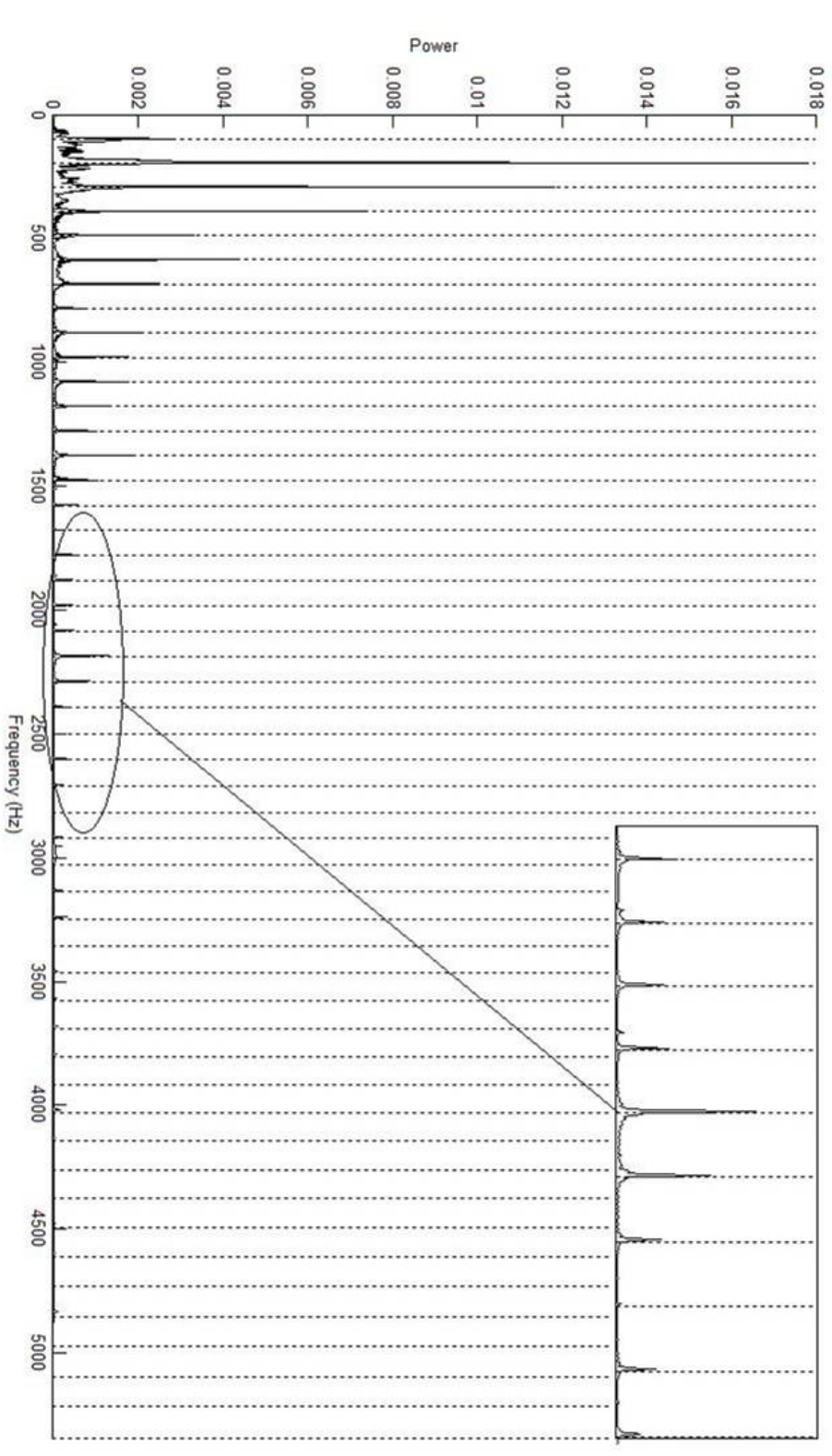


Fig 6.16 Our prediction on the G2 piano tone inharmonic frequency structure and its real FFT spectrum

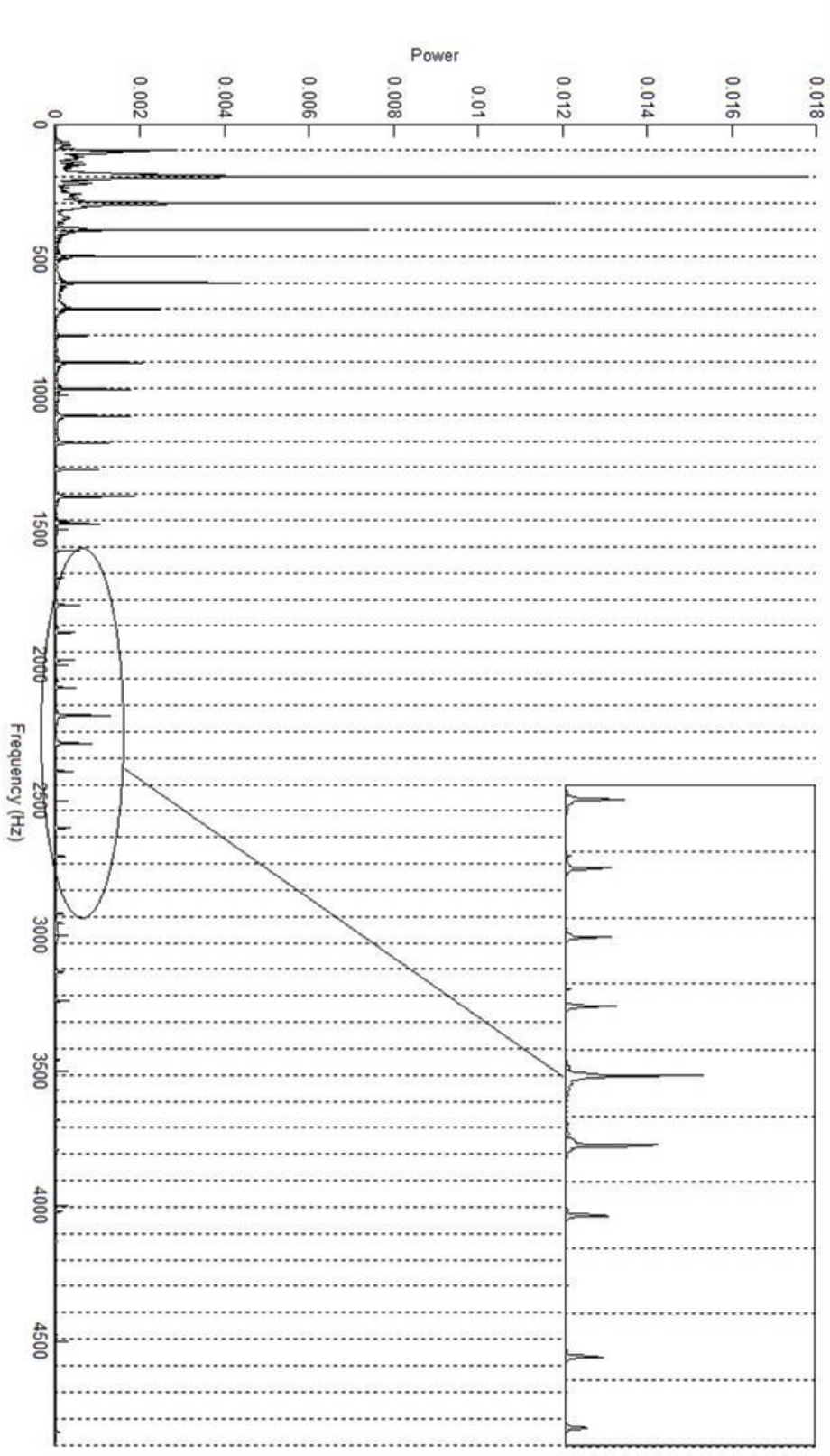


Fig 6.17 The assumed harmonic structure of the G2 piano tone and its real FFT spectrum Note the frequency range roughly from 1700 Hz to 2700 Hz

Measurement of piano tone D3#

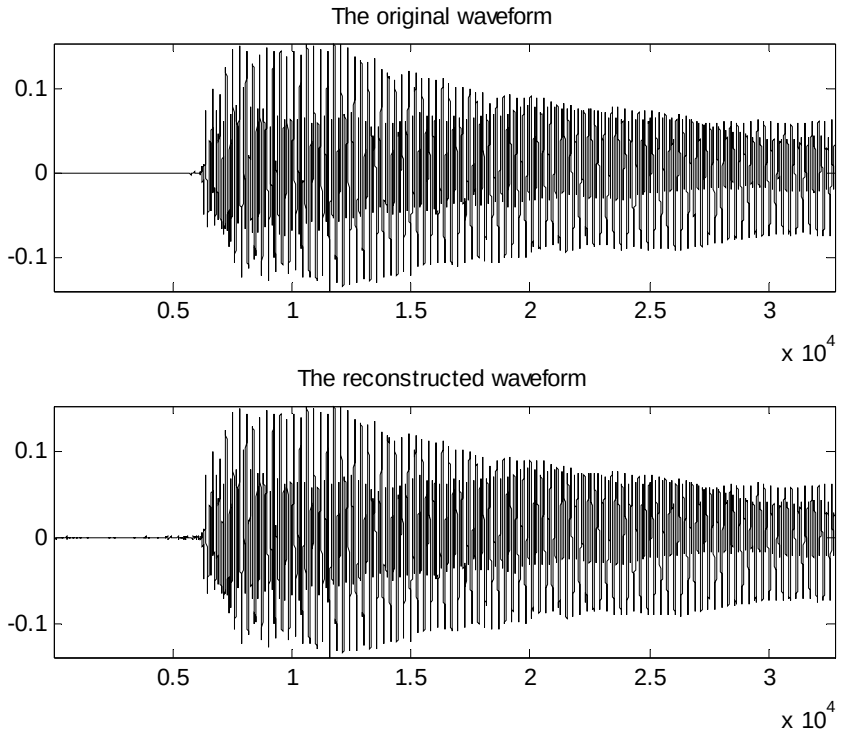
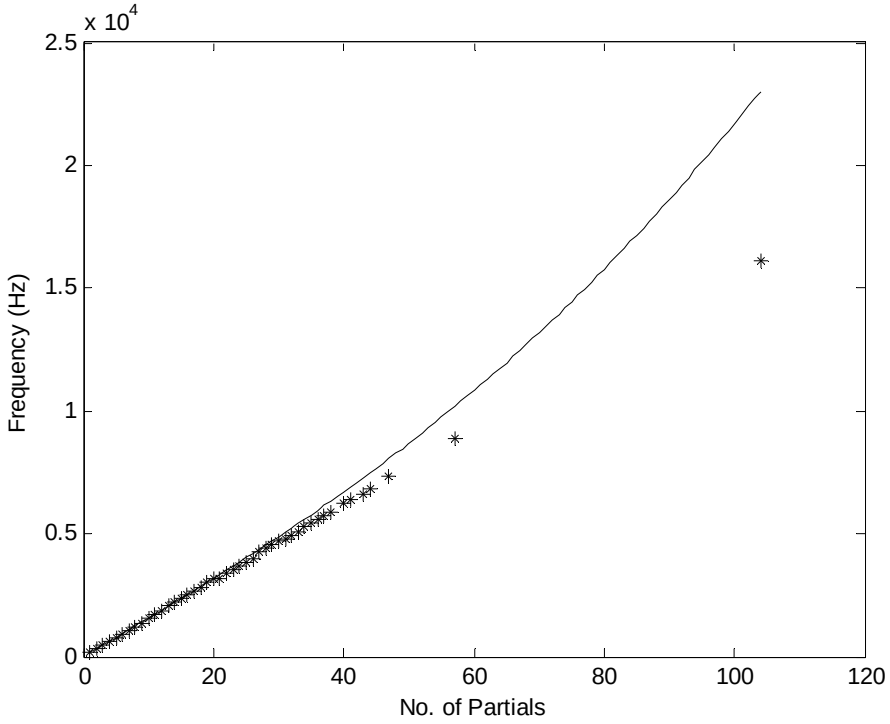


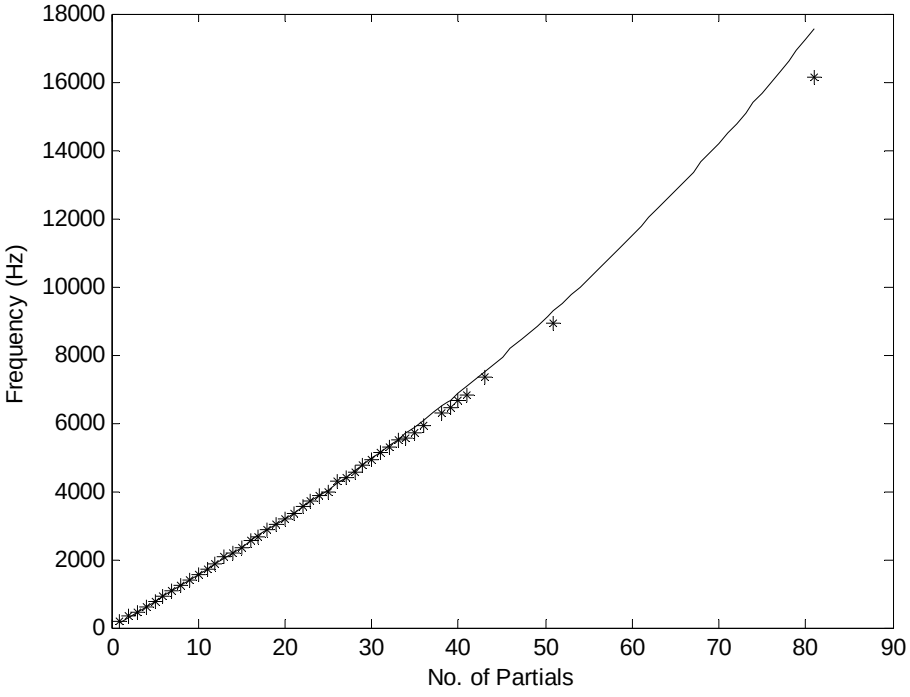
Fig 6.18 Reconstruction of a 32768-point tone D3# sample by *m*=1500 most significant time-frequency blocks

	F	B
Rough Estimate	156.36	0.000093
1st iteration	155.08	0.000145
2nd iteration	154.74	0.000159
3rd iteration	154.69	0.000163
4th iteration	154.64	0.000165
5th iteration	154.63	0.000165

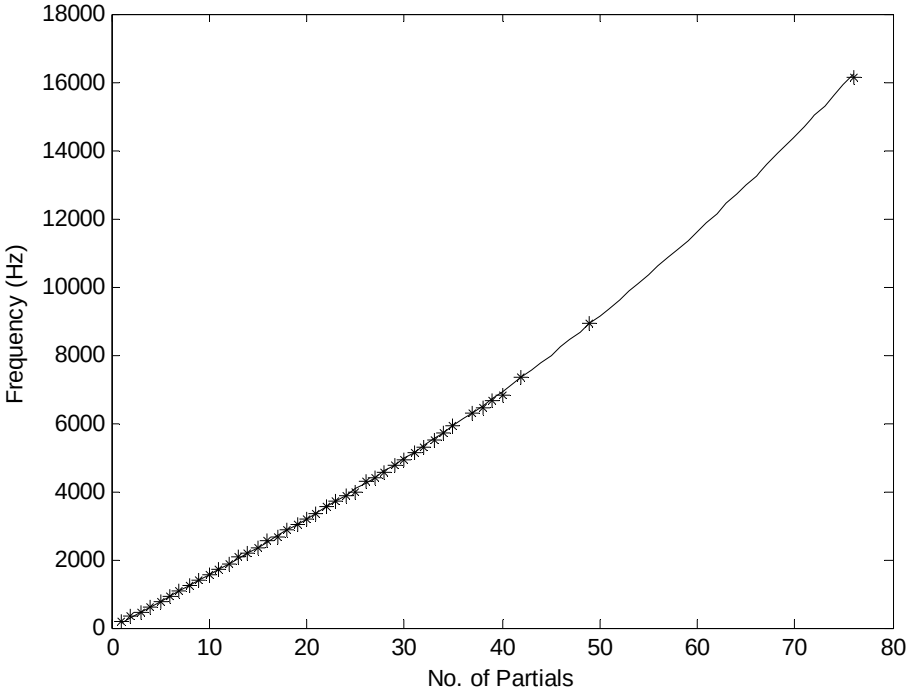
Table 6.8 F_1 and B calculated for the D3# piano tone



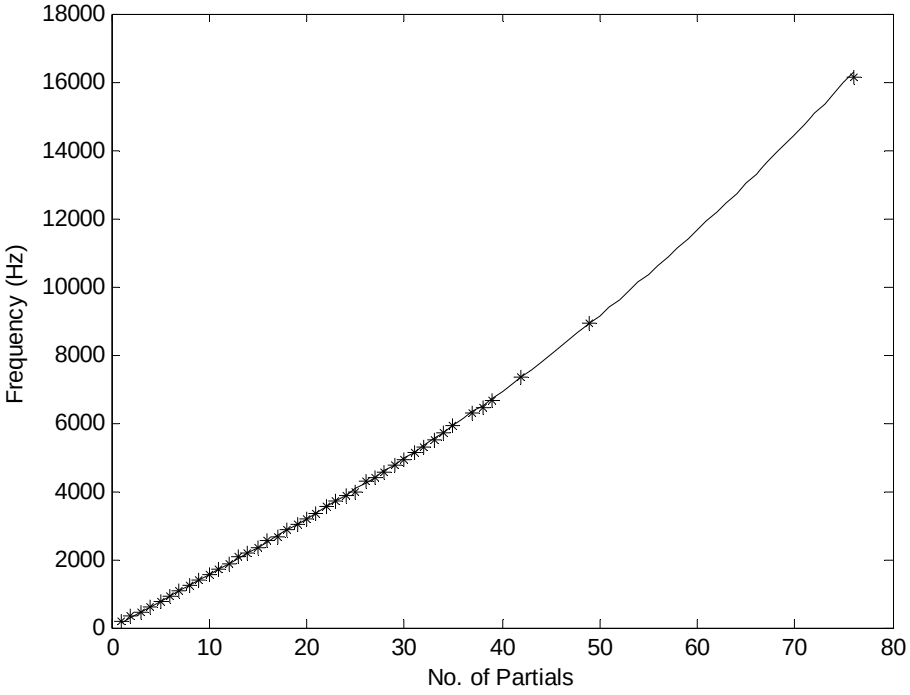
(a) Rough Estimation



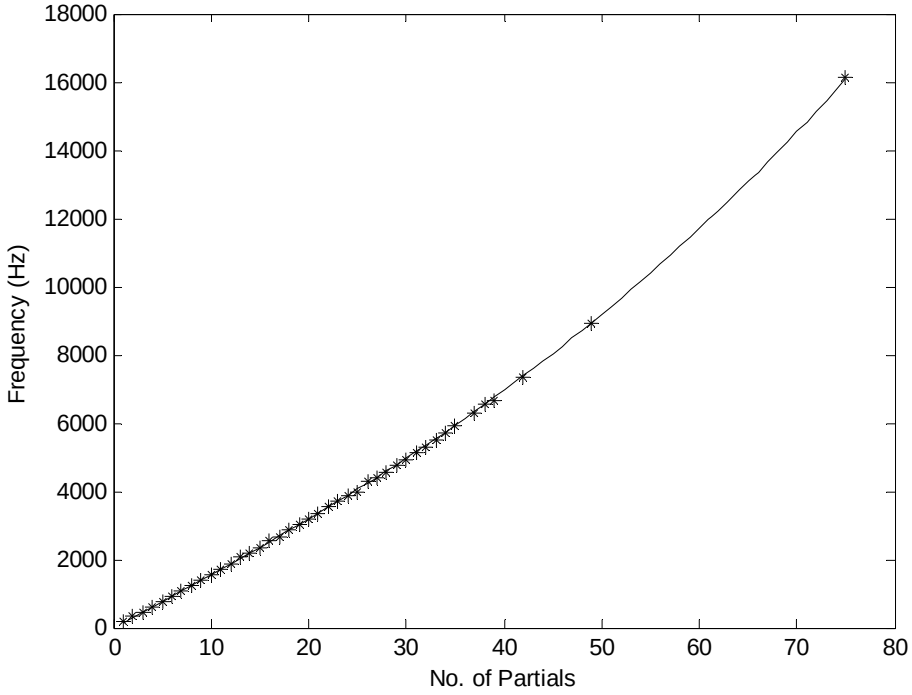
(b) The 1st iteration



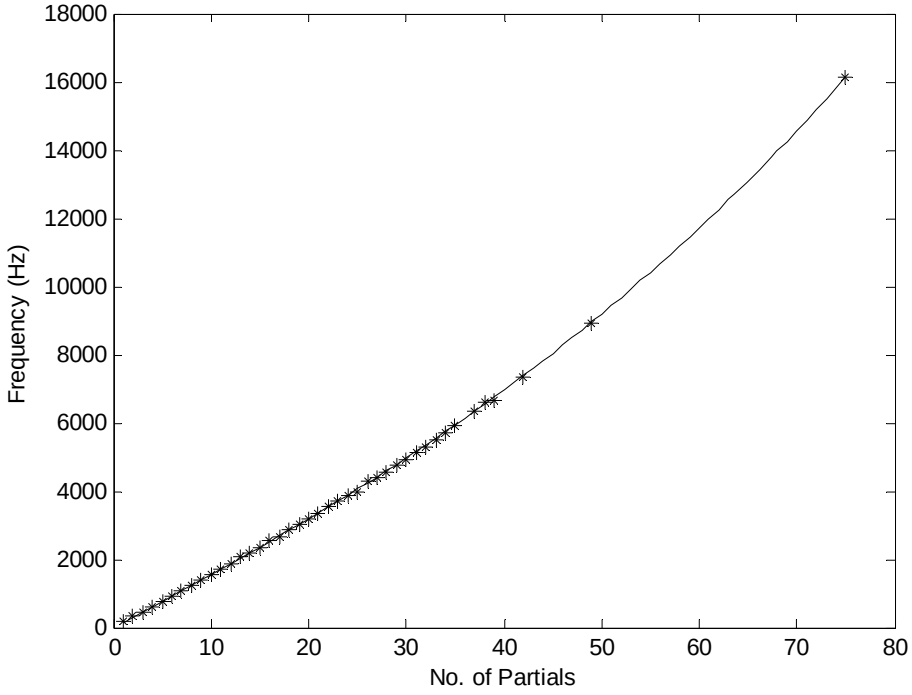
(c) The 2nd iteration



(d) The 3rd iteration



(e) The 4th iteration



(f) The 5th iteration

Fig 6.19 Results of the 6-iteration correction process for the D3# piano tone

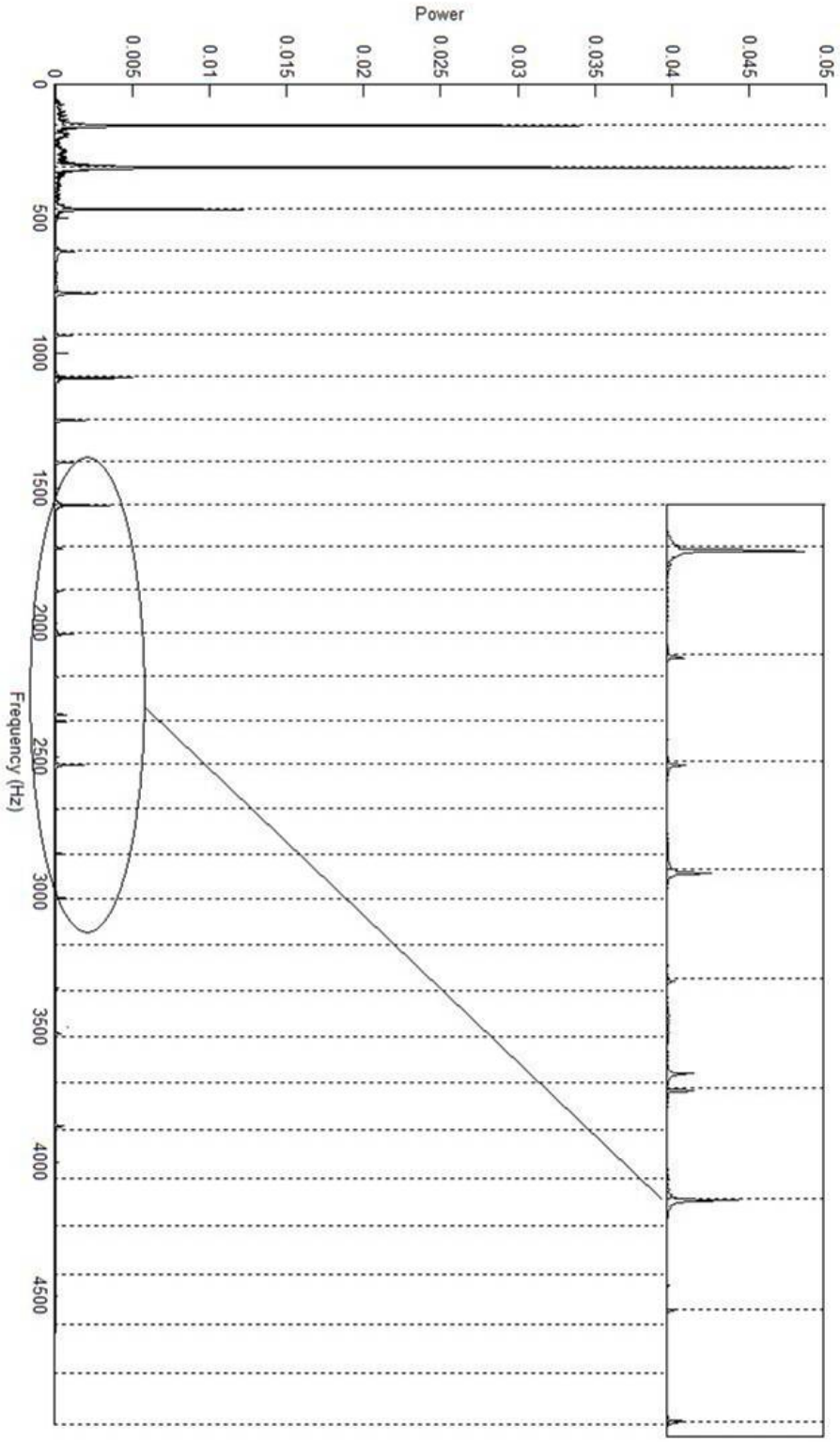


Fig 6.20 Our prediction on the D3# piano tone inharmonic frequency structure and its real FFT spectrum

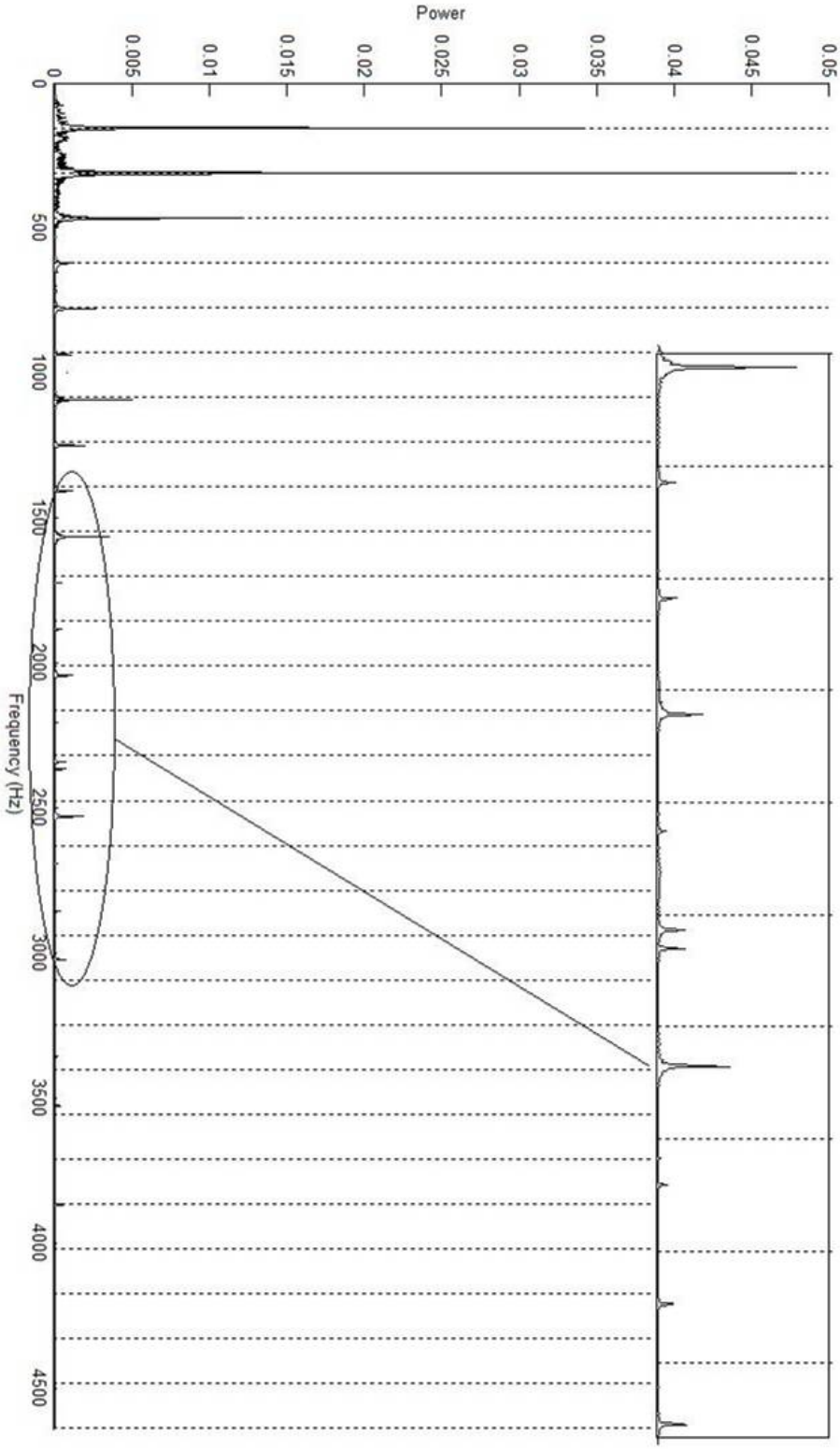


Fig 6.21 The assumed harmonic structure of D3# piano tone and its real FFT spectrum. Note the frequency range roughly from 1500 Hz to 3000 Hz

Measurement of piano tone C4

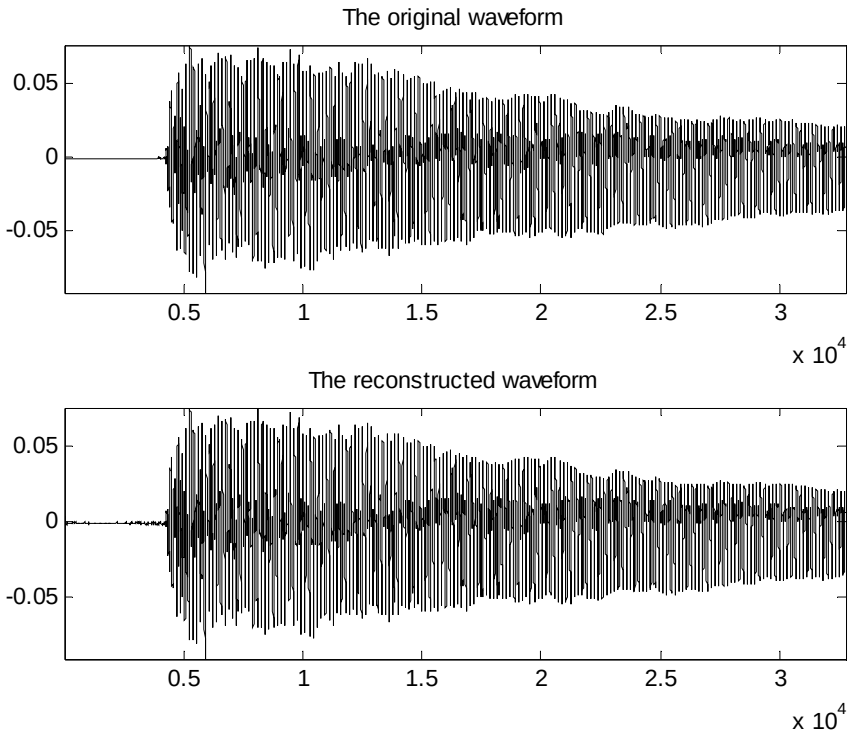
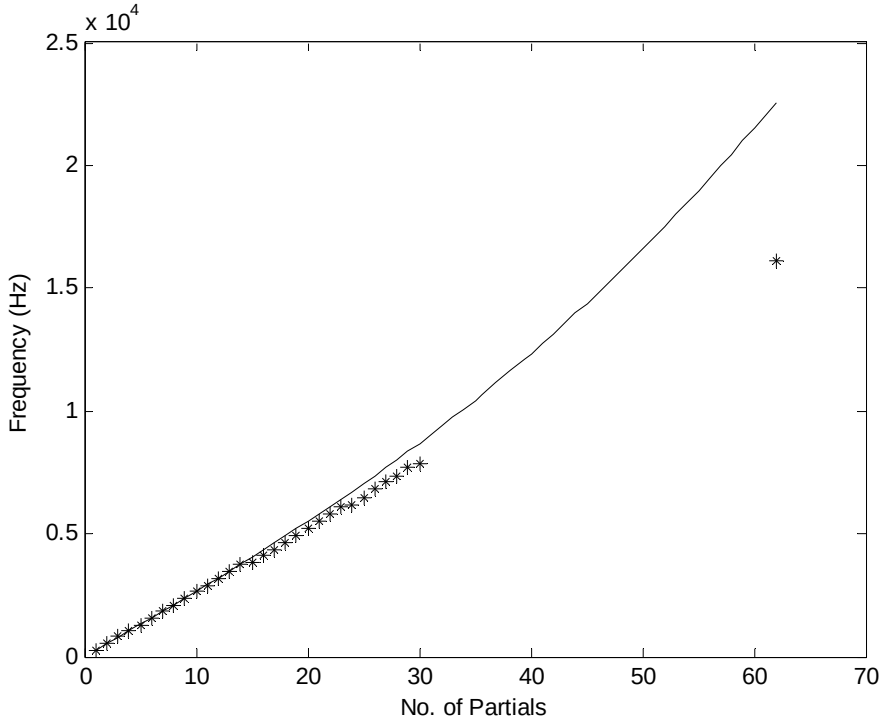


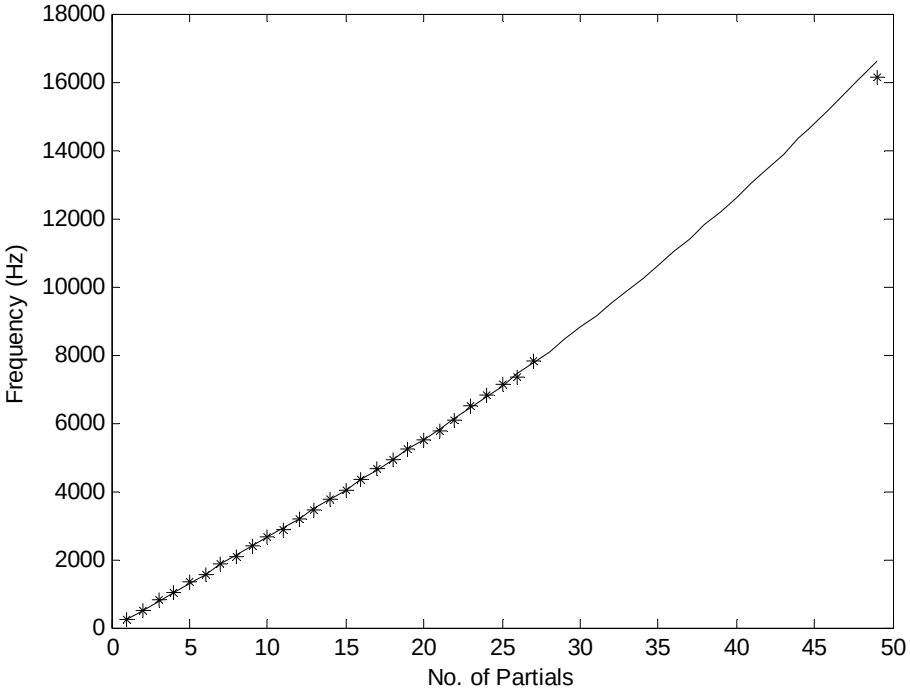
Fig 6.22 Reconstruction of a 32768-point tone C4 sample by *m*=1500 most significant time-frequency blocks

	F	B
Rough Estimate	262.46	0.000240
1st iteration	261.54	0.000284
2nd iteration	261.95	0.000275
3rd iteration	262.01	0.000278

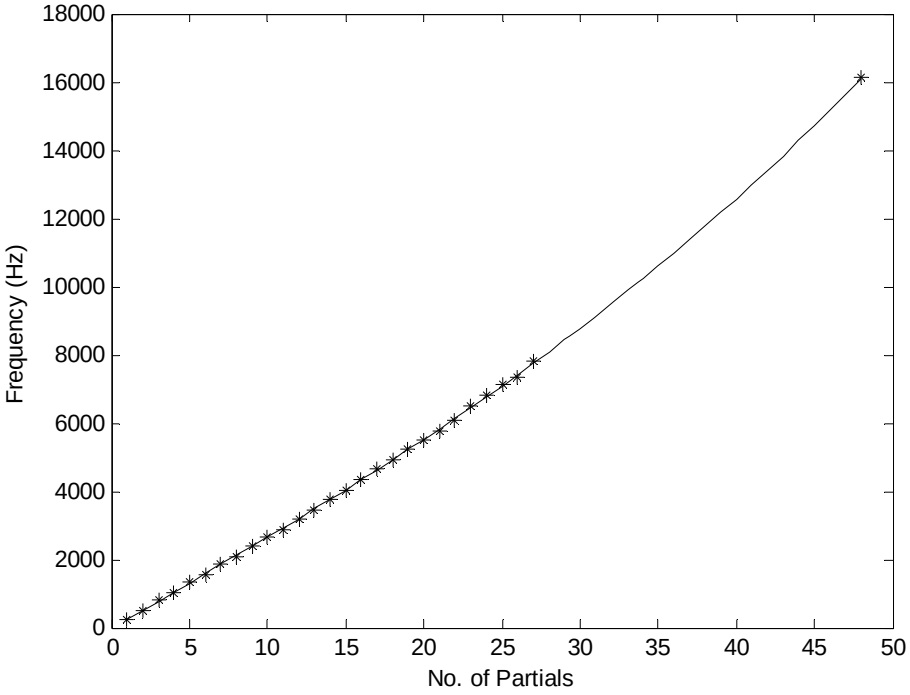
Table 6.9 F_1 and B calculated for the C4 piano tone



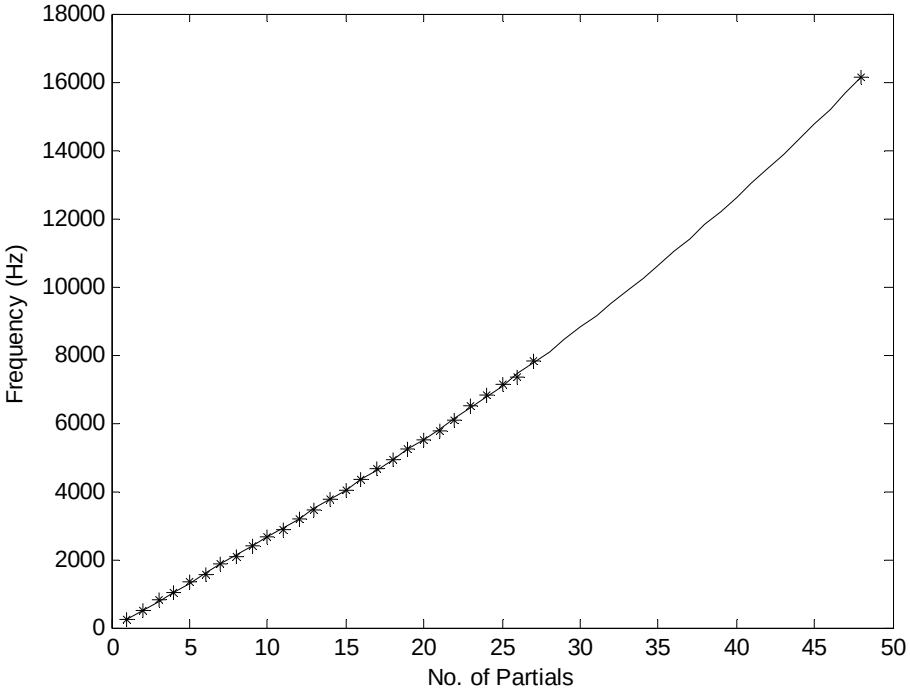
(a) Rough Estimation



(b) The 1st iteration



(c) The 2nd iteration



(d) The 3rd iteration

Fig 6.23 Results of the 3-iteration correction process for the C4 piano tone

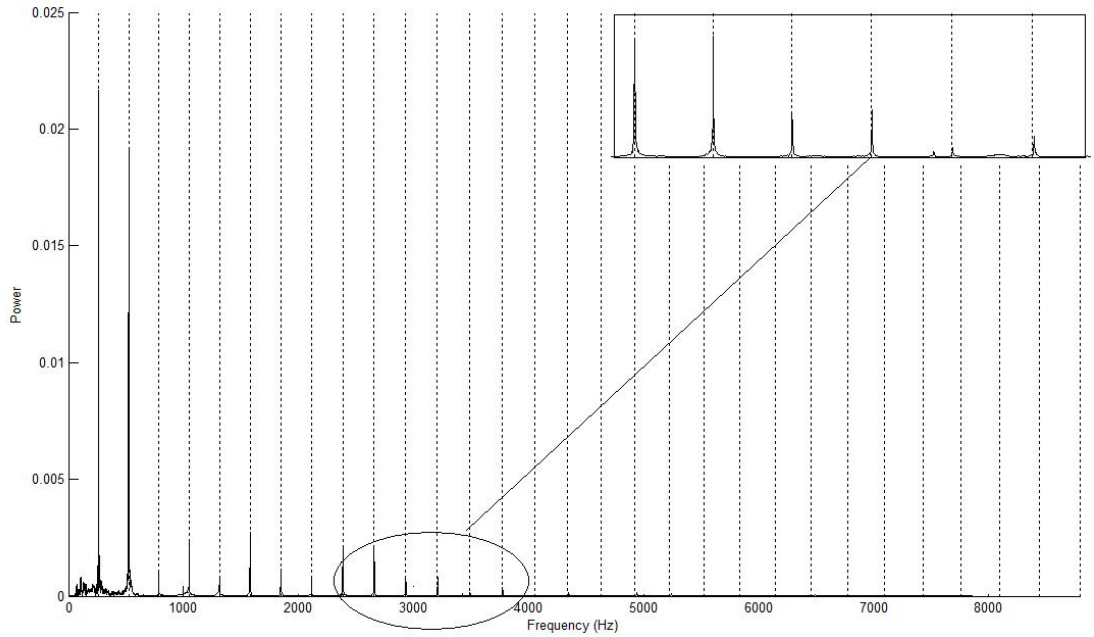


Fig 6.24 Our prediction on the C4 piano tone inharmonic frequency structure and its real FFT spectrum

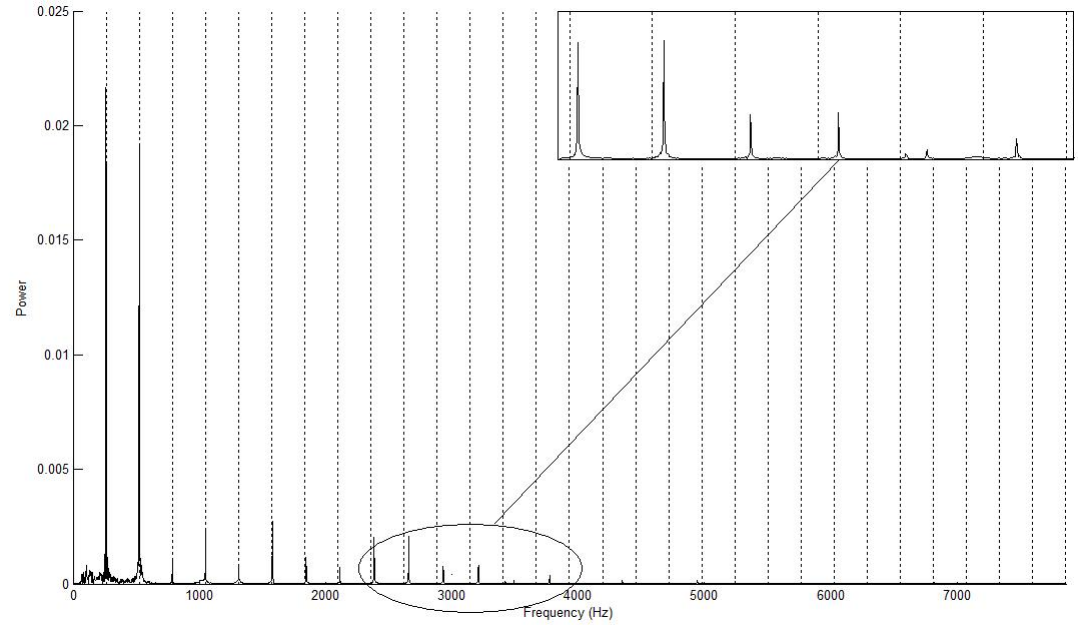


Fig 6.25 The assumed harmonic structure of C4 piano tone and its real FFT spectrum

Note the frequencies around 3000 Hz

Measurement of piano tone A5

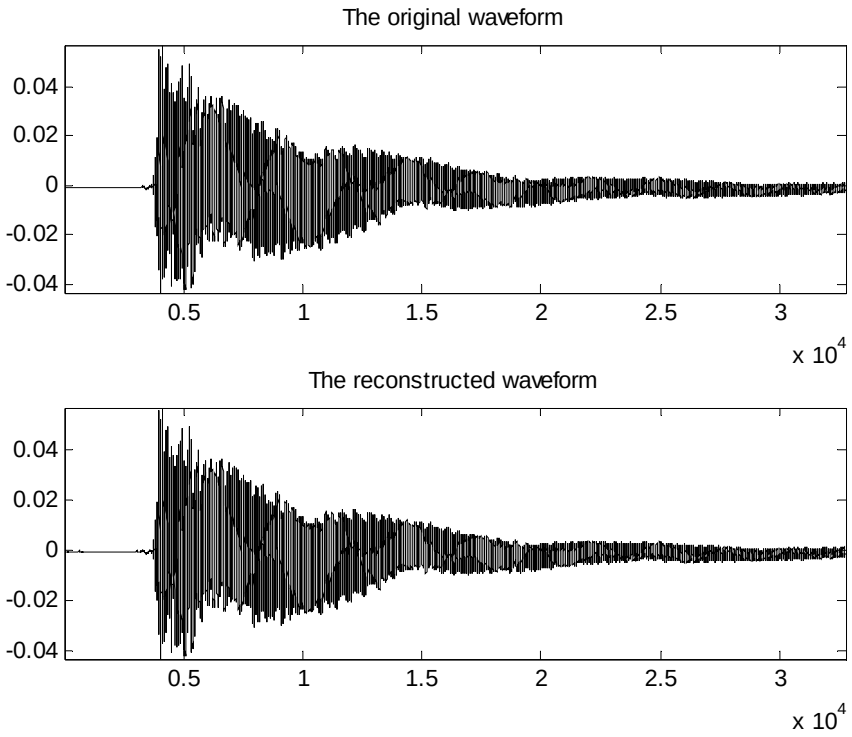
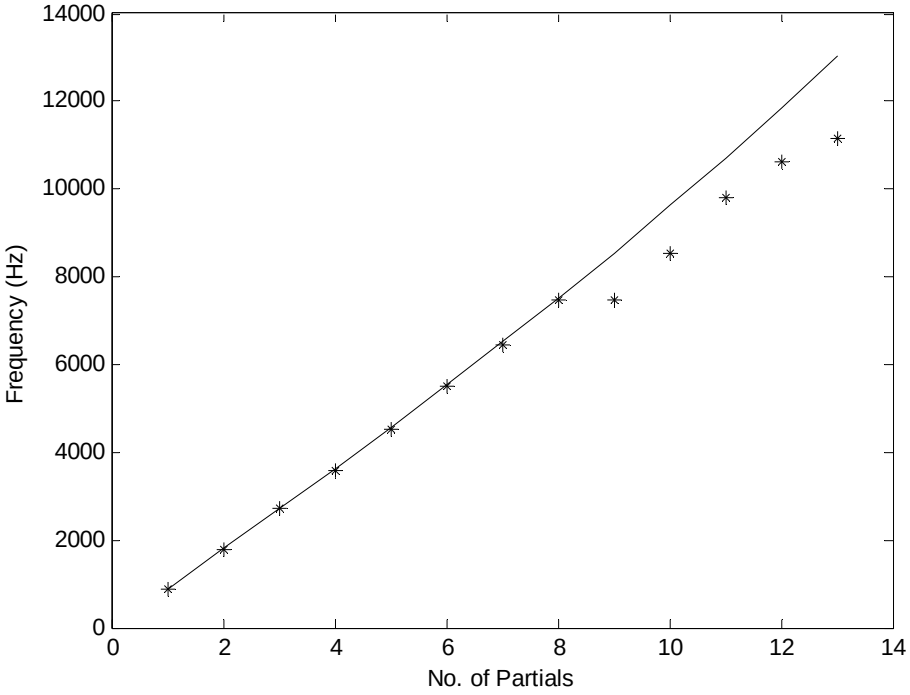


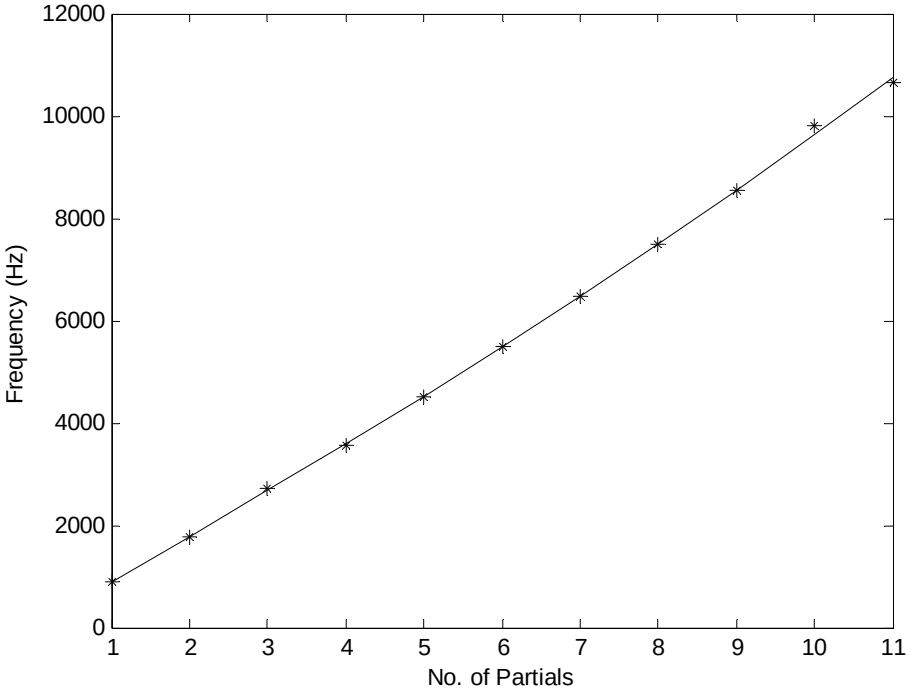
Fig 6.26 Reconstruction of a 16384-point tone A5 sample by m=1500 most significant time-frequency blocks

	F	B
Rough Estimate	897.96	0.001477
1st iteration	883.63	0.001868
2nd iteration	883.62	0.001867

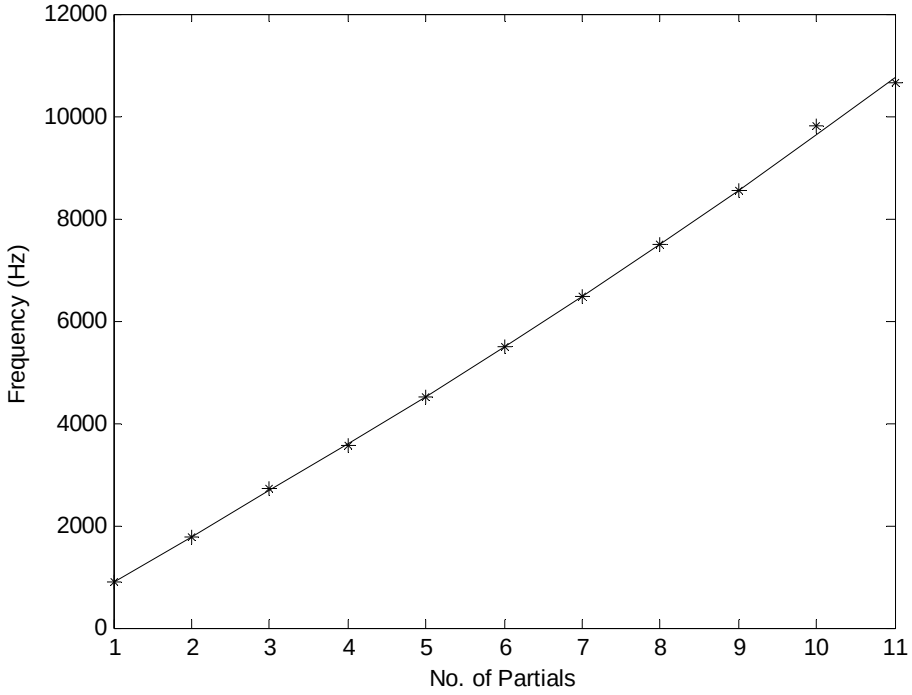
Table 6.10 F_1 and B calculated for the A5 piano tone



(a) Rough Estimation



(b) The 1st Iteraion



(c) The 2nd iteration

Fig 6.27 Results of the 2-iteration correction process for the A5 piano tone

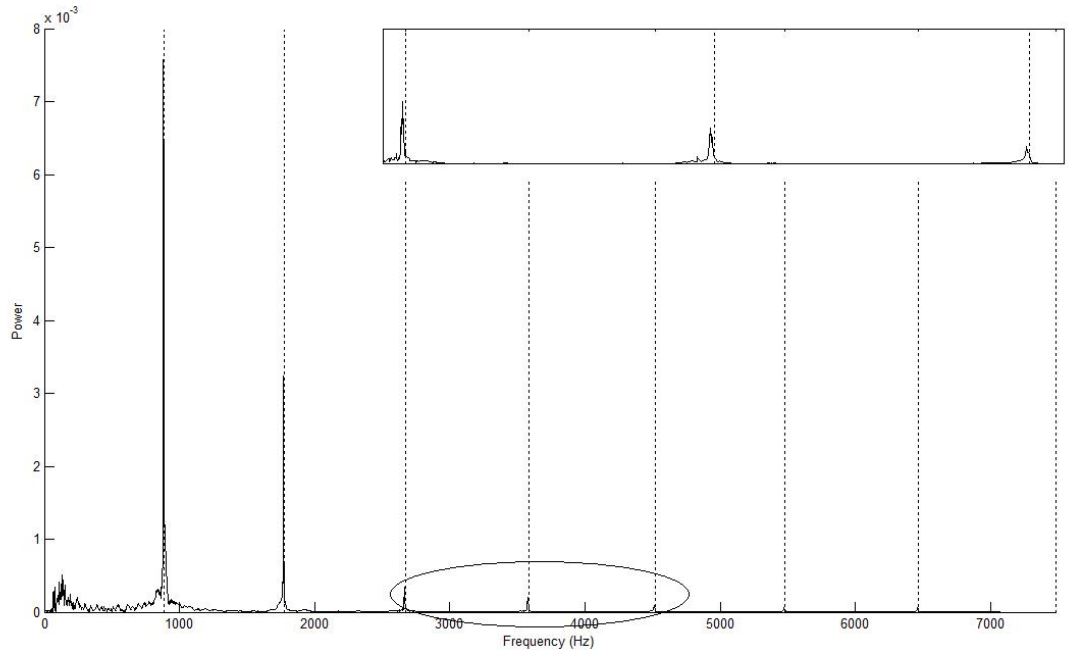


Fig 6.28 Our prediction on A5 piano tone inharmonic frequency structure and its real FFT spectrum

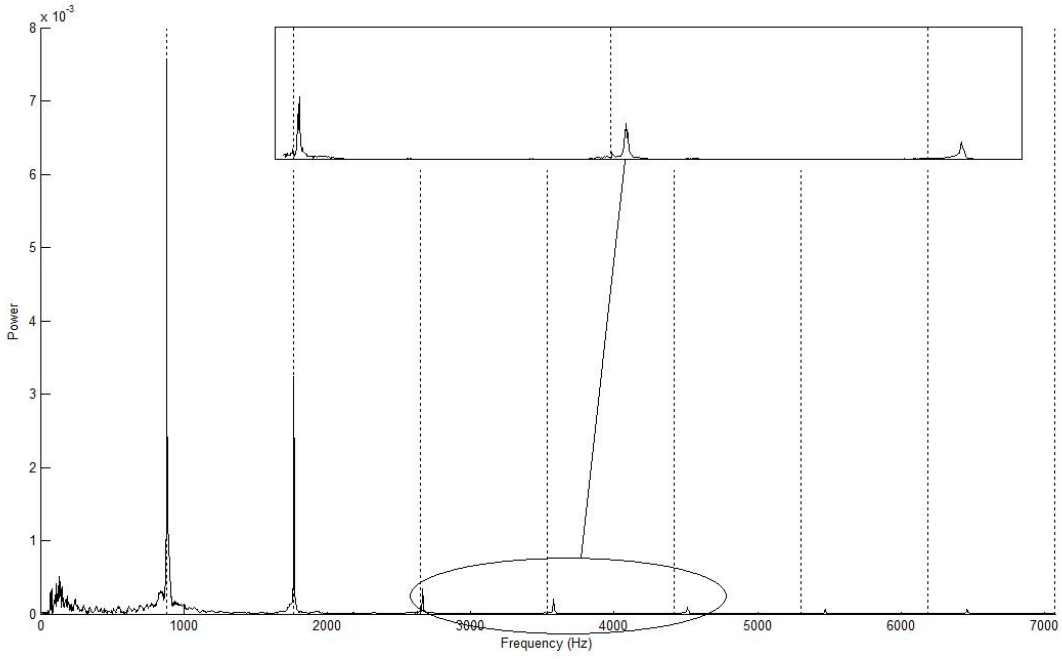


Fig 6.29 The assumed harmonic structure of A5 piano tone and its real FFT spectrum

Note the frequencies around 4000 Hz

The inharmonicity coefficients which we have calculated for some piano tones are shown in the following figure.

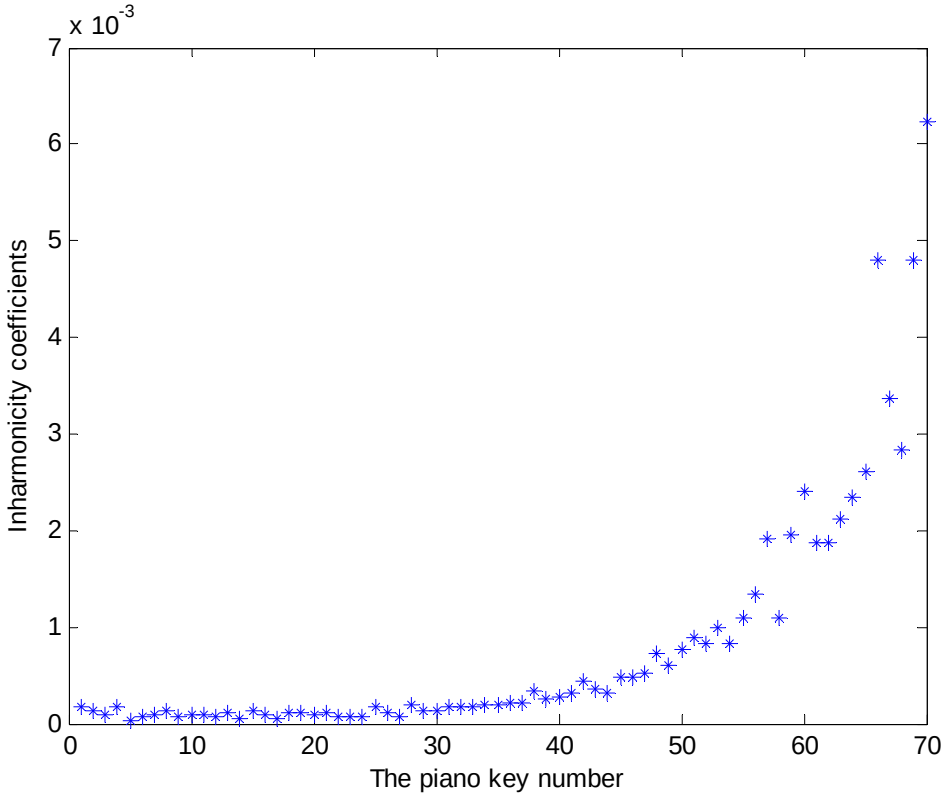


Fig 6.30 Estimated inharmonicity coefficients for some piano tones

Generally speaking, the inharmonicity coefficients increase with the fundamental frequency of piano tones. But the rate of increase with frequency in the inharmonicity coefficients for higher frequencies is high compared to the small rate of increase for lower frequencies [81]. The values of inharmonicity coefficients in Fig 6.30 are tabulated in Table 6.12. Each tone's fundamental frequency can be found in the Appendix A.

Galembo [64] has summarized inharmonicity coefficients for some different brands of piano. From the Table 6.11 we can see that for different pianos, the inharmoncity coefficient for the same tone could be very different. Comparing this

table to our results in Fig 6.30 (i.e., A0=182×10^{-6}, E1=125×10^{-6}, A1=106×10^{-6}, E2=112×10^{-6},A2=158×10^{-6}) we find our results are a little similar to that of a Steinway C piano, although results derived from different pianos may not be directly compared.

Note	Steinway D		Steinway C		Nordiska		Straud	
	Manual	Calculation	Manual	Calculation	Manual	Calculation	Manual	Calculation
A0	160	154	190	186	345	338	568	558
E1	065	059	091	090	207	198	263	277
A1	057	058	092	091	160	173	214	215
E2	068	067	114	112	142	131	235	233
A2	085	079	131	116	154	134	202	181

Table 6.11 The inharmonicity coefficients estimated by Galembo [64] unit:10^{-6}

Key No.	Tone	Inharmonicity	Key No.	Tone	Inharmonicity	Key No.	Tone	Inharmoicity
1	A0	0.000182	2	A0#	0.000125	3	B0	9.2e-005
4	C1	0.000166	5	C#1	3.8e-005	6	D1	7.5e-005
7	D#1	9.3e-005	8	E1	0.000125	9	F1	6.9e-005
10	F1#	8.2e-005	11	G1	8.5e-005	12	G1#	8.1e-005
13	A1	0.000106	14	A1#	4.9e-005	15	B1	0.000132
16	C2	9.3e-005	17	C2#	4.1e-005	18	D2	0.00011
19	D2#	0.000111	20	E2	0.000112	21	F2	8.2e-005
22	F2#	7.2e-005	23	G2	7.9e-005	24	G2#	6.2e-005
25	A2	0.000158	26	A2#	0.000112	27	B2	7.7e-005
28	C3	0.000187	29	C3#	0.000124	30	D3	0.000127
31	D3#	0.000165	32	E3	0.000177	33	F3	0.000182
34	F3#	0.000197	35	G3	0.000193	36	G3#	0.000221
37	A3	0.000219	38	A3#	0.000346	39	B3	0.000255

40	C4	0.000278	41	C4#	0.000312	42	D4	0.000441
43	D4#	0.000359	44	E4	0.000325	45	F4	0.000473
46	F4#	0.000479	47	G4	0.000529	48	G4#	0.000721
49	A4	0.0006	50	A4#	0.000768	51	B4	0.000888
52	C5	0.000823	53	C5#	0.00099	54	D5	0.00083
55	D5#	0.001089	56	E5	0.001337	57	F5	0.001915
58	F5#	0.001104	59	G5	0.00196	60	G5#	0.0024
61	A5	0.001867	62	A5#	0.001881	63	B5	0.002128
64	C6	0.002349	65	C6#	0.002607	66	D6	0.0048
67	D6#	0.003372	68	E6	0.002844	69	F6	0.004792
70	F6#	0.006234						

Table 6.12 The values of some piano tones' inharmonicity coefficients

Chapter 7 Conclusions and Suggestions for Future Work

7.1 Conclusions

The piano is one of the most important instruments used in western music, classical, pop, jazz and other genres. It produces sound by striking metal strings with felt covered hammers. The sound production mechanism is very complicated. However from another perspective, it is these complications that account for a large part of the interesting and well-loved nature of the piano sound.

In order to understand the complex nature of piano tones, some subtle effects that

can affect the sound production of these tones have been investigated in this dissertation. When the hammer strikes on the strings, the interaction of the hammer with the strings will surely strongly determine the sound's characteristics. This initial stage of the piano tone during which this interaction occurs is usually referred to as the onset transient or onset attack.

In this dissertation, we have explored the onset transients and inharmonicity of piano tones by wavelet-based techniques. Our findings have confirmed that the use of wavelets is an effective tool for analyzing piano tones.

Summing up, these findings are listed as follows.

In Chapter 3 where the multiresolution analysis was applied, we measured the onset durations of all 88 piano tones. Generally speaking, the onset transient durations of bass piano tones are around 100ms or greater, and for treble tones, they are much less, falling to about 10ms to 20ms. The onset transient durations of mid-range tones fall between that of the bass and treble tones. We also provided a possible explanation for why onset durations fall by a decreasing exponential-like trend with increasing frequency.

In our multiresolution analysis, we did not use a traditional inverse discrete wavelet transform to restore the signal where subbands are inherently organized. Each subband was instead reconstructed separately (or independently) from the other subbands. That means we are able to obtain the signal waveform associated with each subband. The whole process has been described in Fig 3.11. Although these subbands are separately reconstructed, summing the waveforms of these subbands results in a

perfect recovery of the original signal. This has been verified in experiments in Chapter 3.

The output of multiresolution analysis is a number of waveforms corresponding to a series of subbands which do not overlap each other and cover the whole frequency range from 0 Hz up to 22,050 Hz. The waveform of the lowest frequency band is the energy envelope of the signal. In this energy envelope, there is a surprising negative dip immediately before a large peak. The coefficient of the wavelet transform at this time point is expected to be very large and thus the wavelet basis is enlarged in amplitude to match the envelope peak. But since a wavelet basis must be an oscillatory function (see in Chapter 2), at the same time the positive peak of the wavelet basis is going to match the envelope peak, the negative peak of the wavelet basis leaves a negative dip at the start of the tone. Nevertheless, the negative dip seems to violate common sense that energy cannot take a negative value. This negative dip can actually be complemented by positive humps in the waveforms of upper subbands at the same time position.

Our results further indicate that this negative dip accurately points to the beginning of the piano tone and thus can be used as the starting point of that tone. Since the peak of a signal's waveform can be easily measured, the time interval between the negative dip and the peak is considered as the duration of the onset transient, according to the second definition described in Chapter 3.

In Section 4.2 of Chapter 4, where wavelet packet analysis was applied, we have analyzed the frequency band structure of onset transients for piano tones of various

fundamental frequencies. Our analysis may explain why bass piano tones are often perceived as having greater inharmonicity than mid-range and treble piano tones.

In comparison with the DWT, the wavelets packet transform (WPT) also iteratively subjects the *detail* coefficients to the basic decomposition unit, which could result in many possible tree structures. The number of different possible WPT trees can be very large. Using the Shannon entropy function as the cost function, Coifman has designed an algorithm to efficiently select the 'best' tree to represent any particular signal. According to Coifman's idea, the 'best' tree should have the lowest Shannon entropy, which means that the signal's energy is most efficiently represented.

The results of WPT are visualized in the form of the time-frequency plane whose pattern is determined by the resultant tree structure in the WPT. The time-frequency plane is an important concept and was extensively utilized in the remainder of this dissertation.

In Section 4.3 of Chapter 4, where local cosine bases analysis was applied, we have described the whole temporal process for the establishment of the harmonics in the time-frequency plane.

In Section 4.4 of Chapter 4, where matching pursuit analysis was applied, we have found that the large changes during the initial phase of the onset transient affect the fundamental frequency and lower harmonics more profoundly than the higher harmonics.

From the discussion on the time-frequency planes in Chapter 4 (including WPT

analysis, local cosine analysis, and matching pursuit analysis) we can draw a conclusion that is close to Palmer's opinion that during the onset transient of a piano tone, lower partials play a very important role, having much of the waveform's energy and undergoing great changes both in frequency and waveform. The peak amplitude is also primarily determined by the first few (i.e., the lowest frequency) spectral components.

These conclusions drawn from Chapter 4 provide a reasonable explanation as to why piano tones can be greatly compressed by wavelet packet transformation, as shown in Chapter 5. This large compression means that with only a relatively small number of time-frequency blocks (or time-frequency atoms) we can obtain an accurate representation of a piano tone's time-frequency behavior, as shown by our reconstruction of several piano tones in Chapter 5. The discussions in Chapter 5 greatly facilitate our measurements of inharmonicity coefficients in Chapter 6 by reducing the computational load to a viable level.

From the review in Chapter 1, we know that inharmonicity is another important feature in piano sounds. The root cause of inharmonicity is the fact that real strings inherently have nonzero bending stiffness.

In Chapter 6, we applied wavelet-based techniques to the measurement of the piano's inharmonicity coefficient. Using wavelet packet transform (WPT), the time-frequency plane can show how frequencies vary with time. Each time-frequency block in the plane represents a wave component which is the product of a coefficient with its associated wavelet basis. Because the energy of piano tones is concentrated

mainly in the lower harmonics (as shown in this dissertation), only a small proportion of all the time-frequency blocks are sufficient to accurately reconstruct the original waveform.

Each wave component is obtained by wavelet impulse synthesis and classified into a particular partial in terms of a series of analysis frequencies, thus allowing the estimation of the partial frequencies. After eliminating the 'partial shift' effect by a correction process, fundamental frequency and inharmonicity coefficients were accurately measured. The calculated results agree with FFT spectra of piano tones.

7.2 Suggestions for future work

The results in this research suggest several possibilities for future research:

- Onset detection and localization is very useful in a variety of applications in analyzing and indexing musical signals. Real music signals are a noisy polyphonic signals (i.e., multiple sound objects presented together at a given time). Thus, localizing a simple piano tone as discussed in Chapter 3 is not enough. Research may be required to extend wavelet multiresolution analysis to onset detection polyphonic signals. In the majority of existing onset detection approaches, various peak-picking algorithms are used with the help of detection functions. Without focusing on peaks, our method pays attention to negative dips in the waveform of the approximation subband resulting from wavelet multiresolution analysis on the signals' energy envelope. Therefore integrating

the wavelet multiresolution analysis into existing onset detection approaches might improve the accuracy.

- In Chapter 4 where wavelet time-frequency planes for the onset transients and the stationary part of piano tones are presented, the type of wavelet bases are Daubechies bases. Although Daubechies bases presented give satisfying results, other types of wavelet bases may disclose more properties of piano tones in terms of the time-frequency plane. Hence, more types of wavelet bases need to be tried for the purpose of comparison.
- The wavelet impulse synthesis in Chapter 5 and Chapter 6 has shown that the original signal can be additively reconstructed by a much smaller number of blocks selected from the total number of time-frequency blocks. In our discussion, the time-frequency blocks are selected according to the values of coefficients they are associated with. However, this may bring potential problems. Some partials that have weak amplitudes may have strong influence on the estimation of inharmonicity coefficients. Because of the weak amplitudes, they may not be selected and therefore ignored in the measurement. An improved method to select the appropriate blocks is necessary.

References

1. Benade, A.H., *Fundamentals of musical acoustics*. Second ed. 1990, New York: Dover Publications, INC.
2. Roads, C., *The Computer Music Tutorial*. 1995: The MIT Press.
3. Chowning.J, *The synthesis of complex audio spectra by means of frequency modulation.* Journal of the Audio Engineering Society, 1973. **21**(7): p. 526-534.
4. F.R.Moore, *Table lookup noise for sinusoidal digital oscillators.* Computer Music Journal, 1977. **1**(2): p. 26-29.
5. Beauchamp.J, A.Horner, and L.Haken, *Methods for multiple wavetable synthesis of musical instrument tones.* Journal of the Audio Engineering Society, 1993. **41**(5): p. 336-356
6. Cadoz, C. and A. Luciani. *Processes, Sound Synthesis Models and the Computer Designed for Musical Creation*. in *ICMC*. 1984.
7. Lansky, P., *Digital mixing and editing*. 1982: Princeton: Godfrey Winham Laboratory, Department of Music, Princeton University.
8. Freed, A., *MacMix: recording, mixing, and signal processing on a personal compute*, in *Music and Digital Technology*, J. Strawn, Editor. 1978, Audio Engineering Society: New York.
9. D.Jaffe and L.Boynton, *An overview of the sound and music kits for the NeXT computer.* Computer Music Journal, 1989. **13**(2): p. 48-55.
10. Seashore, C.E., *The Psychology of Music. III. The Quality of Tone: (1) Timbre.* Music Educators Journal, 1936. **23**(1): p. 24-26.
11. Helmholtz, H.v., *On the sensations of tone as a physiological basis for the theory of music*. 1954, New York: Dover Publications.
12. Bartholomew, W.T., *Acoustics of music*. 1942, New York: Prentice-Hall.
13. Corso, E.L.S.a.J.F., *Timbre cues and the identification of musical instruments.* Journal of the Acoustical Society of America, 1964. **36**(11): p. 2021-2026.
14. G.Stork, R.E.B.a.D., *The physics of sound*. 2nd ed. 1995: Prentice Hall.
15. Grey, J.M., *Multidimensional perceptual scaling of musical timbres.* J.Acoust.Soc.Am, 1977. **61**: p. 1270-1277.
16. J, V. and R.A. Rasch, *The perceptual onset of musical tones.* Perception and Psychophysics 1981. **29**: p. 323-335.
17. Lee, M.P.a.A.R., *Dispersion of waves in piano strings.* Journal of the Acoustical Society of America, 1980. **83**(1): p. 305-317.
18. R.W.Peters, B.C.J.M.B.R.G.a., *Relative dominance of individual partials in determining the pitch of complex tones.* Journal of the Acoustical Society of America, 1985. **77**(5): p. 1853-1860.
19. E.D.BlackHam, *The physics of the piano.* Scientific American, 1965. **213**(6): p. 88-96.
20. Mather, R., *Evaluation of a method for separating digitized duet signals.*

Journal of Audio Engineering Society, 1990. **38**(12): p. 956-979.
21. Luce.D.A. and Clark.M., *Physical Correlates of Brass-Instrument Tones.* J.Acoust.Soc.Am, 1967. **42**(6): p. 1232-1243.
22. Luce.D.A., *Dynamic Spectrum Changes of Orchestral Instruments.* Journal of Audio Engineering Society, 1975. **23**(7): p. 565-568.
23. Beauchamp, J.W., *Time-variant spectra of violion tones.* J.Acoust.Soc.Am, 1974. **56**(3): p. 995-1004.
24. Beauchamp, J.W., *Analysis and Synthesis of Cornet Tones Using Nonlinear Interharmonic Relationships.* Journal of Audio Engineering Society, 1975. **23**(10): p. 778-795.
25. Risset, J.C. and M.V. Mathews, *Analysis of Musical-Instrument Tones.* Physics Today, 1969. **22**(2): p. 23-30.
26. Anderson, B.E. and W.J. Strong, *The effect of inharmonic partials on pitch of piano tones.* J.Acoust.Soc.Am, 2005. **117**(5): p. 3268-3272.
27. Mathews, M., J.Miller, and E.David, *Pitch synchronous analysis of voiced sounds.* Journal of Audio Engineering Society of America, 1961. **33**: p. 179-186.
28. Freeman, M.D., *Analysisof musical instrument tones.* Journal of the Acoustical Society of America, 1967. **41**: p. 793-806.
29. Moorer, J.A., *The optimum comb method of pitch period analysis of continuous digitized speech.* AIM-207.Standford: Stanford Artificial Intelligence Laboratory, 1973.
30. Beauchamp, J., *Data reduction and resynthesis of connected solo passages using frequency, amplitude and brightness detection and the nonlear synthesis technique.* Proceedings of the 1981 International computer music Conference, 1981: p. 316-323.
31. Portnoff, M., *Implementation of the digital phase vocoder using the fast Fourier transform.* IEEE Transactions On Acoustics, Speech and Signal Processing, 1976. **24**(3): p. 243-248.
32. Moore, F.R., *An introduction to the mathematics of digital signal processing. Part 1: algebra, trigonometry, and the most beautiful formula in mathematics.* Computer Music Journal, 1978a. **2**(1): p. 38-47.
33. Moore, F.R., *An introduction to the mathematics of digital signal processing. Part 2: sampling, transforms, and digital filtering.* Computer Music Journal, 1978b. **2**(2): p. 38-60.
34. Delson, M., *The phase vocoder: a tutorial.* Computer Music Journal, 1986. **10**(4): p. 14-27.
35. Flanagan, J.L. and R.M. Golden, *Phase Vocoder.* Bell System Technical J, 1966. **10**(4): p. 1493-1509.
36. McAulay, R.J. and T.F. Quatieri, *Speech Analysis/Synthesis Based on a Sinusoidal Representation.* IEEE Transactions On Acoustics, Speech and Signal Processing, 1986. **ASSP-34**(4): p. 744-754.
37. Maher, R. and J.W. Beauchamp, *An investigation of vocal vibrato for synthesis.* Applied Acoustics, 1990. **30**: p. 219-245.

38. Brown, J.C., *Calculation of a constant Q spectral transform.* J.Acoust.Soc.Am, 1991. **89**(1): p. 425-434.
39. Beauchamp, J.W., *Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music*. 2007: Springer.
40. Brown, J.C. and M.S. Puckette, *An efficient algorithm for the calculation of a constant Q transform.* J.Acoust.Soc.Am, 1992. **92**(5): p. 2698-2701.
41. Mallat, S., *A theory for multiresolution signal decomposition: the wavelet representation.* IEEE Pattern Anal. and Machine Intell., 1989. **11**(7): p. 674-693.
42. Kronland-Martinet, R., *The wavelet transform for the analysis, synthesis, and processing of speech and music sound.* Computer Music Journal, 1988. **12**(4): p. 11-20.
43. Evangelista, G., *Pitch-synchronous wavelet representations of speech and music signals.* IEEE Transactions On Signal Processing, 1993. **41**(12): p. 3313-3330.
44. Berger, J., R.R. Coifman, and M.J. Goldberg, *Removing noise from music using local trigonometric bases and wavelet packets.* Journal of Audio Engineering Society, 1994. **42**(10): p. 808-818.
45. Wannamaker, R.A. and E.R. Vrscay, *Fractal Wavelet Compression of Audio Signals.* Journal of Audio Engineering Society, 1997. **45**(7/8): p. 540-553.
46. Arfib, D., *Analysis, transformation, and resynthesis of musical sounds with the help of a time-frequency representation*, in *Representations of Musical Signals*, G. De Poli, A. Piccialli, and C. Roads, Editors. 1991, The MIT Press. p. 87-118.
47. Hiller, L.R., P, *Synthesizing Musical Sounds by Solving the Wave Equation for Vibrating Objects.* Journal of the Audio Engineering Society, 1971. **19**(6): p. 462-470.
48. Karplus, K.S., A, *Digital synthesis of plucked-string and drum timbres.* 1983. **7**(2): p. 43-55.
49. III, J.O.S., *Physical Modeling using digital waveguides.* Computer Music Journal, 1992. **16**(4): p. 79-91.
50. A.Horner, *Wavetable matching synthesis of dynamic instruments with genetic algorithms.* Journal of the Audio Engineering Society, 1995. **43**(11): p. 916-931.
51. Palamin, J.-P.P., P.; Ronveaux, A, *A Method of Generating and Controlling Musical Asymmetrical Spectra.* Journal of the Audio Engineering Society, 1988. **36**(9): p. 671-685.
52. Tan, B.T.G.a.G., S. L, *Real-Time Implementation of Asymmetrical Frequency-Modulation Synthesis.* Journal of the Audio Engineering Society, 1993. **41**(5): p. 357-363.
53. Schottstaedt, B., *The Simulation of Natural Instrument Tones Using Frequency Modulation with a Complex Modulating Wave.* Computer Music Journal, 1977. **1**(4): p. 46-50.
54. LeBrun, *A Derivation of the Spectrum of FM with a Complex Modulating*

Wave. Computer Music Journal, 1977. **1**(4): p. 51-52.
55. Tan, B.T.G., Gan, S. L., Lim, S. M. and Tang, S. H., *Real-Time Implementation of Double frequency Modulation (DFM) Synthesis.* Journal of the Audio Engineering Society, 1994. **42**(11): p. 918-926.
56. Askenfelt, A. and E.V. Jasson, *From Touch to Vibrations III: String motions and spectra.* Journal of the Acoustical Society of America, 1993. **93**(4): p. 2181-2196.
57. Conklin, H.A., *Design and Tone in the Mechanoacoustic Piano Part I. Piano Hammers and Tonal Effects.* Journal of the Acoustical Society of America, 1996. **99**(6): p. 3286-3296.
58. Fletcher, N.H. and T.D. Rossing, *The Physics of Musical Instruments*. 1998, New York: Springer-Verlag.
59. Blackham, E.D., *The physics of the piano.* Sci.Am., 1965. **213**(6): p. 88-96.
60. Fletcher, H., E.D. Blackham, and R. Stratton, *Quality of piano tones.* J.Acoust.Soc.Am, 1962. **34**: p. 749-761.
61. Rasch, A. and V. Heetvelt, *String inharmonicity and piano tuning.* Music Percept.,, 1985. **3**: p. 171-190.
62. Lattard, J., *Influence of inharmonicity on the tuning of piano-Measurements and mathematical simulation.* J.Acoust.Soc.Am, 1993. **94**: p. 46-53.
63. Jarvelainen, H., V. Valimaki, and M. Karjalainen. *Audibility of inharmonicity in string instrument sounds, and implications, to digital sound synthesis.* in *Preceedings of the international computer music conference Beijing, China*. 1999.
64. Galembo, A. and A. Askenfelt, *Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano.* IEEE Transactions on speech and audio processing, 1999. **7**(2): p. 197-203.
65. Rauhala, J., H.-M. Lehtonen, and V. Välimäki, *Fast automatic inharmonicity estimation algorithm* J.Acoust.Soc.Am Express letters, 2007. **121**(5).
66. Askenfelt, A. and A. Galembo, *Study of the spectral inharmonicity of Musical Sound.* Acoust. Phys, 2000. **46**(2): p. 121-132.
67. Klapuri, A., *Multiple fundamental frequency estimation based on harmonicity and spectral smoothness.* IEEE Transactions on Speech Audio Process, 2003. **11**(6): p. 184-194.
68. Boggess, A. and F. J.Narcowich, *A First Course in Wavelets with Fourier Analysis*. 2001: Prentice Hall.
69. Daubechies, I., *Ten lectures on wavelets*. 1992, Philadelphia: Society for Industrial and Applied Mathematics.
70. Keiler, F., *Analysis of transient musical sounds by auto-regressive modeling.* Porc. of the 6th Int. Conference on Digital Audio Effects, 2003.
71. Jensen, A. and A. La Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet*. 2001: Springer.
72. Coifman, R.R. and M.V. Wickerhauser, *Entropy-based algorithms for best basis selection.* IEEE Transactions On Information Theory, 1992. **38**(2): p. 713-718.

73. D.Donoho, *WAVELAB 802*. 1999, Standford University.
74. Winckel, F., *Music, Sound and Sensation: A Modern Exposition*. 1967: Dover Publications, INC.
75. Mallat, S., *A Wavelet Tour of Signal Processing*. 1997: Academic Press.
76. Mallat, S. and S. Zhang, *Matching Pursuits with Time-Frequency Dictionaries.* IEEE Transactions On Signal Processing, 1993. **41**(12): p. 3397-3415.
77. Palmer, C. and J.C. Brown, *Investigations in the amplitude of sounded piano tones.* J.Acoust.Soc.Am, 1991. **90**(1): p. 62-66.
78. Fletcher, H., *Normal vibration frequencies of a stiff piano string.* J.Acoust.Soc.Am, 1964. **36**(1): p. 203-209.
79. Galembo, A. and A. Askenfelt, *Perceptual relevance of inharmonicity and spectral envelope in the piano bass range.* Acta Acustica united with Acustica, 2004. **90**: p. 528-536.
80. Wang, E., *Application of Wavelets to Onset Transients and Inharmonicity of Piano Tones.* Journal of Audio Engineering Society, 2008. **56**(5).
81. H. A. Conklin, J., *Piano design factors --- their influence on tone and acoustical performance*, in *Five Lectures on the Acoustics of the Piano*, A. Askenfelt, Editor. 1990.

Publication

Wang, E. and B.T.G Tan, *Application of Wavelets to Onset Transients and Inharmonicity of Piano Tones.* Journal of Audio Engineering Society, 2008. **56**(5).

Appendix A

Key	name	Frequency (Hz)	Key	name	Frequency (Hz)
88	C8	4186.01	44	E4	329.628
87	B7	3951.07	43	D4#	311.127
86	A7#	3729.31	42	D4	293.665
85	A7	3520.00	41	C4#	277.183
84	G7#	3322.44	40	C4 (Middle C)	261.626
83	G7	3135.96	39	B3	246.942
82	F7#	2959.96	38	A3#	233.082
81	F7	2793.83	37	A3	220.000
80	E7	2637.02	36	G3#	207.652
79	D7#	2489.02	35	G3	195.998
78	D7	2349.32	34	F3#	184.997
77	C7#	2217.46	33	F3	174.614

76	C7 (Double high C)	2093.00	32	E3	164.814
75	B6	1975.53	31	D3#	155.563
74	A6#	1864.66	30	D3	146.832
73	A6	1760.00	29	C3#	138.591
72	G6#	1661.22	28	C3 (Low C)	130.813
71	G6	1567.98	27	B2	123.471
70	F6#	1479.98	26	A2#	116.541
69	F6	1396.91	25	A2	110.000
68	E6	1318.51	24	G2#	103.826
67	D6#	1244.51	23	G2	97.9989
66	D6	1174.66	22	F2#	92.4986
65	C6#	1108.73	21	F2	87.3071
64	C6 (Soprano C)	1046.50	20	E2	82.4069
63	B5	987.767	19	D2#	77.7817

62	A5#	932.328	18	D2	73.4162
61	A5	880.000	17	C2#	69.2957
60	G5#	830.609	16	C2 (Deep C)	65.4064
59	G5	783.991	15	B1	61.7354
58	F5#	739.989	14	A1#	58.2705
57	F5	698.456	13	A1	55.0000
56	E5	659.255	12	G1#	51.9130
55	D5#	622.254	11	G1	48.9995
54	D5	587.330	10	F1#	46.2493
53	C5#	554.365	9	F1	43.6536
52	C5 (Tenor C)	523.251	8	E1	41.2035
51	B4	493.883	7	D1#	38.8909
50	A4#	466.164	6	D1	36.7081
49	A4 (A440)	440.000	5	C1#	34.6479
48	G4#	415.305	4	C1 (Pedal C)	32.7032

47	G4	391.995	3	B0	30.8677
46	F4#	369.994	2	A0#	29.1353
45	F4	349.228	1	A0	27.5000