RECOGNITION OF NONSTATIONARY SIGNALS WITH PARTICULAR REFERENCE TO THE PIANO

Ana Cristina Pereira Rosa Pascoalinho Carrasco

A thesis submitted for the degree of

Doctor of Philosophy

Supervisor

Prof. J.K. Hammond

University of Southampton
Faculty of Engineering and Applied Science
Institute of Sound and Vibration Research
Sponsored by FCT (Portugal) and FSE

September 2002

ACKNOWLEDGEMENTS

I wish to thank my supervisor Professor Joe Hammond for his supervision inestimable guidance.

I am grateful to Dr. Paul White for his advices.

I am also grateful to the staff of the ISVR for their help.

Finally, I would like to express my gratitude to my parents, brother and my niece and nephew for all the support during this research.

CONTENTS

Glossary 6

CHAPTER 1 **Introduction** 8

CHAPTER 2 **Automatic music transcription: a review** 12

2.1 The problem of automatic music transcription 12
2.2 The early history of polyphonic automatic music transcription 14
2.3 The psychoacoustics and signal processing approach to music transcription 15
2.4 The signal processing approach 16
2.5 The choice of KLE 17
2.6 Conclusion 18

CHAPTER 3 **The piano** 20

3.1 The morphology of the piano 20
3.1.1 Description of the piano 21
3.2 Time histories and spectra of piano notes 24
3.3 Piano note production 32
3.4 Modelling of a piano note as a nonstationary signal 36

CHAPTER 4 **The Karhunen-Loève Expansion (KLE)** 39

Introduction 39
4.1 The Karhunen-Loève expansion (KLE) 40
4.2 The discrete version of the KLE 44
4.3 How the KLE is applied in practice 46
4.4 The application of the KLE to nonstationary signals 47
4.4.1 The nonstationary autocorrelation function 47
4.4.2 Theoretical nonstationary autocorrelation function of a simulated signal 49

4.4.3 Nonstationary autocorrelation matrix estimated by ensemble averaging for modelled piano signal 51
4.4.4 The eigen decomposition of the nonstationary autocorrelation matrices 53
4.5 Nonstationary versus stationary data 55
4.6 The KLE of a Wiener process 59

CHAPTER 5 **The Karhunen-Loève Expansion (KLE) and the Singular Value Decomposition (SVD)** 62
5.1 The SVD Theorem 62
5.1.1 Singular Value Decomposition 63
5.2 The relation between SVD and KLE 66
5.3 How SVD is applied in practice and its relation with KLE 68
5.4 Examples of KLE and SVD similar results for simulated signals 69
5.5 Relation between the SVD, KLE and the Discrete Fourier transform 73
5.6 Relation between the SVD, KLE and the Discrete Fourier transform 76

CHAPTER 6 **The application of SVD to simulated piano signals** 78
Introduction 77
6.1 The representation of a nonstationary sine wave in singular value space 77
6.2 Factors that influence the shape of the singular vectors of a nonstationary sine wave 84
6.2.1 The initial phase 85
6.2.2 The frequency 87
6.3 The modelled piano signal and the influence the variables on the SVD 89
6.3.1 Influence of the damping on the modelled signal 92
6.3.2 Influence of the frequency on the modelled signal 94
6.3.3 The influence of the delay in the modelled signal 95
6.4 SVD of nonstationary signals 97
6.4.1 Uniformly modulated noise 98
6.5 Conclusions 99

CHAPTER 7 **SVD and recognition process of simulated piano notes** 100
Introduction 100
7.1 Recognition of modelled piano notes 100
7.1.1 The recognition process 102
7.2 The projection of a single note with only one

frequency component on SVD basis sets 103
7.3 Projections using delay 108
7.4 Projection of an octave on the SVD basis sets of the component notes 111
7.5 The norm of the projections 118
7.5.1 The norm of the projections of a simulated signal in a SVD basis set 119
7.6 Recognition of single notes with more than one frequency component 120
7.7 Conclusions 129

CHAPTER 8 **Application of SVD and a recognition process to real piano notes** 132

Introduction 132
8.1 The recognition process 132
8.2 The SVD of single piano notes 136
8.2.1 The SVD of a piano note with all the events aligned 136
8.2.2 The SVD of a piano note with delays between events 141
8.3 Recognition of a single note 144
8.4 Recognition of octaves 149
8.5 Recognition of notes separated by an interval of a tone 157
8.6 Recognition of notes in chords 158
8.6.1 Major chords with three notes 159
8.6.2 Major chords with four notes 165
8.6.3 The seventh chords with four notes 169
8.7 Conclusion: The performance of the method 172

CHAPTER 9 **Conclusion** 176

9.1 Summary of results 176
9.2 Possible future work 178

Annex 1 Pitch identification with an auditory model 180
Annex 2 Pitch identification by summary autocorrelation function 187
Annex 3 Proof of the Karhunen-Loève expansion 191
Annex 4 The KLE and SVD of stationary and nonstationary signals 194
Annex 5 The modulation of the singular vectors of a nonstationary sine wave 196
Annex 6 Equal Temperament 199
Annex 7 Data acquisition 201

References 202
Bibliography 204

Chord

Set of musical notes that are played together.

Interval

Musical distance between two notes.

Harmonic series

Set of musical notes whose frequency is a integer multiple of the fundamental.

Major Chord

Set of three notes played simultaneously taken form the harmonic series. One of the notes is the fundamental and the other two are separated by intervals of 3rds.

Monophonic Music

Music based on a single melody or single line of sound. Monophony is typically performed by a single vocalist or instrumentalist, or by a group of vocalists/instrumentalists performing in unison.

Note

A signal produced by a musical instrument composed of a fundamental frequency and all the harmonics.

Octave

Set of two notes played simultaneously. Also, interval between two notes whose consecutive notes labelled with the same name and whose frequencies of the fundamental, f1 and f2, component are related by f2=2*f1.

Polyphonic Music

Music based on two or more concurrent independent melodies or lines of sound. Polyphonic can be performed by a single instrument (e.g. piano) or by a group of instruments.

Pitch

An attribute of auditory sensation by which sounds may be ordered from low to

high (Rossing,1990).

Seventh Chord

Set of four notes played simultaneously taken form the harmonic series. One of the notes is the fundamental and the other two are separated by intervals of 3rds.

Timbre

An attribute of auditory sensation by which two sounds with the same loudness and pitch can be judge dissimilar (Rossing,1990).

CHAPTER 1

INTRODUCTION

Since 1970 the automatic transcription of music has started to attract interest. This is a multidisciplinary task linking knowledge of different areas including music, psychoacoustics and signal processing.

The transcription of instruments playing a single musical line, such as a flute playing a solo has already been achieved by many systems with great accuracy. There are on the market several different software programs that perform music transcription with satisfactory results. When we refer to music transcription generally we mean to write a score, the corresponding note pitches, from a live or recorded performance. Musical scores contain different information: the notes, the tempo of each note, the tonality, the instruments playing, dynamics and other information that it is not written and it is there that the musicians put their unique interpretation. The systems of automatic music transcription that we refer to generally indicate the notes, the duration of each one and sometimes the beat. That is generally enough for the musicians to play the music.

When there is more than one instrument playing the music is more complex and the transcription becomes a very difficult task. Even when we have only one instrument playing, such as the piano, where several notes can be played at the same time the transcription is very difficult. Music where more than one musical line is played at the same time is called polyphonic. It can be played by a single instrument such as the piano and the guitar, or by several instruments as in the case of an orchestra. There is not much

research work in this area. There are several attempts to perform polyphonic music transcription and the results are very far from satisfactory.

One of the great problems in the automatic transcription of polyphonic music is the recognition of the instruments that are being played in a musical piece. Another is to recognize notes that have coincident harmonics. In many cases some of the notes played have all their harmonics coincident with other notes. This last case is the most persistent in automatic transcription. From the earlier research works the detection of notes with coincident harmonics has seemed a practically impossible task. This problem is very interesting because it implies the *separation* of notes with coincident harmonics. All the approaches to solve this problem have been based on Fourier analysis.

In this thesis we analyse the possible implementation of a transform for automatic music transcription that it is not Fourier based. The purpose of this research is to attempt to solve one of the oldest problems in the automatic music transcription history namely the recognition of the notes when separated by an octave.

In this chapter we provide a brief introduction to the motivation for the work and an outline of the structure of the thesis. In chapter two we give a review of automatic music transcription research and the different approaches. From analysis of the different work we conclude that the Karhunen-Loève Expansion (KLE) offers a very promising approach for the purpose. The KLE was previously only used in musical instrument synthesis (not including the piano) and never in automatic music transcription.

The instrument used to perform the research on polyphonic music transcription was the piano. In chapter three we do a general description of the piano morphology and an analysis of the piano sounds. We also present a simple model for the piano notes for a first approach to the use of the KLE in music transcription. The piano notes are modelled as nonstationary stochastic transients.

In chapter four we describe of the KLE and how it is applied in practice to stationary signals. Generally the KLE is applied to stationary signals and weakly stationary signals. We apply the KLE to our simulated nonstationary signals estimating the nonstationary autocorrelation function theoretically and also by ensemble averaging. We present results showing the difference between the KLE of nonstationary signals using the stationary and the nonstationary autocorrelations.

In chapter five another mathematical transform the singular value decomposition (SVD) is analysed. The interest in SVD for a possible application to this research is based on its more convenient formalism than that of the KLE. For nonstationary data

represented by an ensemble it is more convenient to use SVD since it is applied directly to the time histories. We establish that the resulting singular vectors and eigen vectors of both transforms of a signal are similar.

In chapter six we apply SVD to the simulated piano signals and analyse how the different parameters of the signals influence the resulting decomposition. The piano notes are simulated as stochastic transient signals. For each signal simulating a piano note we generate a number of possible events and apply SVD to the time history matrices.

In this thesis KLE and SVD, generally applied to stationary or weakly stationary signals, are applied to nonstationary signals. We used a different approach for KLE and SVD for the nonstationary case. To determine the KLE of a nonstationary signal we calculate the theoretical or estimate by ensemble average the nonstationary autocorrelation function. For SVD we use the time histories matrices composed by possible events of the signal. The results show that the characteristics of the nonstationary signals are preserved in the first singular vectors which is not the case for stationary signals.

In chapter seven we develop a recognition process, using SVD, for simulated notes that can be extrapolated to the real piano notes. The process is applied to single simulated notes and to octaves to analyse the performance of the method. For each simulated note we determine the SVD. The resulting orthonormal singular vectors define a new space of each signal and are the reference for that note in the recognition process. A possible realisation of the signal represented in this new space uses fewer vectors than for any other signal. For the recognition process we project a signal on the spaces defined by the singular vectors of the simulated piano notes. These projections are analysed to extract information about a possible recognition of the signals and application to real piano notes.

In chapter eight the recognition process developed on the previous chapter is applied to real piano notes. First the process is used in the recognition of single piano notes and then octaves and finally chords composed of three and four notes. For the real case we recorded a number of possible realisations of each note and determined the SVD. The singular vectors were used as reference for the recognition of notes in the piano signals. Those signals were projected on the basis sets defined by the singular vectors and according to the analysis of the results we seek to determine the presence of a note in the signals.

Chapter nine has the conclusions of the research and considerations about future work.

CHAPTER 2

AUTOMATIC MUSIC TRANSCRIPTION: A REVIEW

2.1 The problem of automatic music transcription

In the last few decades the automatic transcription of polyphonic music has been attracting interest. Automatic music transcription is a multidisciplinary task since it involves such areas as psychoacoustics, music theory and signal processing.

We can define automatic music transcription as a process where a musical piece is converted into a musical score. The information written in a conventional musical score includes, as an example, the musical notes played by an instrument or several instruments, the indication of which instruments are playing, the duration of each note, the measure, the dynamics, the way of playing a note or a group of notes. There is other subtle information involved on the score that is not directly written. That information is usually transmitted between musicians. Even with all the information in a score the degree of liberty in the interpretation of the score is high. On the shelves of the music stores we can see a great number of interpretations of the same musical piece and each one has its unique characteristics such that in many cases is possible to identify the interpreter.

Why is there interest in automatic music transcription? In what is called classical music all the musical pieces are performed over a written score. The written score allows

a good reproduction of all the intentions of the composers and a new interpretation of the musicians. We cannot forget that the score of one composition is nothing more than a symbolic representation of the intentions of the composer. The existence of a score provides a way of studying and analysing the music and provides a tool for a new interpretation of the score. Folk music, world music and most recent styles of music such as rock, pop and jazz have no tradition of a written score. The case of jazz is a bit different since the themes are written as the harmonic structure but not the improvisations. The existence of a score of those styles of music would help the theoretical study of the music and thousands of professional and amateur musicians that would like to play the music of the most famous artists and groups, but do not have enough skill or time to perform the transcription of the music. To show how the interest in the transcription of music has grown, on the internet there are now at least 10 pages related to people offering the transcription of music at approximately 100$(USD) per song. The existence of a program performing an automatic music transcription would indeed be very useful.

When we are referring to automatic music transcription we are, at this stage, searching mostly for a process that is able to detect the duration and the notes played in a musical piece. Before referring to the research performed on automatic music transcription we must separate music in two categories: monophonic music, where only one musical voice is played and polyphonic music where more than one musical voice is played by one single instrument or several instruments. One example of monophonic music is when we have a single instrument playing one note at time (e.g. flute, oboe). The greater part of the music played is polyphonic and an example of polyphonic music that we can give is an orchestra playing. A single instrument can also play polyphonic music. That is the case of the guitar and piano where is possible to play several notes at the same time.

Automatic music transcription in the case of monophonic music has achieved good results and there are many software programs in the market with very satisfactory results. In the case of polyphonic music automatic transcription is only in the early stages. A trained musician can easily identify the instruments playing in an orchestra and the respective lines they are playing. When trying to perform that automatically the great problem is to identify the number of instruments playing or the musical voices, the different type of instrument and the notes played simultaneously with coincident harmonics. The last case is one of the great problems in note identification. A musical

interval whose notes have coincident frequency components is the octave. An interval of an octave, in musical terms, is defined as the distance between two consecutive notes with the same name, whose frequency relation is 2:1. The automatic music transcription performed till now fails to separate these two signals.

2.2 The early history of polyphonic automatic music transcription

As the automatic transcription of polyphonic is a difficult task that it is not even near to being solved there are several different approaches to trying to solve the problem. We can divide the research into two groups: one group includes psychoacoustics and signal processing techniques and the other only includes signal processing. Both of them insert some musical knowledge or theory. The approach to the transcription is different in these two fields of research. In the first one the bases of all the transcription process is the simulation of the ear mechanism for sound analysis and in the second the musical signals are directly analysed with signal processing techniques.

The first research on automatic music transcription was performed by Moorer in 1977. The research consisted of the transcription of polyphonic music with two voices played by two different instruments that should not cross, meaning that the fundamental frequency of one instrument should never be greater than the one of the other instrument. So the track of each of the instruments was not a problem in this system since the musical lines of the instruments had different frequency ranges. Moorer used band pass filters organised in a comb harmonic structure. The frequency of the fundamentals was estimated by autocorrelation. The system was able to detect all the notes played since there were no intervals where the higher notes had the fundamental frequency coincident with a harmonic of the lower one. The octave is one of these intervals and Moorer's system was not able to detect the notes separated by this interval. In this first work on automatic music transcription the problem of the detection of the notes in an octave arose that today it is not quite solved.

In 1990, Maher performed polyphonic music transcription of a duet. The two instruments of the duet where monophonic and their musical lines where not allowed to

cross. The technique used to analyse the musical signal of the duet was the short time Fourier transform (STFT). The results show that the system for transcription worked satisfactorily subject to the restrictions imposed but again the detection of notes with most of the harmonics coincident with the harmonics of other voice failed. The detection of notes with coincident harmonics with another note has been a problem since the first research on polyphonic music transcription.

2.3 The psychoacoustics and signal processing approach to music transcription

The majority of this approach to music transcription is very much concerned with the ear mechanism. The research is based on psychoacoustics theories that try to explain how the ear works when a sound reaches it and how we perceive it. Those theories are used as a basis for the models development. Those models are built using signal processing techniques.

Meddis and Hewitt (1991) proposed a model with a processing sequence where the first stages simulate the behaviour of the outer and middle ear. One of the most important stages simulates the basilar membrane mechanism using a filter bank. The output of this stage is called the cochleagram. After the basilar membrane processing the next stage produces the autocorrelograms for the different channels that represent the nerve fibres of the middle ear. The autocorrelograms are aggregated to produce a summary autocorrelation function. According to Meddis and Hewitt (1991) the autocorrelation function contains the necessary information for the purpose of simulation the human listener. In annex 1 we simulated the stage of the basilar membrane according to Meddis and Hewitt's work. In annex 2, we present the autocorrelograms resulting form the output of the filter bank representation the basilar membrane. The model was tested with a musical major C chord with three notes. The summary autocorrelation function did not give information about the notes of the chord but presented a peak corresponding to note that has the three notes in the harmonic series that is the C two octaves below the C note of the chord. This model needs a great amount of ear model theory and its application needs a great amount of computation.

Martin (1996) again used the autocorrelograms and the summary autocorrelation function as a basis for music transcription. The problems that Meddis and Hewitt had are still present since the signal processing method used was the same. Martin added to the model a blackboard control structure that included a knowledge music source. The blackboard structure includes different knowledge to perform the transcription. The first notes of a Bach's chorale where identified but the system was not able to recognise more notes due lack of musical knowledge on the blackboard. The recognition relies mostly on the musical knowledge included on the blackboard. Since there are many musical styles, and new ones will be created the knowledge source has to be very large and complex due to all the amount of information needed. According to these two researches the correlation methods were able to determine the root of the chord that corresponds to the first note of the musical harmonic series. Some experiments with chords should have been tested to sustain this assumption.

In both researches the method used was the autocorrelation. The results seem to indicate that autocorrelation methods by themselves do not recognise the notes in a chord but they detect the note that is the fundamental of the harmonic series of the chord. There are more chords that have exactly the same note as the fundamental and the results do not seem to indicate that it is possible to distinguish between them.

2.4 The Signal Processing Approach

The Fourier transform is the most used technique used for pitch detection. The problem with the Fourier techniques is that widely they cannot separate signals with harmonics with exactly the same frequency since the Fourier transform is a linear transform where a signal is decomposed on the linear summation of sines and cosines.

The two early works on automatic music transcription mentioned in section 2.2 are in the category of the signal processing approach. Kashino et al (1995) used the blackboard structure in the polyphonic music transcription. The signal processing technique used was the sinusoidal analysis. He determined the partials in the signal and then determined the notes and chords. In this blackboard he integrated a priori knowledge in human hearing and music. He also introduced a Bayesian network to determine the transitions of the chords.

Klapuri (2001) presents a music transcription system that could be applied to real audio recordings. The multipitch estimator is the principal algorithm of the system. This algorithm first estimates the pitch through the analysis of the harmonic components in the spectrum. The spectrum of a detected sound is determined and subtracted from the total spectrum. The process was tested with MIDI data and there is an example of what Klapuri considers a relatively well transcribed song. The drums are not represented in the score. One of the problems found is the detection of a long duration note. The same note can be detected several times.

Walmsley (1999) et al, analysed harmonic signals with a Bayesian modelling framework with a priori information about the musical signals. The data is separated in frames where it is assumed as stationary. The data in each frame is modelled as the sum of a number of musical notes. They assume that the parameters, such as pitch, of the musical signals are highly correlated between frames. The parameters are estimated using Markov chain Monte Carlo (MCMC) methods. The results present errors in the detection of octaves and fifths. Some chords are detected successfully with this model.

An interesting study was performed by Stapleton & Bass (1988). This research was not directly related to automatic music transcription. The subject was the synthesis of musical tones. They used a correlation based technique: the Karhunen-Loève expansion (KLE) to synthesise musical notes of different instruments. The purpose was to replace the sinusoids of the Fourier transform by this expansion basis set of orthogonal functions of the signal for musical instruments synthesis. They took segments from the sustain regime with characteristics of a steady state of the signals. They did not apply the KLE to the transient part of the signals corresponding to the attack. The synthesis of the instruments (e.g. French horn, bassoon) according to Stapleton and Bass was of very high level of natural quality using a small number of vectors of the basis since the energy is mostly concentrated on the first vectors of the basis set (we note that the piano was not included in the research of Stapleton and Bass). The interest in this expansion for music transcription was related to the offer of a different basis set from the Fourier transform. In this way some of the typical problems of music transcription, such as the detection of the notes separated by an interval of an octave, could be solved.

2.5 The choice of KLE

As we have seen, since the beginning of the automatic music transcription, one of the great problems of music transcription is the detection of signals with coincident harmonics. All the research generally has problems when transcribing music with intervals such as the octaves where one of the notes has all the harmonics coincident with the lower frequency (fundamental) note. The approaches for transcription even for the psychoacoustics case are generally Fourier based. Since in the Fourier transform the resulting basis is a set of sine and cosines, the separation of two signals where one of them has all its harmonics coincident with the other signal is not possible. The results of the presented researches fail in the detection of notes with coincident harmonics.

The existence of an expansion that offers a basis set not based on sine and cosine could be an approach to try to solve this problem. The KLE is not a Fourier based transform. As we will show later in chapter 5, the Fourier transform is a special case of the KLE. The basis set of the Fourier space is always the same for all signals: sine waves (or cosines). In the KLE the basis sets are defined by the signal itself through the eigenvectors of the autocorrelation function of the signal. Each signal has a characteristic basis set. We will analyse if this characteristic of the expansion is suitable for application to musical signals and automatic music transcription.

Another important feature is that the KLE can also be applied to stationary and nonstationary signals. By definition of the KLE it can be applied to any signal with finite second order moment, so a transient signal has a KLE. In our research we will analyse piano signals that will be treated as nonstationary.

These two features of the expansion, the basis set defined by the signal itself and the applicability to nonstationary signals makes it attractive from the point of view of automatic music transcription.

2.6 Conclusions

The automatic transcription of polyphonic music is until now far from satisfactory results. There are many variables involved, such as the different instruments playing and

the correct identification, notes with coincident harmonics, that makes the task difficult. Trained musicians are able to perform the transcription of polyphonic music. Automatic systems until now are far behind the human ear.

CHAPTER 3

THE PIANO

3.1 The morphology of the piano

The piano is defined in musicology as a percussion instrument. This classification has its basis in the way a piano note is produced. When one plays a piano note a key that is connected to a hammer is depressed. This hammer strikes a string or a set of strings that transmit the vibration to a soundboard.

The piano that we know today has a history of approximately three hundred years. It is far from the first version presented by Bartolomeo Cristofori at the beginning of the XVIII century. But the generic idea of the percussion of the strings continues. Previous keyboard instruments, such as the harpsichord, generated the sound through the use of a plectrum to pluck the string. The player was not able to control the dynamics of each note played since the key action mechanism was not sensitive to the dynamics applied on the key. With this new instrument, the piano, the strings were struck instead of plucked and the resulting sound could have different dynamics depending on the force that the musician used on the keys. Through the years many modifications were applied to the initial piano and the sound of the instrument today is quite different. Even now there are innovations in the piano construction trying to improve the quality of the sound and the resulting performances.

Since the first piano was built it became very popular. Pianos are associated with nearly all the cultures and have been incorporated in many music styles from classical to pop and jazz. One of the reasons for this popularity is that almost any kind of music for almost any situation can be performed on the piano. Big orchestra scores can be easily reduced for a piano performance, which shows its versatility. One important characteristic of the piano is that any number of notes can be played at the same time. In music an instrument with this characteristic is called polyphonic. This characteristic makes the piano a suitable instrument for our research and for that it was chosen for the present work.

3.1.1 Description of the piano

When we speak about a piano we generally think about the concert piano. There are at least three standard sizes of concert pianos and there is another smaller one, the upright piano, that we have all seen being played in a 'western' film. The mechanism is quite similar for all the pianos. Here we will briefly describe the concert pianos morphology mentioning the most important parts of the mechanism.

The case

The case is the external part of the piano that contains all the components of the piano with the exception of the pedals. It is build of laminated wood and finished with varnish or polyurethane. Inside the case we have the keys, the action mechanism (that includes the hammers and dampers), the strings and the soundboard.

The keys

The number of the piano keys has changed over time. Nowadays the number of keys is 88, which corresponds approximately to 7 octaves, from A 27.5Hz to C4186Hz.

The action mechanism

This mechanism includes the keys, the hammers and dampers and other small pieces. When the key is pressed it transmits the energy to a hammer (fig.3.1). This

hammer strikes the corresponding string or a group of strings (in the middle frequency range and treble part of the piano a set of three strings corresponds to each note). At the same time the key is pressed the damper, which is over the string, lifts to allow the string or strings to vibrate. When the key returns to the rest position the damper rests again on the string stopping the vibration. The action mechanism cuts the connection between the hammer and the key before the hammer hits the strings.

The action is very important because it responds to the force applied by the pianist resulting in final sounds with different dynamics.

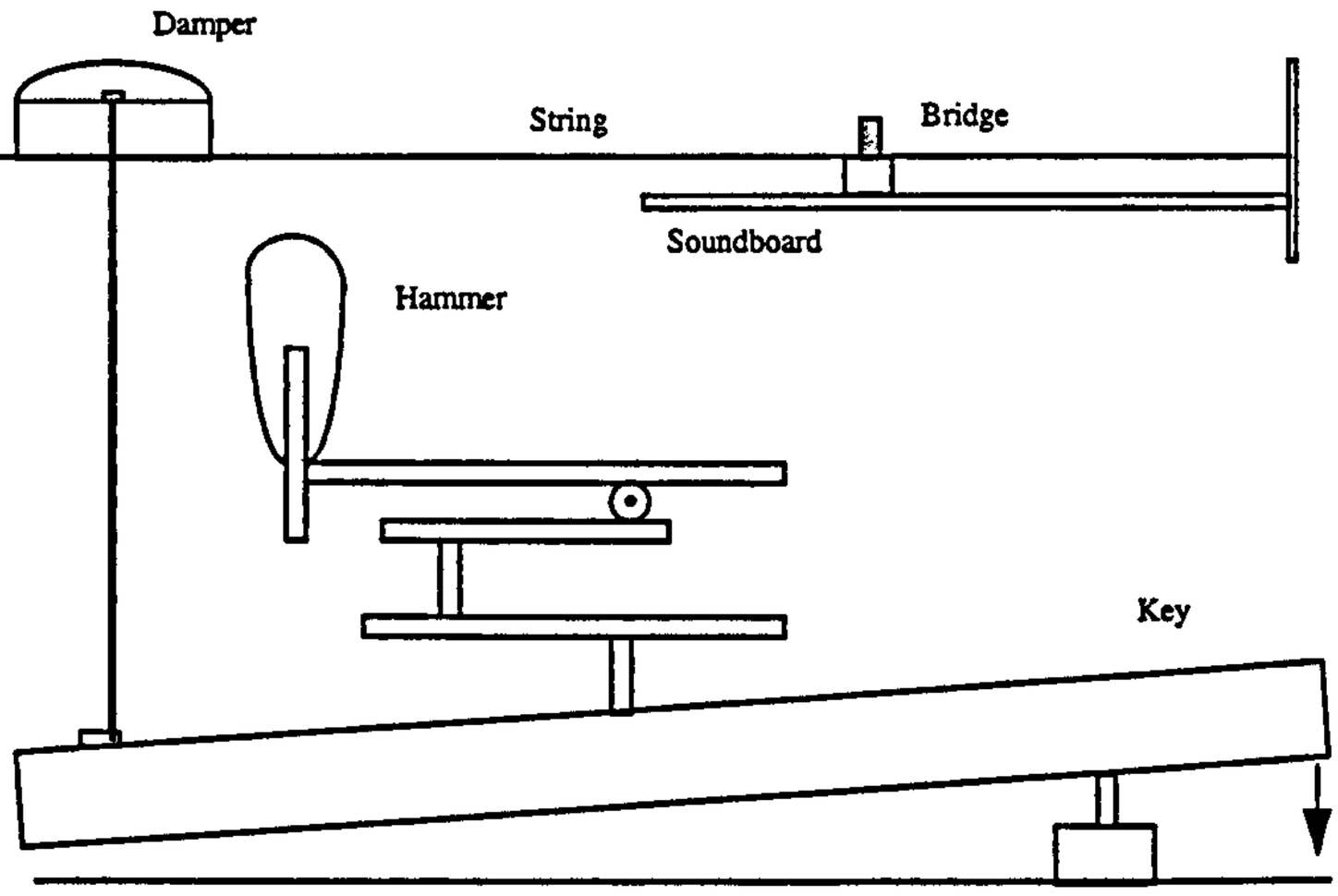


Fig 3.1 The piano action

The hammers

The hammers are very important pieces of the piano since they are at the origin of the string vibration. The physical characteristics of the hammers influence the way they interact with the strings according to the force applied on the keys.

The hammers are made of wood and covered with a woollen felt. The woollen felt of the hammer has a nonlinear behaviour (Hall, D. & Askenfelt 1988). When the hammer reaches the strings with higher impact velocities the felt becomes stiffer. Since the hammers hit different strings with different masses and radii they have different sizes and masses. Their masses vary form 12g to 6g from bass to treble. In the middle range the mass of the hammers is similar to the mass of the strings. Only in the bass range is

the mass of the hammers lighter than the strings (that implies a small decay of the time history of the string).

The strings

The piano strings are made of steel and there are 200 strings in a piano. This means, remembering that there are only 88 keys, that for some notes we have more than one string. The number of strings hit by the hammer change through the frequency range of the piano. In the bass range there is only one string per note. In the low middle range there are two strings and for the rest of frequency range we have three strings. This doubling and tripling of strings has the purpose of raising the sound level produced. If there were only one string the resulting final sound level would not be satisfactory for the final musical output.

The strings in the bass have a greater length to produce lower frequencies. Since the size of the case is limited for convenience, the bass strings are thicker and wrapped round with another copper string. In this way the mass of the string increases and with the same length, it is possible to produce lower frequencies.

All the strings are connected to a soundboard through a piece of wood, the bridge. The vibrations of the string or strings are transmitted to the soundboard, through the mechanical coupling at the bridge. The sound produced is then amplified. These string ends are not rigid supports and the effective length of the string can change while vibrating.

The soundboard

The soundboard is made of wood and lies under the strings. It has a concave shape to hold the tension of the strings. As was mentioned the vibration of the strings is passed to the soundboard through the bridge and it amplifies the sound produced by the strings.

The pedals

The piano has two pedals: *una corda* and *sostenuto*. The *una corda* pedal (left pedal) in the grand pianos move all the mechanism of the keys and hammers a few millimetres to the right so the hammer only strikes a few strings of each note. The *sostenuto* pedal (right pedal) releases all the dampers of the notes. All the strings have a damper over them that does not allow their vibration. When pressing the *sostenuto* pedal

the rests are lifted and the string can continue to vibrate even when the key goes to the initial position of rest.

3.2 Time histories and spectra of piano notes

We present some time histories and spectra of piano notes played on a grand piano with different ranges of the piano scale. All the notes where digitally recorded from a small grand piano with a mono microphone with sampling frequency of 44100Hz (see annex 7 for a more detailed description of data acquisition).

Since the dynamic is a characteristic of the piano we show the time history of the same note played with different musical dynamics: *piano*, *mezzoforte* and *fortissimo*. In figure 3.2 three time histories are presented for the same note whose fundamental frequency is 98Hz. The musical denomination of the note is G and it is in the bass range of the piano scale. We will refer to this note as G98Hz. The three time histories of the note G98Hz correspond to the note played with three different dynamics: *piano*, *mezzoforte* and *fortissimo*.

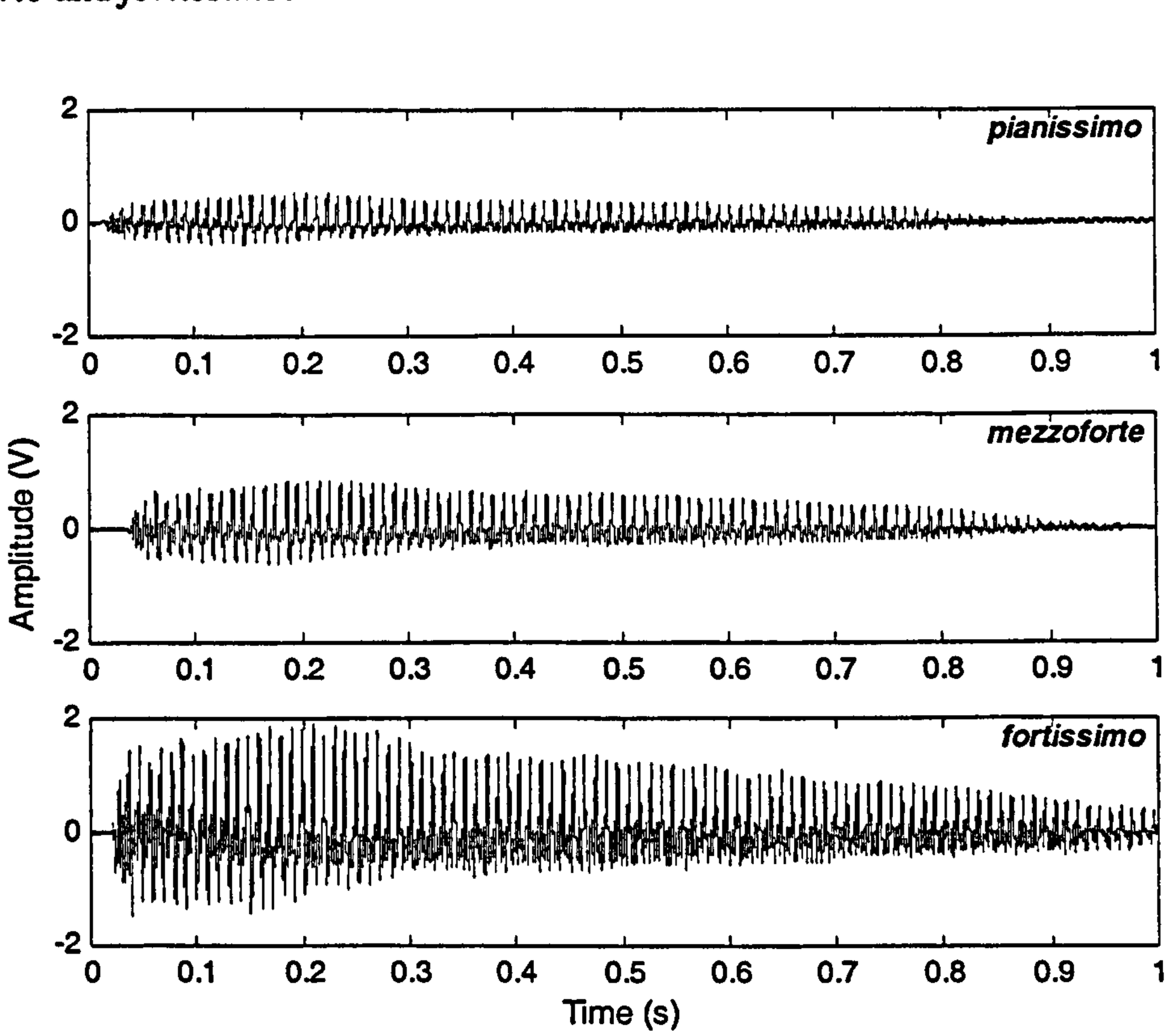


Fig 3.2 Time history of piano note G98Hz played with three different dynamics: *piano*, *mezzoforte* and *fortissimo*.

The envelopes of each of the three events seem quite similar except of course the values of the amplitude that are greater for the *fortissimo* case. The events can be divided in three parts. The first few milliseconds (approximately 0.004ms) correspond to the attack. The next 0.2s (approximately) the signal reaches the maximum value of amplitude and finally it decays. The rate of the decay for this note G98Hz is quite small.

The spectra of those time histories (figs. 3.3, 3.4 and 3.5) show the differences between the frequency components of the signals. The Fourier transform was applied to 1 s of the signals in figure 3.2, and they were zero padded to 88200 samples to obtain a frequency spacing resolution of 0.5 Hz (the sampling frequency of the signals was 44100Hz). Only the spectrum of the note played in fortissimo should present some window effects since it was truncated before it ceased. Since the components above 5000Hz are not very significant and for reasons of better graph scaling we only consider frequencies up to 5000Hz. It is possible to see immediately that the number of frequency components increases from *piano* to *fortissimo*. This means that the *fortissimo* version of the note is not an amplified version of the *piano*. The relation between the magnitudes of the components changes also. Another feature that can be observed is that for the *fortissimo* there is a greater number of nonharmonic components specially present above the 1600Hz.

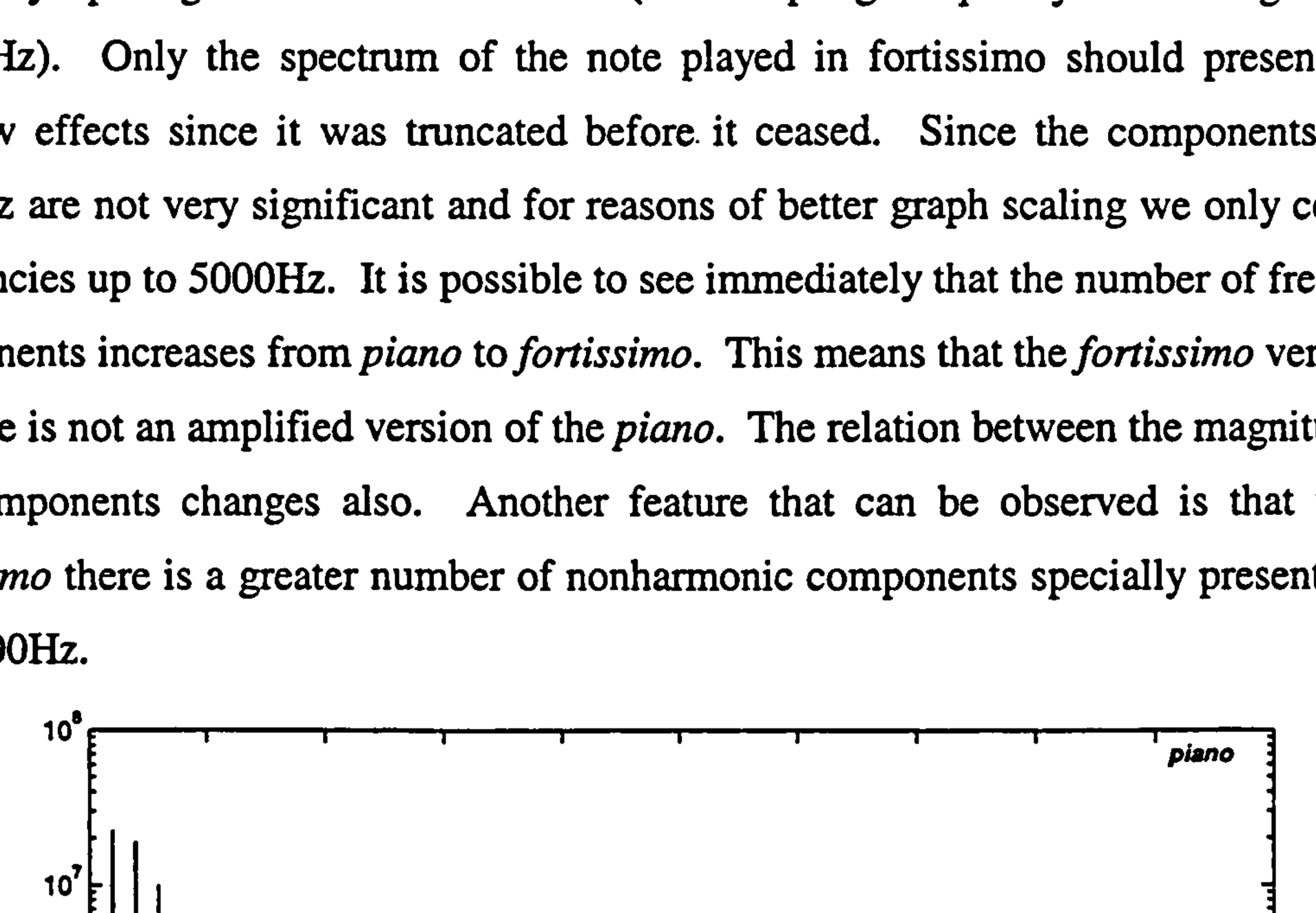


Fig 3.3 Spectrum of piano note G98Hz played with *piano* dynamics (FT 88200points)

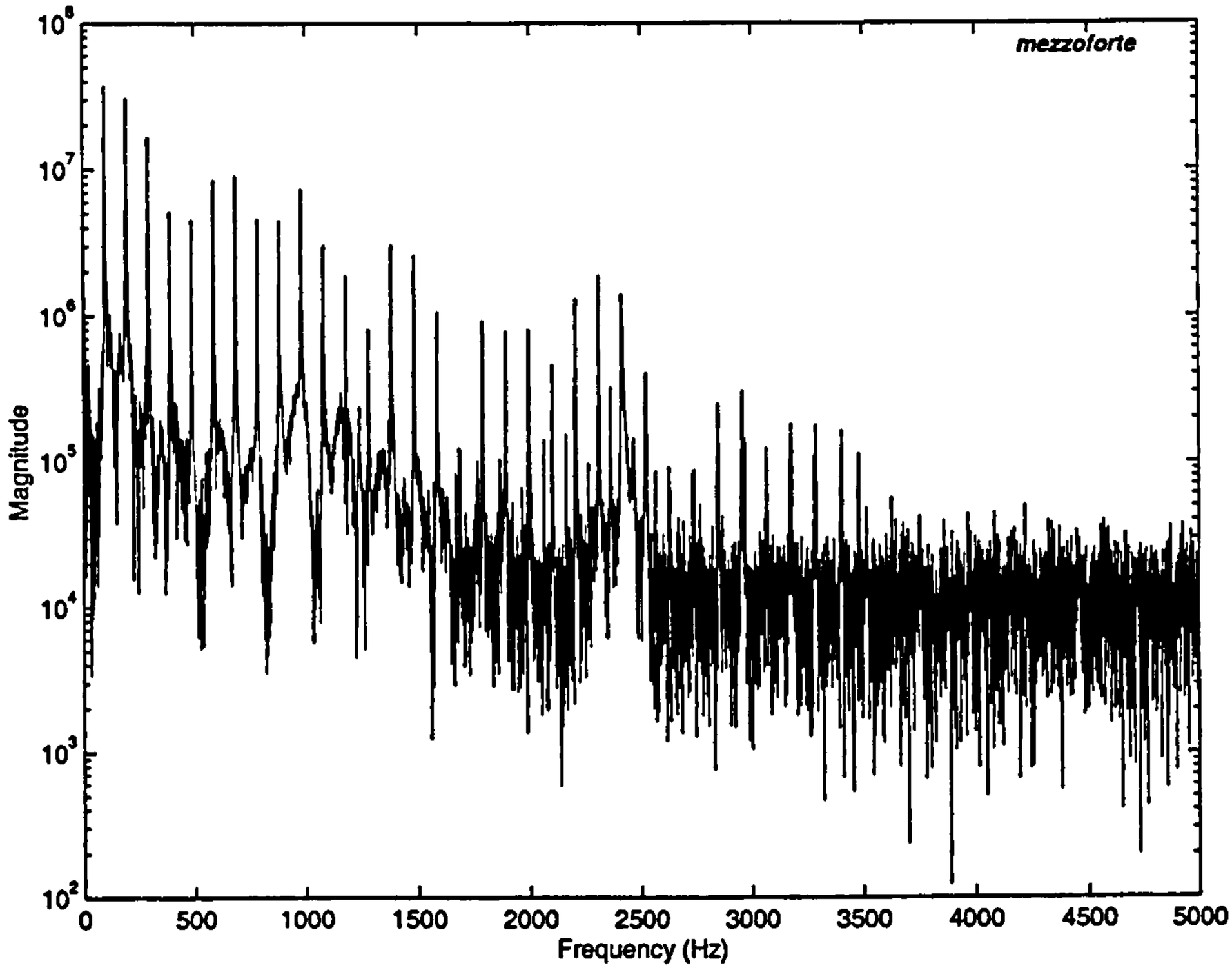


Fig 3.4 Spectrum of piano note G98Hz played with *mezzoforte* dynamics (FT 88200points).

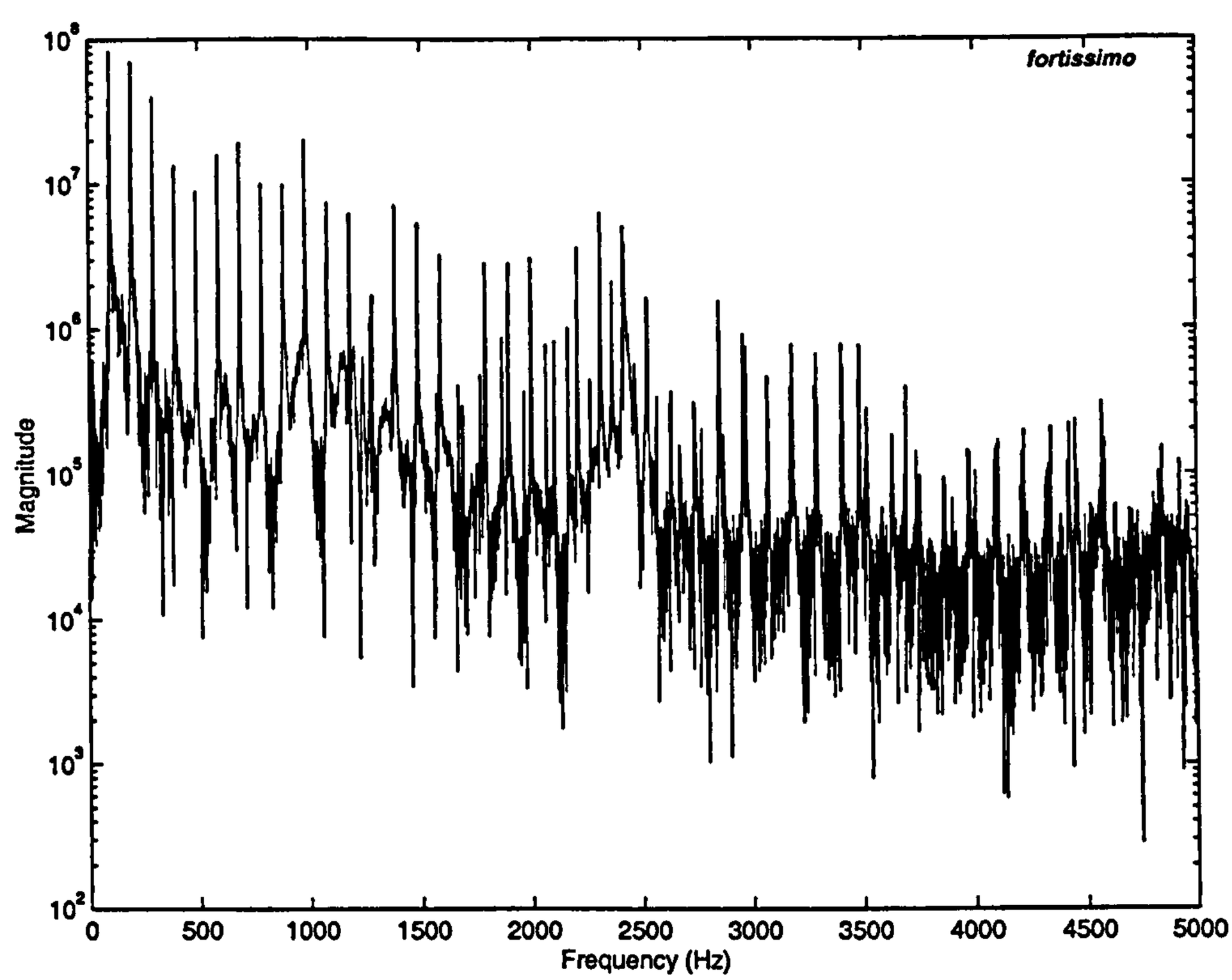


Fig 3.5 Spectrum of piano note G98Hz played with *fortissimo* dynamics (FT 88200points)

For a better perception of the time history corresponding to the attack we have in graphs of fig.3.6 a detail of the note G98Hz. Only the first few milliseconds corresponding to the attack are shown. The attack part of the signal represents the time history of the interaction between the hammer and the string or strings. The *piano* and *mezzoforte* versions of the note G98Hz have very similar features in the attack portion of the signal. The precise beginning of the attack, in the *piano* graph, seems somehow difficult to define due to smaller signal to noise ratio. The recordings were performed in a room with low ambient noise so the signal to noise ratio does not allow a very good determination of that point. In the *fortissimo* time history is not a problem and it is possible to easily determine the beginning of the attack which gives a lead as where to consider that point in the case of *piano* time history. The *fortissimo* attack seems to present a different pattern with a series of small ripples that do not seem to be an amplified version of the other two time histories. This indicates the presence of nonlinearity in the interaction between the hammer and the strings.

The spectra of the different attacks were taken for the initial 0.0068s of the signals, corresponding to the beginning of the signal till the end of the first raising slope of the time history. The 0.0068ms of the attack correspond to 300 samples. That part of the signal was windowed with a Hann window of 300 points. The signals were zero padded to 44100 samples. The frequency components (see figure3.7, 3.8 and 3.9) are different for the three cases.

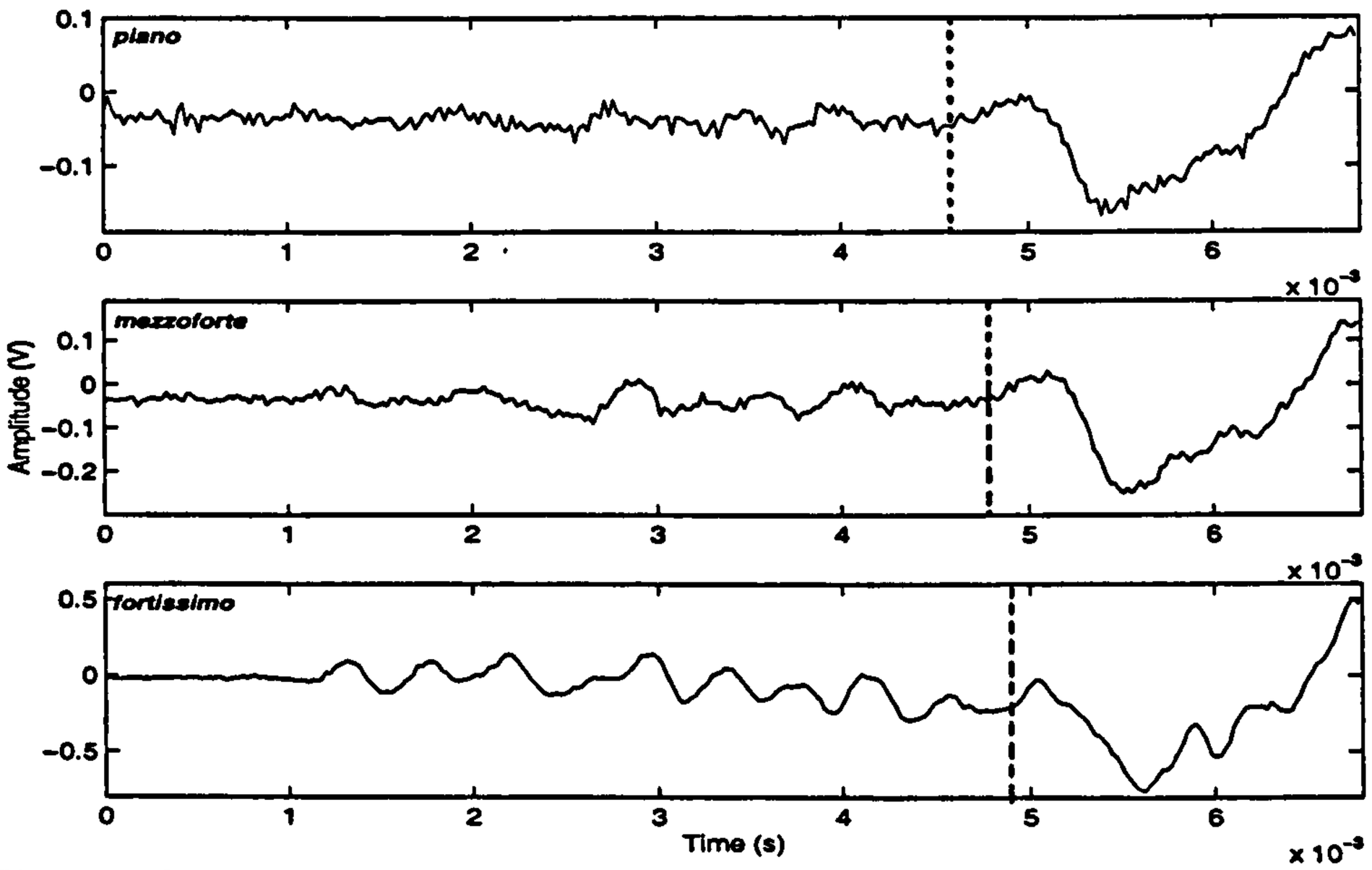


Fig 3.6 Same as figure 3.2 with detail of the first milliseconds of the signal. The dotted lines represents the beginning of the attack.

The great part of the energy of the attack is concentrated on the frequency range under 1000 Hz. The attack part of the note G98Hz does not have any frequency component defined at 98Hz. There has not time for the vibration modes to develop. The frequency components of the three attacks are not harmonics. The most noticeable difference between the three spectra is the presence of a defined peak at *fortissimo* at 166Hz that it is not present in the other cases.

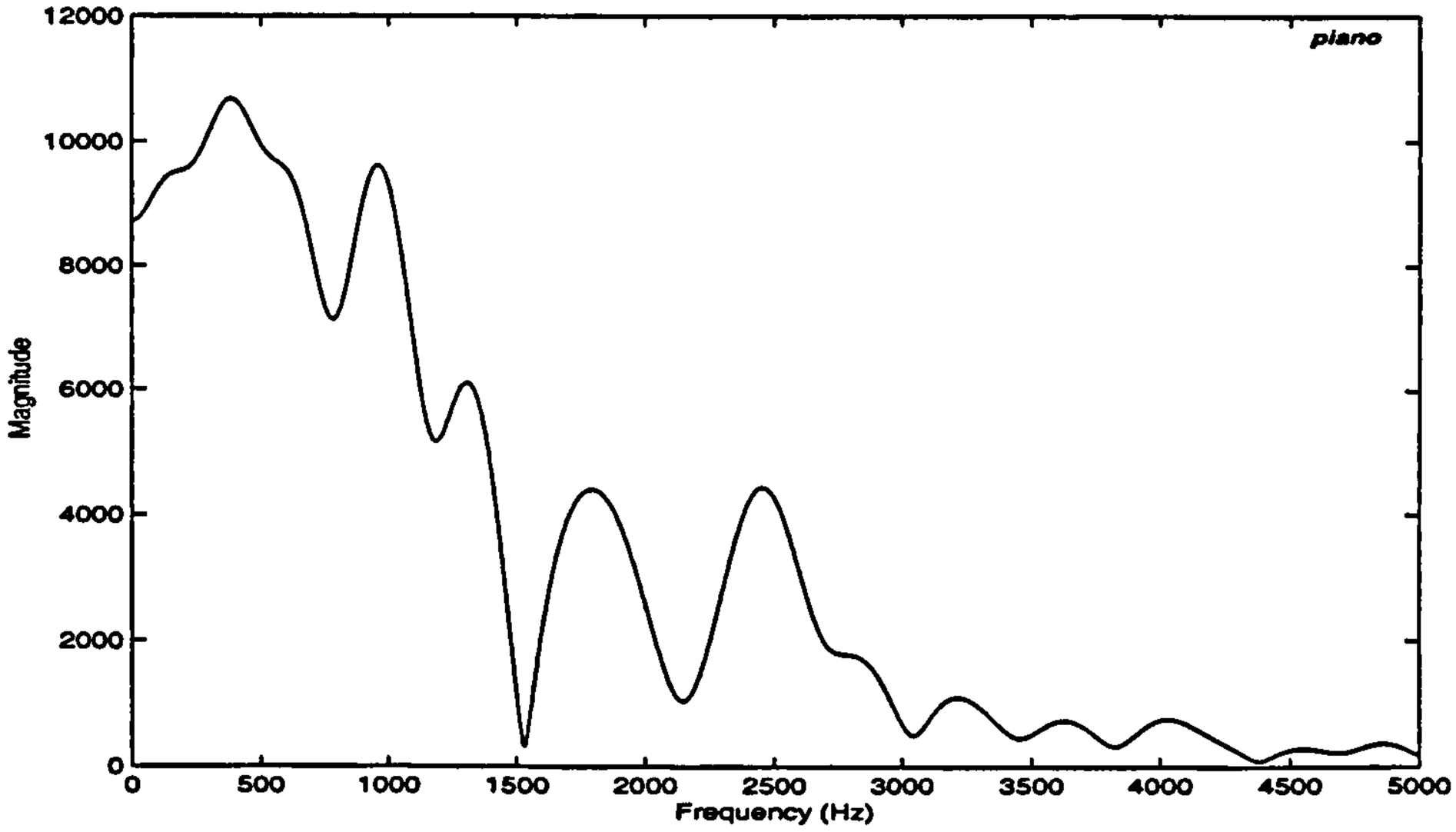


Fig 3.7 Spectrum of note G98Hz played with *piano* dynamics. The FT (44100 points) was taken over the first 0.0068s corresponding to the attack. The attack was windowed with a hanning window of 300 points.

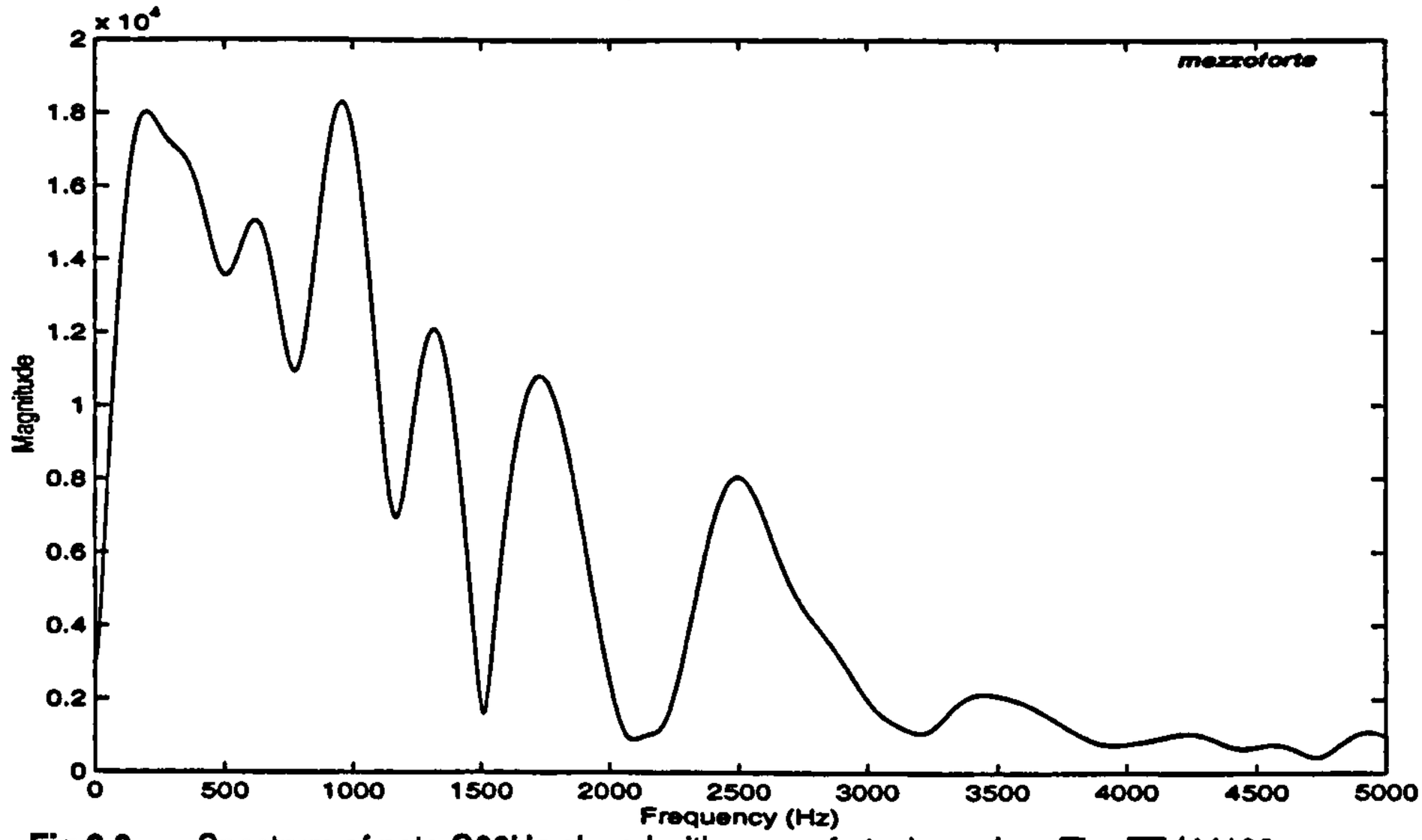


Fig 3.8 Spectrum of note G98Hz played with *mezzoforte* dynamics. The FT (44100 points) was taken over the first 0.0068s corresponding to the attack. The attack was windowed with a Hann window of 300 points.

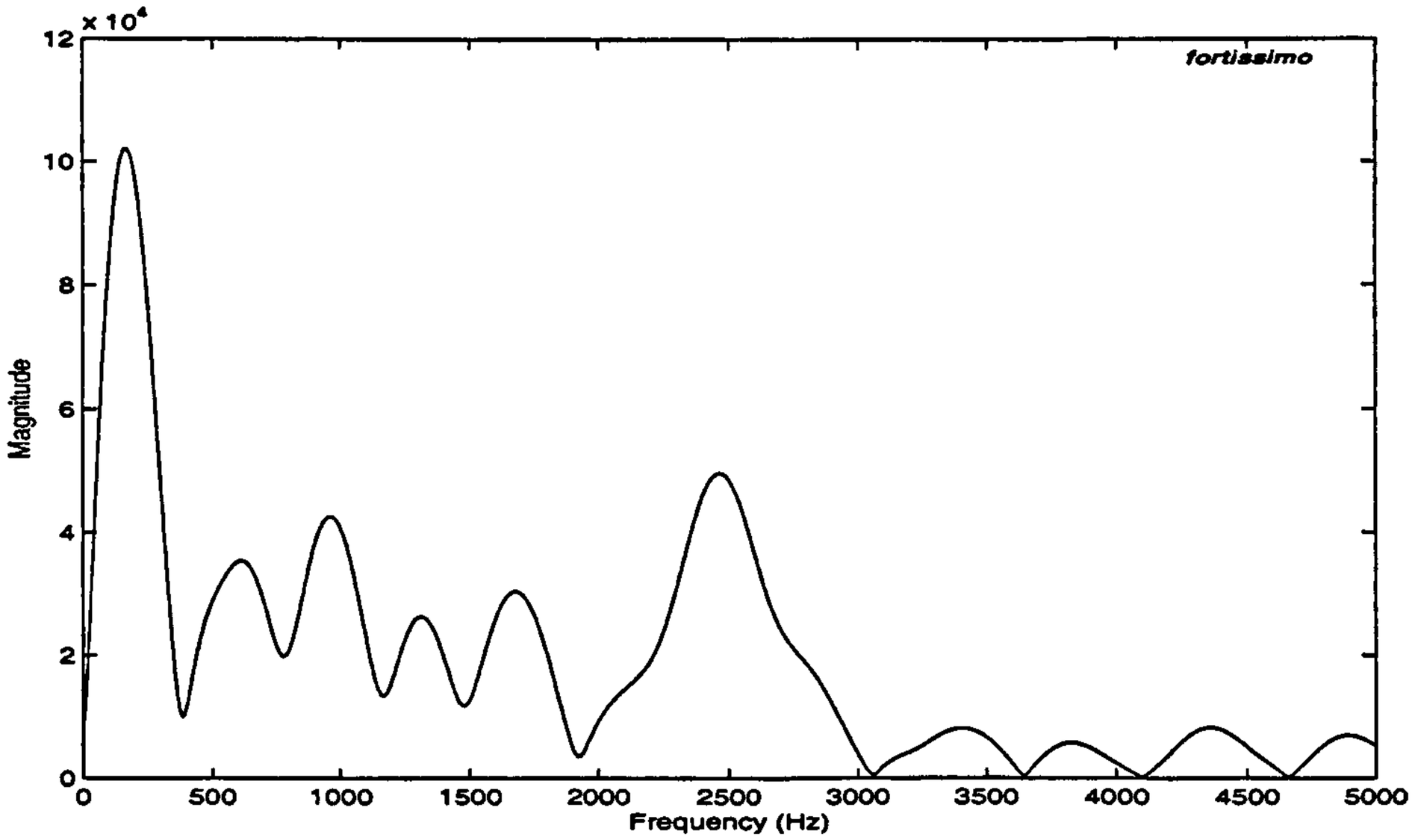


Fig 3.9 Spectrum of note G98Hz played with *fortissimo* dynamics. The FT (44100 points) was taken over the first 0.0068s corresponding to the attack. The attack was windowed with a Hann window of 300 points.

The value of the first frequency component of the attack becomes smaller from *piano* to *fortissimo*. The last significant peak has a frequency around 2470Hz with small variations for the three graphs.

The different frequency components in the attack and the presence of more frequency components, harmonics and nonharmonics, at high frequencies in the case of *fortissimo* leads to the nonstationarity of the piano notes. Each time a note is played with different dynamics they have different spectra. After the attack the time histories become quasi periodic which should correspond to the free vibration of the string.

Since the signal are nonstationary a time frequency analysis may be useful and the spectrograms of the three versions of the signal G98Hz show the difference of the number of harmonics and how they vary in time. The highest frequency components have a shorter duration and the great majority are modulated.

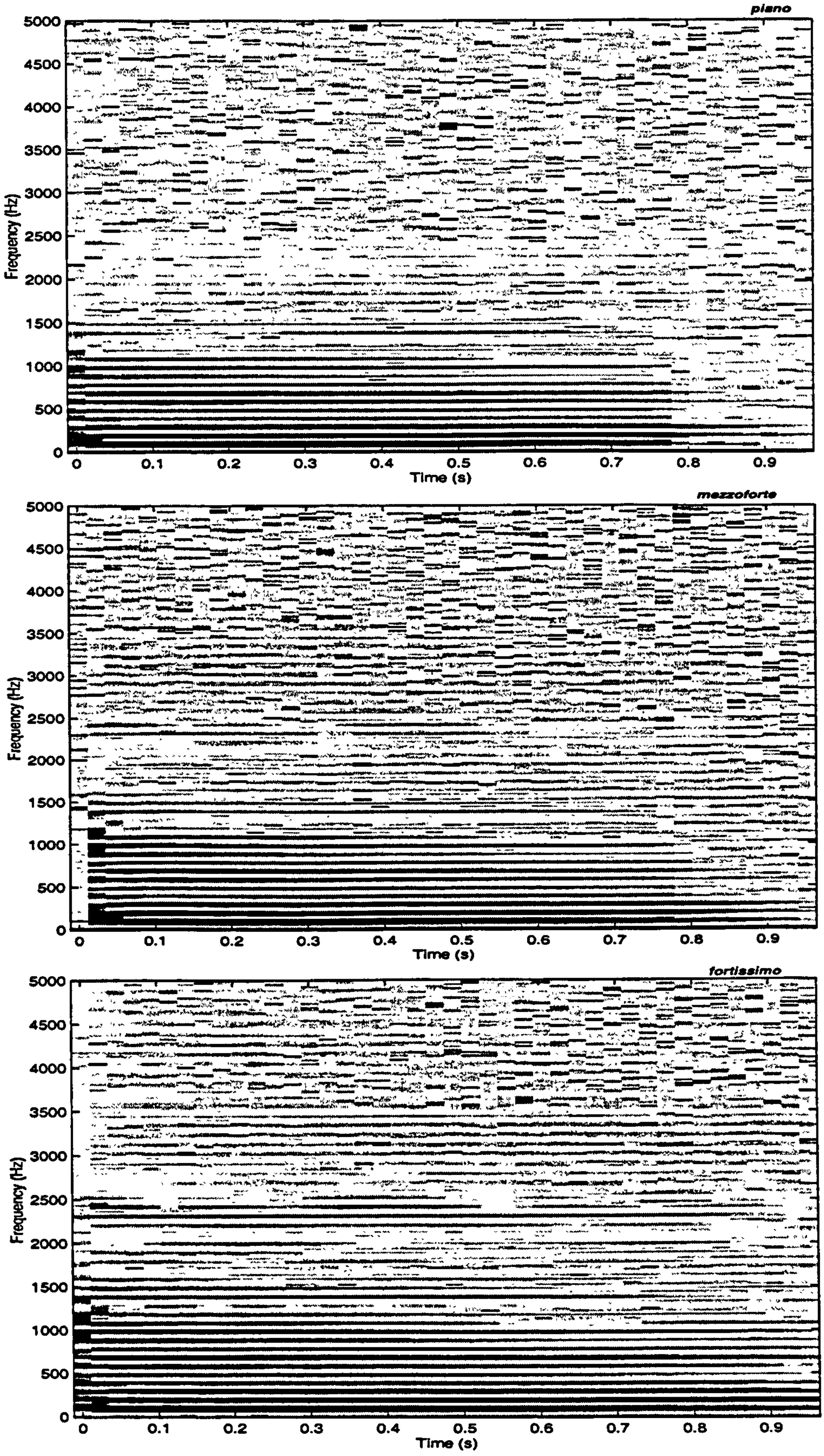


Fig 3.10 Spectrograms of the piano note G98Hz played with the different dynamics of *piano*, *mezzoforte* and *fortissimo*. (FFT 2048 points, 1024 overlap points).

The spectrograms were performed over 1s of each version of G98Hz. The signals were split into segments of 2048 samples each. The window used was the Hann window applied to that same length of the segments. The overlapping between segments was of 1024 samples.

Now we will see the time histories and respective spectra and spectrograms of the same note G but played in different consecutive octaves: G98Hz, G196Hz and G392Hz (the events analysed were played with *fortissimo* dynamics). The envelopes of the notes (fig.3.11) are different presenting distinct decays (the number of strings played by the hammer of G98Hz is two while for the other G notes there are three). The spectra (determined over 1s of each signal) and the spectrograms (determined with the same parameters of the previous spectrograms) show that the number of frequency components becomes smaller for the G note with the highest frequency and the relation between the magnitudes of the components is different for each case (see fig3.2 for the G98Hz *fortissimo* and fig.3.11 for the other two notes G196Hz and G392Hz). Because of the different characteristics of the strings and hammers along the piano range these differences in the spectra should be expected.

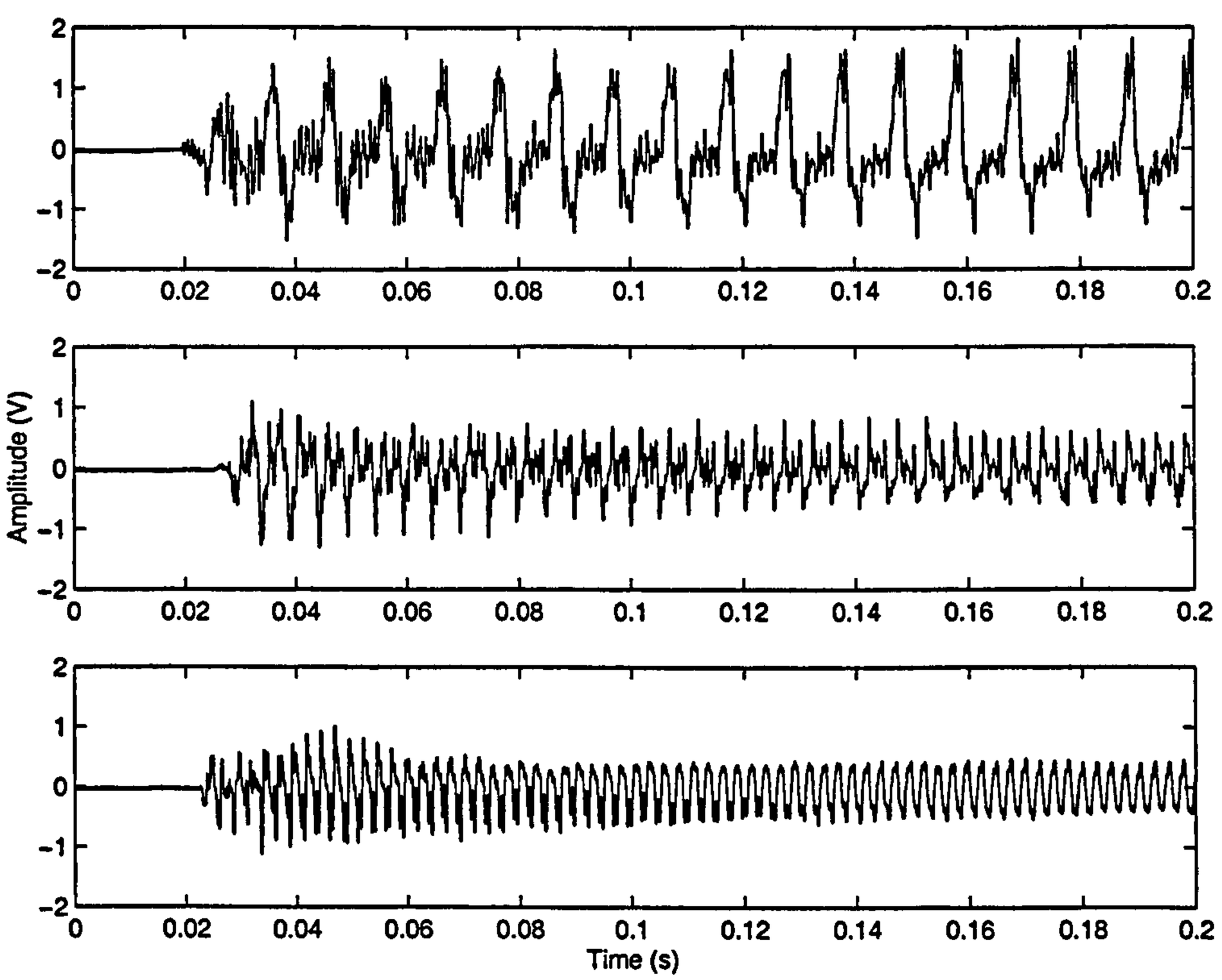


Fig 3.11 Time histories of the piano notes G98Hz, G196Hz and G392Hz.

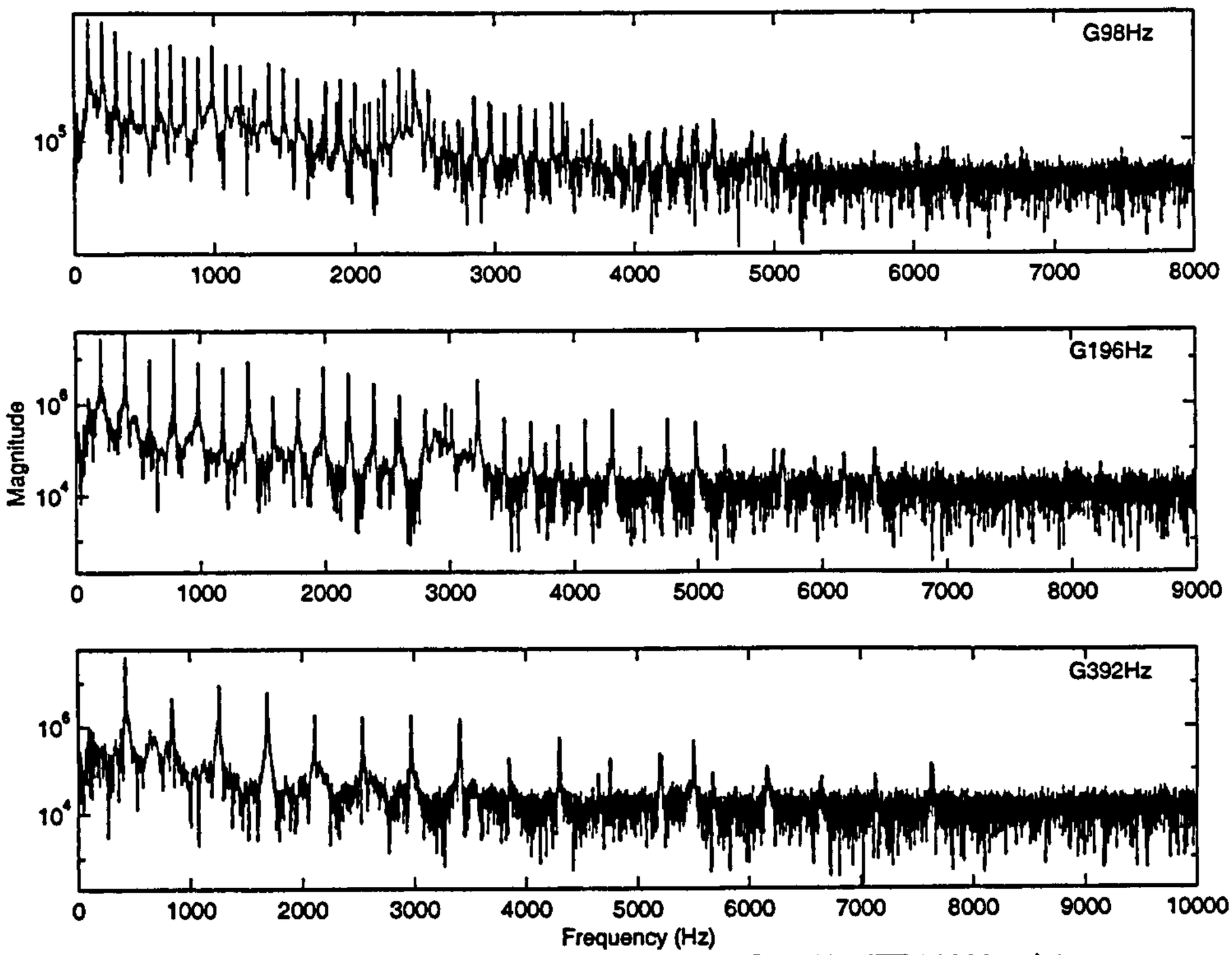


Fig 3.12 Spectra of the piano notes G98Hz, G196Hz and G392Hz. (FT 88200 points, resolution of 0.5 Hz).

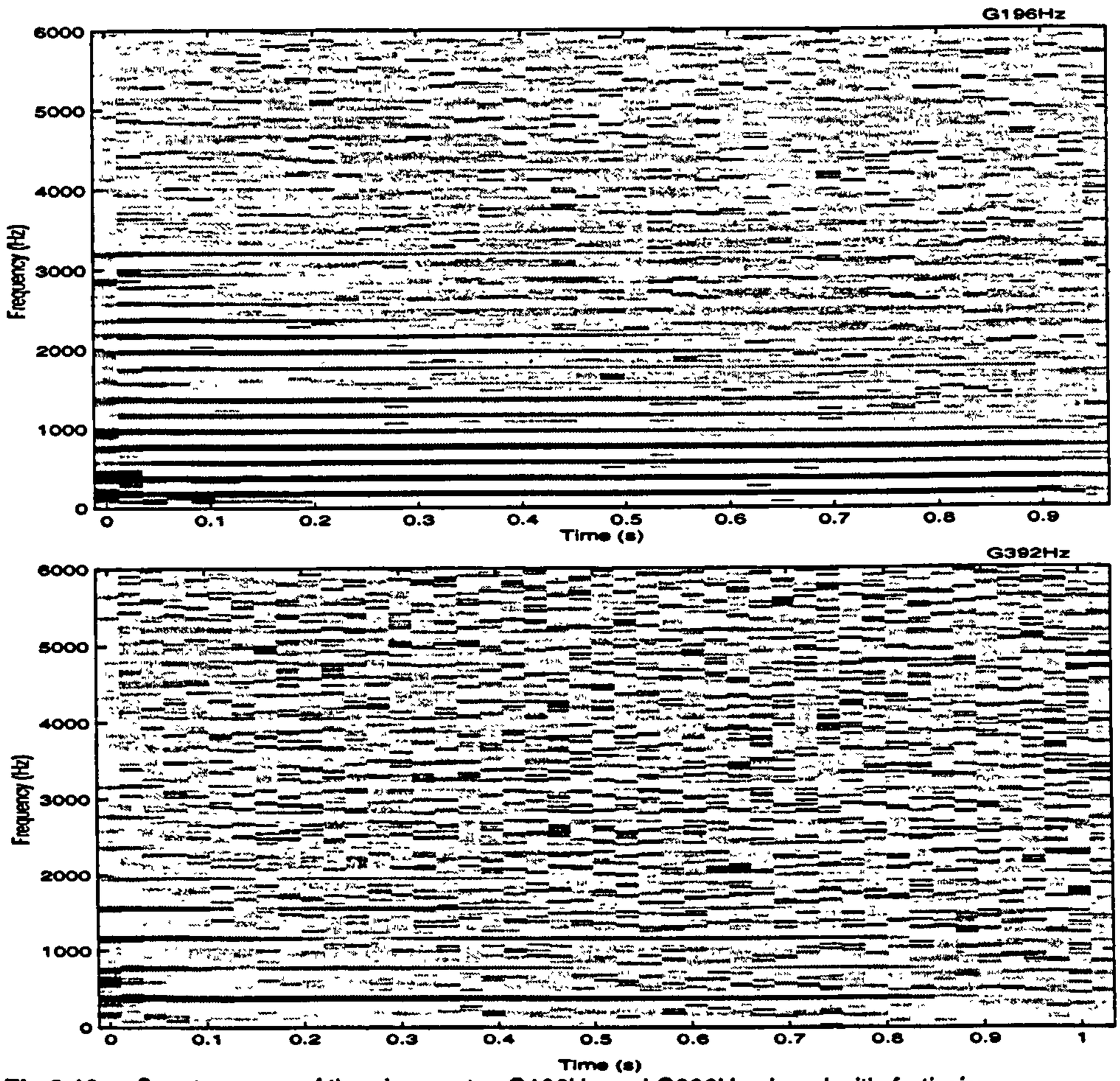


Fig 3.13 Spectrograms of the piano notes G196Hz and G392Hz played with *fortissimo* dynamics. (FFT 2048 points, 1024 overlap points).

The spectrum of the note G98Hz presents a significant number of inharmonic components above the 1500Hz. The other notes have a smaller number of inharmonic frequency components.

In next section we will describe the piano note production mechanism in more detail and try to establish a relation between that and the time histories, spectra and spectrogram presented here.

3.3 Piano note production

The production of a piano sound or note can be considered as consisting of three phases:

1- Application of a force on the key.
2- The hammer-string interaction.
3- String vibration and transmission to the soundboard

1- Application of a force on the key

The external action to the piano system that produces a piano sound is the application of a force on the piano keys.

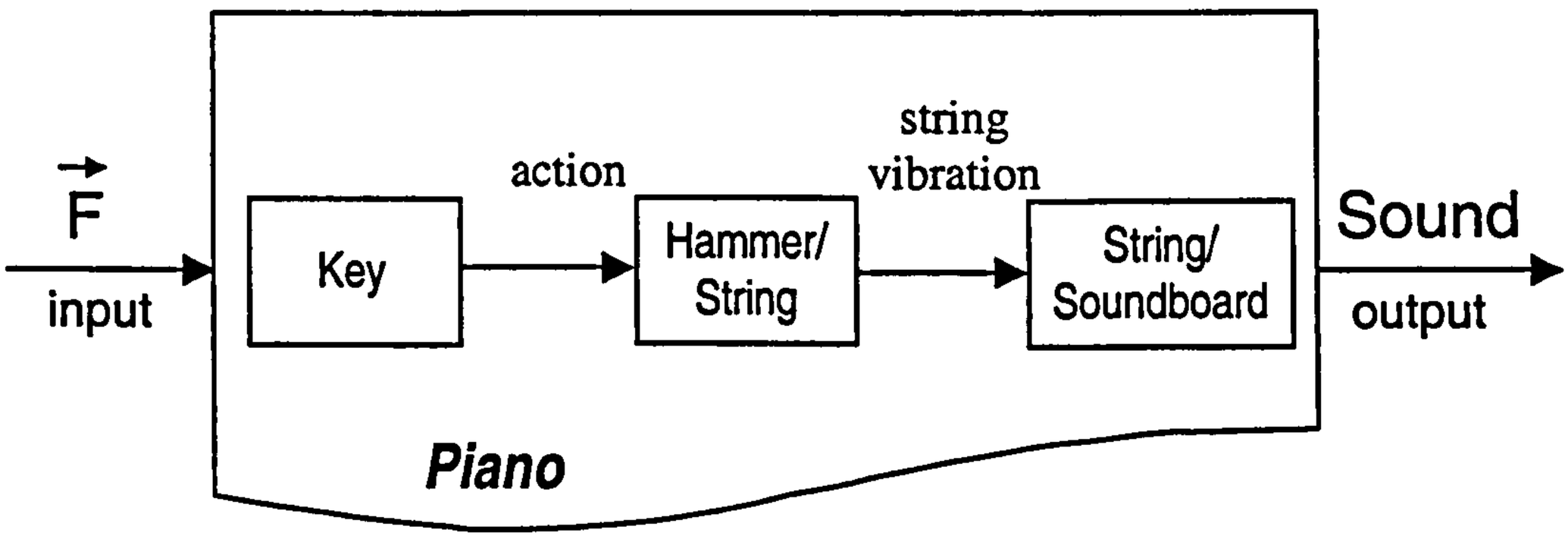


Fig 3.15 Piano sound production.

The piano was the first key instrument where the resulting dynamics of the produced sound was related to the force applied by the musician. The great pianists are recognised by their own style of playing. They spend several years developing what is called the touch, meaning the way they produce a note or a set of notes. Generally pianists say that according to the way they strike a key the sound of the note produced changes and has its own 'colour'. This assumption has generated a great controversy between the pianists and the scientists that have studied the piano mechanism and note production. According to some earlier studies (Jeans, 1937) nothing affects the time history of the piano notes but the dynamics. Independently of the type of touch used by the pianist, as the *legato* or *staccato*, the dynamics is the only factor that influences the final time history. The hammer strikes the string or strings in the same way and only the force is different. More recent studies (Askenfelt, 1990) note that the acceleration and the velocity of the hammer can change depending on the type of touch. They present graphs and values that seem to prove these changes of values that can induce to an indirect control of the hammer string interaction by the pianist due to different touches.

The control of the interaction is indirect as Askenfelt (1990) verified that the force applied to the key, due to the mechanism of the action always ceases to act over the hammer before the hammer strikes the strings. This is due to the intrinsic morphology of the piano. When the pianist plays the note he does not have any direct control over the hammer at the exact moment when it strikes the strings. The pianist controls the acceleration and velocity of the hammer.

After the key has been played the only control that the pianist has over the note produced is the duration keeping the key pressed. (Another factor that a pianist can change is the reinforcement of the harmonics of the note using the *sostenuto* pedal.)

The controversy about the control of the pianist over the way the hammer strikes the strings, depending on the touch, continues since there are no conclusive and definitive studies.

2- The hammer-string interaction

Since Helmholtz the interaction between the hammer and the string has been studied. Recently Hall (1986, 1992), Askenfelt and Jansson (1990, 1991,1993) have performed more detailed studies about the piano and the hammer string interaction.

The hammer strikes the strings due to the piano action that is driven by the force applied by the pianist on a key. The pianist only controls the action of the piano through the keys but not the hammer string contact directly. This interaction defines the time history of the piano notes. There are several factors that make the time histories and spectra of piano notes in different ranges have different characteristics although for the human ear, any note in any range, is easily recognised as a "piano note". That means, that although many factors contribute to a different response for each note the timbre of the instrument is perfectly recognisable. In the graphs of note G98Hz the spectrum changes with the dynamics but it continues to be recognised as a piano note. The same notes G played in different octaves have different numbers of harmonics and a different relation between their magnitudes and they still are recognised as piano notes. To characterise this interaction we have to define some properties of the hammer and of the strings.

The piano hammers are built of wood and covered with a compressed felt fabric. This fabric has a nonlinear stiffness when the hammer strikes the strings with different velocities (Hall, D.E. 1992). For higher velocities the nonlinear stiffness makes the hammer harder. For the *fortissimo* dynamics the hammer reaches the strings with a higher velocity. The time of contact becomes then smaller (Akenfelt, 1990) and that should imply a greater number of harmonics. We saw that in the spectra of the note G98Hz played with different dynamics: at *fortissimo* the number of harmonics present in the spectra increases. The time of contact between the hammer and the string also varies through the piano frequency range (Askenfeld, 1990) and can reach 4ms in the bass and 0.5ms in the treble. These values indicate that for the bass range the period of the fundamental of the note produced is greater and for the treble is smaller. Then the hammer in the treble can hit the string more than once during the attack. This time of contact is very important because it defines the decay of the note produced. We could have more harmonics for the treble notes since there is a very small time of contact. But if we compare that time with the values of the period of the notes produced the period for the treble is smaller than the time of contact which does not allow the modes to set so fast. So the number of harmonics is smaller for the treble.

The striking of a string or strings by a hammer together with the amplification by the soundboard characterizes the sound produced by the piano.

3-String vibration

We saw that the interaction between the hammer and the string is basically nonlinear. The time history of the piano notes (see fig.3.7) show that the attack part of the signal corresponds to a forced vibration of the strings. After the attack the modes of vibration of the string are set and we have a linear process. The linear response of the string is a result of the nonlinear process of the attack. In the same graph we can see that the *fortissimo* attack has new features that do not correspond to an amplification of the *piano* case.

The inhamonicity present on the spectrum of the note G98Hz is due to the presence of stiffness in the string. When the string is displaced from the rest position the restoring force due to the string tension adds with the force due to the string stiffness increasing the frequency of the harmonics. Another factor contributing to the inharmonicity are the end supports of the string that are not perfectly rigid producing slight variations in the length of the string.

In the piano note production the most influential factor in the final sound is the striking of the strings by the hammer.

3.4 Modelling of a piano note as a nonstationary signal

After the previous considerations about piano notes, we can now build a simple model of the notes for a simplification of the research as a first approach. As noted above the most influential factor in the note production is the string excitation by the hammer through a stroke. The attack is nonlinear since it presents different frequency components according to the force applied on the hammer. Each time a note is played with different dynamics we have different values for the different variables that are present in the note production. The free vibration of the string is influenced by these initial factors in the form of inharmonic components, more or less harmonic components

and change in the relation between their magnitudes and modulation. This leads to consider the piano note production as a stochastic process.

Initially for a simplification of the model we are going to consider only one frequency component. Another variable that we have to consider is the decay of the note. In the example of the different G in different octaves the decays seem different. That is due to the presence of a different number of harmonics with different time of raising and decay (including modulation). For simplification purposes we consider that the decay of the notes is exponential.

We are considering a transient signal with only one frequency component. Each event has some randomness since in the real note production the time histories and spectra have variations. The decay also varies for each event and for simplification reasons we will consider as exponential. The modelled piano note can be represented by the equation:

$$x_i(t) = A_i e^{-\alpha_i t} \sin(2\pi f_i t + \phi_i) \tag{3.1}$$

where A_i is the amplitude, α_i the damping, f_i the frequency, ϕ_i the phase and t the time.

Each event represented by $x_i(t)$ is a transient signal where the amplitude, damping, angular frequency are parameters that will vary randomly from realisation to realisation. We shall assign same probability density function to all the parameters. For convenience each of these parameters A_i, α_i, f_i and ϕ_i is considered to be drawn from independent uniform probability distributions. The randomness of these variables simulates the nonstationarity of the real piano for each event of the modelled piano notes.

The graphical representation of a modelled piano note with only one frequency component of 10 Hz, amplitude 1 and 0.001 damping and initial zero phase, is in fig.3.16.

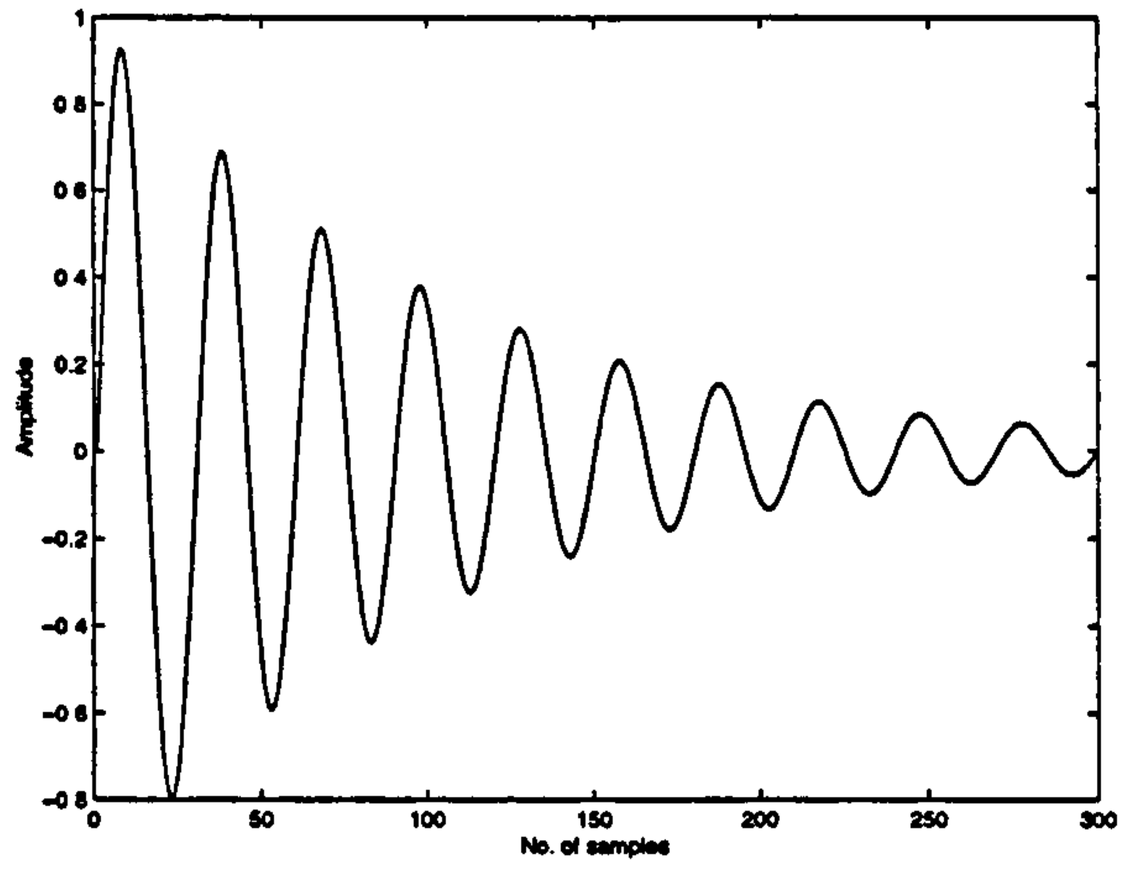


Fig 3.16 Event of a simulated piano note determined by equation 3.1, with frequency value of 10Hz, amplitude value of 1, damping value of 0.001 and zero initial phase.

For the many studied cases we will not consider the phase ϕ_i. The possible realisations of the signal are going to be considered as starting all with zero phase. As we will see later another variable will be considered: the delay between the realisations. The frequencies used in the simulations will be smaller compared with the real piano note frequencies (the lowest frequency of a piano note is 27.5Hz) for a better analysis of data.

This simplified model of the real piano note will be used as a starting point for the real piano note analysis. A previous analysis will be applied to the model and then we will analyse the real piano notes.

CHAPTER 4

THE KARHUNEN-LOÈVE EXPANSION

Introduction

The Karhunen-Loève Expansion is a mathematical technique used in the treatment of stochastic processes. This expansion was established by Karhunen in 1946 and Loève in 1955. This expansion is an orthonormal expansion and represents any stochastic signal as a linear combination of uncorrelated functions of a basis set. It can be applied to any stationary or nonstationary stochastic signal since there are no restrictions in the theorem of the expansion. This is very important for this research since we are considering the piano notes as transient stochastic processes. The second order statistics of the signal define the basis of the expansion through its eigen decomposition. The signal itself selects the optimal new basis set which is a particular feature of this expansion.

We note that the study entitled Principal Component Analysis (PCA) widely used in statistics is essentially the same as the KLE.

4.1 The Karhunen-Loève expansion (KLE)

If we have a stochastic process represented by $X_\mu(t)$, with μ corresponding to a possible realisation of the process (t may discrete or continuous) it can be expanded under certain conditions with the KLE. The application of the KLE to a random process produces a new representation of the process in terms of a product between a set of orthogonal random variables and orthonormal (the functions are orthogonal and have norm equal to one) functions.

A starting point for the KLE is Mercer's theorem. The characterisation of stochastic signals uses the second order statistical properties. The theorem, by Mercer (1909), states that any second order functions, under certain conditions, can be expressed as a linear combination of the eigenfunctions weighted by the eigen values of the correlation function. We will see that the correlation function satisfies the conditions of the Mercer's theorem and so can be written as a linear combination of the product of the eigenfunctions.

We will now state the Mercer's theorem also known as the spectral theorem.

Mercer's Theorem

Any Hermitian symmetric and square integrable function $R(t_1,t_2)$ with $a \le t_1 \le b$ and $a \le t_2 \le b$ that is nonnegative, i.e, satisfying the condition

$$\int_a^b \int_a^b R(t_1,t_2)x(t_1)x(t_2)dt_1dt_2 \ge 0 \tag{4.1}$$

can be expressed as:

$$R(t_1,t_2) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(t_1)\Phi_i(t_2) \tag{4.2}$$

where $\{\Phi_i\}$ and $\{\lambda_i\}$are the eigenfunctions and eigen values of the function $R(t_1,t_2)$.

Mercer's theorem establishes that under certain conditions the correlation function can have a spectral decomposition. A continuous function is now represented by the linear combination of the products of discrete eigenfunctions.

Covariance, or correlation functions satisfy the conditions of Mercer's theorem, i.e. they are Hermitian (or symmetric if the process is real) and nonnegative as summarised in the following propositions.

Consider a stochastic process $X(t)$ defined on the interval $a \le t_1 \le b$ and $a \le t_2 \le b$ let $R(t_1,t_2)$denote the autocorrelation function and:

$$R(t_1,t_2) = E\left[X(t_1)X^*(t_2)\right] \tag{4.3}$$

Proposition 1 (symmetry). Every covariance function, R, is Hermitian, i.e. $R(t_1,t_2) = R^*(t_2,t_1)$, ($R^*$is the complex conjugate). For a real valued process it is symmetric, i.e. $R(t_1,t_2) = R(t_2,t_1)$

Proposition 2 (nonnegativity). Every covariance function, R, is nonnegative definite, i.e. for any function $x(t)$ we have: $\iint R(t_1,t_2)x(t_1)x(t_2)dt_1dt_2 \ge 0$.

We are now in a position to state the KLE namely that a stochastic process can be expanded as a linear combination of eigen functions and eigen values of the autocorrelation function of the process.

The Karhunen-Loève Expansion (Continuous time)

An ensemble of random functions, $X_\mu(t)$, representing a continuous process with a μ number of possible realisations of the signal and $t \in [a,b]$, can be represented as (Newman,1996):

$$X_\mu(t) = \lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t) \tag{4.4}$$

where $\Phi_n(t)$ are a set of orthonormal eigen functions of the covariance function $R(t_1, t_2)$ satisfying:

$$\int_a^b \Phi_m(t)\Phi_n(t)dt = \delta_{mn} \tag{4.5}$$

The coefficients $c_n(\mu)$ satisfy,

$$c_n(\mu) = \left(\sqrt{\lambda}_n\right)^{-1} \int_a^b \Phi_n(t) X_\mu(t) dt \tag{4.6}$$

and

$$E\left[c_n(\mu)c_m(\mu)\right] = \delta_{nm} \tag{4.7}$$

The correlation matrix satisfies the equation:

$$\int_a^b R(t_1, t_2)\Phi_n(t_1)dt_1 = \lambda_n \Phi_n(t_2) \qquad n = 1,2,\ldots \tag{4.8}$$

The equation (4.4) is the KL expansion (see annex 3 for the proof of KLE). The random signal $X_\mu(t)$ is represented as a linear combination of orthonormal functions (eq.4.5) weighted by random variables that are uncorrelated (eq.4.7).

The relation represented by equation 4.8 shows the eigen decomposition of the covariance matrix where the Φ_n are the eigen functions and λ_n are the eigen values of the covariance R.

The KLE allows the approximate representation of a signal in terms of a finite orthonormal basis set whose vectors are the eigen functions of the correlation function and that representation is optimal since equation (4.8) minimizes the mean square error of the finite representation of the signal. The squared error of the finite representation of the signal is given by:

$$e^2 = \left(\sum_{n=1}^{\infty} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t) - \sum_{n=1}^{N} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t) \right)^2 = \left(\sum_{n=N+1}^{\infty} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t) \right)^2 \qquad (4.9)$$

The eigen values of the KLE are related with the energy associated with each eigen vector. If we determine the power of a signal $X(t)$ expressed in terms of the KLE:

$$E\left[X(t) X^*(t) \right] = R(t,t) = E\left[\sum_{n}^{\infty} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t) \sum_{m}^{\infty} \sqrt{\lambda_m} c_m^*(\mu) \Phi_m^*(t) dt \right]$$

$$= \sum_{n}^{\infty} \lambda_n \left| \Phi_n(t) \right|^2 \qquad (4.10)$$

The total energy of the signal is then given by:

$$\int_a^b \sum_{n}^{\infty} \lambda_n \left| \Phi_n(t) \right|^2 dt = \sum_{n}^{\infty} \lambda_n \int_a^b \left| \Phi_n(t) \right|^2 dt = \sum_{n}^{\infty} \lambda_n \qquad (4.11)$$

The sum of the eigen values of the autocorrelation function of $X(t)$ gives the total energy of the signal.

4.2 The discrete version of the KLE

In this thesis we use both continuous and discrete time. Continuous time is used in some analytic descriptions and discrete time is needed for our computational implementation.

For a discrete process an ensemble of random generalised functions $X_\mu(t_k)$, where t_k refers to discrete time and μ is a possible realisation of the signal, can be represented by:

$$X_\mu(t_k) = \lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(\mu)(\Phi_n)_\mu \tag{4.12}$$

where $\{\Phi_n\}$ are a set of orthonormal eigen vectors of the covariance matrix R whose elements are given by $(R)_{ij} = E\left[X_\mu(t_i) X_\mu^*(t_j)\right]$ satisfying,

$$\sum_{i=1}^{\infty} (\Phi_m)_i (\Phi_n)_i = \delta_{mn} \tag{4.13}$$

The coefficients $b_n(\mu)$ are given by:

$$b_n(\mu) = \left(\sqrt{\lambda}_n\right)^{-1} \sum_{i=1}^{\infty} (\Phi_n)_\mu X_i(t_k) \tag{4.14}$$

and

$$E[b_n b_m] = \delta_{mn} \tag{4.15}$$

The correlation matrix satisfies the equation

$$R\Phi_n = \lambda_n \Phi_n \qquad n = 1,2,\ldots \tag{4.16}$$

Equation (4.12) is the Karhunen-Loève expansion. In this expansion a possible realisation of a signal is expressed as a linear combination of functions that are the eigen vectors of the correlation matrix of the signal.

The correlation matrix has to satisfy the conditions of the spectral theorem for the expansion to be possible.

Spectral theorem Let $X(m \times m)$ be a real symmetric (or Hermitian) matrix with eigen values $\lambda_1,\ldots,\lambda_m$; X can be written as $X = U\Lambda U^H$, where $\Lambda = diag(\lambda_1,\ldots,\lambda_m)$ and U is the unitary $(m \times m)$ matrix whose columns are orthonormal eigen vectors $v_1,\ldots,v_m$ of X, associated with $\lambda_1,\ldots,\lambda_m$. In other words:

$$X = \sum_{i=1}^{m} \lambda_i v_i v_i^H \tag{4.17}$$

The spectral theorem is the equivalent of the Mercer's theorem in the discrete domain. The covariance is symmetric and nonnegative as was stated for the continuous case. The same applies for the discrete case and the covariance matrix satisfies the conditions of the spectral theorem.

4.3 How the KLE is applied in practice

In practice the KLE is applied to stationary and weakly stationary random signals in the same way. If we have a stationary (or weakly stationary) signal $x(t)$, the first step to determine the KLE expansion is to estimate the stationary autocorrelation of the signal. The values of the autocorrelation are taken from the zero lag value up to some chosen maximum lag. To determine the basis of the signal in the KL space we have to estimate the eigen values of the matrix. The autocorrelation function of a stationary signal is a vector from which we construct the autocorrelation matrix satisfying the conditions of the spectral theorem to make possible the required eigen decomposition. The form of the matrix is Toeplitz. A finite order Toeplitz matrix can be defined as:

$$T = \begin{bmatrix} t_0 & t_1 & \cdots & t_n \\ t_1 & t_0 & t_1 & \vdots \\ \vdots & t_1 & \ddots & \\ t_n & \cdots & \cdots & t_o \end{bmatrix} \tag{4.18}$$

The positive definite Toeplitz matrix satisfies the conditions of the spectral theorem.

The elements of the matrix are given by $t_n = \overline{x(t_i)x(t_{i+n})}$.

The next step is to determine the eigen values and eigen vectors of the autocorrelation Toeplitz matrix. The set of eigen vectors is the new basis set of the signal for the KLE and the eigen values are related to the energy of each of those vectors. The KL coefficients are calculated according to the equations (4.6) and (4.14) that correspond to projections of the signal onto the vectors of the new basis set weighted by the eigen values.

Summarising, the procedures for KLE applied to stationary or weakly stationary signals are:

1-Estimation of the autocorrelation function (the values taken are form the zero lag to the chosen maximum lag).

2-The estimate is used to build a square matrix of Toeplitz form.

3-The eigen decomposition is applied over the autocorrelation Toeplitz matrix.

4-The eigen vectors of the matrix are the vectors of the basis of the KLE and the eigen values are the corresponding energy associated with each vector.

4.4 The application of KLE to nonstationary signals

One of the aims in this research is the application of the KLE to nonstationary signals. The KLE can be applied to nonstationary signals since the nonstationary autocorrelation function has the properties required by the expansion. We shall establish the steps to apply the KLE to such signals using the modelled piano note expressed in equation 3.1.

For our model it is possible to determine the theoretical expression for the autocorrelation for some relatively simple cases. For more complex signals the nonstationary autocorrelation generally has to be estimated by ensemble averaging.

We shall verify if, for our model, the ensemble averaging autocorrelation matrices (we use the discrete form) can replace the theoretical without significant error.

4.4.1 The nonstationary autocorrelation function

The nonstationary autocorrelation of a random nonstationary signal $\{x(t)\}$ is for a pair of times t_1 and t_2 of $x(t_1)$ and $x(t_2)$ expressed as:

$$R_x(t_1,t_2) = E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty}\int x_1 x_2 p(x_1,t_1;x_2,t_2)dx_1 dx_2 \tag{4.19}$$

where $p(x_1,t_1;x_2,t_2)$ is the joint second order nonstationary probability density function of $x(t_1)$ and $x(t_2)$, defined as:

$$p(x_1,t_1;x_2,t_2)=\lim_{\substack{\Delta x_1\to 0\\ \Delta x_2\to 0}}\frac{\text{Prob}[x_1<x(t_1)\le x_1+\Delta x_1\cap x_2<x(t_2)\le x_2+\Delta x_2]}{\Delta x_1\Delta x_2} \tag{4.20}$$

The autocorrelation varies with both times t_1 and t_2 instead of only the difference t_2-t_1 as in the stationary case.

As a first step we are interested in applying the KLE to the piano signal represented in equation (3.1). The autocorrelation can be calculated theoretically from equation (4.19) or it can be can be estimated by ensemble averaging where there is the need to have an ensemble of possible realisations of the signal. For our signal we consider that the amplitude, damping and frequency are random variables that are statistically independent of each other:

$$p(A,\alpha,\omega)=p(A)p(\alpha)p(\omega) \tag{4.21}$$

and for convenience we assume to be discrete in an interval with a uniform distribution over some chosen range. We are interested to verify if for our modelled signal the ensemble averaging can replace the theoretical correlation.

Suppose that X is the ensemble of possible realisations of the discrete signal (each row of the matrix is one member of the ensemble):

$$X=\begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_k)\\ x_2(t_1) & x_2(t_2) & \cdots & \vdots\\ \vdots & \cdots & \ddots & \vdots\\ x_n(t_1) & \cdots & \cdots & x_n(t_k)\end{bmatrix} \tag{4.22}$$

The ensemble average estimator for autocorrelation of X is computed as:

$$\hat{R}_x(t_1,t_2)=\frac{1}{n}\sum_{i=0}^{n}x_i(t_1)x_i(t_2) \tag{4.23}$$

This is equivalent to determine the matrix product (i.e. if X is a row matrix):

$$R=X^TX \tag{4.24}$$

to within a scale factor. The matrix X^T is the transpose of X (in the matrix product we use the transpose matrix because we are working with real data. For complex data the transpose matrix is replaced by the transpose of the complex conjugate). This matrix product corresponds to the autocorrelation of the signal.

4.4.2 Theoretical nonstationary autocorrelation function simple model of a piano signal

The estimation of the theoretical nonstationary autocorrelation function was applied to our modelled piano note. For this calculation of the theoretical autocorrelation function we shall assume that the phase as zero and the amplitude A_i is assumed unity. The other variables of equation 3.1, corresponding to the modelled signal, α_i and ω_i were drawn randomly from intervals with uniform distribution. The corresponding density functions are $p(\alpha)$ and $p(\omega)$. These density functions, for a variable D is expressed as:

$$p(D)=(D_1-D_2)^{-1} \qquad D_1 \le D \le D_2 \tag{4.25}$$

$$=0 \qquad D_1 > D > D_2$$

The theoretical expression of the autocorrelation function of the modelled piano signal determined by equation 4.19 is given then by:

$$R_x(t_1,t_2)=\int_{\alpha_1}^{\alpha_2} e^{-\alpha(t_1+t_2)}p(\alpha)d\alpha\int_{\omega_1}^{\omega_2}\sin(\omega t_1)\sin(\omega t_2)\,p(\omega)d\omega \tag{4.26}$$

And the final expression is:

$$\begin{aligned}R_x(t_1,t_2)=&\frac{1}{2}\left[\frac{e^{-\alpha_2(t_1+t_2)}-e^{-\alpha_1(t_1+t_2)}}{t_1+t_2}\right]\times\\&\times\left[\frac{\sin\omega_2(t_1-t_2)-\sin\omega_1(t_1-t_2)}{t_1-t_2}-\frac{\sin\omega_2(t_1+t_2)-\sin\omega_1(t_1+t_2)}{t_1+t_2}\right]\end{aligned} \tag{4.27}$$

This expression was computed considering that the amplitude always was unity. The damping varied in the interval $[0.1,0.3]$ and the frequency in the interval $[9.5,10.5]$Hz.

In figure 4.1 we have one possible realisation of the signal corresponding to those intervals of variation. The resulting autocorrelation is represented in the 3D graph of figure 4.2. The graph of the autocorrelation is a detail of the first 0.5s and it can be seen clearly that the values decrease from the zero lag until the 0.5s lag. The diagonals of the matrix have different values and decrease. The structure of this autocorrelation matrix is not Toeplitz in contrast to the case of a stationary autocorrelation matrix.

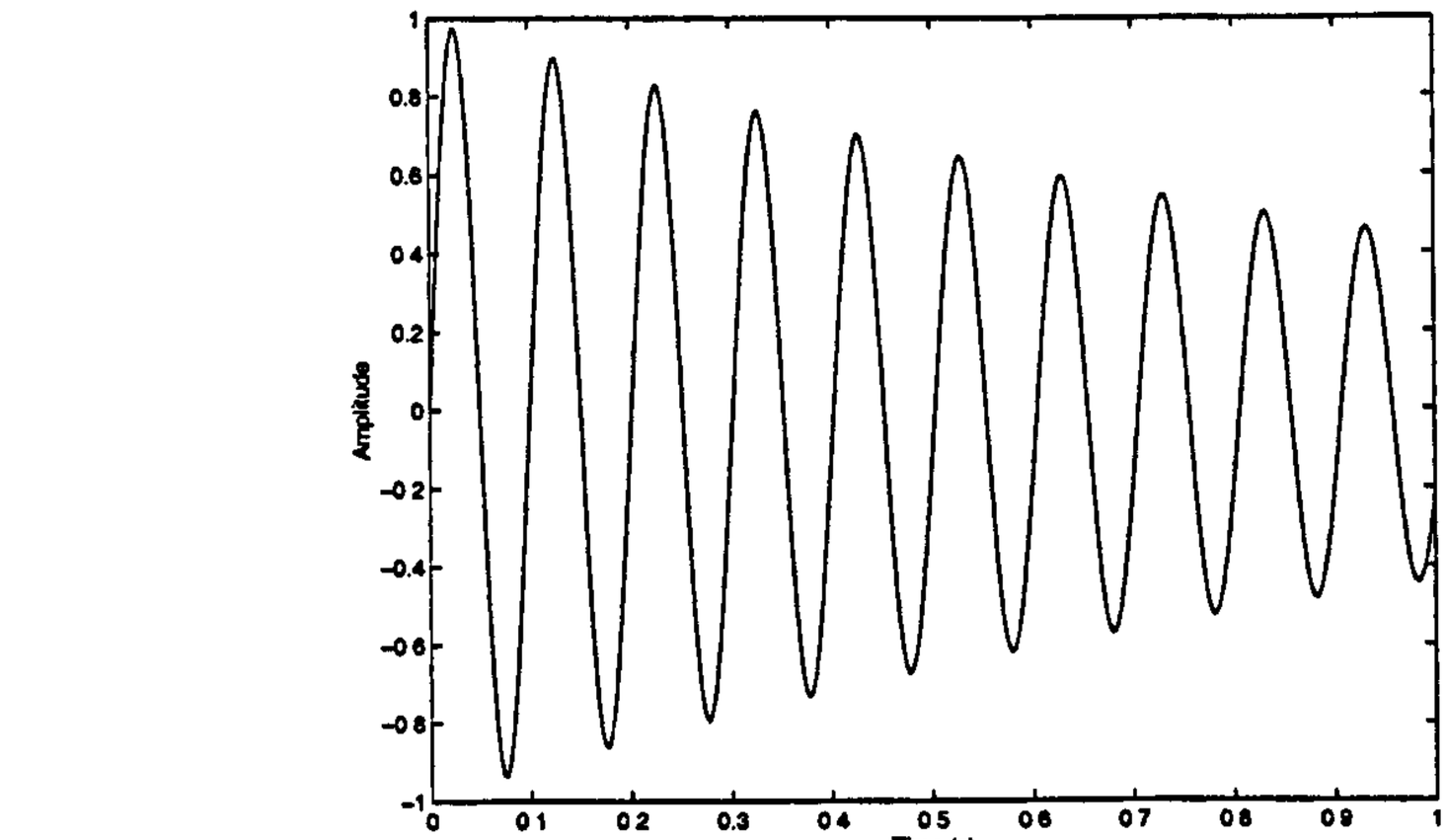


Fig 4.1 Possible event of the signal with unity amplitude and damping taken from the interval $[-0.3,-0.1]$ and frequency form the interval $[9.5,10.5]$Hz.

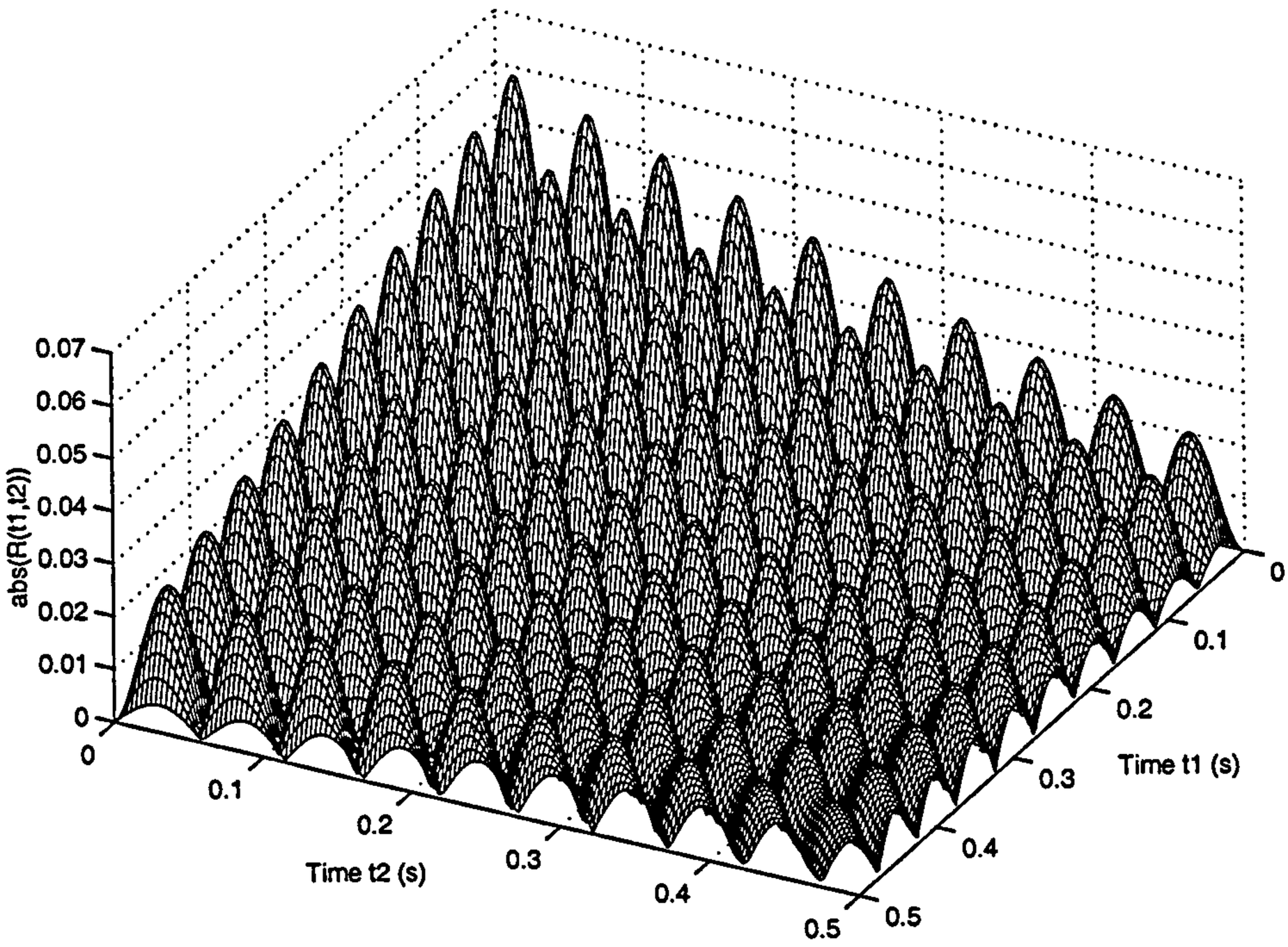


Fig 4.2 Absolute values of the theoretical autocorrelation matrix.

4.4.3 Nonstationary autocorrelation matrix estimated by ensemble averaging for modelled piano signal

To determine the autocorrelation function by ensemble averaging we generate a set of 100 possible events with 300 samples each. The realisations are created by drawing the frequency and damping from uniform distributions. The intervals of

variation of the frequency and damping were exactly the same as for the theoretical case. We chose the example of an underdetermined system (smaller number of events than number of samples) since, as we will see later, for the case of the real piano notes we will be working with underdetermined systems. The time history matrix of the signal was built with those 100 events (we choose to build a matrix whose rows are the time histories). The. autocorrelation function of the signal was estimated by ensemble averaging using the expression (4.24) which corresponds to a matrix multiplication of one matrix by its transpose.

The resulting matrix has the same structure as the theoretical matrix (see fig.4.3). We are interested to verify in more detail if the results are similar. For that we will address the diagonals of those matrices (fig.4.4) that correspond to zero lag. It is possible to see that there is a small difference in the later elements of the diagonals. In the rest of the samples the values are practically coincident. The elements of the last lines and columns present a slight difference from the theoretical. The ensemble averaging is only an approximation to the theoretical one and the difference of the results from the number of events used. For this case the approximation between the two autocorrelations is considered satisfactory and the ensemble averaged autocorrelation can replace the theoretical.

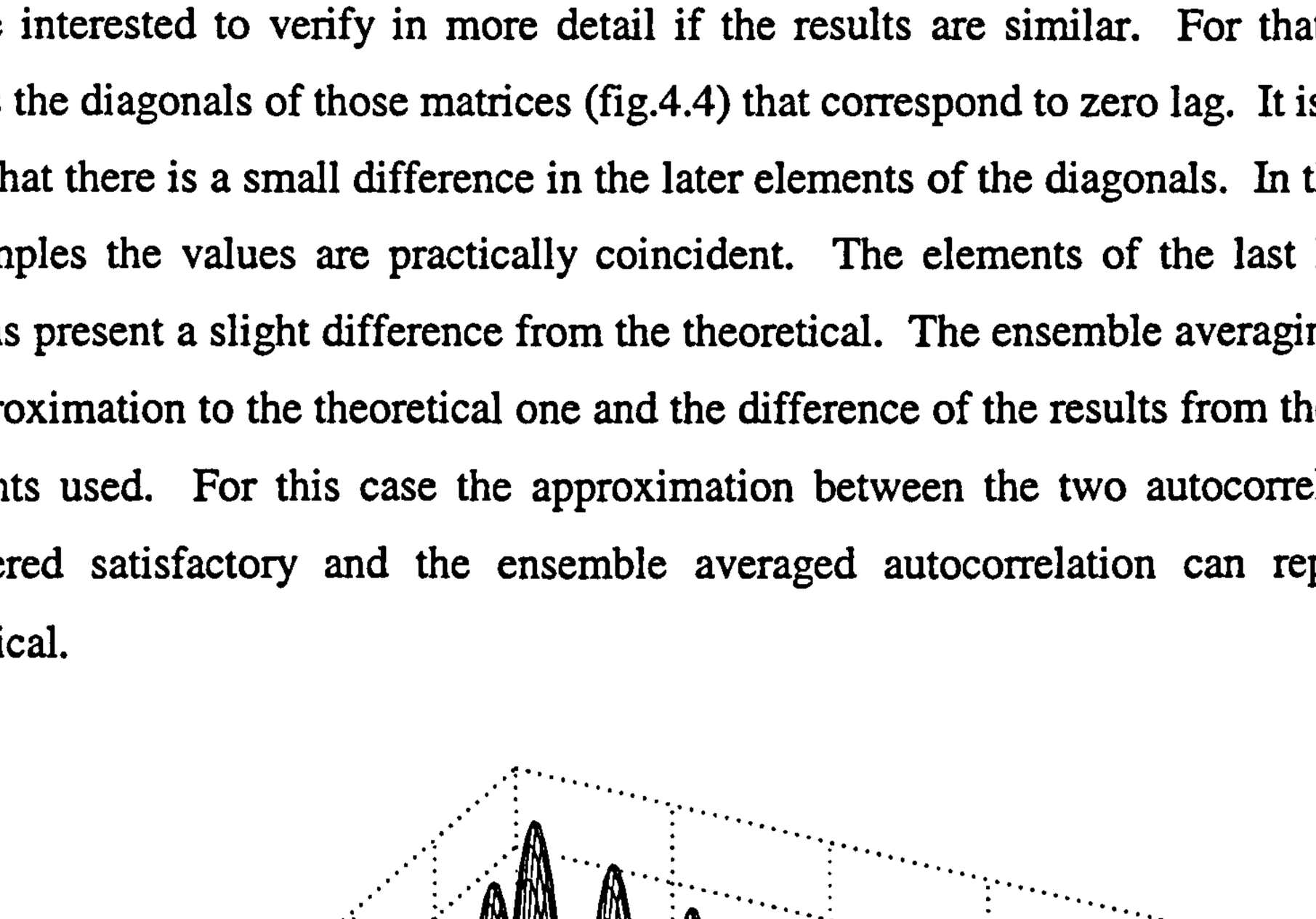


Fig 4.3 Absolute values of the ensemble averaged autocorrelation matrix.

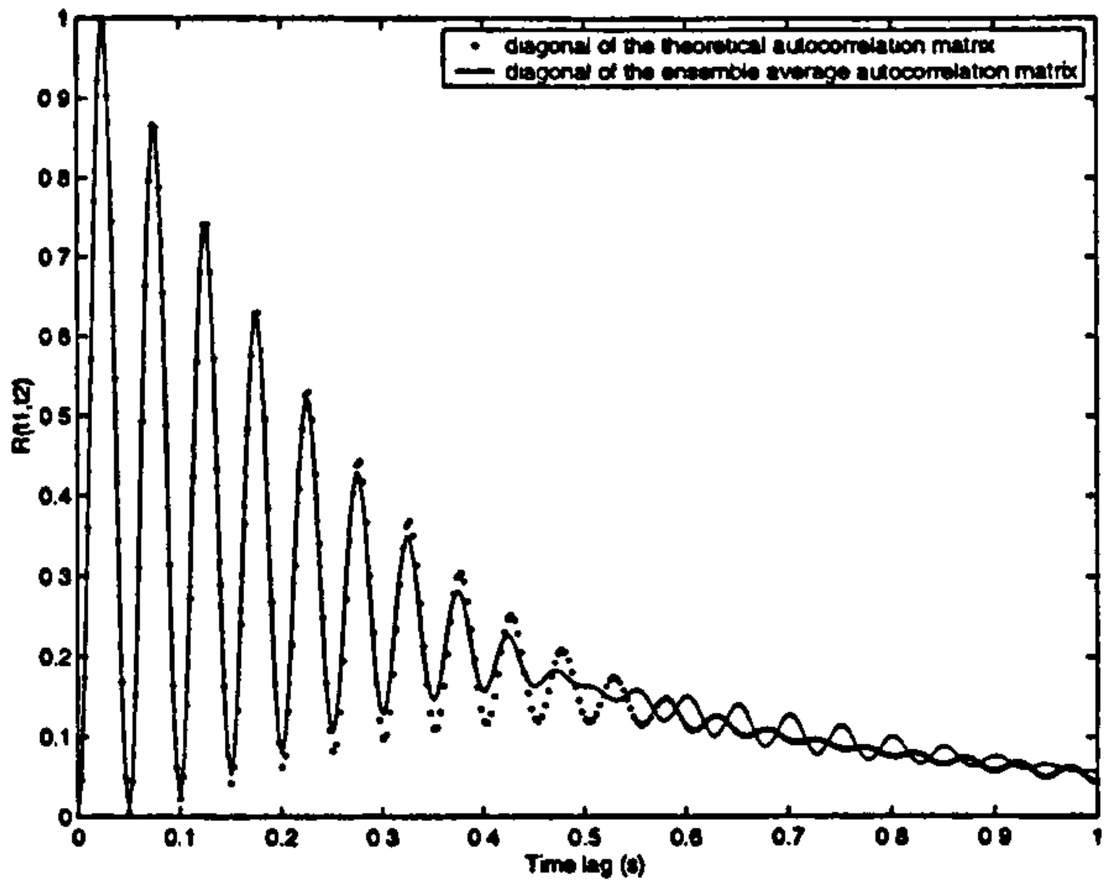


Fig 4.4 Diagonals of the autocorrelation matrices.

There are further factors that should be analyzed namely the off diagonal terms to ensure that the theoretical autocorrelation could be replaced with a small margin of error. We shall do this indirectly by verifying whether the resulting eigen vectors and values of the KLE are significantly affected by this small difference or not. In the next section we will compare the result of the eigen decomposition of these matrices.

4.4.4 The eigen decomposition of the nonstationary autocorrelation matrices

Now we will eigen decompose both matrices to see if the small difference interferes significantly in the resulting eigen vectors and eigen values.

The number of significant eigen values (i.e. those significantly non zero) is the same for both cases (fig.4.5). The eigen values for each case were divided by the first value. This way we can see if the relation between the values were changed. The relation between the eigen values changes for the second value.

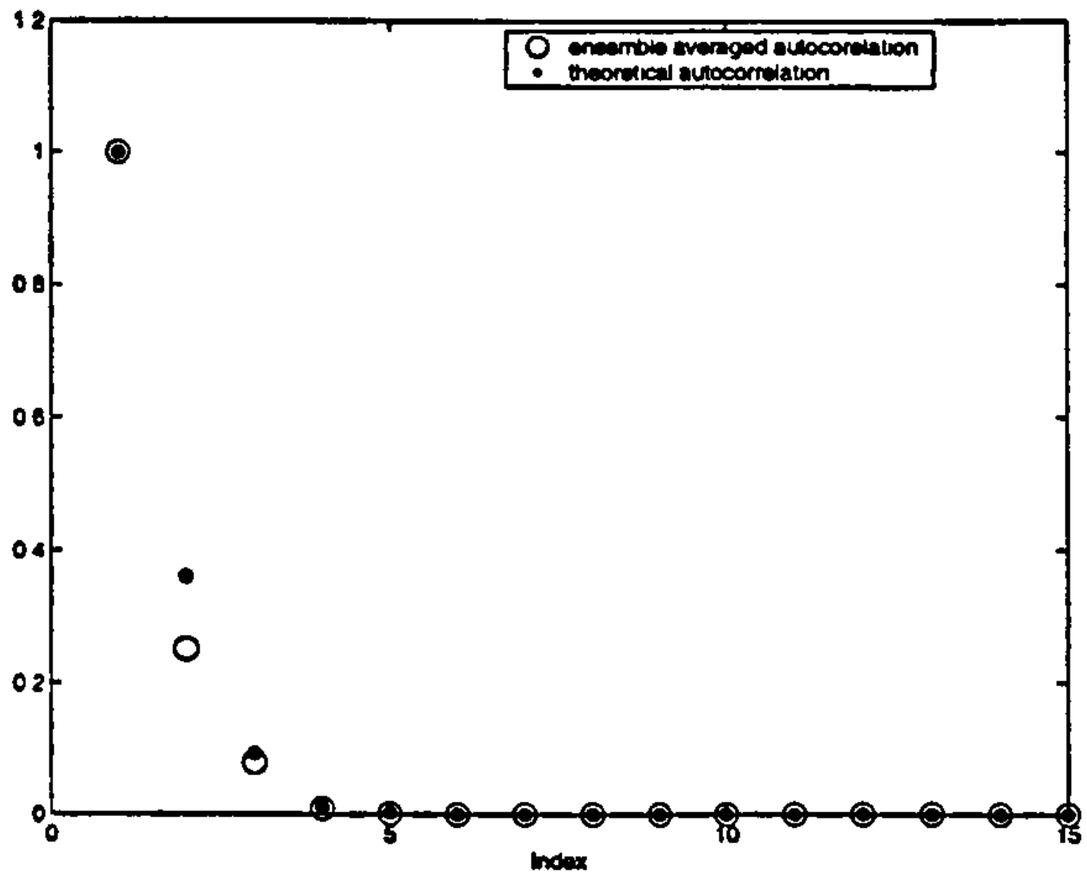


Fig 4.5 Eigen values of the theoretical and ensemble averaged autocorrelation matrices.

We also considered the corresponding first three eigen vectors as presented in fig.4.6 and 4.7 and observed that there are small differences. On the first vector those differences are in the latter samples. The second eigen vectors are quite similar, although there is a difference in the eigen values, and on the third the difference is now in the early samples.

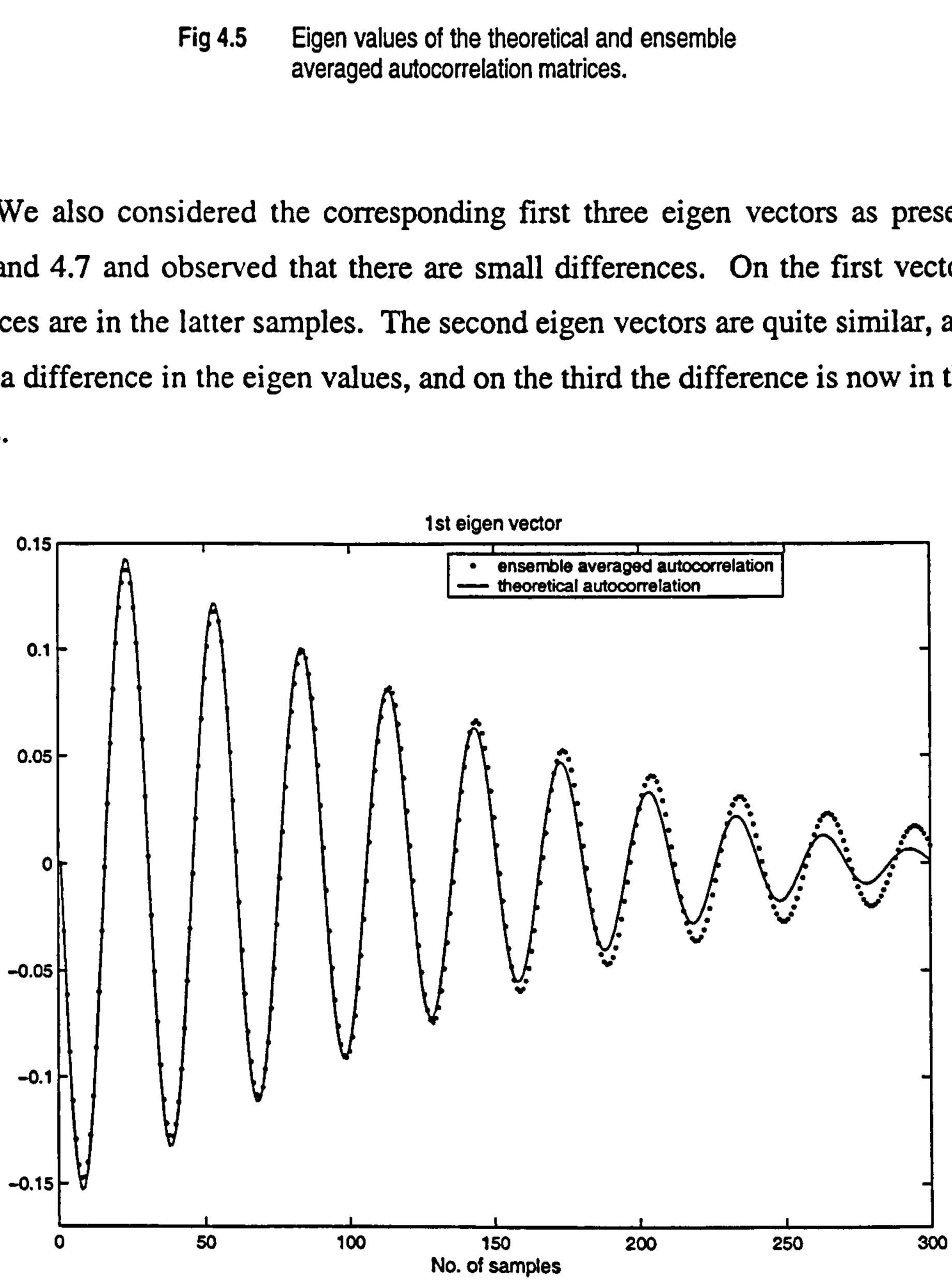


Fig 4.6 First eigen vectors of the autocorrelation matrices

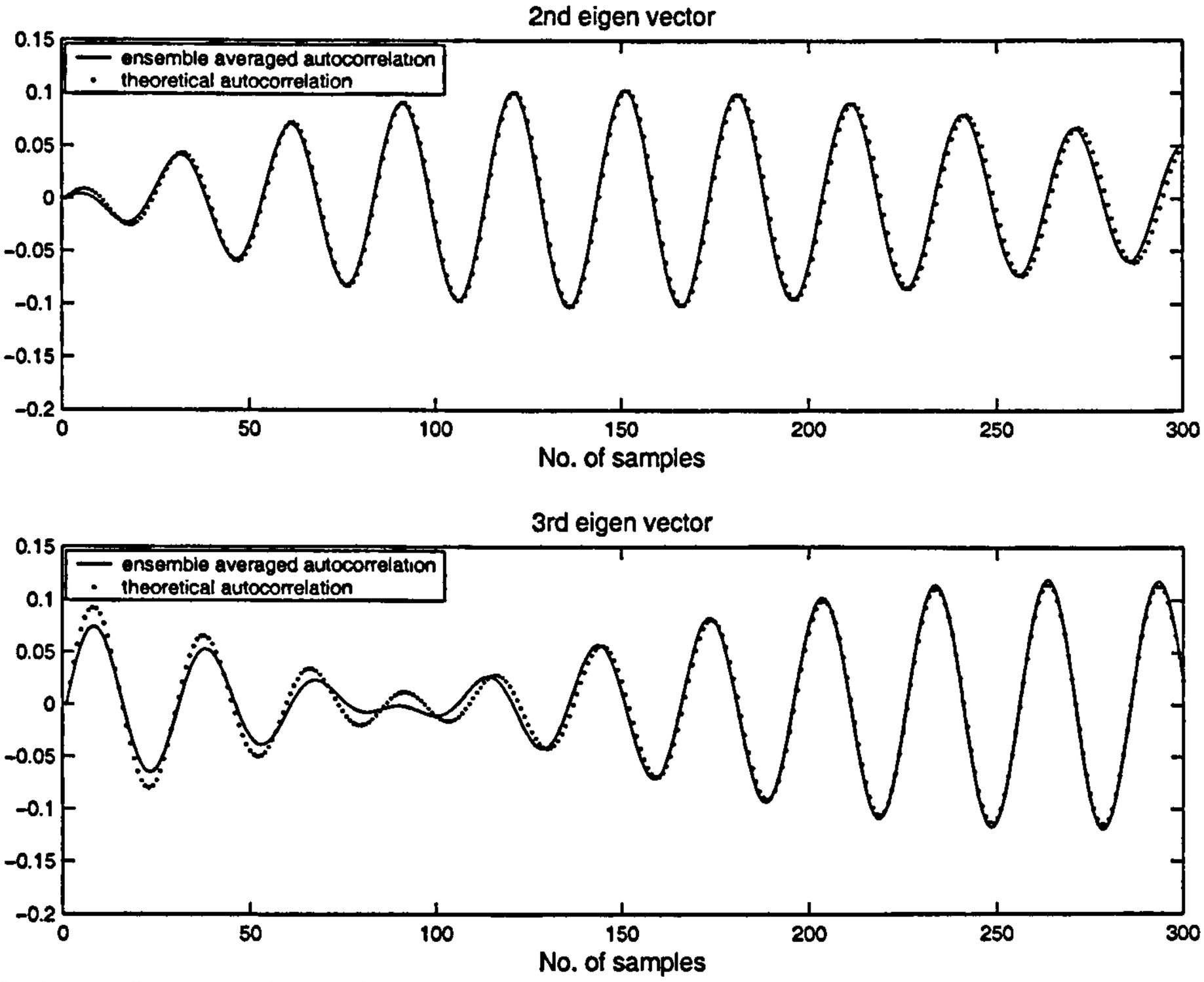


Fig 4.7 Second and third eigen vectors of the simulated signal for the stationary case.

Whilst these examples are empirical they provide evidence that the decompositions of the autocorrelations do not present differences that can be considered important for our research. Accordingly we will use the ensemble averaging approach with some confidence.

4.5 Nonstationary versus stationary data

We saw that the KLE can be applied to nonstationary signals using the nonstationary autocorrelation function. We will see how the KLE behaves when we use only one possible realisation of a random nonstationary signal and apply the stationary procedures.

The treatment of the signal as stationary or nonstationary is very important for the final result of the KLE. The estimated autocorrelation functions are different for each case that will result in different eigen values and eigen vectors. We will see how the vectors completely change their shape when we use different autocorrelation functions.

The signal used to compare both cases is our modelled piano note represented in equation (3.1). For this signal we calculated the stationary and nonstationary autocorrelation functions. The values the stationary autocorrelation $R(m)$ are estimated by time averaging as:

$$R(m) = E\left[x(n)x(n+m)\right] = \frac{1}{N}\sum_{n=1}^{N-m} x(n)x(n+m) \tag{4.28}$$

The stationary autocorrelation matrix is built up using the estimated autocorrelation function computed as in equation (4.23). The autocorrelation matrix is a Toeplitz matrix. The nonstationary autocorrelation matrix was determined by ensemble averaging using a number of possible events of the signal. (The nonstationary autocorrelation matrix was determined also using the theoretical equation (4.20)).

Then the KLE was applied to each case. In the stationary case the event of the signal used to determinate the eigen values is represented in figure 4.8. The resulting expansion presents the most significant eigen values in pairs (fig.4.9). The corresponding eigen vectors for each pair have similar shapes (fig.4.10). The two first vectors have the shape of modulated sine waves. Those vectors do not relate in form to any of the original transient characteristics of the signal.

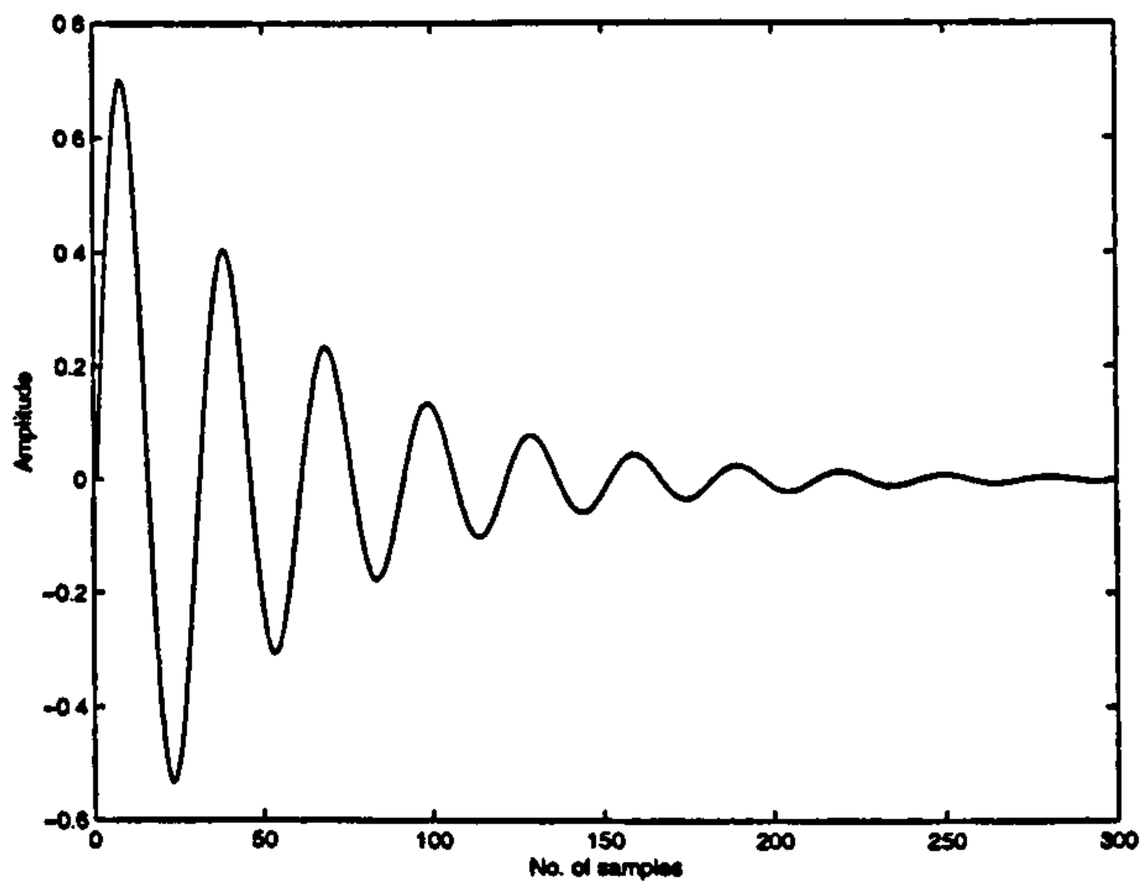


Fig 4.8 Event of a simulated note with frequency varying in the interval $[9.5,10.5]$ Hz (centre frequency of 10Hz).

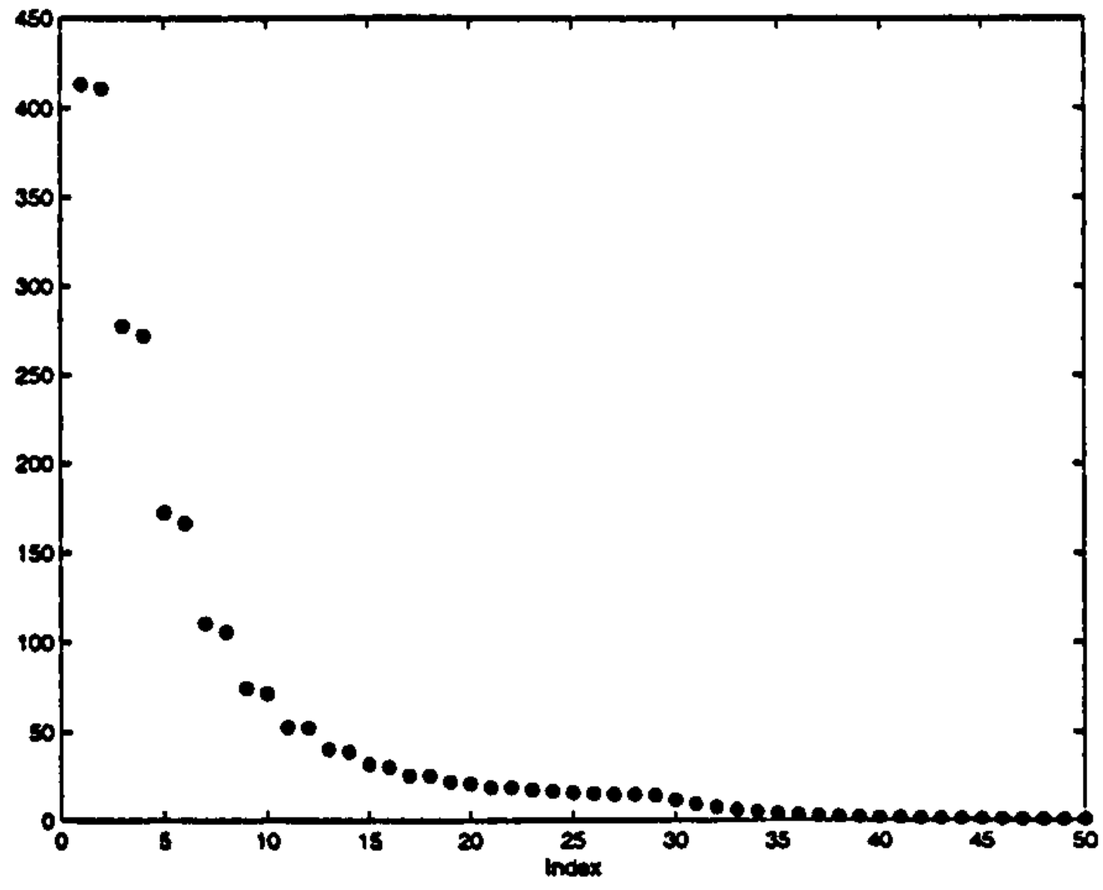


Fig 4.9 Eigen values for simulated signal for the stationary case.

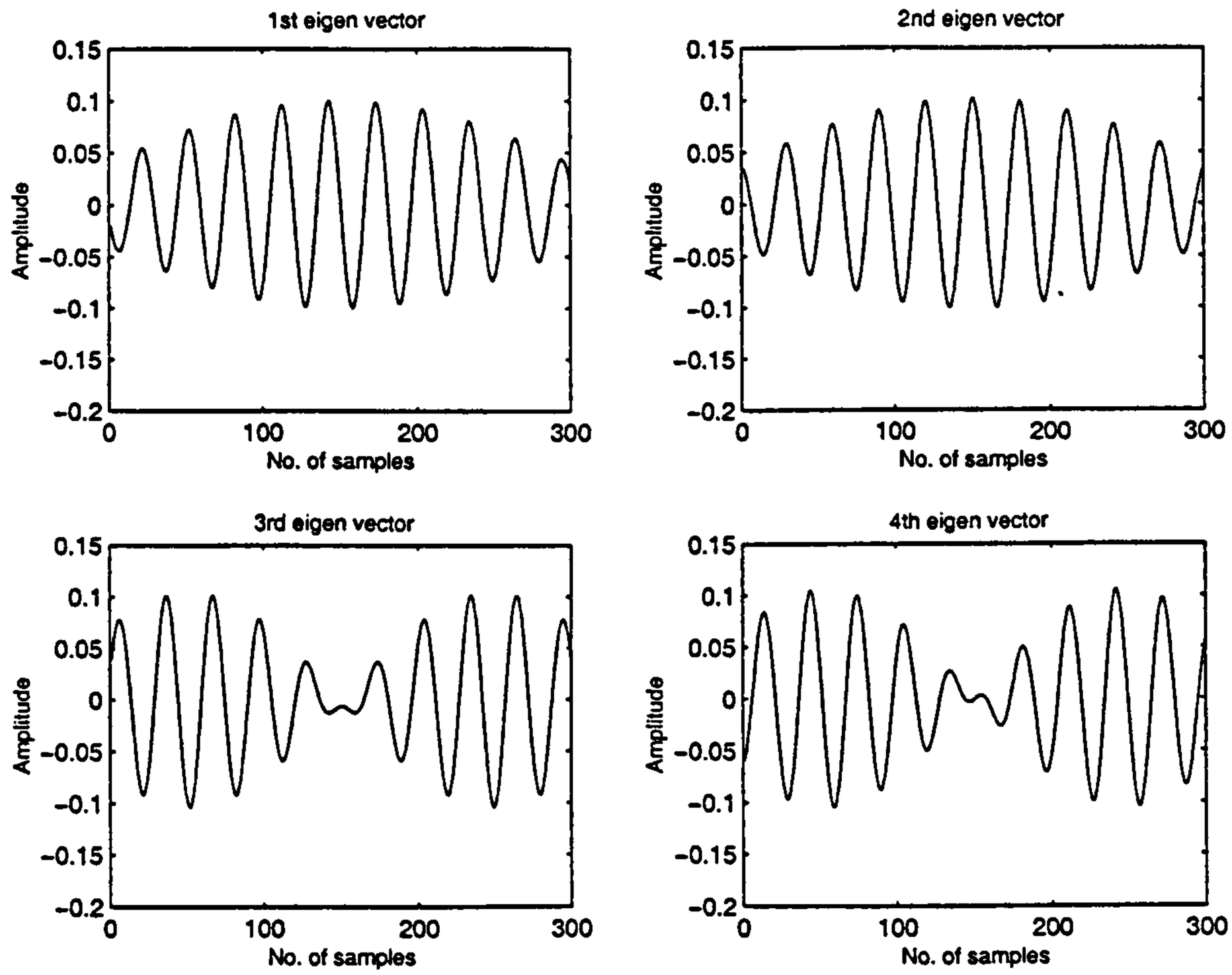


Fig 4.10 First four eigen vectors of the simulated signal for the stationary case.

For the nonstationary case the relation between the eigen values changes completely. The eigen values are not in pairs and the first one has a very high value compared with the other ones. This means that there is more information about the original signal in one vector, the first vector in the nonstationary case. The new spaces

defined by the singular vectors of are different according to the way the signal is treated as either stationary or nonstationary.

The first singular vector (fig.4.12) has a shape very similar to the original signal. All the transient characteristics are present in this vector. The other vectors present modulation and the 3rd and 4th vectors appear to be a version of the 2nd vector moved to the right.

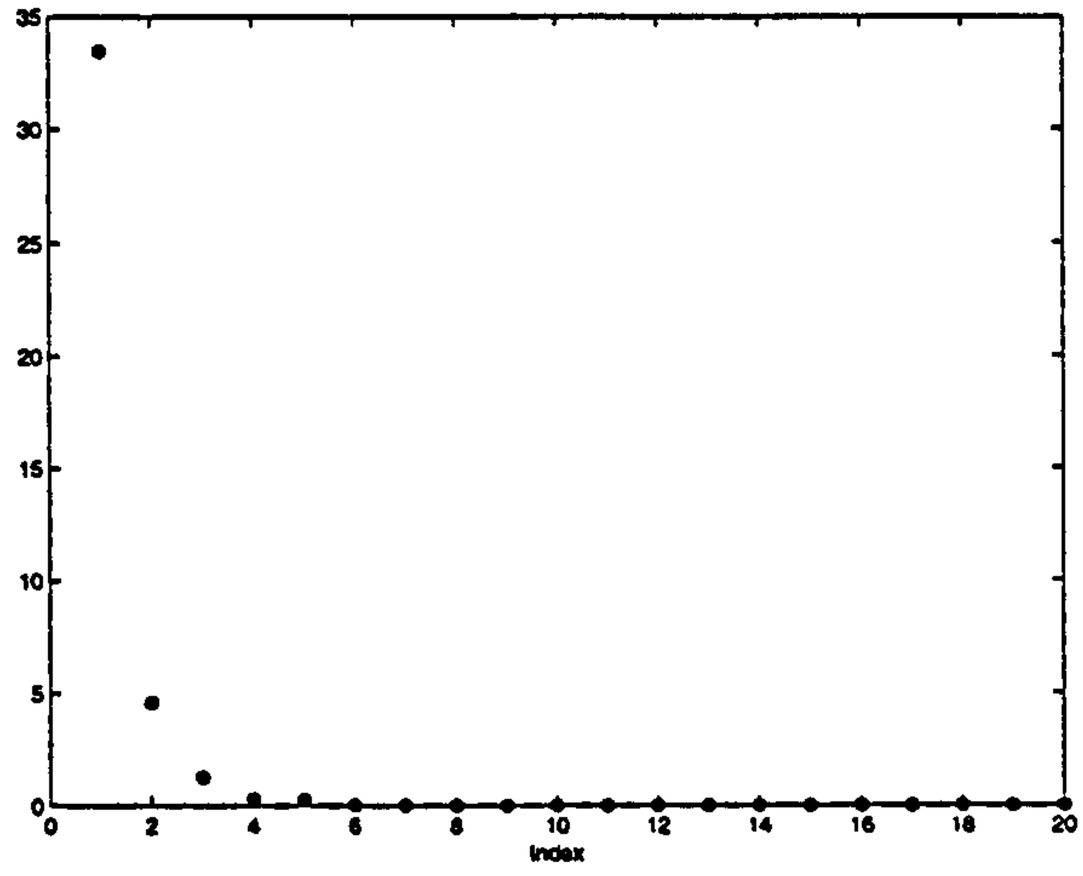


Fig 4.11 Eigen values of the same simulated signal for the nonstationary case.

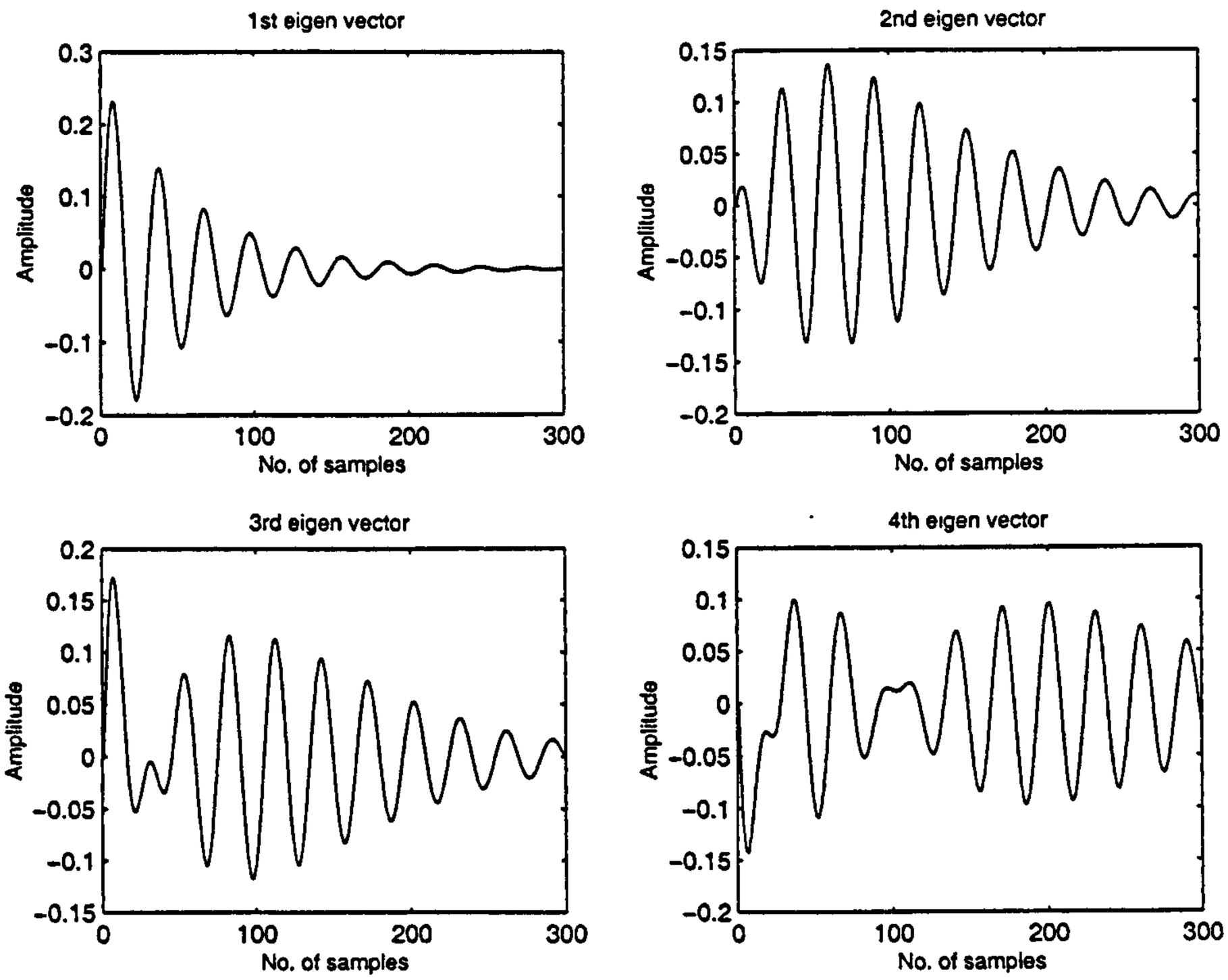


Fig 4.12 First four eigen vectors for the simulated signal in the nonsntationary case.

The reason for the difference in the resulting eigen vectors for the stationary and nonstationary case is that the autocorrelation matrices have different structure although they are both symmetric. In the stationary case all the elements of each of the diagonals of the Toeplitz matrix have the same value, i.e. $x_{ij} = x_{(i+1)(j+1)}$, and for the nonstationary autocorrelation matrix the diagonal elements have different values (see fig.4.3).

The results of the KLE show that it is highly dependant on the form of the autocorrelation matrix. Thus our assumptions as to the nature of the data and the computation of the autocorrelation matrix are crucial.

In next chapter we will see how the shape of the vectors in the nonstationary case can vary depending on the different parameters.

4.6 The KLE of a Wiener process

Since this research is concerned with nonstationary piano signals it is instructive to examine other nonstationary signals for which we can write the theoretical autocorrelation function. One such process is the Wiener process (also known as Brownian motion). This process, w, can be expressed as:

$$\frac{dw}{dt} = \xi \tag{4.29}$$

where ξ is a white noise zero mean Gaussian process, i.e.:

$$E\left[\xi(t_1)\xi(t_2)\right] = \delta(t_1 - t_2) \tag{4.30}$$

From the solution,

$$w(t)=\int_0^t \xi(t')dt' \tag{4.31}$$

it is possible to show that the correlation of the process $w(t)$ is:

$$R_w(t_1,t_2)=\min(t_1,t_2) \tag{4.32}$$

for any t_1 and $t_2 \in [0,\infty]$.

This is a nonstationary process where the variance grows linearly namely:

$$\mathrm{var}\left[w(t)\right]=t \tag{4.33}$$

We define a correlation matrix $R_w(300\times300)$ with the variables t_1 and t_2 taking integer values in the interval $[1,300]$. The KLE of this stochastic process has a very interesting result. Despite the nonstationary nature of the process the eigen vectors are sine waves (see fig.4.14) with increasing frequency for the higher order eigen vectors. There are 8 significant positive eigen values indicating that the correlation matrix has 8 linear independent lines or rows.

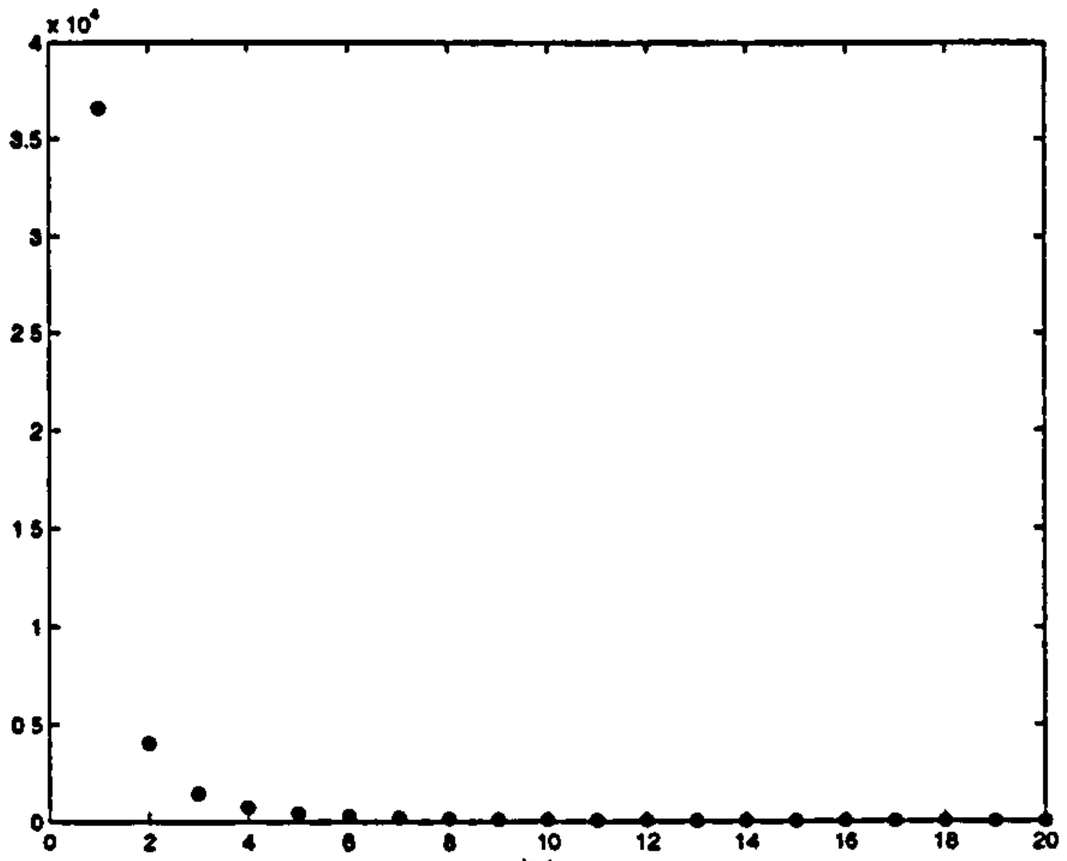


Fig 4.13 First 20 eigen values of a Wiener process.

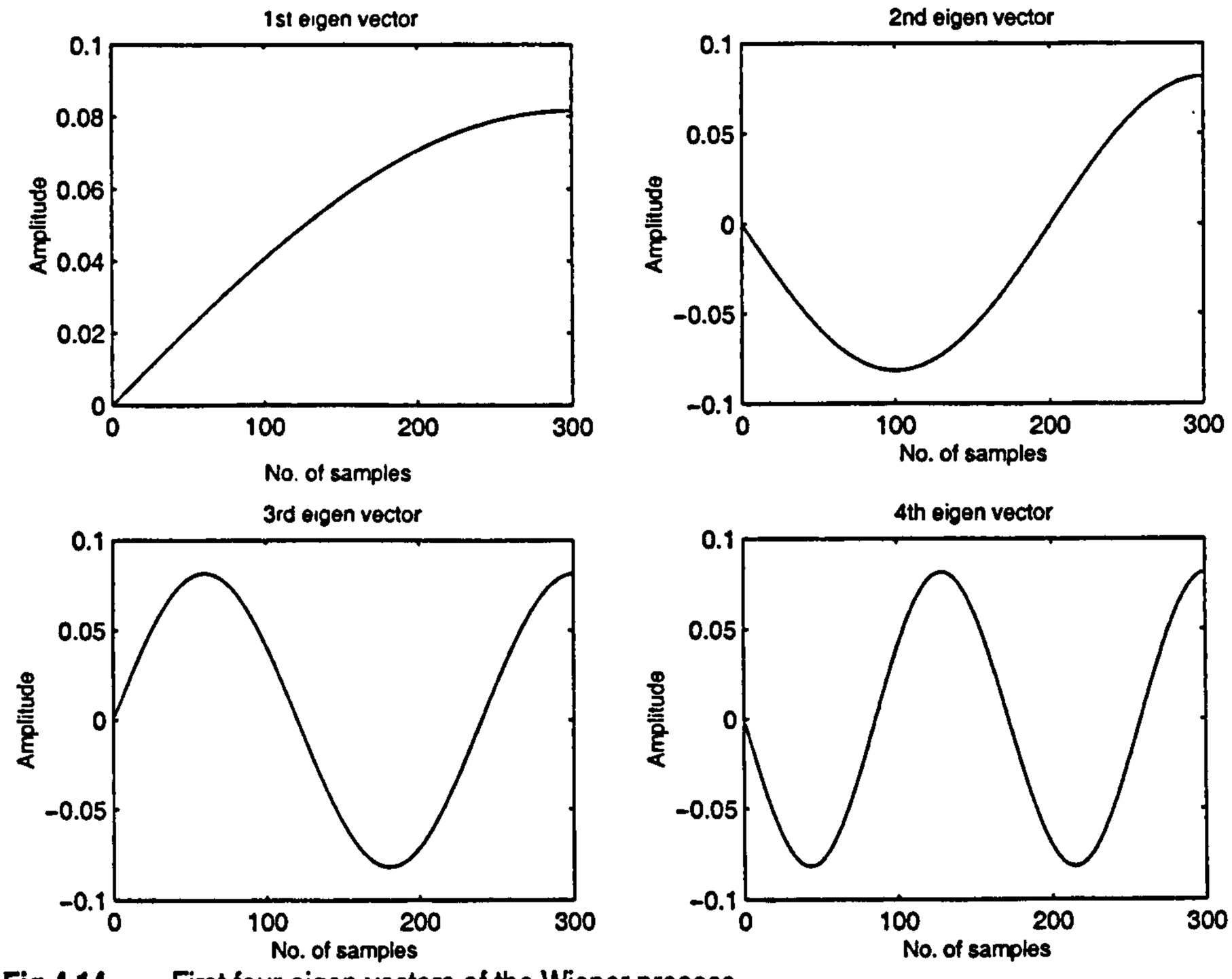


Fig 4.14 First four eigen vectors of the Wiener process.

So the Wiener process, is represented in the KLE space, defined by the eigen vectors of the correlation matrix, in terms of sine functions. A nonstationary process whose variance is a function of time that grows linearly has a KLE decomposition whose basis sets are sine wave functions whose frequency increases from vector to vector.

CHAPTER 5

THE KARHUNEN-LOÈVE EXPANSION (KLE) AND SINGULAR VALUE DECOMPOSITION (SVD)

5.1 The SVD Theorem

Singular value decomposition (SVD), as with the KLE, is a mathematical method used in matrix decomposition similar to the KLE. The interest in SVD for a possible application to this research is based on its more convenient formalism, with particular reference for the case of nonstationary data. The KLE and SVD, although similar, operate over different statistical moments of the signal in the sense that for KLE we need to calculate the autocorrelation function and for SVD we simply need to form the time history matrix. SVD becomes very convenient for nonstationary data characterised by an ensemble since the second order statistics required for the KLE are not computed.

We will demonstrate the equivalence between SVD and KLE in later sections.

5.1.1 Singular Value Decomposition

Any matrix $A(m\times n)$ can be represented in the form of a product of three matrices defined as,

$$A = U\Sigma V^{H} \tag{5.1}$$

where $U(m\times m)$ and $V(n\times n)$ are orthonormal matrices. The matrix V^{H} (Hermitian matrix) corresponds to the complex conjugate of the transpose of the matrix V. If the rank of A is $k=\min(m,n)$, the matrix $\Sigma(m\times n)$ has k nonnegative diagonal elements arranged in descending order and we refer to them as:

$$\Sigma = diag(\sigma_1,\sigma_2,\ldots,\sigma_k) \tag{5.2}$$

with real $\sigma_i \geq 0, \quad i=1,2,...,k$. These elements of the diagonal matrix are the singular values of the matrix A. The columns of matrix U and V are defined as the left and right singular vectors of matrix A. For real matrices V^{H} can be replaced by the transpose V^{T}.

The important fact about this theorem is that it states that any $(m\times n)$ has a singular values decomposition. The matrices are decomposed as the product of three matrices, two of them are orthonormal corresponding to the left and right singular vectors and one is diagonal whose elements are the singular values of the matrix. We give a practical example of the SVD. Suppose that we have a row matrix $A(3\times 4)$ representing an underdetermined system (we choose the underdetermined matrix $m<n$ since we will be using these systems in our research for the real piano notes) whose SVD results in twopositive singular values σ_1 and σ_2. If the right singular vectors are represented by $\vec{r}_{ij}$ and $\vec{l}_{ij}$ then we can represent the SVD of A as:

$$A=\begin{bmatrix}\vec{l}_{i1} & \vec{l}_{i2} & \vec{l}_{i3}\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\vec{r}_{i1} & \vec{r}_{i2} & \vec{r}_{i3} & \vec{r}_{i4}\end{bmatrix}^T \tag{5.3}$$

The left singular vectors matrix is a (3×3) column matrix with three singular vectors. The right singular values matrix is a (4×4) column matrix with four singular vectors (this matrix is transposed so the vectors will be are represented in rows). Both matrices represent an orthonormal basis set one in the $\mathbf{R}^3$ space and the other in the $\mathbf{R}^4$ space.

Any row i of the matrix A, a_{ij}, representing an event of the system is in terms of the SVD written as:

$$a_{ij}=\sum_k l_{ik}\sigma_{kk}r_{jk} \qquad i=1,2,...,n \quad j=1,2,...,m \tag{5.4}$$

The diagonal matrix of the singular values always has the dimension of the A matrix to allow the multiplication of the matrices. The number of positive singular values indicates the rank of the matrix. In our example the rank of A is 2 (the rank of the left and right singular vectors matrix is full rank). This means that there are two linearly independent rows in matrix A. With only two positive singular values we have the right singular vectors $\vec{r}_{i1}$ and $\vec{r}_{i2}$ defining the right singular space of A and the left singular vectors $\vec{l}_{i1}$ and $\vec{l}_{i2}$ defining the left singular space of A. The other vectors of each matrix define the null space of $A\ (3\times3)$. The null space of the matrix $A\ (3\times3)$ is composed of the vectors that satisfy the equation:

$$A\vec{v}=0 \tag{5.5}$$

Generally when there are no zero singular values in the diagonal matrix, the null space of the right singular vector matrix has $n-m$ singular vectors (in the example we should have $4-3=1$ singular vectors in the null space). If any of the singular values is zero then the vectors associated with that vector belong also to the null space. We can suppress the vectors of the null space, since they do not contribute to the resulting A matrix, and rewrite A as:

$$A=\begin{bmatrix}\vec{l}_{i1} & \vec{l}_{i2}\end{bmatrix}\begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}\vec{r}_{i1} & \vec{r}_{i2}\end{bmatrix}^T \tag{5.6}$$

One vector of the left singular vector matrix is omitted as are two vectors of the right singular matrix. The equation 5.4 represents the "economy size" version of the SVD. We note that the vector of the left singular vectors is only removed since there is a zero singular value in the diagonal matrix.

Suppose that matrix A (3×3) represents a signal without the presence of noise. If white noise is added to the signal the rank of the matrix becomes full. The noise introduces linear independency between the events. In this way all the zero singular values will have now a small positive value. In equation 5.3 the diagonal matrix will have one more singular value σ_3. The singular values corresponding to the presence of noise are smaller than the other ones. The noise component on the most significant vectors and is very small since the white noise is uncorrelated with the signal. If the signal has another type of noise then it is possible to have a significant amount of noise in the vectors.

For signals with noise, SVD can be used to remove the noise from the signal. The signal can be reconstructed omitting the singular values and vectors corresponding to the noise.

5.2 The relation between SVD and KLE

The KLE and SVD are two similar techniques and we are interested to verify under which conditions the KLE can be replaced by SVD with similar results. In particular we shall investigate in which practical cases the both KLE and SVD can give similar resulting basis sets.

The singular vectors and the eigen vectors of a matrix can be related. For a matrix $A(m \times n)$ the nonnegative square roots of the eigen values of the matrix product $A^H A$, if $m \geq n$, and of AA^H, if $m \leq n$, are the singular values of A (Lütkepohl,1996).

If a matrix A, represented by equation 5.1, is multiplied by its transpose A^T (left side and right side), the products will be:

$$A^T A = \left(U\Sigma V^T\right)^T \left(U\Sigma V^T\right) = V\Sigma^2 V^T \tag{5.7}$$

and,

$$AA^T = \left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T = U\Sigma^2 U^T \tag{5.8}$$

since the matrices U and V are orthonormal and Σ a diagonal matrix.

The product of a matrix by its transpose (left or right product) that are represented by its SVD, results in a product of three matrices. The resulting products have the same left and right singular vectors matrices and the diagonal matrix (the same for both products) continues to be diagonal but is now a squared matrix (Σ^2 is a $(m \times m)$ or a $(n \times n)$ matrix) and the elements are the squares of the elements of matrix Σ. These decompositions of a matrix as the product of three matrices one diagonal and the left right matrices are the same since we have the same matrices.

So equations (5.7) and (5.8), state that the matrix product of a matrix by its transpose, or vice versa, can be eigen decomposed since the left and right matrices are the same. The matrix of the right singular vectors, V, is also the matrix corresponding to the eigen vectors (equation (5.3)). We have the same for matrix U in equation 5.4. This means that the singular vectors of the matrix A are equivalent to the eigen vectors of the matrix product. Since the products correspond to an eigen decomposition the matrix, Σ^2, is a diagonal matrix whose elements are the eigen values of those products and correspond to the square of the elements of Σ. If λ_i are the eigen values of the matrix products and λ_i and the singular values of A are σ_i then:

$$\lambda_i = \sigma_i^2 \qquad i = 1,2,...,n \tag{5.9}$$

These two techniques, KLE and SVD are equivalent according to equations (5.7) and (5.8) for any matrix $A(m \times n)$. The SVD is applied over any matrix $A(m \times n)$ and the KLE over the matrix product, $A^T A$ or AA^T. Both decompositions of these products have similar vectors (right or left) and the singular values are the positive square roots of the eigen values. For real data this means that SVD is similar to KLE when for a stochastic signal SVD is applied over the time history matrix and KLE over its autocorrelation matrix determined by the product, $A^T A$ or AA^T, depending if the signal is represented respectively by a row or column matrix.

Now we are interested to see how in practice for real data these two techniques are similar. We saw (eq. (4.24)) that the product of two matrices $X^T X$ or XX^T (apart from a scaling factor) is by definition, for real data, the autocorrelation function. Following the results in equations 5.7 and 5.8 the SVD when applied directly over the time history matrix of a stochastic signal has similar vectors to the KLE of the autocorrelation function.

5.3 How SVD is applied in practice and its relation to KLE

We will see how SVD is applied for stationary and nonstationary processes and how the results can be related to the KLE of the same processes.

For stationary processes only one realisation of the signal is used. In order to apply the SVD to stationary stochastic signals there is the need to build a time history matrix. This matrix can be built in two ways: one is to use a circulant structure and other is to use different windowed parts of the signal (in this case a window with length smaller than the signal, moves along the signal) that we can call the delay version. In the first case all the same samples of the signal are present in all rows and in the second case we introduce new samples.

In nonstationary processes the signal should be defined when possible by an ensemble of possible realisations of the signal. Since SVD does not impose any conditions to the matrices that can be decomposed we can apply it directly over the time history matrix containing the possible realisations of the signal.

As previously mentioned when the KLE is applied to stationary stochastic signals we use the autocorrelation function to build a square Toeplitz matrix and then determine the eigen decomposition. This matrix is not determined by the product of the matrices of the time histories since there is only one event. The resulting decomposition of the autocorrelation of the Toeplitz matrix does not have similar vectors to the SVD of the signal. The time history for the SVD of stationary signals is built using the circulant or the delays structure. The Toeplitz autocorrelation matrix does not correspond to the product of these time histories.

As we mentioned previously for nonstationary signals the KLE is similar to the SVD since the autocorrelation matrix is determined by the product of the time histories. In this case SVD can replace the KLE with similar results since equation 5.7 is verified.

In annex 4 we have a table with description of how both SVD and KLE are applied to stationary and nonstationary signals.

5.4 Examples of KLE and SVD for simulated signals

The next examples will show that in practice that the results of the KLE and SVD for nonstationary signals are similar. The modelled piano notes represented in equation 3.1, was used to determine the SVD and the KLE. A number of possible events of the signal were randomly generated with varying frequency, damping and amplitudes. The SVD was applied directly over the time history matrix of the nonstationary signal. For the KLE we used the nonstationary autocorrelation matrix estimated by the product $R = X^T X$, considering the matrix of the time data, X, as a row matrix.

The signal analysed was the same used in chapter 4, section 4.4. The singular values and eigen values represented in fig.5.1 and 5.2 are related according equation (5.9). The results could not be exactly related as by the equation due to small error introduced by the algorithms.

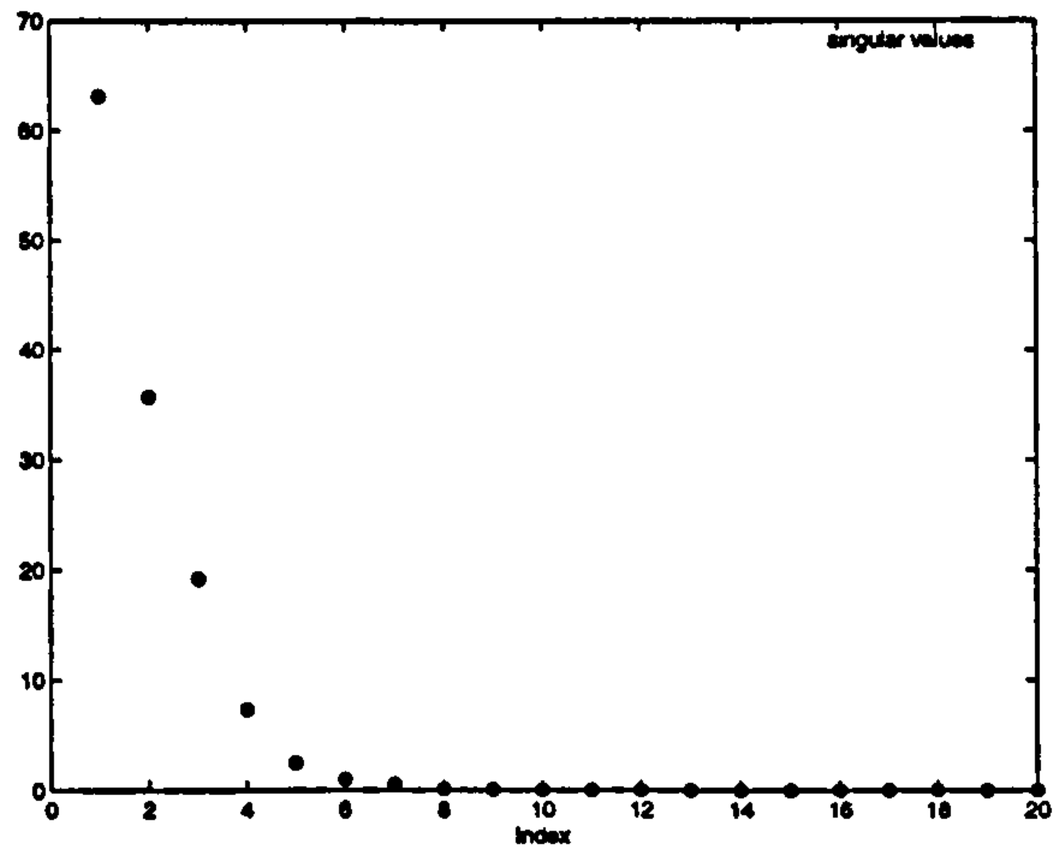


Fig 5.1 The singular values of the time history matrix of the modelled piano signal with centre frequency of 10Hz.

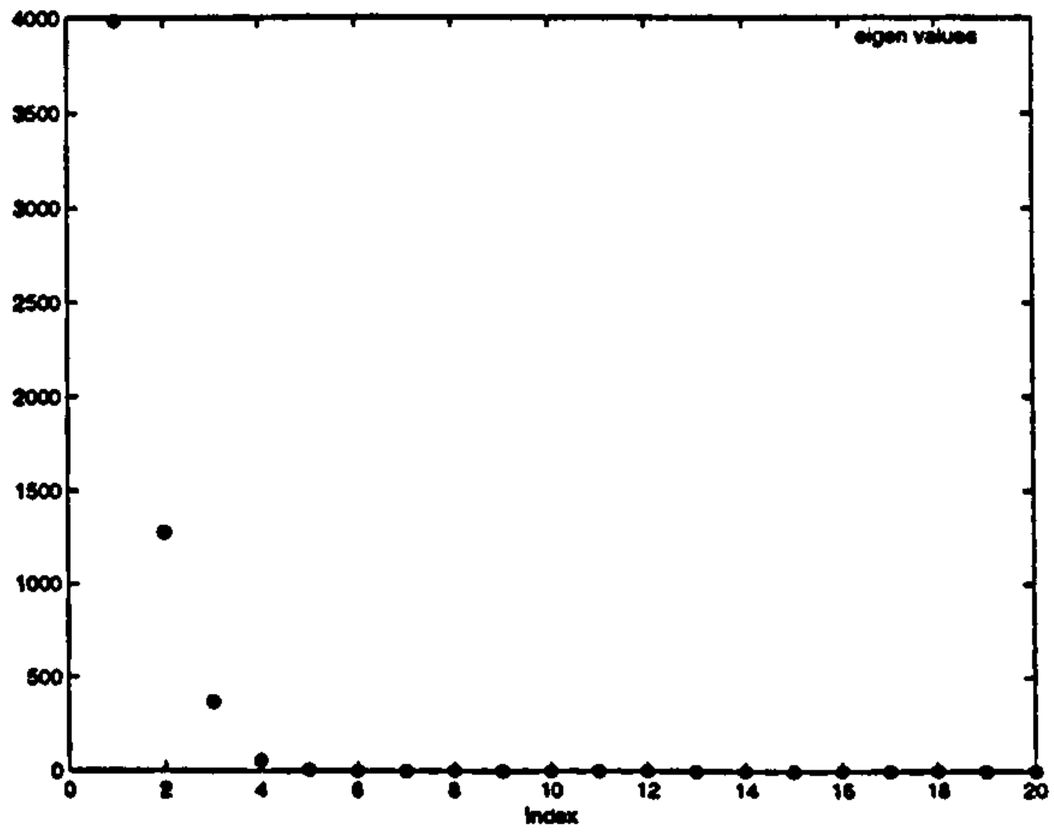


Fig 5.2 The eigen values of the nonstationary autocorrelation matrix of the modelled piano signal with centre frequency of 10Hz.

The left singular vectors of the time history matrix and the eigen vectors of the autocorrelation matrix (see figs 5.3 and 5.4) for the nonstationary case have similar shapes as was expected according to equation 5.7. But they are not exactly the same. It is possible to see that the shapes of some singular vectors have differences concerning the polarity of the vectors. The third singular vector of the SVD of our signal has symmetric values in relation to the third eigen vector. The values are the same but have opposite sign. The first vectors are the most important since they have most significant values associated with them.

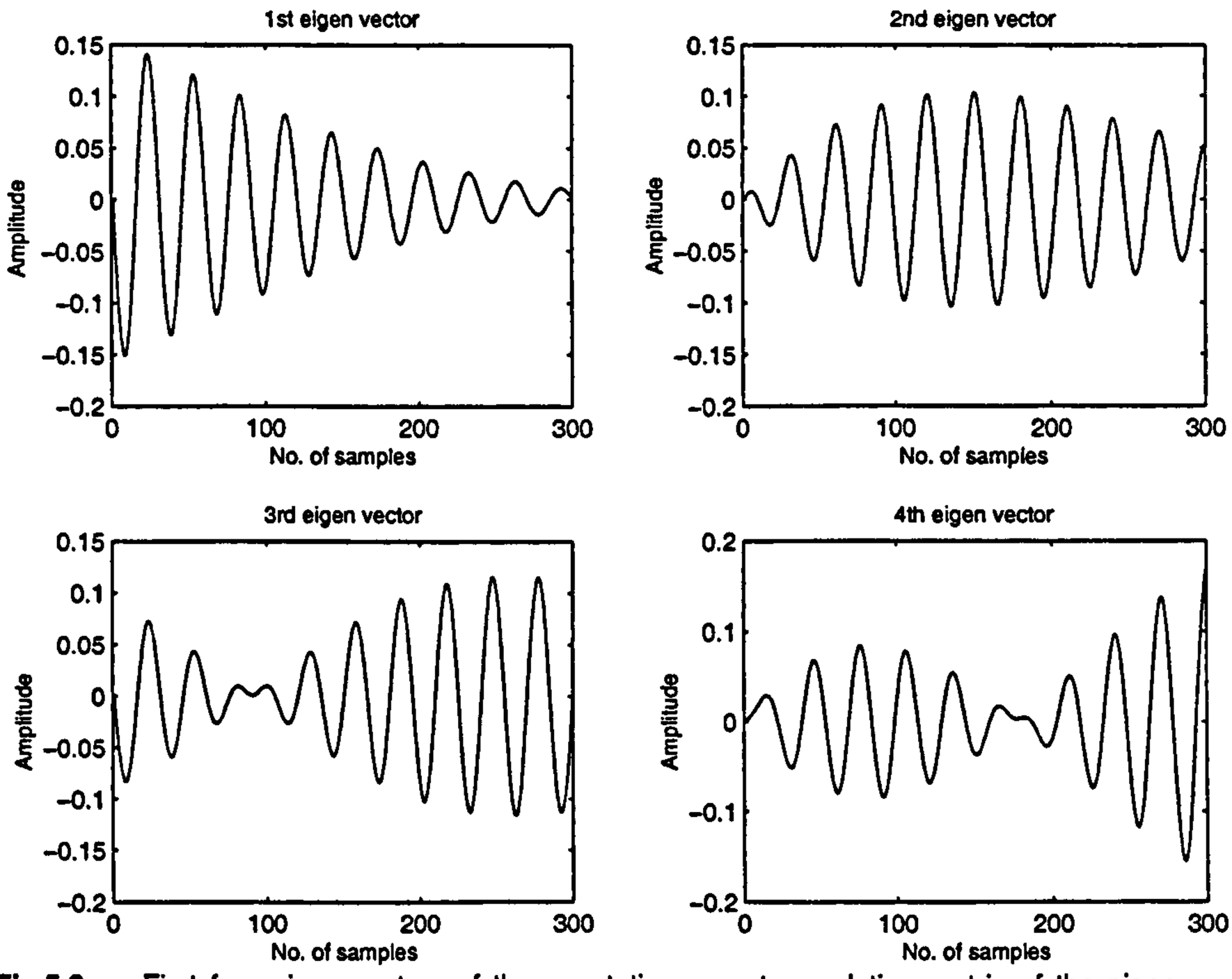


Fig 5.3 First four eigen vectors of the nonstationary autocorrelation matrix of the piano modelled signal with centre frequency of 10 Hz.

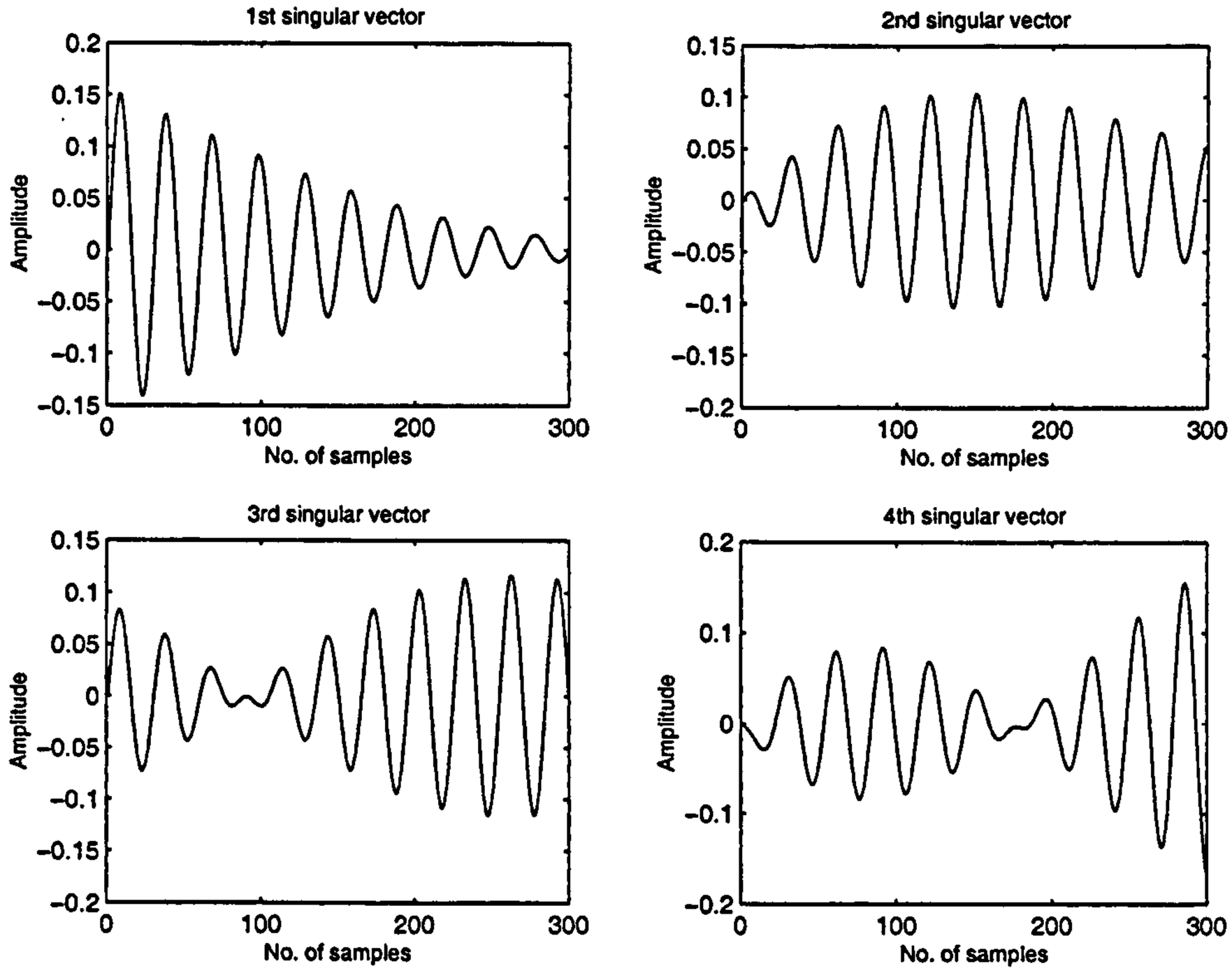


Fig 5.4 First four singular vectors of the time history matrix of the piano modelled signal with centre frequency of 10 Hz.

These equivalences continue until the singular values become very small. For example if we analyse the 23rd and 24th singular vectors (fig.5. 5) we see that there are no similarities (the corresponding singular and eigen values are very small, in the order of 10^{-15}). The 23rd singular vector is still a modulated sine wave while the 23rd eigen vector represents a noisy signal. The 24th singular vector is no longer a modulated sine wave.

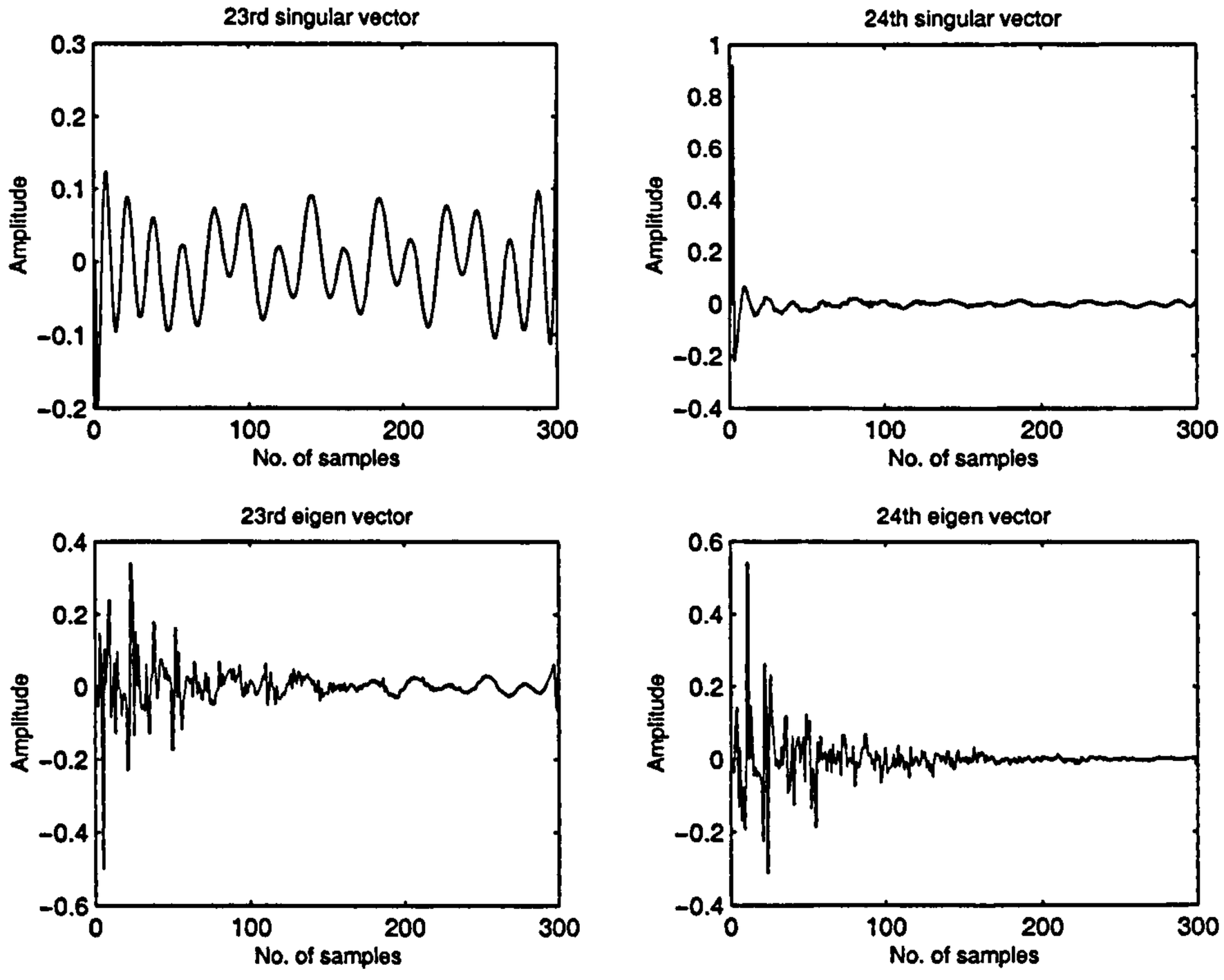


Fig 5.5 The 23rd and 24th singular vectors and eigen vectors of the time history matrix of the piano modelled signal with centre frequency of 10 Hz.

We saw (theoretically) that the singular vectors of the time history and the eigen vectors of the autocorrelation of a nonstationary process are similar as stated in equation 5.7 (and 5.8), since the matrix of the vectores is the same for both cases.

These practical results of the KLE and SVD of a nonstationary signal expressed by equation 3.1, confirm that for the nonstationary case equations (5.7) and (5.8) are verified. More precisely for the most significant values the corresponding vectors are similar. However there are singular vectors with opposite signs to the similar eigen vectors although the absolute values are the same (figs. 5.3 and 5.4). After the 20th vectors the singular and eigen vectors are not similar. The values associated with these vectors are very small. The results of the KLE and SVD of a nonstationary signal depend on the algorithms used to determine the eigen values and vectors and the singular values and vectors. The use of the squared value in KLE (due to the correlation) of the data leads to less precision in the determination of the eigen values.

The differences between the results of the KLE and SVD for the same signals are not significant for our research. Our purpose is to recognise piano notes in a musical

signal projecting that signal in a space of a note defined by the singular vectors. We will use the norm form measuring the presence of notes in a signal and this will not be significantly affected by the differences described.

5.5 Relation between the SVD, KLE and the Discrete Fourier transform

The discrete Fourier transform is related to the KLE and SVD. Both KLE and SVD are mathematical methods where a matrix can be decomposed as the product of three matrices. We saw that the matrices for the case of KLE have to obey certain properties. For the case of the SVD any matrix can be decomposed.

There is a special class of matrices, which decomposition corresponds to the discrete Fourier transform. These matrices are the circulant matrices and are a special case of Toeplitz matrices. A circulant matrix, C, is defined only by the first row (or column). All the other elements are obtained by shifting the around the first row (or column).

$$C = \begin{bmatrix} c_0 & c_1 & \cdots & c_n \\ c_n & c_0 & & c_{n-1} \\ \vdots & & \ddots & \\ c_{n-1} & \cdots & & c_0 \end{bmatrix} \tag{5.10}$$

The circulant matrix can be decomposed as the product (Sjostrom,1996):

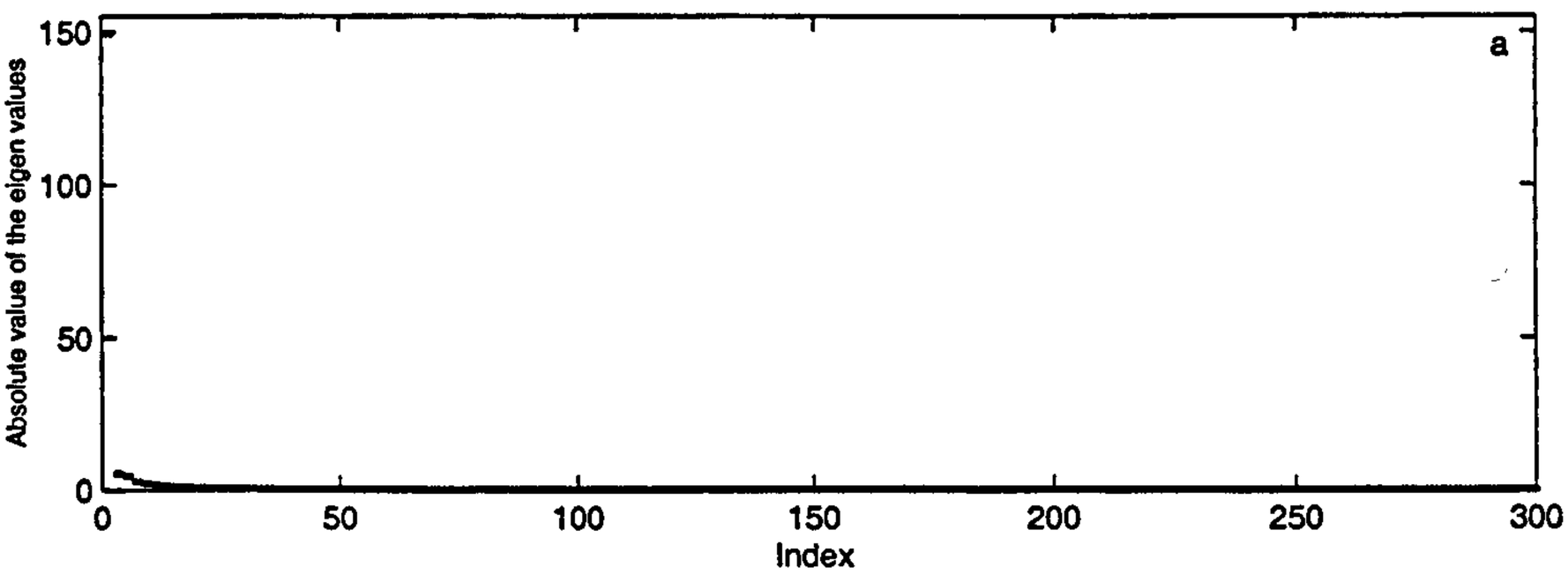
$$C = F\Sigma F^T \tag{5.11}$$

This corresponds to the eigen decomposition of the matrix, where F is the unitary Fourier transform matrix,

$$F = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & W^1 & \cdots & \cdots & W^{(n-1)} \\ \vdots & \vdots & W^4 & \cdots & W^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W^{(n-1)} & W^{2(n-1)} & \cdots & W^{(n-1)(n-1)} \end{bmatrix} \tag{5.12}$$

with $W = e^{-2\pi i/n}$, (i.e. correspond to the eigenvectors of the C matrix), F^H its transpose of the conjugate and Σ is a diagonal matrix (correspond to the eigen values of C) which corresponds to the squared magnitude of the discrete Fourier transform of the first row (or column) of C. Any matrix that can be expressed in the form presented in equation (5.11) is circulant (M. Gray, 2000) and all of the matrices have the same set of eigen vectors. Any signal represented by a circulant matrix has always the same set of eigen vectors corresponding to the discrete Fourier transform basis set.

If we build a time history of a random signal and apply the eigen decomposition (as in equation 5.10) the resulting matrices, F will correspond to the unitary Fourier transform matrix and Σ is a diagonal matrix equivalent to the discrete Fourier transform of the first line of the matrix (corresponding to the signal). We can see the results in a practical example. A single sine wave with frequency of 10 Hz, with 300 samples and sampled at 300 Hz was used to built a circulant matrix. In fig.5.6 we have the absolute values of the eigen values of the circulant matrix. The values of the discrete Fourier transform and eigen values of the circulant matrix are similar. If we sort the discrete Fourier transform values (fig 5.7) by decreasing order the graphs will be identical, demonstrating in practice that the eigen values of a circulant matrix are similar to the DFT coefficients (small differences in the values are due to the algorithms). The eigen vectors correspond to the columns of the unitary Fourier transform matrix. We note that now the circulant matrix eigen values and vectors are complex.



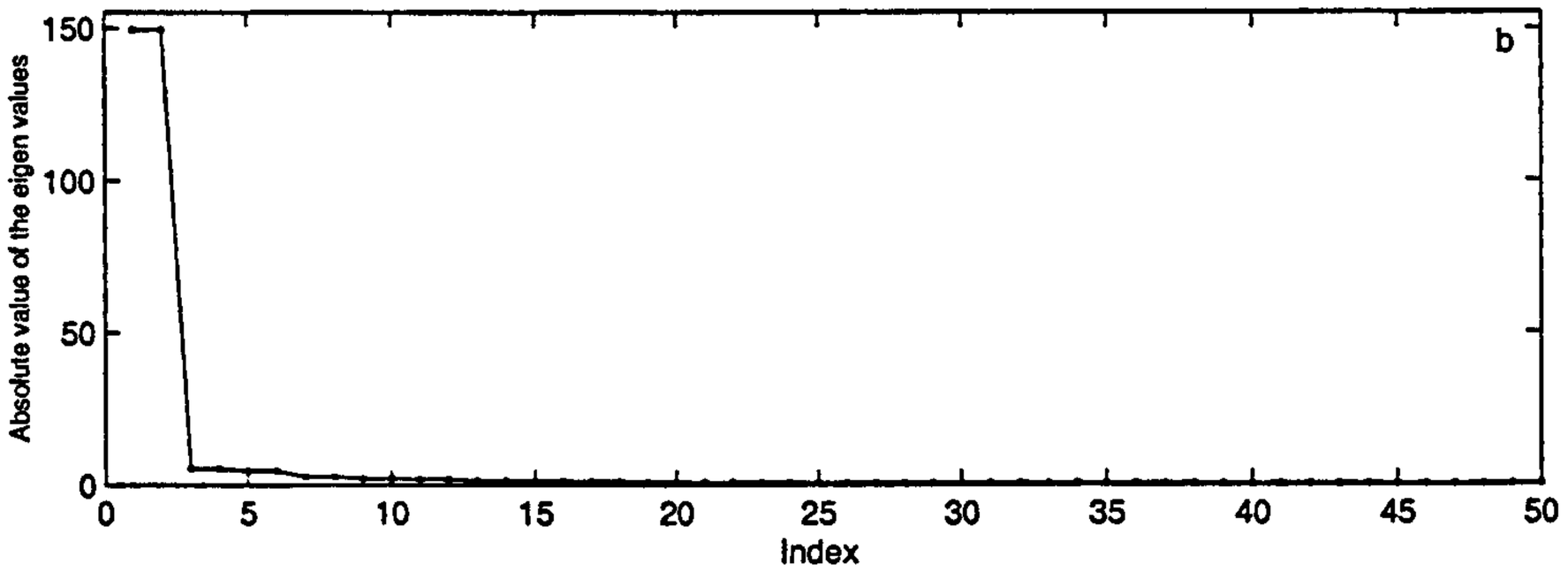
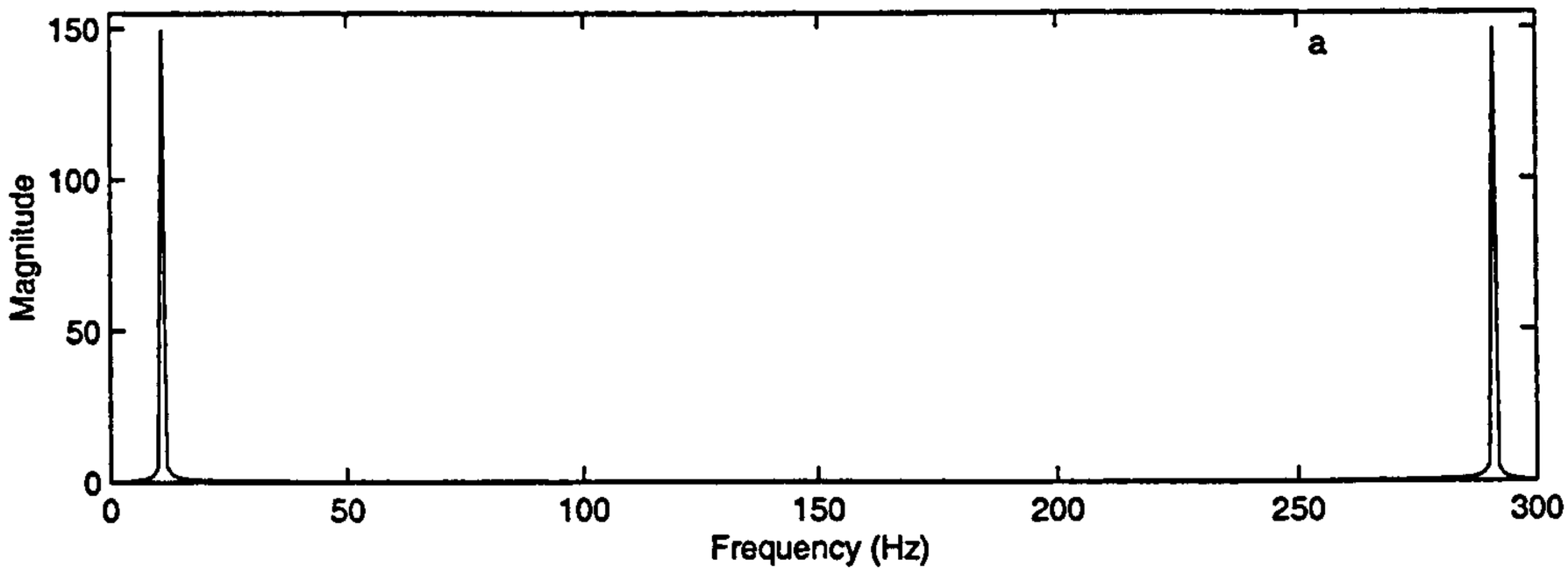


Fig 5.6 Absolute value of the eigen values of a single sine wave signal (10Hz) (a). A detail of the first 50 absolute values (b).



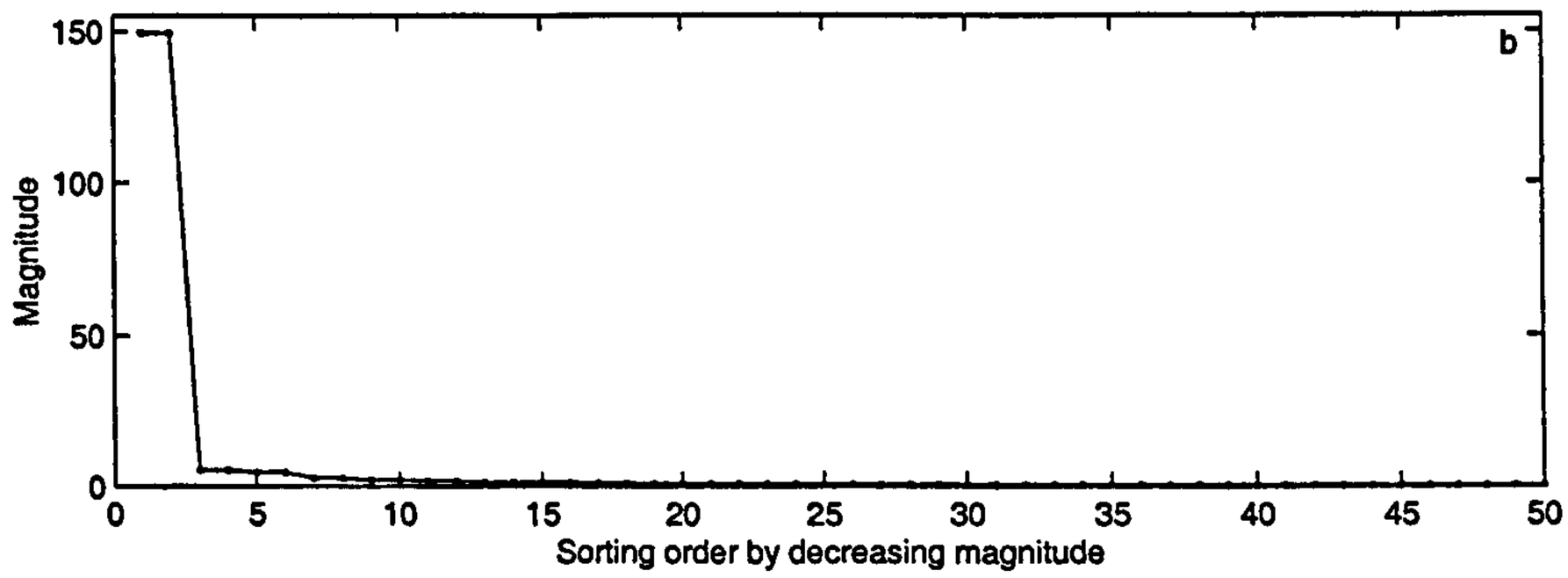


Fig 5.7 Discrete Fourier transform of a sine wave 10Hz (a). The coefficients of the transform sorted by descending order of magnitude (first 50 ordered coefficients)(b).

If we apply the SVD directly over a time history circulant matrix, the resulting singular values are identical to the absolute values of the eigen values and to the DFT. The singular values are real as it should be expected by the theorem of the SVD. There are small differences in the values (e.g. the two first singular values are not exactly the same) that are due to the SVD algorithm. The singular vectors are also real.

The discrete Fourier transform of a signal can be considered as a special case of the eigen decomposition of a square matrix or even more generally a special case of the singular value decomposition. The condition imposed to have the Fourier transform is to build a circulant time history matrix.

5.6 Conclusion

The three transforms presented here are all related. The Fourier transform is a special case of the KLE. While in the KLE the vectors of the expansion are selected by the signal itself in the Fourier case the vectors of the base are always determined to be in the shape of sine and cosine. This selected base of the Fourier transform is achieved in the KLE when the autocorrelation matrix of the signal is circulant. The decomposition of circulant matrices has always the Fourier transform matrix as the eigen vector matrix.

The SVD is an equivalent technique to KLE as expressed in equations 5.7 and 5.8. In practical applications the SVD and KLE of nonstationary stochastic signals have the same resulting basis vectors (the right singular vectors are similar to the eigen vectors. In the examples shown we verified that similarities and there are only some differences in the orientation of the vectors that for some cases can have opposite signals although they have the same values. In our research we are interested in the vectors and we will use them to project our signals. The differences pointed are not significant for our work. The application of the SVD is much simpler, since we will be working with the time history matrix without the explicit need to create second order statistics.

CHAPTER 6

THE APPLICATION OF SVD TO SIMULATED PIANO SIGNALS

Introduction

The aim of this chapter is concerned with the application of SVD to the simulated nonstationary signal expressed by equation 3.1. We will start by analysing the SVD of a simple sine wave and then investigate how the variation of the parameters in the expression 3.1 affects the final results of the SVD. Later in the chapter we consider another type of nonstationary signal namely a uniform modulated process. The aim of this work is to get a good understanding of the use of SVD for the classes of nonstationary signals.

6.1 The representation of a nonstationary sine wave in singular value space

In the previous chapter we saw how the KLE, SVD and the DFT are related. The KLE and SVD are similar for nonstationary signals where the time histories are represented by a number of possible realisations and the autocorrelation is estimated by ensemble averaging. When the autocorrelation matrix or time history matrices are

circulant (and the matrices are square $(m=n)$) the results of the KLE and SVD are similar to the DFT. We have this last similarity between transforms in the case of stationary signals. The structure of the circulant matrix implies that the original signal has to be stationary. For nonstationary signals defined by an ensemble of possible realisations of the signal $(m=n)$ the time histories and autocorrelation matrices are never circulant. Accordingly the decomposition is not similar to the DFT.

The systems that we shall analyse will be underdetermined $(m<n)$. We choose to have an underdetermined system because for the case of the real piano notes we will have a smaller number of events than the number of samples of each realisation. All these conditions guarantee that we will not be using the SVD as a special case of the DFT.

For simplicity we will start to decompose a single sine wave, in which is considered as a nonstationary process. We clarify this notion as follows, suppose that we have an ensemble of sine waves representing a nonstationary random process and each possible realisation is expressed has:

$$x_i(t)=\sin(2\pi f_i t) \tag{6.1}$$

Each possible realisation has a frequency value f_i drawn from a uniform distribution interval $[a,b]\in\mathbf{R}$. We can easily demonstrate that this is a nonstationary process by considering the first and second moments.

The mean value of the process expressed by the ensemble is given by:

$$E[x_i(t)]=\int_a^b x_i(t)p(x_i)dx_i \tag{6.2}$$

where $p(x_i)$ is the probability density function of x_i. For the signal represented by equation 6.1 the mean value is:

$$E\left[x_i(t)\right]=\frac{1}{b-a}\int_a^b \sin\left(2\pi f_i t\right)df_i \tag{6.3}$$

$$=\frac{1}{2\pi(b-a)t}\left[\cos(2\pi at)-\cos(2\pi bt)\right]$$

$$=\frac{1}{\pi(b-a)t}\left[\sin\left((b+a)\pi t\right)\sin\left((b-a)\pi t\right)\right]$$

$$=\sin\left((b+a)\pi t\right)\frac{\sin\left((b-a)\pi t\right)}{(b-a)\pi t} \tag{6.4}$$

If we replace a by $f_c-B/2$ and b by $f_c+B/2$, where f_c is the centre frequency and B the bandwidth of the interval where f_i takes the values then,

$$=\sin\left(2\pi f_c t\right)\frac{\sin\left(B\pi t\right)}{B\pi t} \tag{6.5}$$

The mean value of the nonstationary sine wave is time dependent and the result is a modulated sine wave whose carrier frequency is the centre frequency of the interval and the modulation frequency as the bandwidth value.

The covariance function for a random process is defined as:

$$R(t_1,t_2)=E\left[x_i(t_1)x_i(t_2)\right]=\int x_i(t_1)x_i(t_2)p(x_i)dx_i \tag{6.6}$$

If we replace t_1 by t and t_2 by $t+\tau$ then:

$$R(t_1,t_2)=R(t,t+\tau)=\int_a^b x_i(t)x_i(t+\tau)p(x_i)dx_i \tag{6.7}$$

For stationary processes the correlation depends on τ. If our particular process is nonstationary the autocorrelation should depend on t and τ.

$$R(t,t+\tau)=\frac{1}{b-a}\int_a^b \sin(2\pi f_i t)\sin\left(2\pi f_i(t+\tau)\right)df_i \tag{6.8}$$

$$=\frac{1}{2\pi(b-a)}\left[\frac{\cos\left((a+b)\pi\tau\right)\sin\left((a-b)\pi\tau\right)}{\tau}-\frac{\sin\left[(a+b)\pi(t+\tau)\right]\cos\left[(a-b)\pi(t+\tau)\right]}{t+\tau}\right]$$

(6.9)

$$R(t,t+\tau)=\frac{1}{2\pi(b-a)}\left[R_1(\tau)-R_2(t,t+\tau)\right] \tag{6.10}$$

The autocorrelation function does depend on t and τ confirming that the sine wave defined as in equation 6.1 is a nonstationary process.

We can consider the sine wave as a special case of the simulated signal where the amplitude has a fixed value for all the possible realisations and no damping. Only the frequency takes different values. The time history of the signal was generated using a random set of sine waves. The frequency values were taken from an interval with uniform distribution, with centre frequency of 10Hz, with the size of 0.5Hz. The amplitude of the sine waves was the same for all the events. The time history matrix (row matrix) was composed of 100 events with the length of 300 points (less events than the number of samples in each event, i.e. a $m<n$ matrix) which corresponds to an underdetermined system since the number of events is smaller than the number of samples in each event.

SVD was applied directly to this time history matrix. The results are presented in figure 6.1 and 6.2. The singular values show that the energy is basically concentrated in the first two vectors. These two vectors practically define the signal in the new space defined by the singular vectors basis set with a small error. There are three more singular values different from zero. The sine wave treated as a nonstationary case for the intervals of frequency chosen has 5 singular values different from zero. From the 6^{th} to the 100^{th} values their values are effectively zero. The number of linear independent rows of the system is 5.

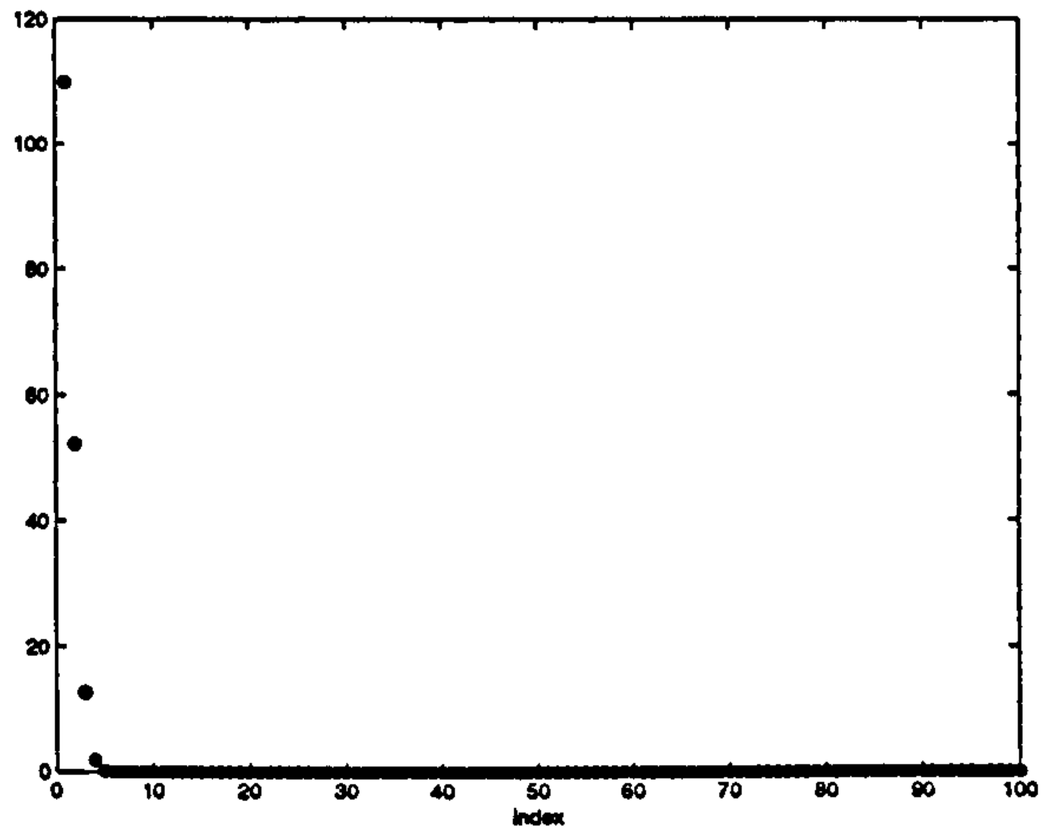


Fig 6.1 The singular values of the time history of a sine wave in the nonstationary case

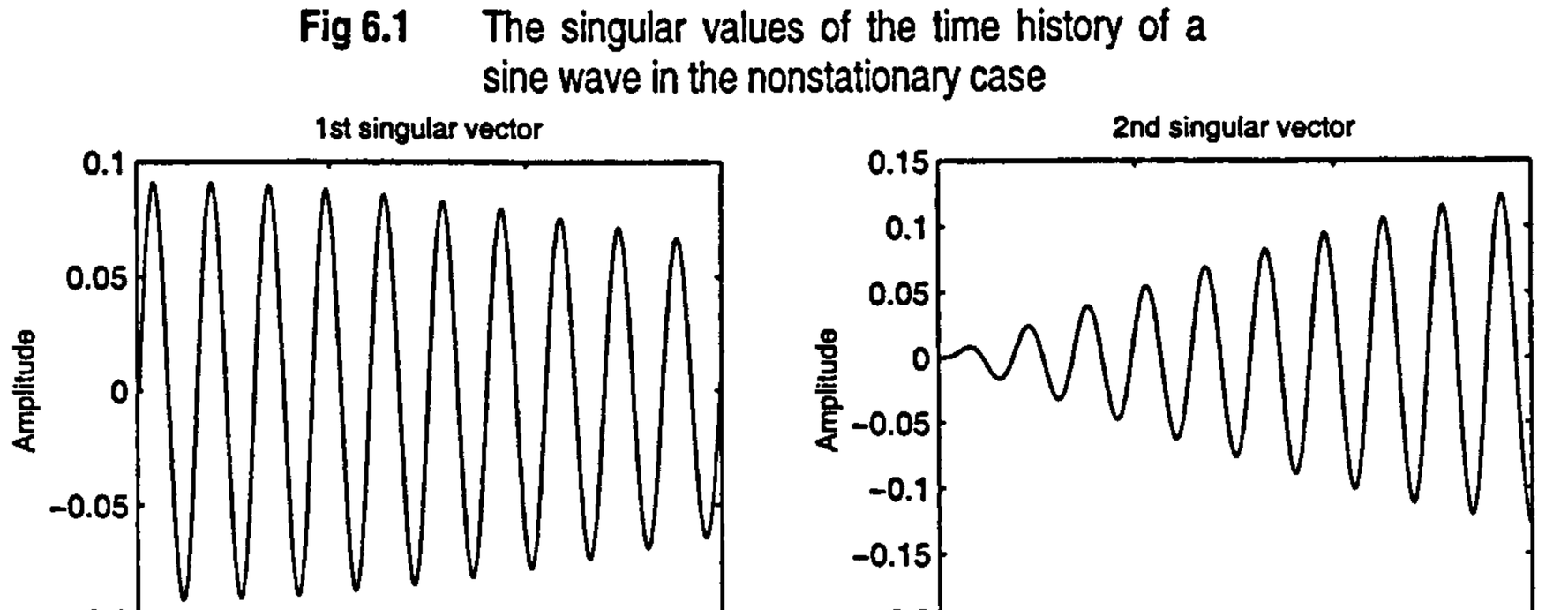


Fig 6.2 First four singular vectors of a time history matrix composed of sine waves with frequencies draw from a interval with centre frequency of 10Hz.

The four singular vectors of the sine wave have the shape of modulated sine waves (we present the first four singular vectors since their corresponding values are the most significant). The envelope of the first vector decreases. It is possible to relate the shape of this vector to the time history matrix. The possible realisations of the signal all start at the same point with the same value namely zero. This corresponds to the first column of the data matrix with the same zero value for each element. The first sample of the first singular vector has the value zero. The various events in the ensemble have different frequencies, which means that each signal has its last point with different amplitude from the other events. There is less correlation between the events for the last samples. The greater concentration of points, with the same value, is situated at the first column of the matrix while the smaller one is at the last column. Due to the difference in the frequency values of each realisation the number of samples in each column with the same value becomes smaller. This is the reason why the shape of the first vector presents a modulation with smaller amplitudes at the last samples of the vector and leads us to think about an average of the events. If we take the sum of the elements of each column of the time history, corresponding to the average of the possible realisations of the signals, we can see that the final shape of the resulting sum is very similar to the first singular vector. There is only a small difference in the latter samples (fig.6.3). This shows that the shape of the first vector is related to the average of the realisations of the signal. This is confirmed by considering the analytic result for the mean (eq.(6.5)).

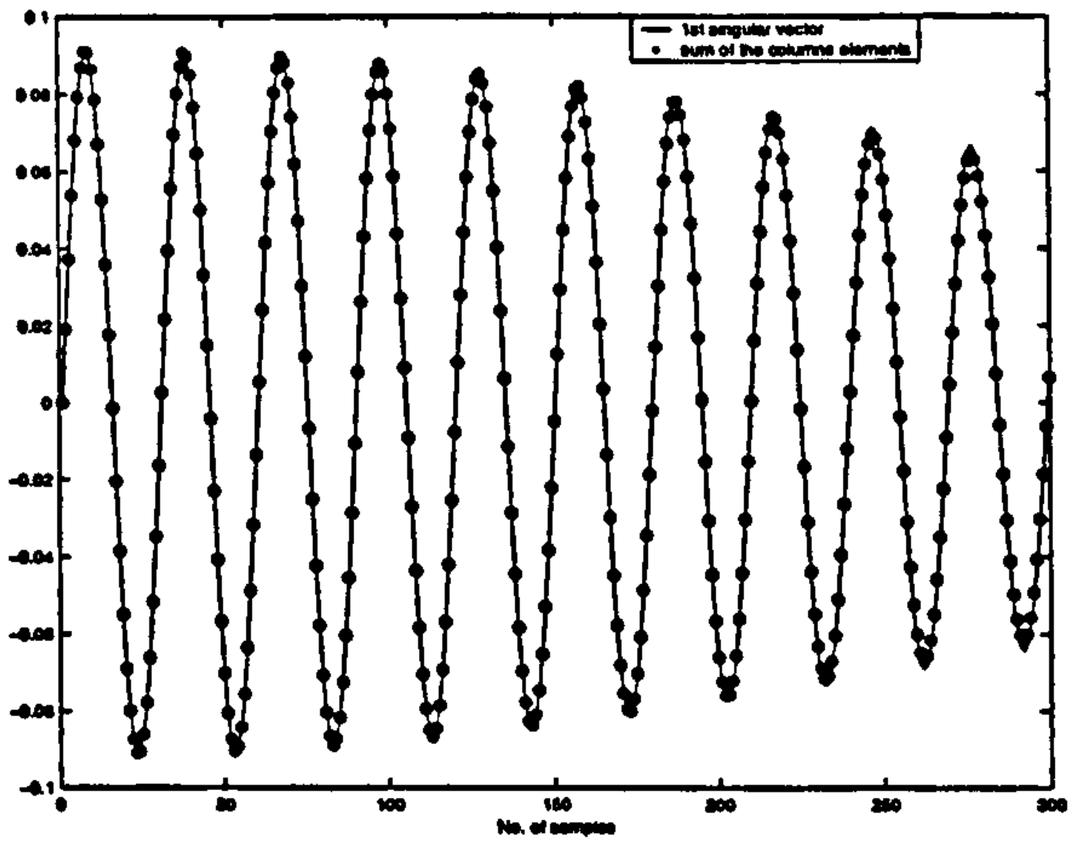


Fig 6.3 First singular vector of the time history of the 10Hz centre frequency sine wave and the sum of the elements of each columns of the time history matrix.

The envelope of the second vector has the opposite behaviour to the first. The maximum values of amplitude in each period increase from the first to the last samples. If we think in terms of signal reconstruction the higher values of the second vector for the last samples compensate the smaller ones on the first vector. (The vectors are not in phase). The other two vectors, 3rd and 4th, have a significant modulation. To seek more information about the vectors we applied the Fourier transform to these first four singular vectors.

The Fourier transform of the two first vectors (fig.6.4) confirms that the vectors are dominated by only one, slowly modulated, frequency component at 10 Hz and 10.25Hz respectively, which is for the first case the centre frequency of the frequency interval with uniform distribution. The other two vectors, 3rd and 4th have two frequency components approximately at frequencies 9.25 and 10.75Hz. The envelope of these two vectors presents a modulation. The modulation of the 3rd vector corresponds to 0.5 Hz since we have practically half of a period. In the spectrum of this vector we have two components on either side of the 10Hz centre frequency of the interval. These two components are due to the amplitude modulation. We can consider this has a ring modulation where the carrier is a 10Hz sine wave and is modulated by a sine wave with frequency of 0.5 Hz.

The theoretical results of the mean value of the nonstationary sine wave (eq.(6.5)) show that it is a modulated sine wave. The practical results of our example confirm the theoretical ones.

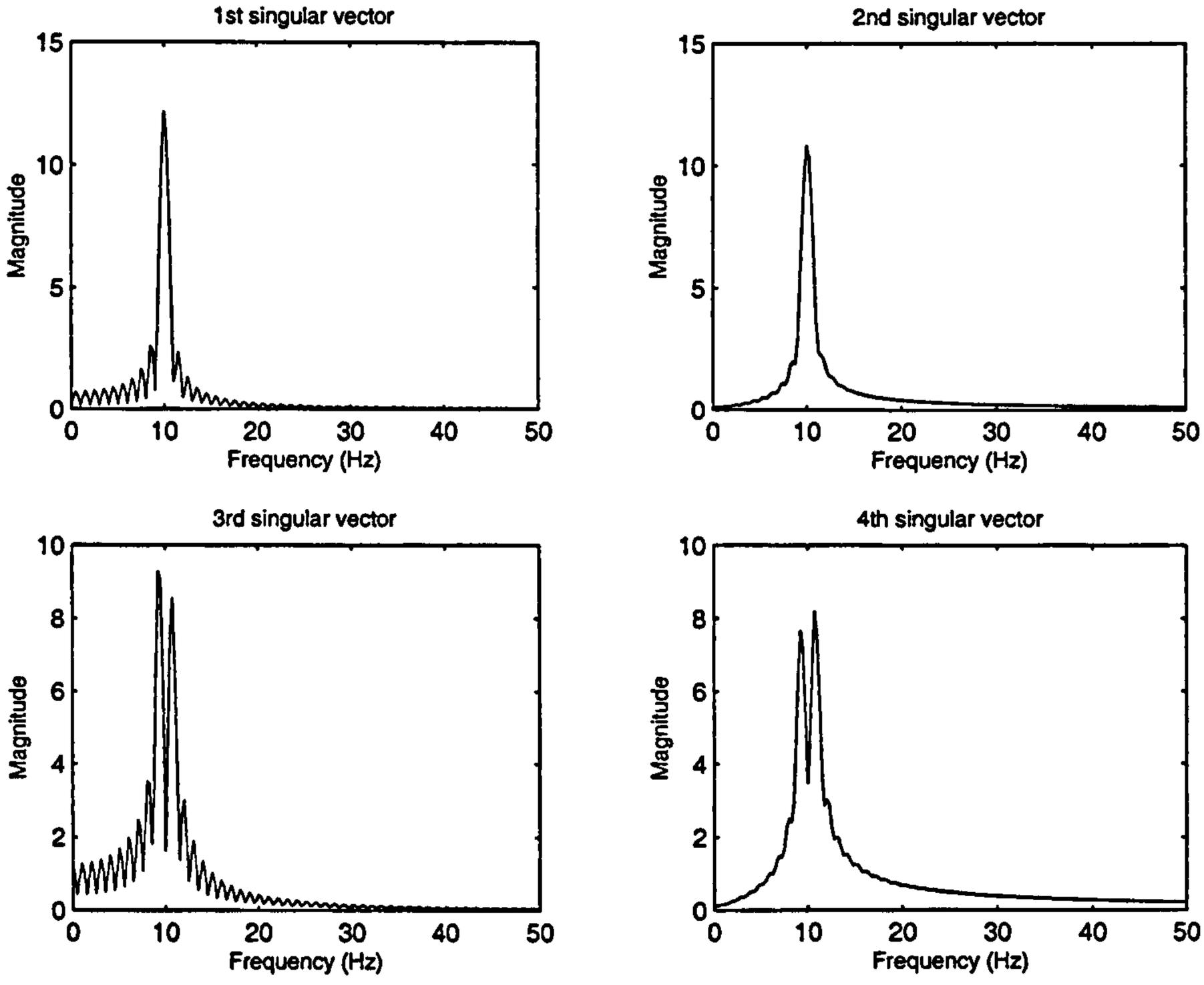


Fig 6.4 The Fourier transform of the first four singular vectors of the single sine wave (FT of 1200 points).

6.2 Factors that influence the shape of the singular vectors of a nonstationary sine wave

We will see how the shape of the vectors and their respective values, for the signal expressed by equation (6.1), change with some variation in parameters. All the previous settings of the previous time history matrices, as to the number of events and the number of samples of each event of the signal, remain exactly the same.

6.2.1 The initial phase

If we introduce a random initial phase to the nonstationary sine wave the greater concentration of points with the same value will not occur at the beginning of the signal. We can say that the initial part of the signal will be as variable as the latter parts. The signals at the 300^{th} sample finish at any point of the period of the sine wave due to the slight variation of the frequency.

For this experiment we introduced a phase value to each event that varied randomly inside the interval $[0,2\pi]$ with a uniform distribution, so that the starting point could be at any point in a period of the sine wave. After applying SVD we can see in figure 6.6, that for the first singular vector the maximum amplitude is shifted to the middle points of the vector. This indicates where there is the most concentration of correlated points with the highest values between the events, as we saw for the single sine wave with the events starting at the same time with the same value.

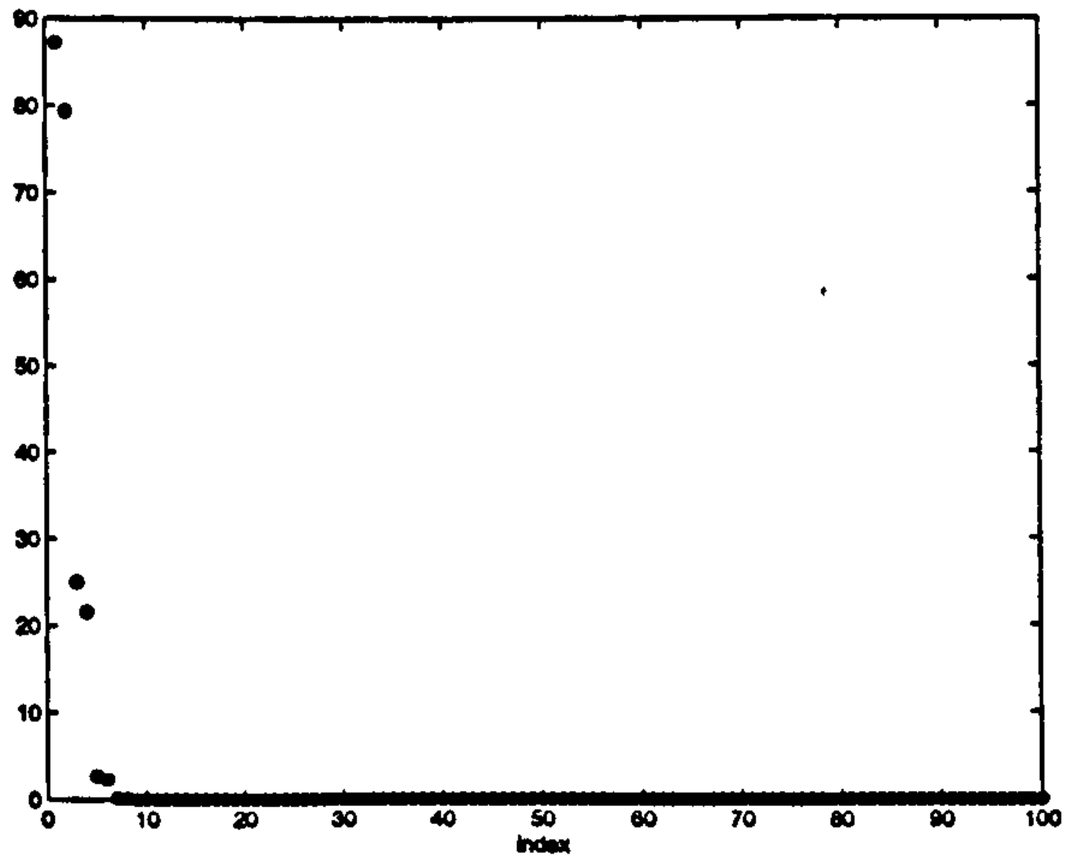


Fig 6.5 The singular values of the time history of a sine wave with random initial phase.

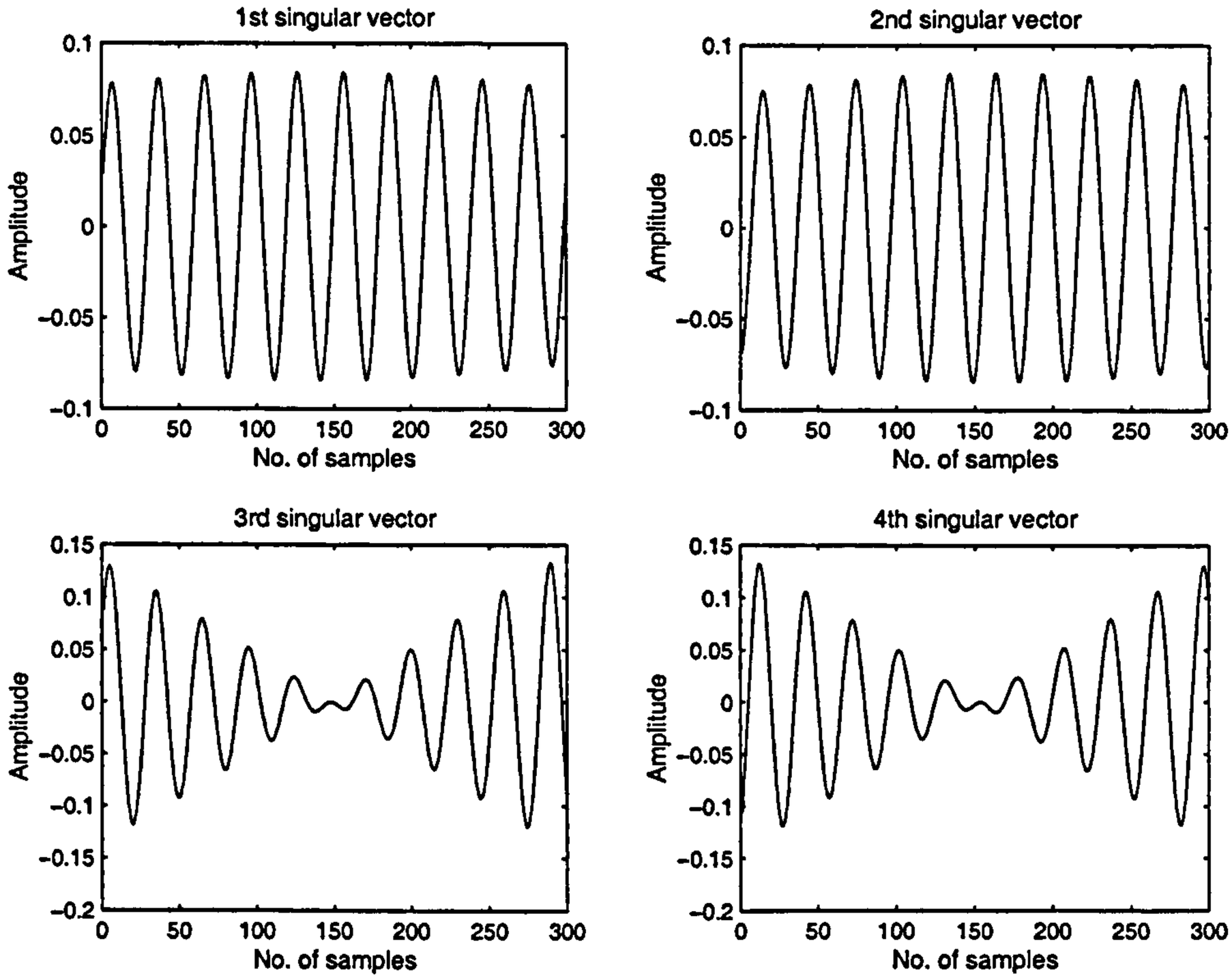


Fig 6.6 First four singular vectors of a sine wave with random initial phase.

In figure 6.5 we can see that number of significant singular values has increased meaning that the introduction of the phase generated more variability on the system. The second singular value has a value close to the first one. The shape of the corresponding vector (fig.6.6) becomes more similar to the first one. The 3rd and 4th values are coupled and the shapes of the corresponding vectors are very similar. This means that the symmetry of the system, for our sine wave, increases with the introduction of the phase.

We can conclude that the phase in the case of single sine waves influences the shape of the singular vectors and we can say that the shape of the vectors together with the singular values can give information where there is more correlation between samples and if the system is more or less symmetric. (When we have a completely symmetric system represented by a square circulant matrix, as we have seen in chapter 5, we have the Fourier transform).

6.2.2 The frequency

In the next case the phase is fixed and the variation of the frequency will be studied. In the previous cases the frequency was randomly taken from an interval with uniform distribution. We will see how the change of the size of the interval will affect the SVD.

For this example we take a greater interval of 1Hz with the same central frequency of 10 HZ. The results of the SVD are in figures 6.7 and 6.8. Comparing with the results when the interval is 0.5 Hz (see figs.6.1 and 6.2), the first value is not so detached from the other values. The second singular value is now much nearer to the first value. There are now six singular values different from zero. This is one more than in the frequency interval of 0.5Hz. The greater size of the interval increases the variability of the system expressed in the SVD by the increase of the number of significant singular values. This result should be expected since the same number of events (100) can take frequency values in a wider interval and are more spread.

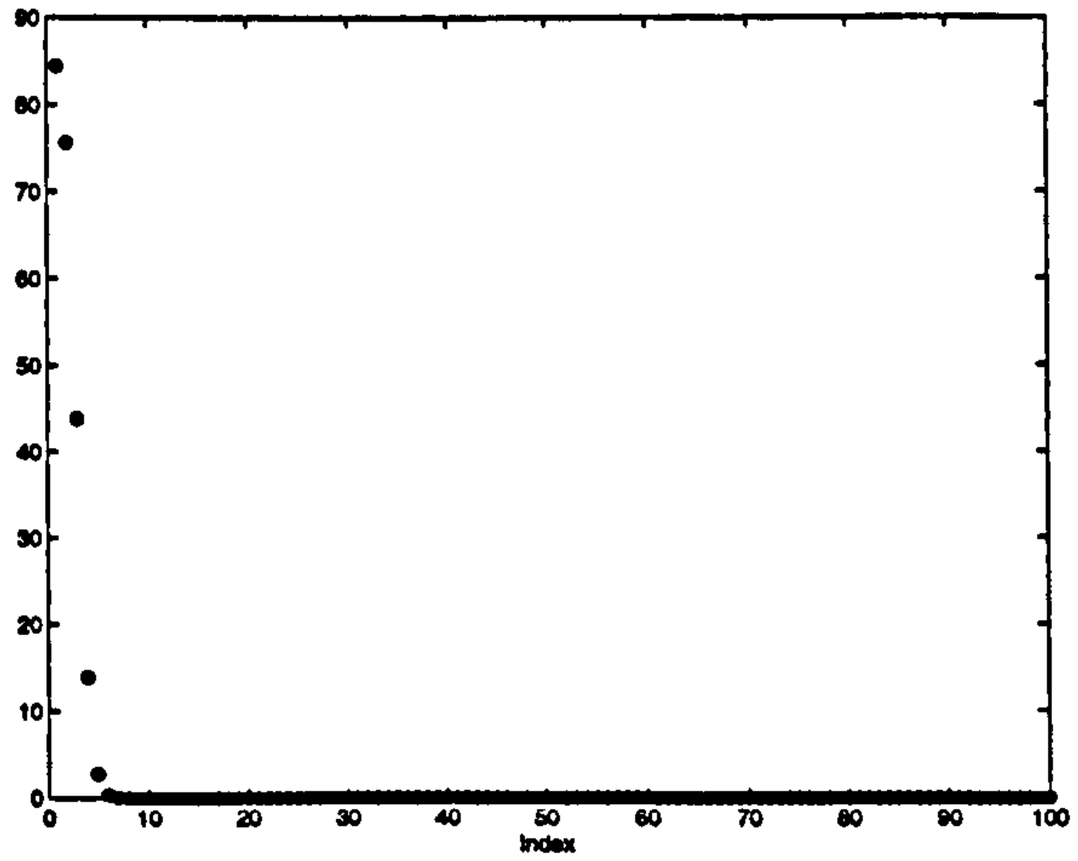


Fig 6.7 The singular values of the time history of a sine wave with frequency interval of 1Hz and centre frequency of 10Hz.

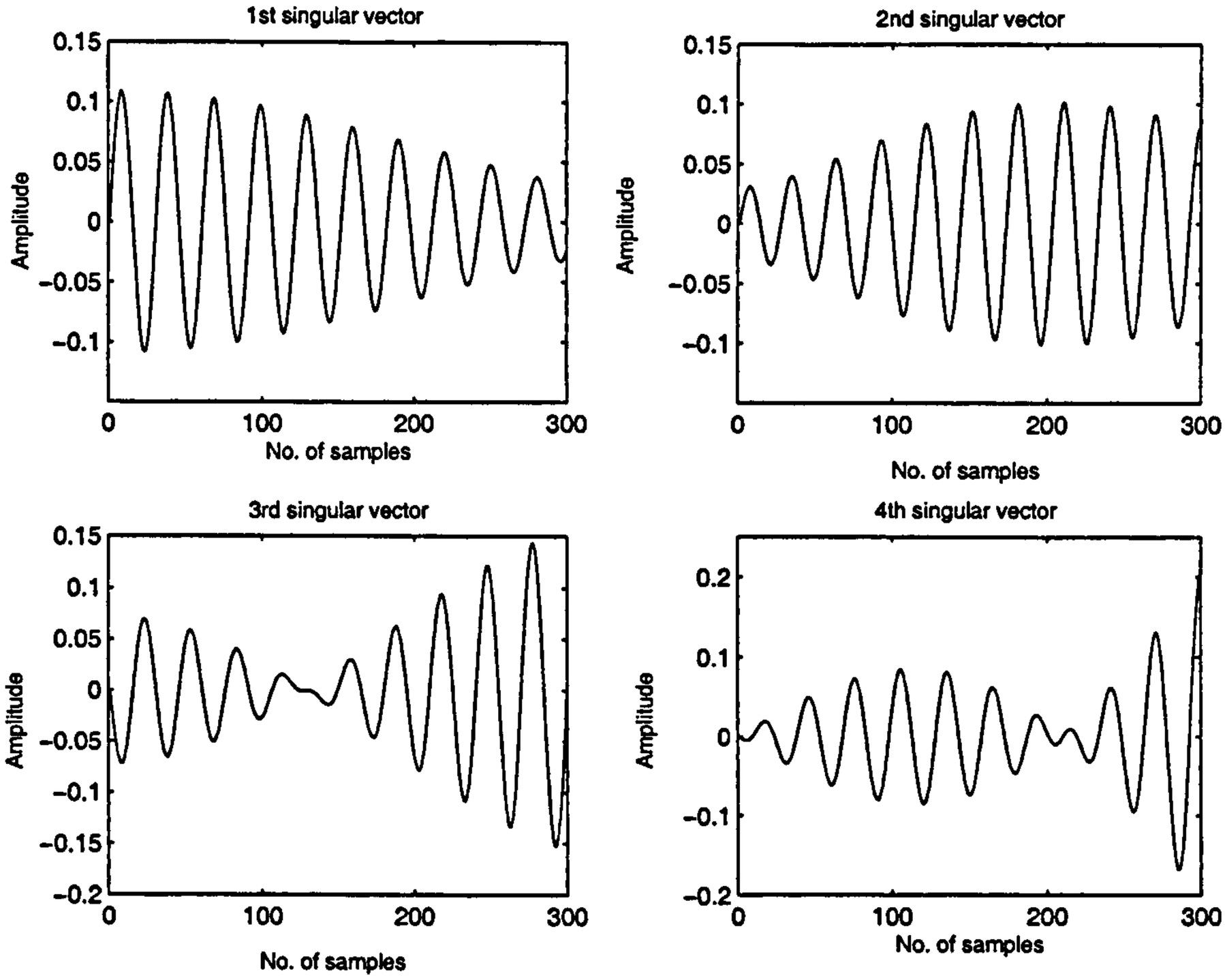


Fig 6.8 Singular vectors of a sine wave with frequency variation interval of 1Hz and centre frequency of 10Hz.

The first vector has now smaller amplitude at the later samples than for the 0.5Hz frequency interval. This is because, as we explained before for the case of the initial phase, the events start all at the same point and since they have small differences at the frequency values for the end of the events there is less coincidence of values.

If the size of the interval increases further, the first singular value will be smaller and the other values will be more significant. The vectors will exhibit more modulation depending where the events are more correlated and have greater amplitudes.

The Fourier transform of these vectors is quite similar to the 0.5Hz interval. The frequency components have the same values as before, 10Hz and 10.23Hz for the 1st and 2nd vectors and 9.25 and 10.75 for the two peaks of the 3rd and 4th vectors. Again we have a splitting in the frequency components of the 3rd and 4th vectors due to the modulation of the singular vectors. The modulation of the vectors is practically the same resulting in the similarity of the frequency values of the split components.

The doubling of the size of the frequency interval and maintaining the number of possible realisations mostly affects relation between the singular values and the number

of significant ones. The shapes of the vectors have only small variations and their frequency components are practically the same.

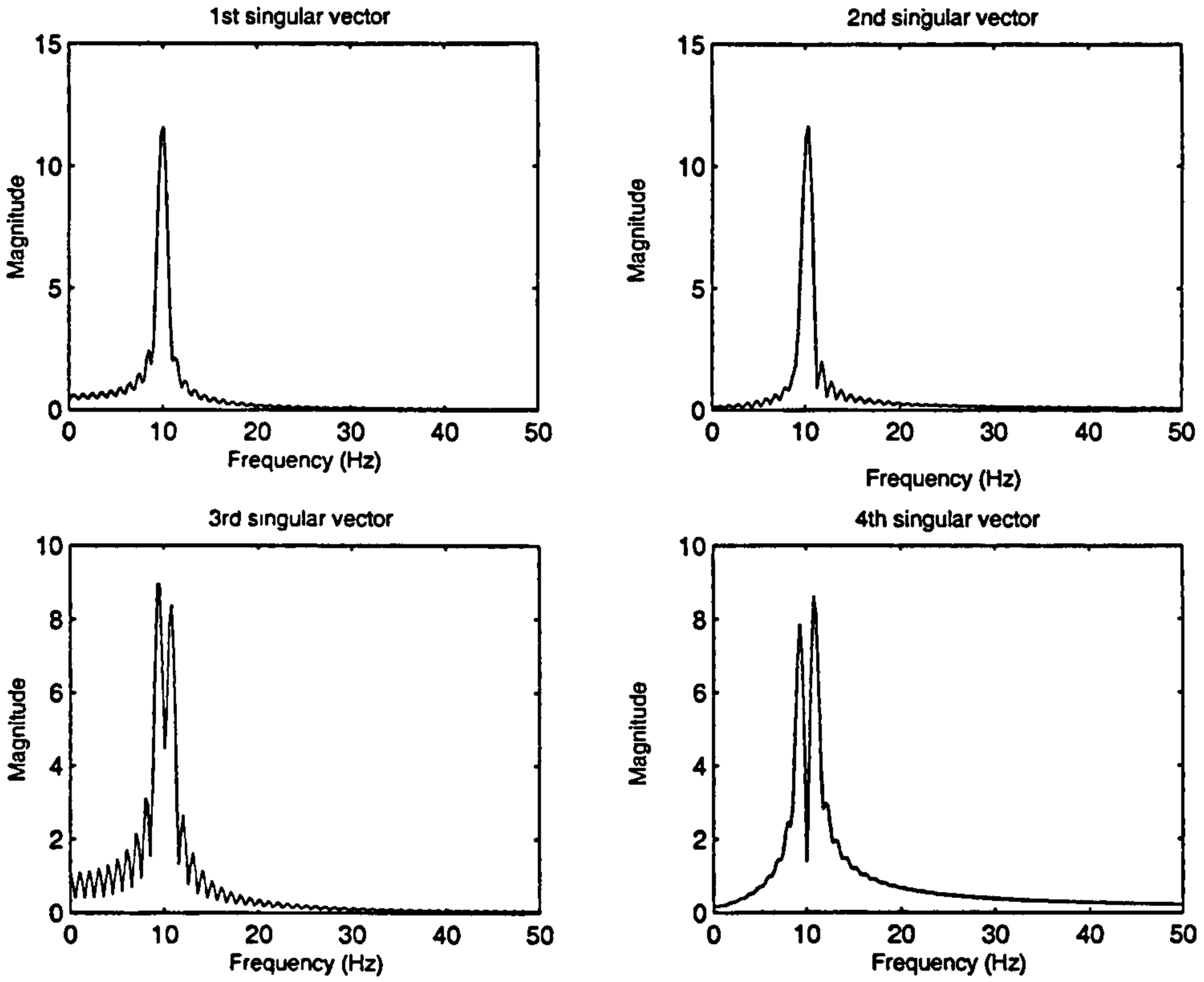


Fig 6.9 Fourier transform of the first four singular vectors of a sine wave with frequency variation interval of 1Hz and centre frequency of 10Hz (FT 1200 points).

6.3 The modelled piano signal and the influence the variables on the SVD

After considering the SVD of a single sine wave we now reconsider the modelled piano signal (expressed by eq.(3.1)). We will consider the same sine wave with 10Hz centre frequency. In the model there are four variables, amplitude A, damping α, frequency f and the initial phase ϕ all drawn randomly from an interval with uniform distribution. The phase will not be considered in this section since for real piano notes we always take the signal form the initial point of the strings excitation by the hammer. Instead the phase will be replaced by the delay between the realisations of the signal. We

will analyse the influence of the variation of these variables one at a time on the shape of the singular vector and singular values.

The SVD of a simulated signal described by equation 3.1 was already presented in previous chapters. In chapter 3 we modelled a piano as a stochastic transient with only one single frequency component. The time history matrix is a row matrix with dimensions (100×300), as in the previous examples of the sine wave. The amplitude A, damping α and frequency f where all drawn randomly from an interval with uniform distribution. All the events were aligned at the starting point of the signal. The SVD was applied to this time history and the resulting singular vectors and singular values were already presented in chapter 4 section 4.5, where we were comparing the application of KLE to nonstationary signals, considering the stationary and nonstationary autocorrelation.

As a reference we first present the SVD of a signal determined by equation (3.1) where all the parameters (amplitude, damping and frequency) vary in the intervals described in table 6.1.

	Frequency interval (Hz)	Centre frequency (Hz)	Amplitude interval	Damping interval
s	[9.71,10.29]	10	[0.75,1]	[0.1,0.3]

Table 6.1 Intervals of variation of the variables of signal *s* expressed by equation 3.1.

The time history of an event of s, the singular values and vectors are presented in the following graphs (figs.6.9, 6.10 and 6.11).

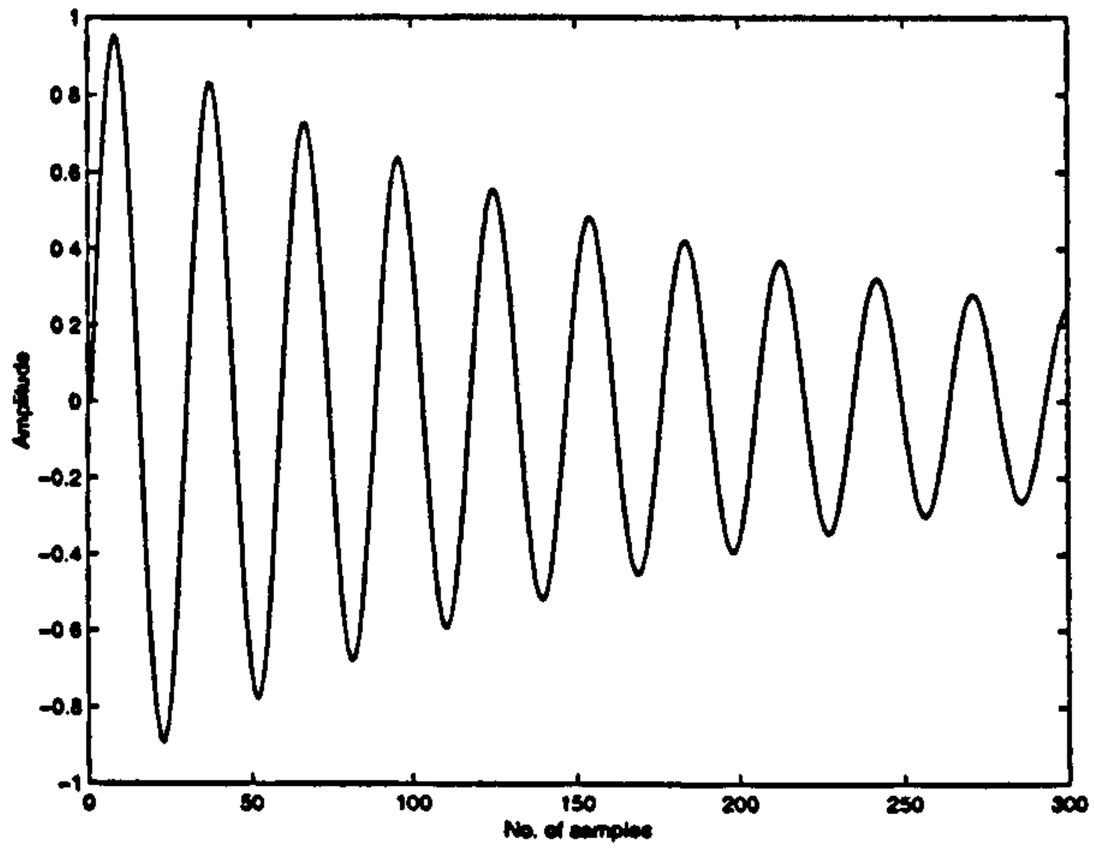


Fig 6.10 The singular values of the time history of an event of signal *s*.

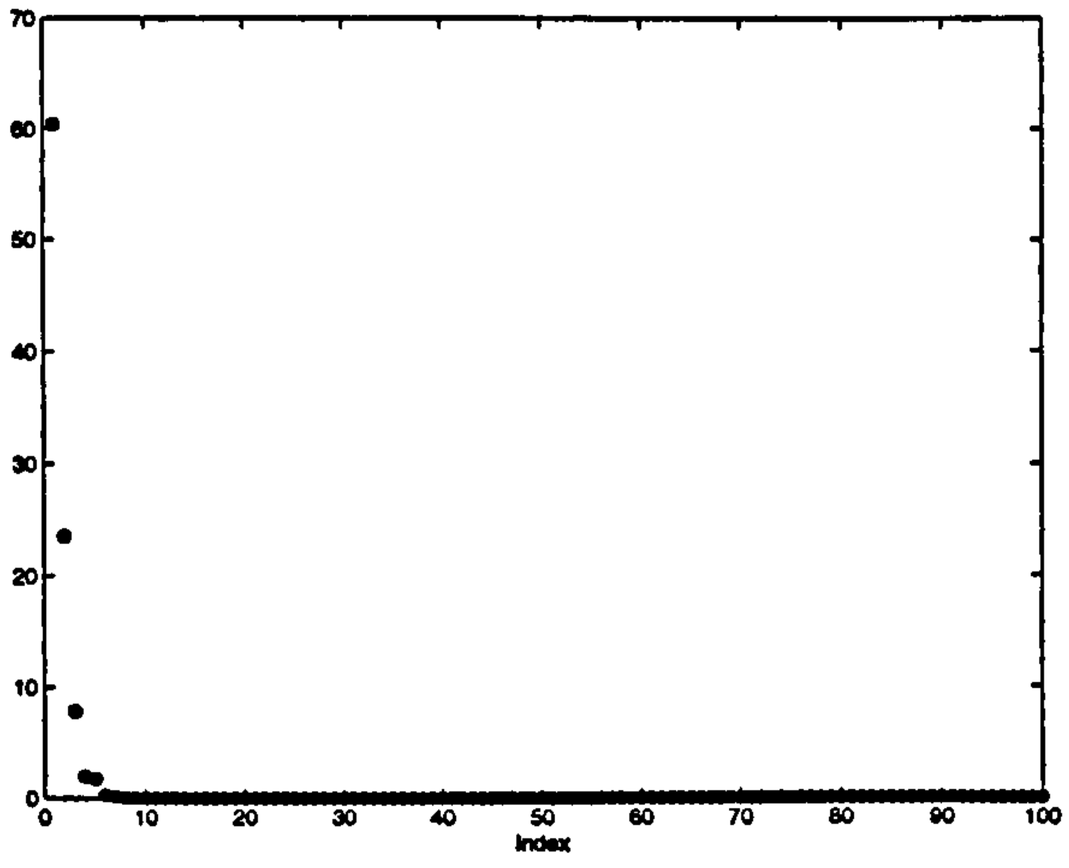


Fig 6.11 The singular values of the time history of signal s.

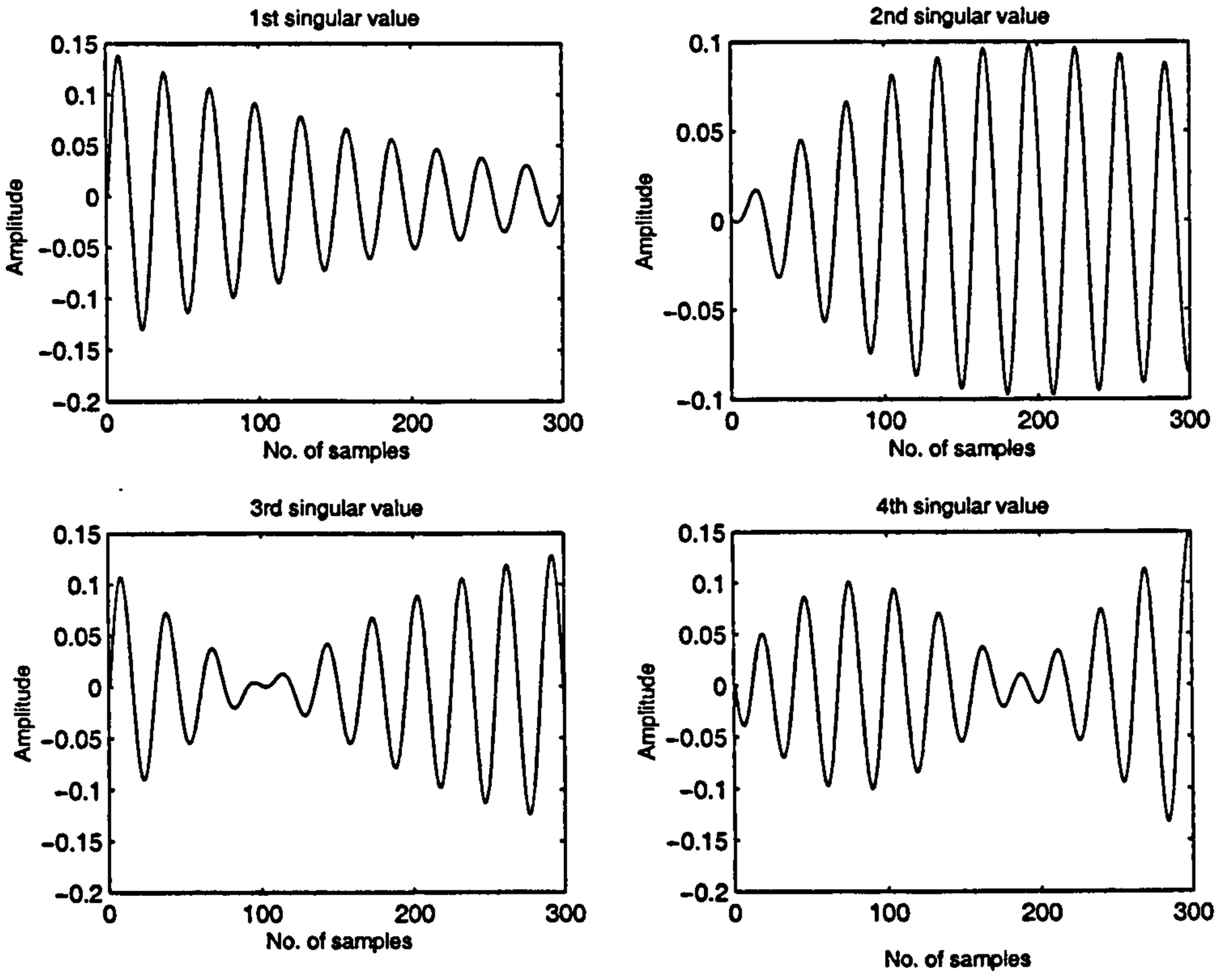


Fig 6.12 Fourier transform of the first four singular vectors of a sine wave with frequency variation interval of 1Hz and centre frequency of 10Hz (FT 1200 points).

6.3.1 Influence of the damping on the modelled signal

The previous example has the amplitude, damping and frequency varying. We will see how the shapes of the vectors change only when one of those parameters is varied. The most important variables are the damping and the frequency. The amplitude by itself, with the other two parameters fixed, will not generate a nonstationary set of events because they are linear combinations of the signal with amplitude one. In this case we previously know that there is only one linearly independent row, which corresponds to a rank one matrix. This way we know that the resulting SVD (the SVD gives information about the rank of the matrix) will have only one singular vector with the corresponding singular value (that corresponds to standard deviation of the matrix).

In this example we take a signal where all the possible events have the same frequency and amplitude values (the amplitude is 1 and the frequency is 10Hz). The damping takes values inside the interval described in table 6.1. When the damping is the only variable in the modelled signal there is one dominant singular value, which corresponds a vector that it is similar to a typical realisation of the signal (see fig.6.14). The other two vectors have very small corresponding values and their shapes are similar to modulated sine waves. The frequency of the modulator seems to increase with the order of the vectors. Comparing these results with the results of the decomposition of signal s we conclude that variation of only the damping in the chosen interval gives rise to a small variability.

Fluctuation in the damping by itself in the modelled signal gives a small variability since the resulting SVD gives effectively only one significant value. (If the size of the interval increases the variability of the system increases but it is not very significant).

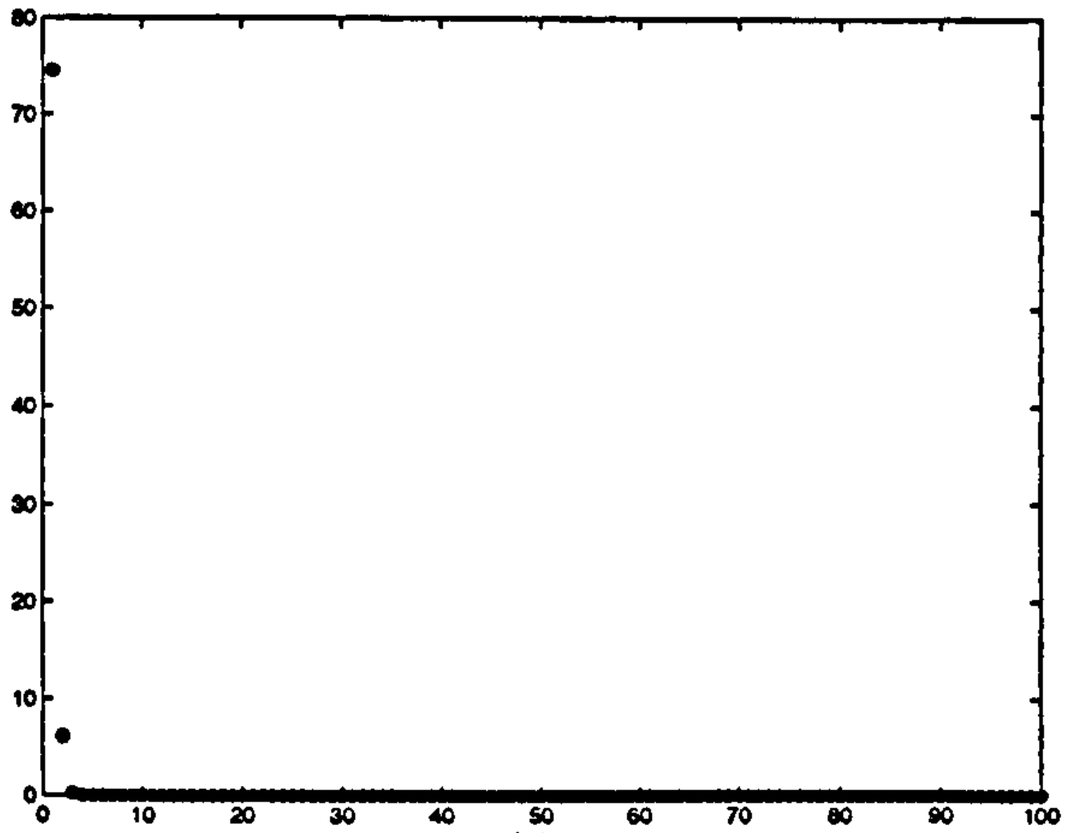


Fig 6.13 The singular values of the time history of the modelled signal with damping varying in the interval [0.1,0.3]and constant amplitude (unity) and frequency of 10Hz.

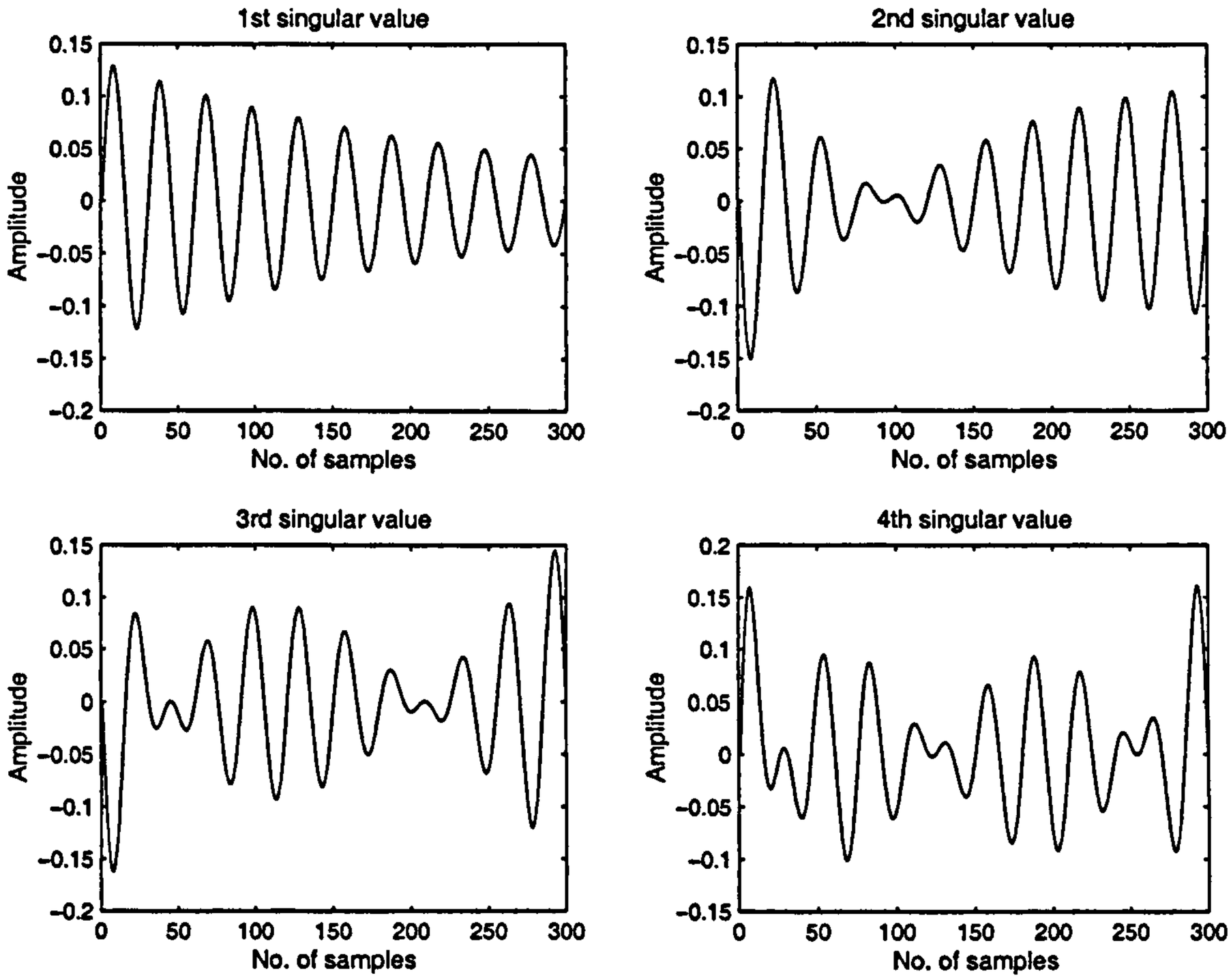


Fig 6.14 First four singular vectors of the time history of the modelled signal with damping varying in the interval [0.1,0.3]and constant amplitude (unity) and frequency of 10Hz.

6.3.2 Influence of the frequency on the modelled signal

Now we will see how the frequency by itself influences the shape of the singular vectors for the case of stochastic transient signals.

For the first example we will change the size of the frequency interval. The centre frequency will be the same (10Hz) and the values of frequency will be taken random from an interval with uniform distribution. In a previous example in chapter 4, section4, we presented the results of the SVD for a signal with the frequency interval of [9.5,10.5]Hz. The size of the interval is now of 4Hz and the central frequency is the same, 10Hz. The results, in fig.6.16, for the singular vectors indicate that the energy is more spread for a greater number of values. This means that the system, defined by the set of possible events of random stochastic transients for a greater frequency interval, has a greater dimension.

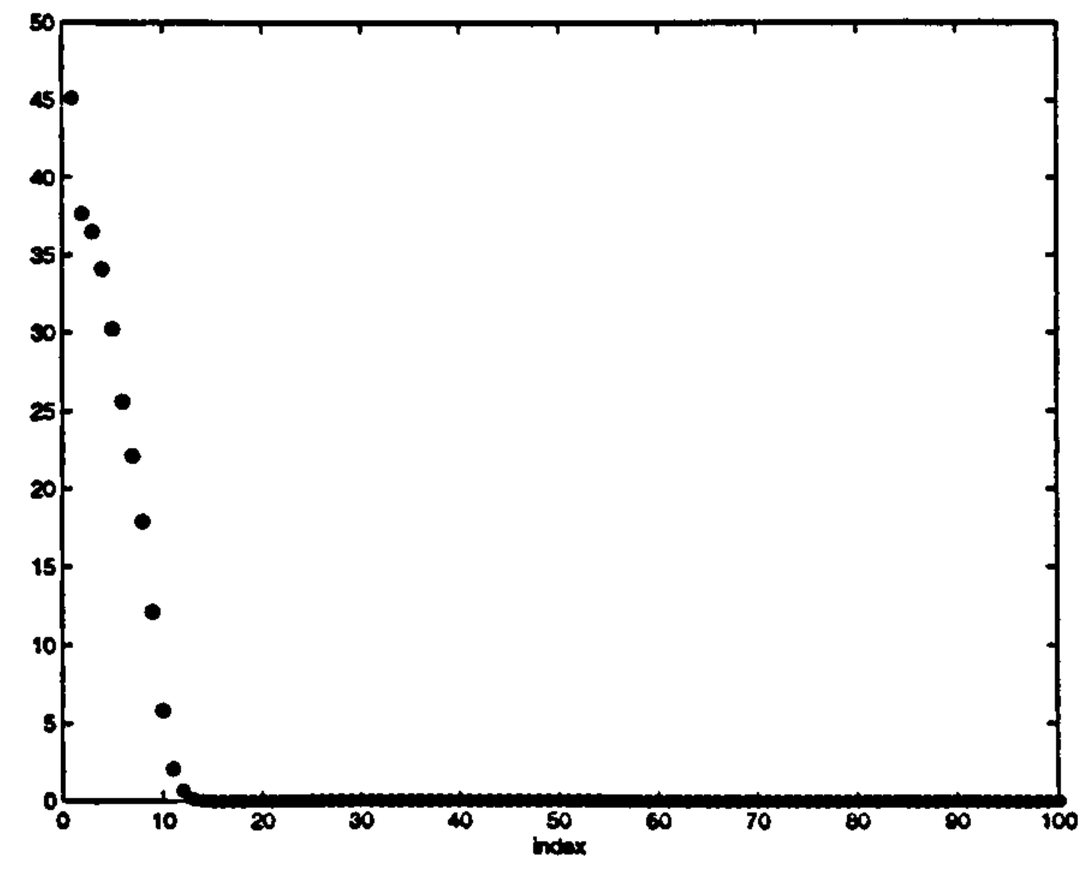


Fig 6.15 The singular values of the time history of the modelled signal with frequency varying in the interval of 4Hz with centre frequency of 10Hz (with constant damping).

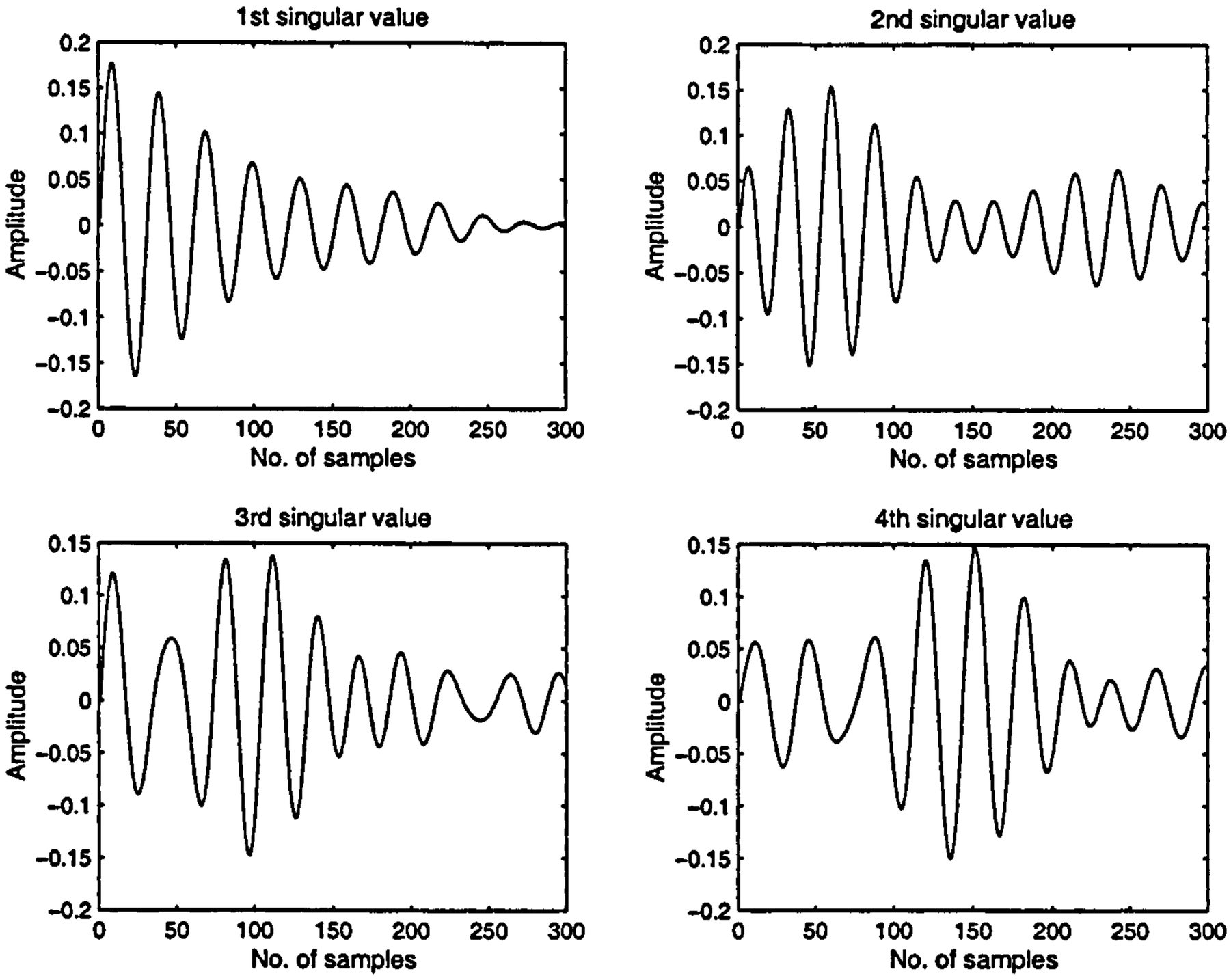


Fig 6.16 First four singular vectors of the time history of the modelled signal with frequency varying in the interval of 4Hz with centre frequency of 10Hz (with constant damping).

These results show that the frequency variation is the most important of all in eq.(3.1) for the variability of the system.

6.3.3 The influence of the delay in the modelled signal

All the experiments performed about the model have the events of the time history matrix starting at the same time. We will see how the presence of delayed signals will affect the vectors. The presence of delay will be very important for the note recognition process, as we will see in later sections. The delay was applied to the model with the same intervals of variation mentioned in section 6.3. The delay was taken randomly from a uniform interval between 0 and 20 samples. The samples, corresponding to the delay, were replaced with zero.

The SVD of the time history matrix of delayed events (see figs.6.17 and 6.18) shows that the most important difference is at the beginning of the singular vectors due to

the presence of samples corresponding to the delay. There are more significant singular values and the second value is much nearer the first. With the presence of delays we are introducing a new variable spreading the energy of the system that will increase the dimension of the space defined by the singular vectors.

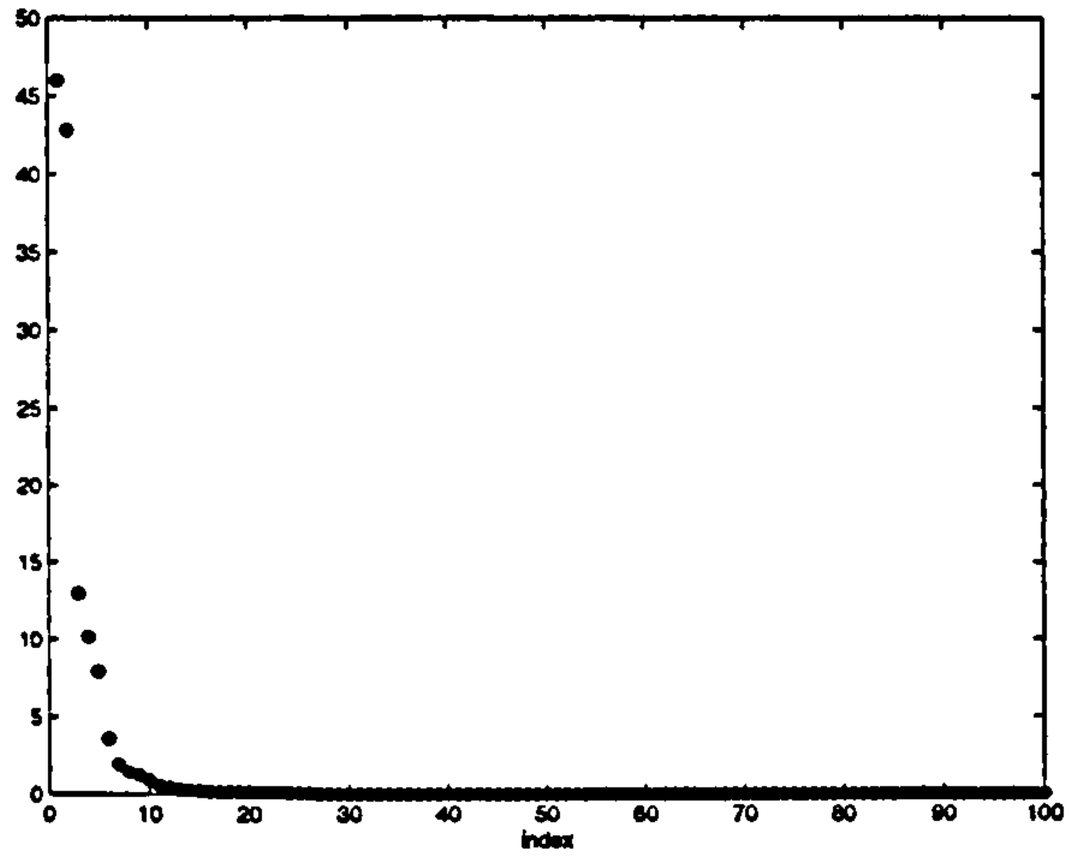


Fig 6.17 The singular values of the time history of the modelled signal with delays between events from 0 to 20 samples.

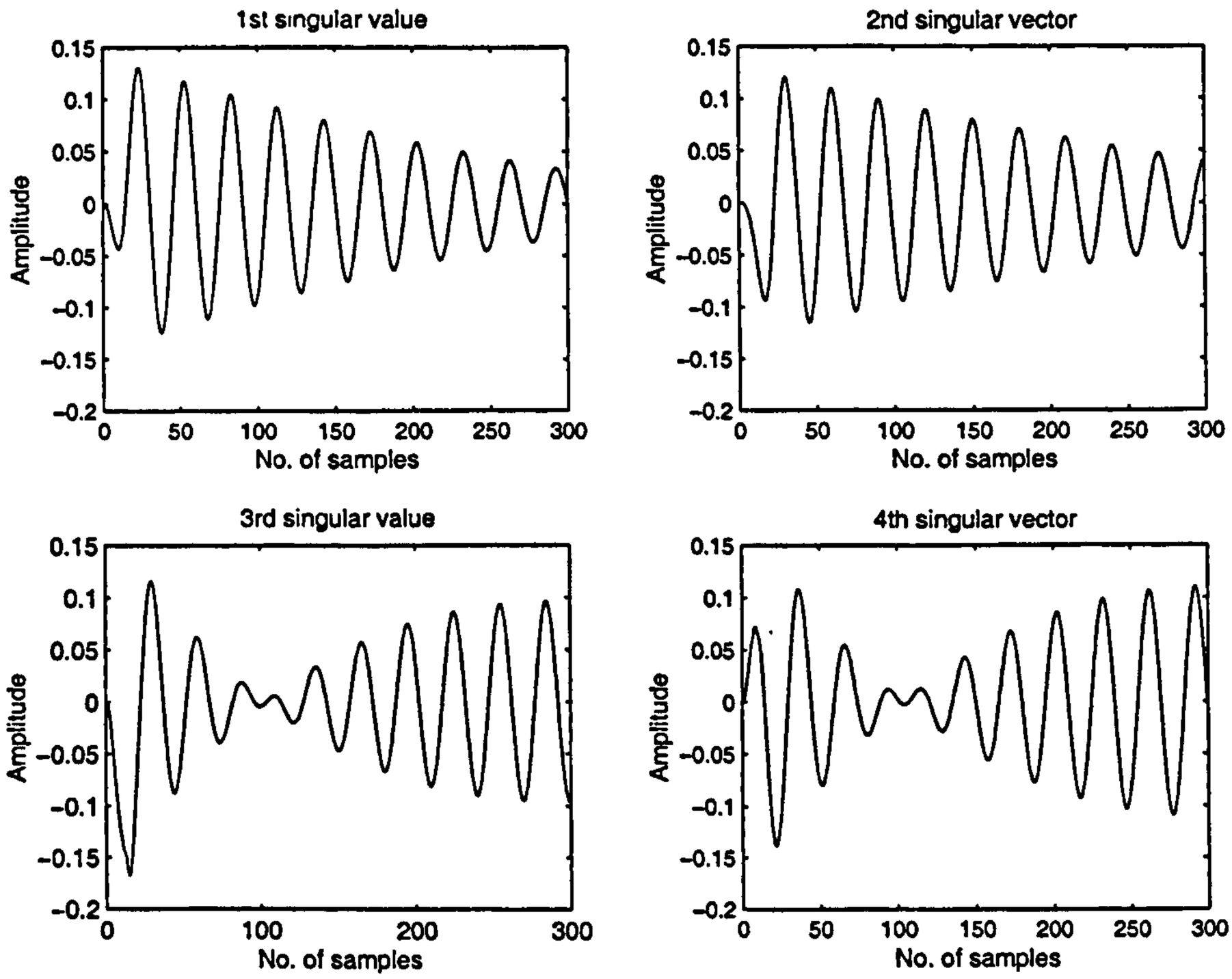


Fig 6.18 First four singular vectors of the time history of the modelled signal with delays between events from 0 to 20 samples.

This case of the nonstationary matrix with delayed events is very interesting because it has a similarity with the stationary case. We saw that when we have a stationary signal we have to build a time history matrix using the delay as a new variable introduced in the system. The delay introduces the variability in the system defined by a stationary signal. In the nonstationary case we do not need to introduce the delay to determine the SVD of the signal. The set of the possible events is enough. But if the events are taken without being aligned we are introducing the delay as a variable on the system. The similarity is that when we introduce the delay, the vectors 1st and 2nd, 3rd and 4th have similar shape although they have corresponding singular values that are not coupled. In the example of chapter 4 section 4.5 we see that the 1st vector is similar to the second an so on and the shapes correspond to sine waves. Now we have the same coupling but the shape of the vectors correspond to modulated signals (those for the first two are similar to a possible event).

If we increase the size of the interval of the delays until the size of the last delay in the stationary case we will approach the results of the stationary case. In this last case the signal was embedded using delayed versions of the signal. In the nonstationary case we are introducing the delay as another variable of the signal drawn randomly for an interval with uniform distribution. The big difference between these two cases is that the first part of the first singular vector has smaller amplitude.

The introduction of delays in the signal events can be very important, as we will see later, for the recognition process of modelled piano notes.

6.4 SVD of other nonstationary signals

In this section we will analyse the SVD of other nonstationary signal and what information can we take about the signal from the decomposition. The case of nonstationary sine wave (with modulated frequency) can be considered as a narrow band signal. For interest we expand to a broadband process.

6.4.1 Uniformly modulated noise

We consider now uniformly modulated noise. We generate a set of white noise events (100 events with 300 samples each) uniformly modulated. Each event is expressed as:

$$x_i = m.w_n = \sin(\omega t).w_n \tag{6.11}$$

where w_n represents the white noise and m the modulation which is the same for all the possible events.

The SVD of the signal shows that the singular values are distributed evenly along a line. For the case of modulated white noise SVD is not able to separate the modulation from the noise. As mentioned in previous chapters if we have additive white noise it is possible to separate the noise from the signal. For other type of noise the SVD does not separate it from the signal.

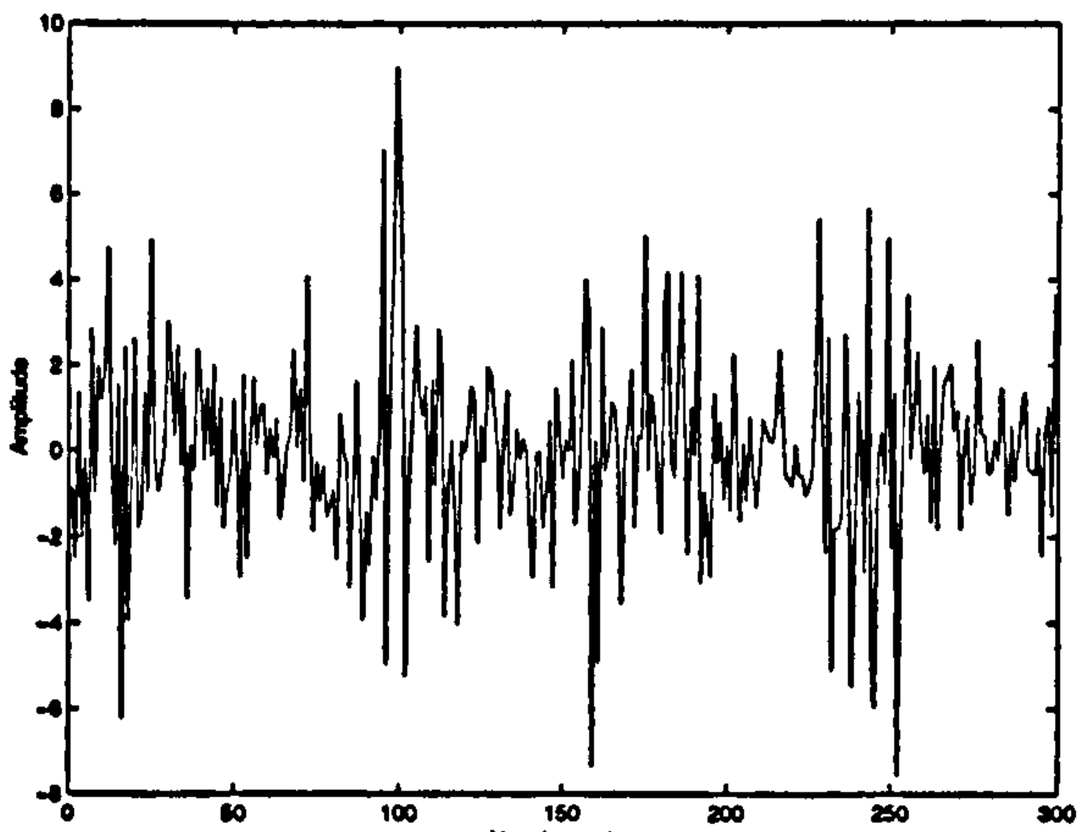


Fig 6.19 Time history of a possible event of the signal representing white noise modulation.

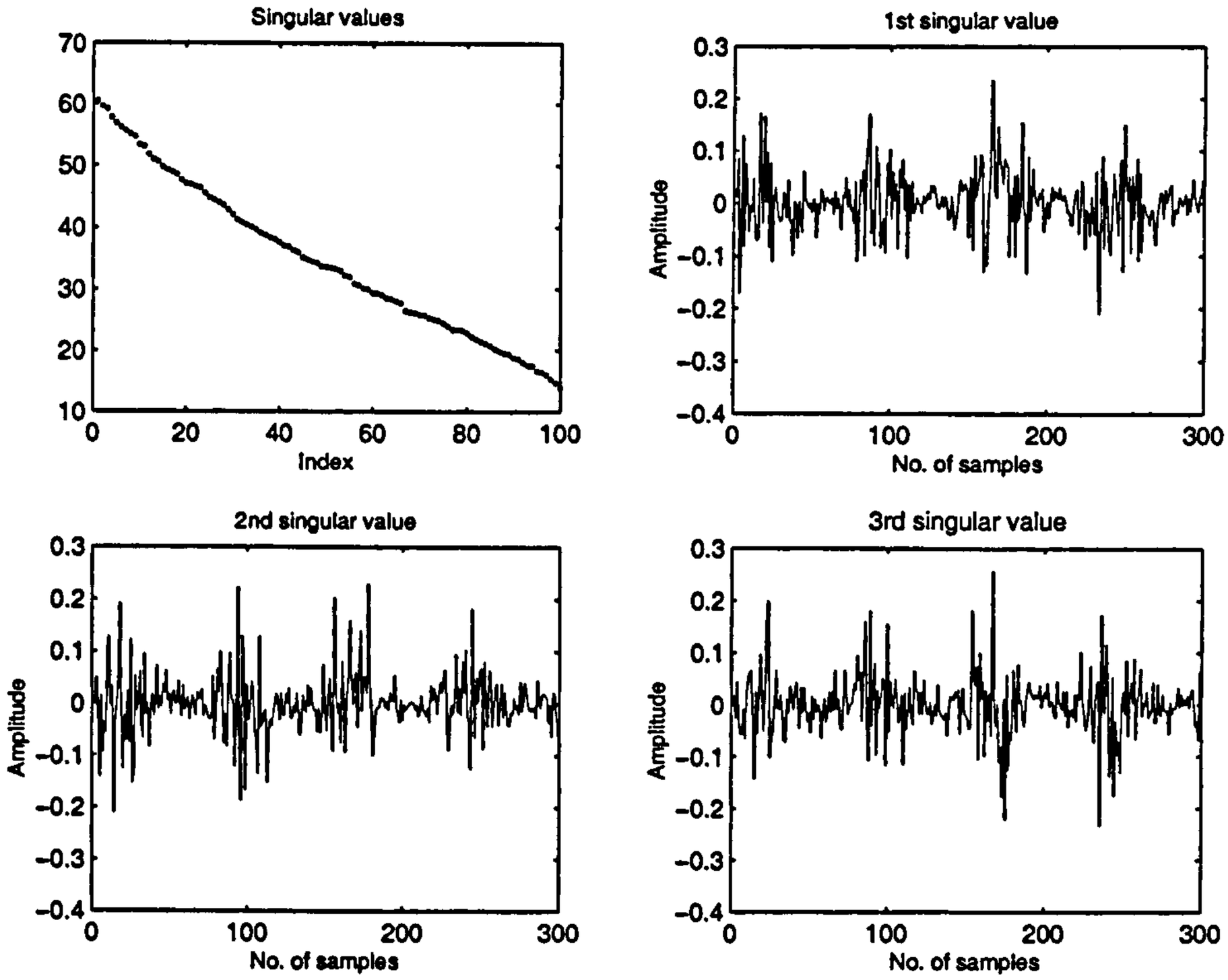


Fig 6.18 The singular values and the first three singular vectors of the time history of the modelled signal with delays between events from 0 to 20 samples.

6.5 Conclusions

The SVD of a sine wave considered as a nonstationary process results in a set vectors that are modulated sine waves. The first vector of the decomposition is similar to the average of the realisations of the time history matrix.

The variation of the parameters of the simulated signal (eq.(3.1)) influences the singular values and vectors of the time history of the signal and the broader the intervals of variation (for the same number of realisations) the higher the number of nonzero singular values with significant values.

The application to a wide band modulated signal shows that the SVD does not separate the noise from the modulation.

CHAPTER 7

SVD AND RECOGNITION PROCESS OF SIMULATED PIANO NOTES

Introduction

The SVD of the simulated piano notes will be used to build a recognition process that can be extrapolated to the real piano notes. The recognition process is applied to simulated notes with one and several frequency components. We apply the process also to simulated octaves to analyse the performance of it in the separation of signals with coincident harmonics.

7.1 Recognition of modelled piano notes

One of the aims of this research is to apply SVD as a technique for piano note recognition in polyphonic music. The aim in this chapter is to develop and test a method for the recognition of simulated piano notes that can be extrapolated to the real case. We need to relate our modelled piano signals to the context of the real piano to define some of the variables. For each real piano note there corresponds a fundamental frequency that is related to a musical pitch. As an example a note with the fundamental frequency at

262Hz corresponds to a C note (middle C in the piano keyboard). The frequency of the fundamental defines the musical pitch of the piano note. Each note is separated from the adjacent ones by an interval of a half tone. The frequency relation between adjacent notes is given by the ratio 1.059 (see annex 6 for a more detailed description). This sets the boundaries for the interval of variation for the frequency. Suppose that to a certain chosen pitch corresponds a certain frequency f_c. This frequency is the central frequency of the interval. The upper and lower limits of the interval should be smaller than a half a tone from f_c to avoid the superimposition with the adjacent notes.

When a piano note is played a few times one after another on the same piano, the variation of the frequency of the fundamental of that note is very small. During a piano solo concert many pianists have the piano tuned during the interval. This means that the frequency of the notes can vary slightly during a performance. The variation is small but for trained musicians can be noticeable. The demanding use of the piano in a performance introduces variations on the fundamental frequency. If a piano is not tuned for long periods of time the deviation of the fundamental frequencies of each note from the norm increases.

These considerations together with the ones mentioned in chapter 3 lead us to choose the values for the limits of the interval of variation for the frequency. The upper and lower limits of the interval of frequency variation for each note are determined by the multiplication and division of the centre frequency corresponds to the value of the tuned note, by a number that represents in music the interval of a quarter of tone (annex 6). This interval is smaller than the one between each note and large enough to contain all the possible variations mentioned above.

The amplitude of the time history of the real piano note changes with the dynamics applied to the key. We decide to take arbitrary values where the maximum value is represented namely by the value one and the minimum is 0.5. The decay of the note also varies with the dynamics applied. The values of the damping where also chosen arbitrarily. We did not consider the phase in this experiment since we will always consider the notes from a zero starting point. Instead we will introduce a delay for each possible realisation of the signal as a new variable.

For a first approach we will consider only a single frequency component for the each note. The signals will then be more complex considering a greater number of frequency components for each realisation.

7.1.1The recognition process

When we are referring to recognition of piano signals we mean the recognition of signals with fundamental frequencies that fall in a certain interval of error that correspond to a certain pitch. A note is similar to another if they have similar frequencies inside that determined interval.

In the case of simulated signals we are interested in developing a recognition process that uses SVD as the principal technique of signal analysis and allows the recognition of signals. SVD when applied to a signal defines a new space where the vectors of the basis are the singular vectors. The representation of a possible realisation of the signal into this new space should use the minimum number of vectors. The recognition method is based on the projection of a signal representing a simulated note in the new spaces defined by the singular vectors. We will analyse those projections searching for a pattern that correspond to the recognition of the note.

This recognition process for simulated piano notes that we want to implement has the following steps:

1- Simulate of a set of time histories with a number of possible events of a signal defined by the frequency representing the pitch.
2- Decompose the time history matrix of the signal using SVD.
3- Project the signal for recognition in the new space defined by the basis set determined by the SVD.
4- Identification of the signals.

This whole process is used to perform the recognition of simulated notes. First we generate time history matrices with possible realisations of a signal. We will use the same time history matrix dimensions used in chapter 6: 100 possible realisations with 300 samples each. As in the examples shown previously each row of the matrix is a realisation of the signal. The frequency is the principal variable that defines a simulated signal associated, in real piano notes, with a pitch. We can say that a signal is similar to another if they have similar fundamental frequency falling inside a certain interval of error.

The time history matrix is then decomposed and the resulting set of the right singular vectors is used as a basis set of the new space where the signal is represented. This will be the "reference" of the signal. These "references" of the single notes, i.e. the singular vectors, will be used to project simulated signals composed of one or more simulated notes (e.g. of the octave that is composed by two notes). We search for similarities between the projected signal and the vectors of the basis set to recognise the notes present in the signal.

On the last step we analyse and treat the data treatment of the projections to extract information about the projected signals.

7.2 The projection of a single note with only one frequency component on SVD basis sets

The projection of simulated notes on a SVD basis set of a single note is part of the proposed recognition process. We shall analyse the projection of some simulated notes with only one frequency component to extract some information for the recognition.

We generate underdetermined systems of 100 possible realisations of a signal and with 300 samples each, some signals according to equation 3.1 and as explained in chapter 4. The frequencies of the simulated signals for the examples are smaller then real piano notes for convenience of representation of the method. We generated two signals s_1, s_2, for which we determined the SVD (their parameters are described in table 7.1.). The frequency relation of the signals is $1:2$ which corresponds in musical terms to an interval of an octave. We use this signals with this relation since we are interested to verify whether the SVD is able to separate signals with coincident harmonics (and the octave is therefore included).

	Frequency interval (Hz)	Centre frequency (Hz)	Amplitude interval	Damping interval
s_1	[9.71,10.29]	10	[0.5,1]	[0.1,0.3]
s_2	[19.43,21]	20	[0.5,0.75]	[0.15,0.35]

Table 7.1 Intervals of variation of the components of the signals s_1 and s_2.

The basis sets of the signals s_1 and s_2 (see fig.71 and 7.2) composed of the right singular vectors (denoted Bs_1 and Bs_2) will be used to project possible events of those two signals and two other signals s_h and s_{12}. The first of these two signals s_h is a single event with frequency relation with the centre frequency of s_1 of 1.059 (i.e. $s_h / s_1 = 1.059$). This note is separated from s_1 by a musical half a tone interval. This is the smallest interval between notes in the western music (see annex 6). The frequency of this single event is 10.59Hz, the amplitude is 1 and the damping is 0.3. We analyse how the projections of the closest signals (in frequency terms) behave. The other signal s_{12} is a signal resulting from the addition of s_1 and s_2, i.e. $s_{12} = s_1 + s_2$. This corresponds to an octave. We introduced this composite signal to see if the projection of that signal in SVD basis sets of single notes has information about the composing single notes.

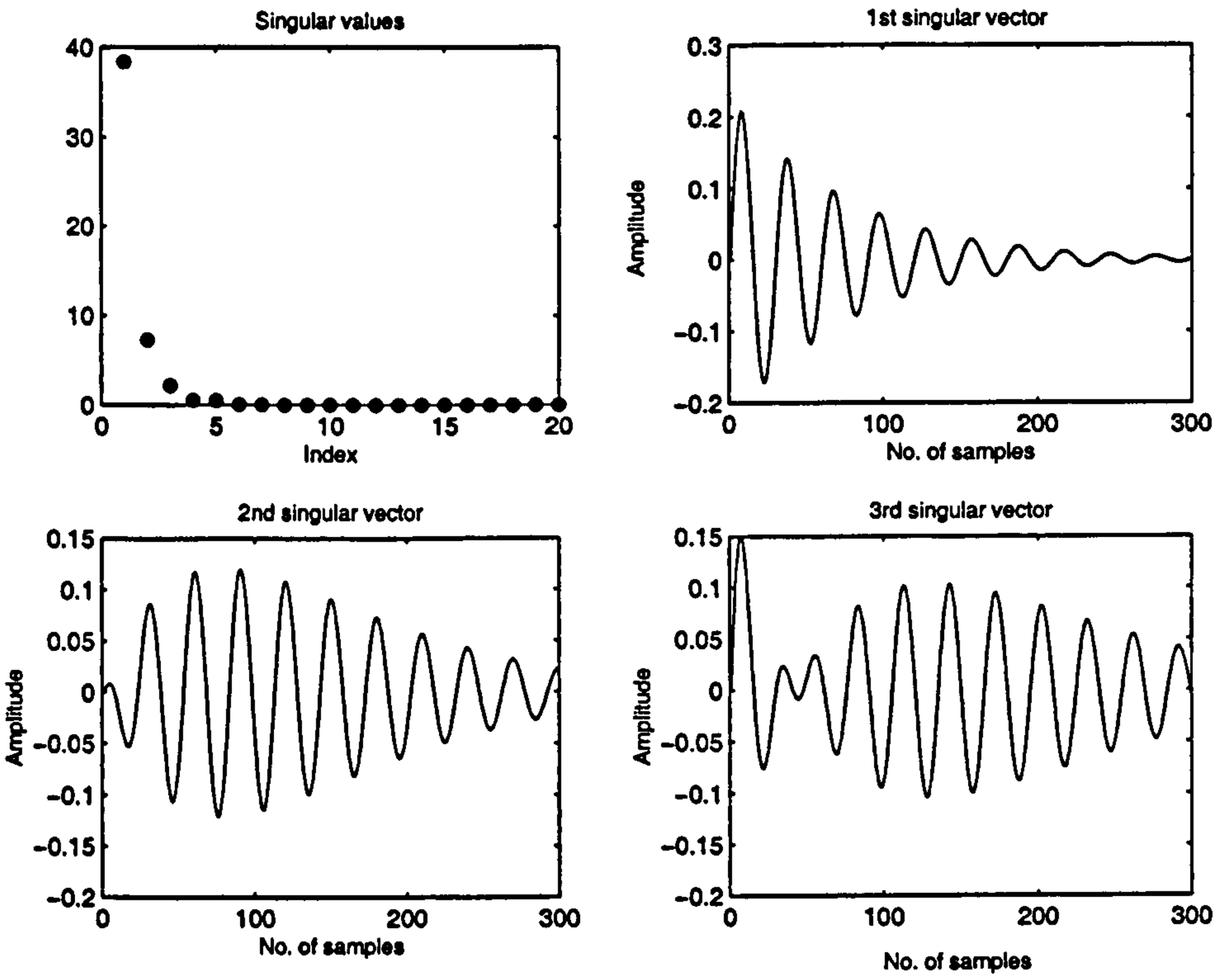


Fig 7.1 SVD of the signal s_1 (10Hz): the singular values and the first 3 singular vectors.

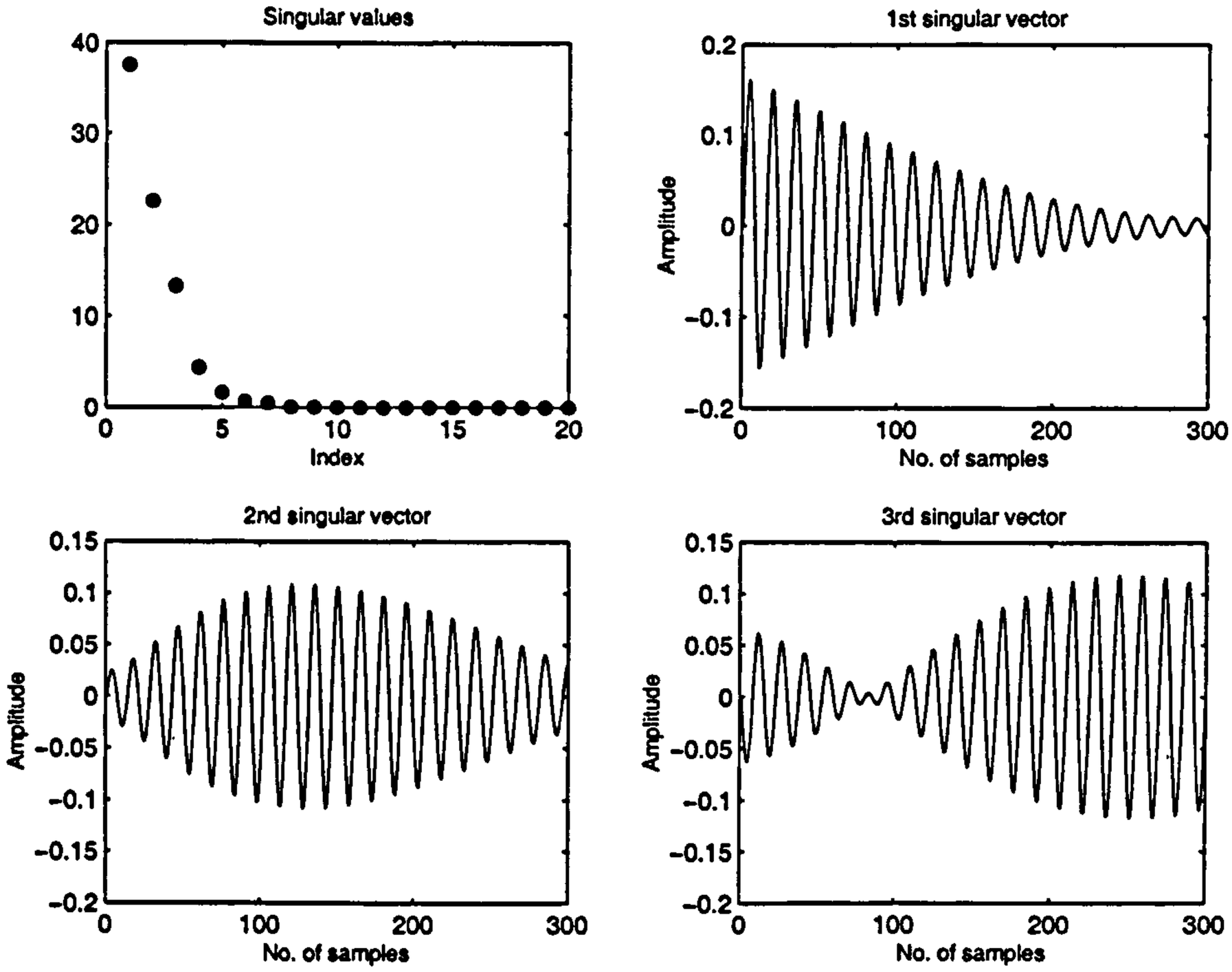


Fig 7.2 SVD of the signal s_2 (20 Hz): the singular values and the first 3 singular vectors.

A possible realisation of notes s_1 and s_2, signals s_h and s_{12} were projected first on the basis Bs_1 of the signal s_1. The number of projections for each signal is 300 since there are 300 singular vectors and the signals are projected on each singular vector.

When a possible realisation of the signal s_1 is projected on its own basis set the first projection, that corresponds to the projection of the signal on the most significant vector, has the maximum value (see fig.7.3). After a few number of projections the values become zero. When the note s_2 is projected on the same basis the pattern of the projections is completely different from the s_1 note. The values seem randomly distributed and there is no preferred vector with a prominent value. The s_h projection values present more than one prominent value on the first projections. Again as in the s_1 case the values become zero after a few projections. In the case of the composed signal s_{12} the first five projection values are very similar to the s_1 projections. The next values are very small and show a random pattern. This similarity with the first projections indicates the presence of the signal s_1 in s_{12}.

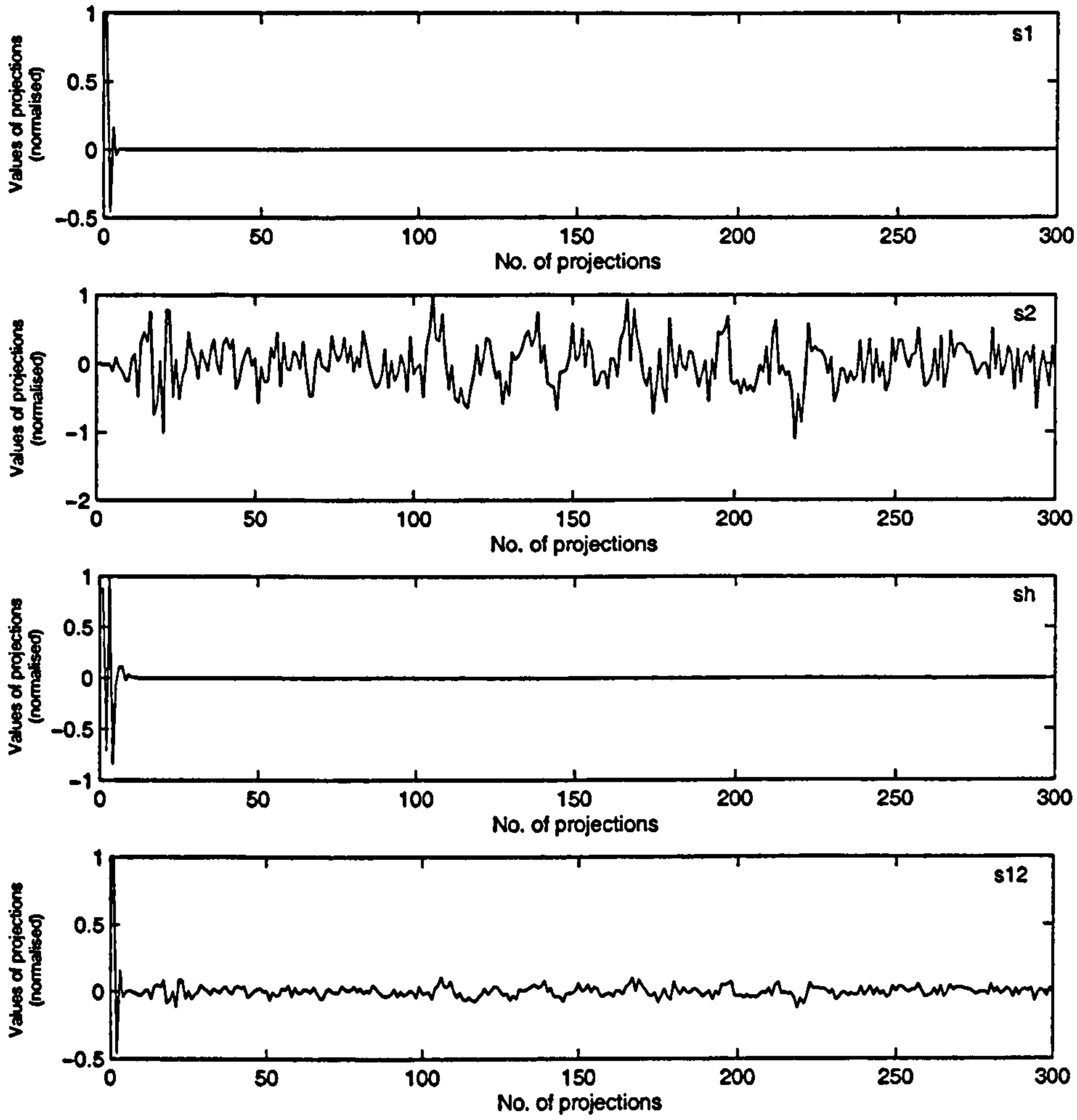


Fig 7.3 Projections of signals possible realisations of the signals s_1, s_2, s_h and s_{12} on the basis set of s_1: Bs_1.

The results show that all the projections present different patterns that can allow the identification of the signal s_1 when projected in its own basis set although the high values on the first projections are not enough since we have also high values for the octave. We need to take in account more projections.

We will perform the same projections of the same signals but now on the basis set of s_2: Bs_2.

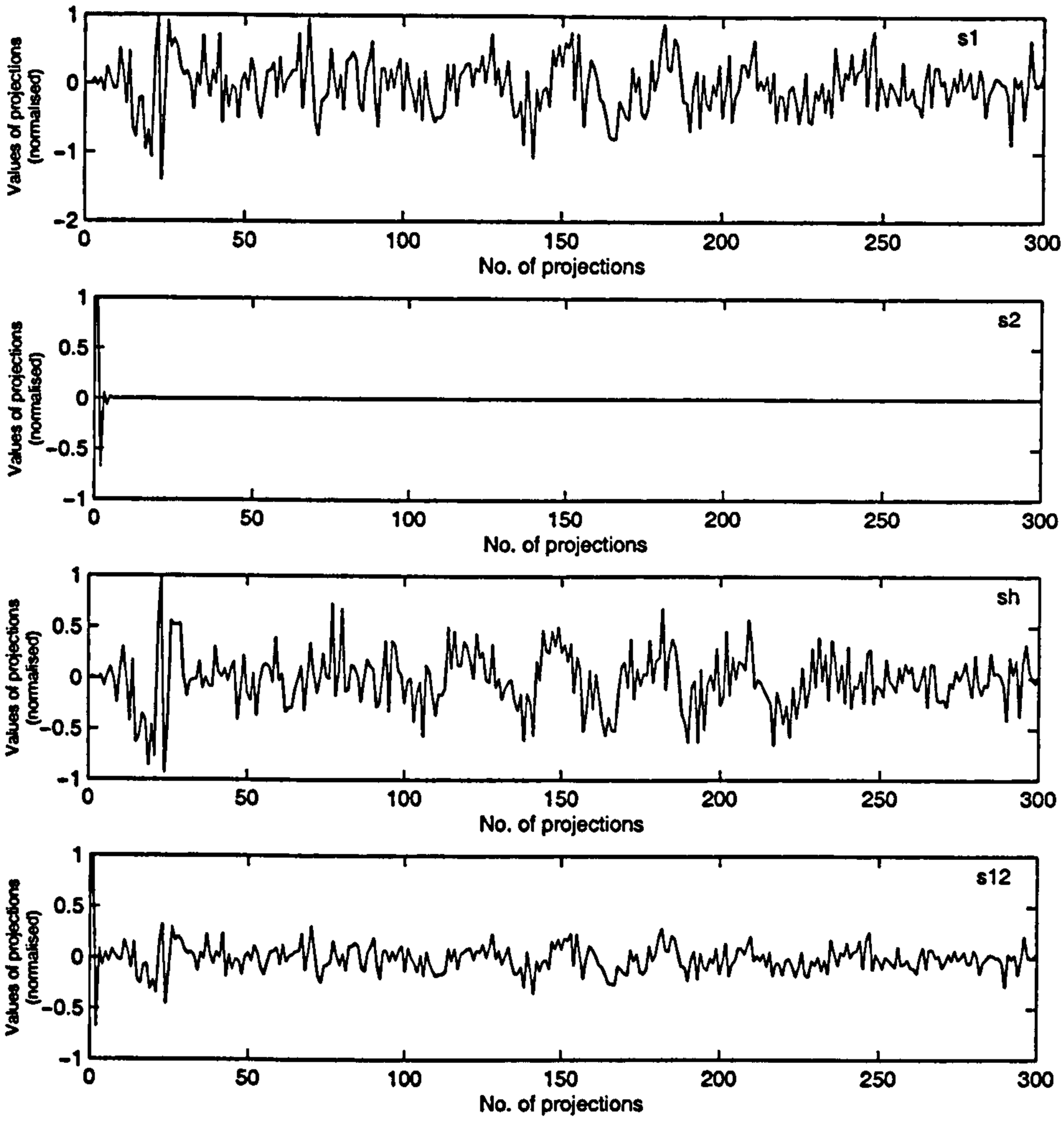


Fig 7.4 Projections of signals possible realisations of signals s_1, s_2 possible realisations, s_h and s_{12} on the basis set of s_2, Bs_2.

The results again reinforce the presence of a maximum value on the first projection when a signal is projected on its own basis set. After a few projections the values become zero. The projections of s_1 and s_h on Bs_2 show a random pattern. The signal s_{12} again has a maximum value on the first projection and the first few values are very similar to s_2 projections. These similarities correspond to the presence of the note s_2 on the signal s_{12}. After the first few projections the values present a random pattern. The identification of the signal s_2 is possible through the analysis of the presented patterns.

The results of the projections show that when a signal is projected onto its own basis set there is generally a maximum value in the first projection and after a few projections the values become zero. The signal s_{12}, that is a summation of two signals, presents the same initial pattern of the projections on Bs_1 and Bs_2 of the notes s_1 and s_2 on their own basis sets. This indicates that SVD is able to "separate" added signals.

7.3 Projections using delay

In the previous experiments all the events of the time history matrix of the original signal were aligned. The same is true for the projected signals that were aligned with the possible events and consequently with the singular vectors. Now we will see if the recognition process generates the same values of projection for the same signal but with a delay. In the first case of the signal s_1 we introduce a small delay of 5 points and then 15 and 30 points. These two delayed versions of s_1 are now projected on the same basis set of the aligned time history matrix. The resulting projections for 5 points of delay show that the pattern of the fig.7.1a is no longer present. We cannot identify the note s_1 with the projections values themselves when we are in the presence of a 5 point delay. If we transpose this to the case of real piano notes this will constitute a problem since we have to detect the beginning of a note and align it with the events of the time history matrix with great accuracy. The detection of the exact point where the attacks begins can have errors of several points. Such errors could lead missing the identification of a note.

In the case of 15 points and 30 points of delay we still have the greatest value of the projections on the first value. If we remember the central frequency of the interval, where the frequency takes values, is 10Hz and corresponds to, more or less, 30 points per period of each event of s_1. So 15 and 30 points of delay correspond approximately to half a period and a whole period of an event. If we see the projections values for the 15 point delay signal, the first value is negative which means that the first vector of the basis set is in opposite phase with the signal. For 30 points they are again in phase and we

have the greatest value for the first projection and it is positive again. In all the three projections the delay does not significantly affect the values of the projections on the last vectors.

If we compare these projections of the same signal s_1 with different delays (the 5 point delay example in fig.7.5) with the projection of the signal s_h without delays in the basis of s_1 (Bs_1) (fig.7.3 s_h) we see that it is not possible to distinguish between s_h as a s_1.

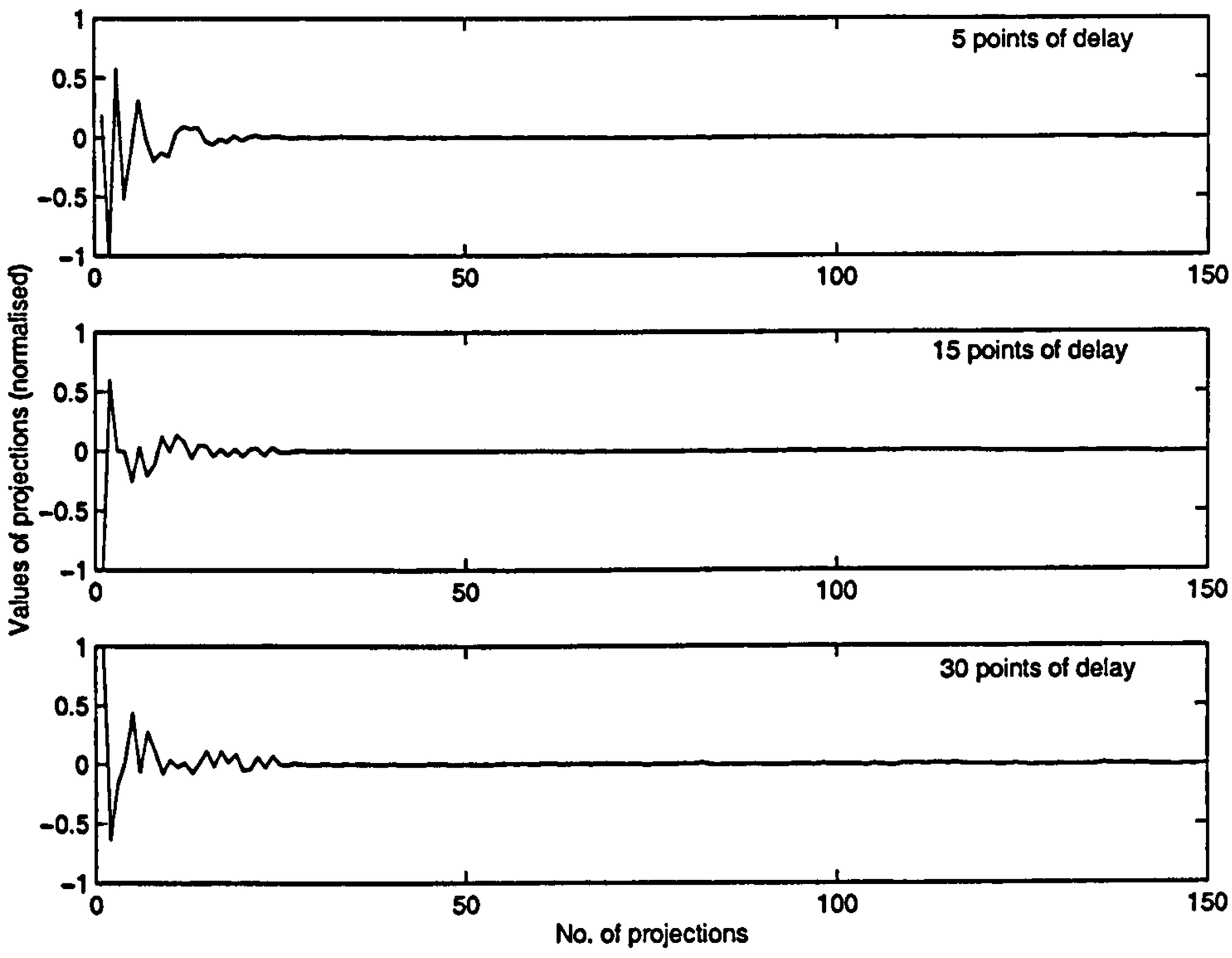


Fig 7.5 Values of the projection of signal s_1 with 5, 15 and 30 points of delay on the basis Bs_1.

If we introduce in the time history matrix of the signal delays in the events as a new variable the problem of the wrong identification can be solved. The delays between events will be a new variable of the system that will introduce a margin of variation for the first samples of the signals. Including signals with delays in the matrix we are including that information in the resulting vectors of the base and now it is possible to identify the without the need of determining the starting point with a great accuracy. The delays will vary between zero and 10 points. We have seen already the results of the

SVD for a matrix with delayed events in the previous chapter. The attack of the singular vectors change with the delay but the other features of the signal remain practically the same. Now we will project the signals s_1 and s_1 with 5 points of delay, s_{1d5}, on this delayed basis set. In fig.7.6 the graphs represent the projections of s_1 and s_{1d5} on the aligned basis set (*a* and *c*) and on the delayed basis set (*b* and *d*). The greatest value of the projections is not present on the first projection of signal s_{1d5} on Bs_1 without delays (fig.7.6.*a*). So we cannot recognise the signal. When we project the same signal on a delayed basis set, (fig.7.6 *b*) we again have the maximum value of the projection on the first value. For the case of the non delayed signal we still are able to recognise the signal on both basis since we still have a maximum value on the first projection (fig.7.6 *c* and *d*).

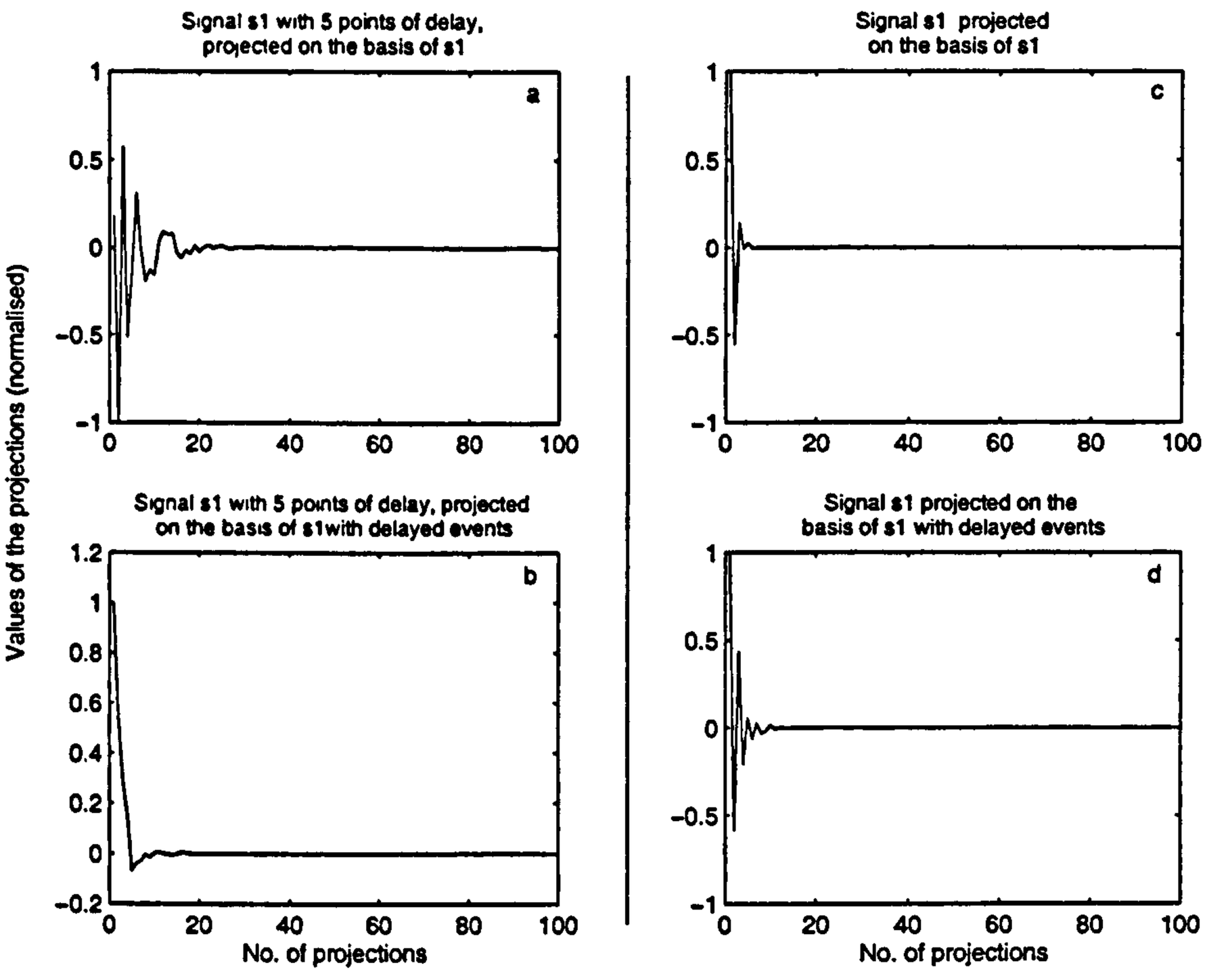


Fig 7.6 Values of the projection of signals s_1 and s_{1d5} (5 points of delay) on basis of s_1, Bs_1 with delayed events.

The projection of the signal half a tone above the centre frequency of the signal s_1 (10Hz) on the basis set of s_1 with delayed events (fig.7.7) presents now distinctive pattern

from the delayed signal s_1. The introduction of delayed possible realisations of the signal in the time history matrix allows a better identification of a delayed event of s_1 and avoids the miss identification of notes with near frequency for simulated signals with one frequency component.

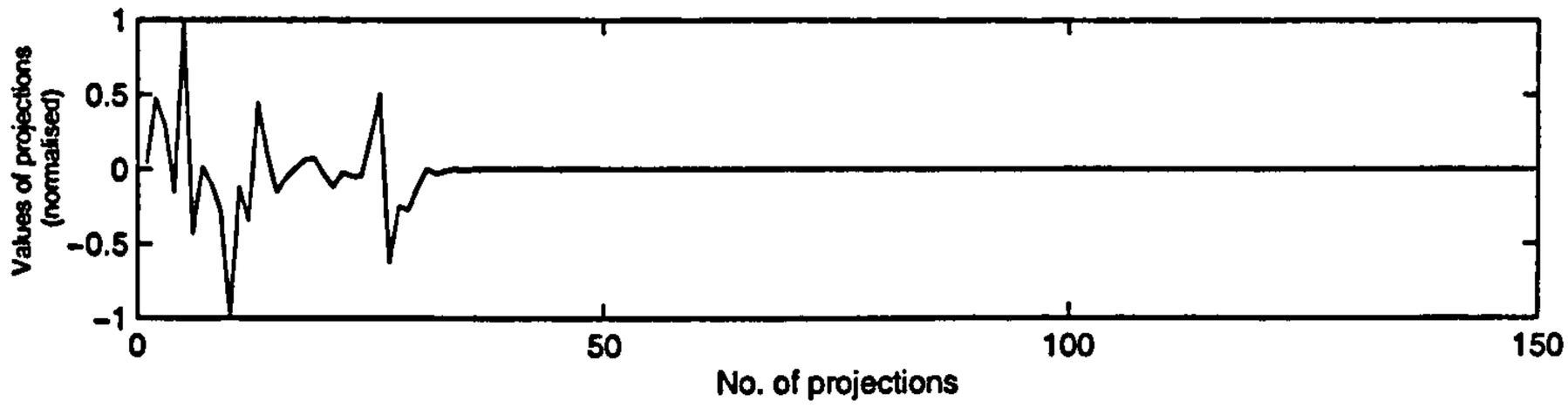


Fig 7.7 Values of the projection of the signal s_h on the basis set Bs_1 with delayed events.

7.4 Projection of an octave on the SVD basis sets of the component notes

In the earlier chapters it was mentioned that one of the great problems in music identification is the separation of notes with coincident harmonics as happens in the musical interval, the octave. That is the case when we have more than one note played together and for instance the note with the higher fundamental frequency has its own components coincident with the components of the other note. Through the Fourier transform this identification is not possible since the notes have coincident harmonics. We will use the process of projections to see if separation and recognition is possible. It is important to verify if this process of recognition using the projections on SVD basis sets of single notes is able to detect the octaves since all the other processes generally fail to do so.

The idea of the identification process is to use the basis sets of single notes as a reference for recognising the notes present in a signal composed of any number of single notes. So we will calculate the SVD basis sets for single notes and then we will project the signals, composed of one or more notes played together, on those basis sets trying to

UNIVERSITY OF SOUTHAMPTON LIBRARY

identify the notes present on that signal. For the identification of the notes present on the octave the signal is going to be projected on the basis sets of those single notes. Summarising we are trying to identify the presence of the notes in a composite signal and or a special feature that able to identify the octave by itself.

In this example we will generate two signals, s_1 and s_2. The first signal has two frequency components and the second one has only one frequency component whose frequency components are drawn from the same interval as the highest frequency component of s_1.

To determine signals with different number of frequency components we again use equation (3.1), that will generate each one of the frequency components of the signals, now represented as:

$$x(t) = \sum_{n=1}^{N} A_n e^{-\alpha_n t} \sin(2\pi f_n t) \tag{7.1}$$

The variable N corresponds to the number of frequency components of the signal. The variables A_n, α_n, f_n represent the same variables as in eq.(3.1) and where drawn from an interval with uniform distribution. For a signal with N frequency components each f_n represents the frequency of each component and the centre frequency of the uniform distribution where they are drawn are integer multiples of the first f_n. We note that the frequency values of the frequency components of a signal are drawn from intervals with the centre frequency related by integer numbers with the centre frequency of the fundamental.

The parameters, of signals s_1 and s_2, are represented at table 7.2 (see time histories on fig.7.8).

	Frequency interval (Hz)	Centre frequency (Hz)	Amplitude interval	Damping interval
s_1 (2 frequency components)	[9.71,10.29]	10	[0.5,1]	[0.1,0.3]
	[19.43,21]	20	[0.5,0.75]	[0.15,0.35]
s_2	[19.43,21]	20	[0.5,0.75]	[0.15,0.35]

Table 7.2 Intervals of variation of the components of the signals s_1 and s_2.

The centre frequencies of the components of signal s_1 have a relation of 2:1. The first harmonic of the signal is similar to the frequency component of the signal s_2. We used exactly the same possible realisations of s_2 as the first harmonic of s_1. These two signals are added to form a new signal s_{12} that corresponds to a simulated octave. The signal s_2 has its frequency components coincident with the harmonics of s_1 (see fig7.9). We introduced a delay for each possible realisation of each signal between 0 and 5 samples (the values were drawn randomly).

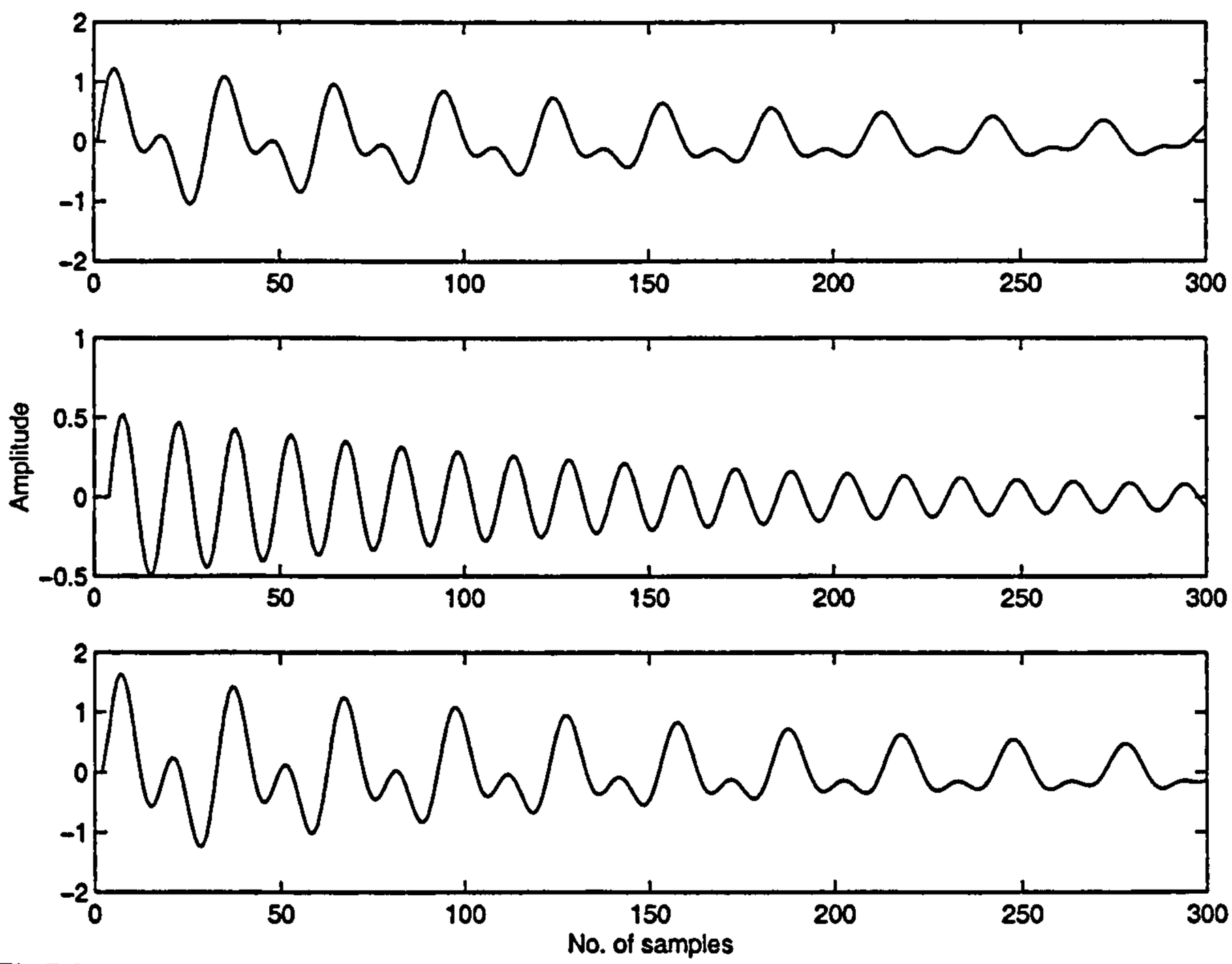


Fig 7.8 Time history of signals s_1 (2 frequency components), s_2 (1 frequency component) and s_{12} (sum of s_1 and s_2).

The SVD of these signals are represented in figs.7.9 and 7.10 with the singular values and the first three singular vectors.

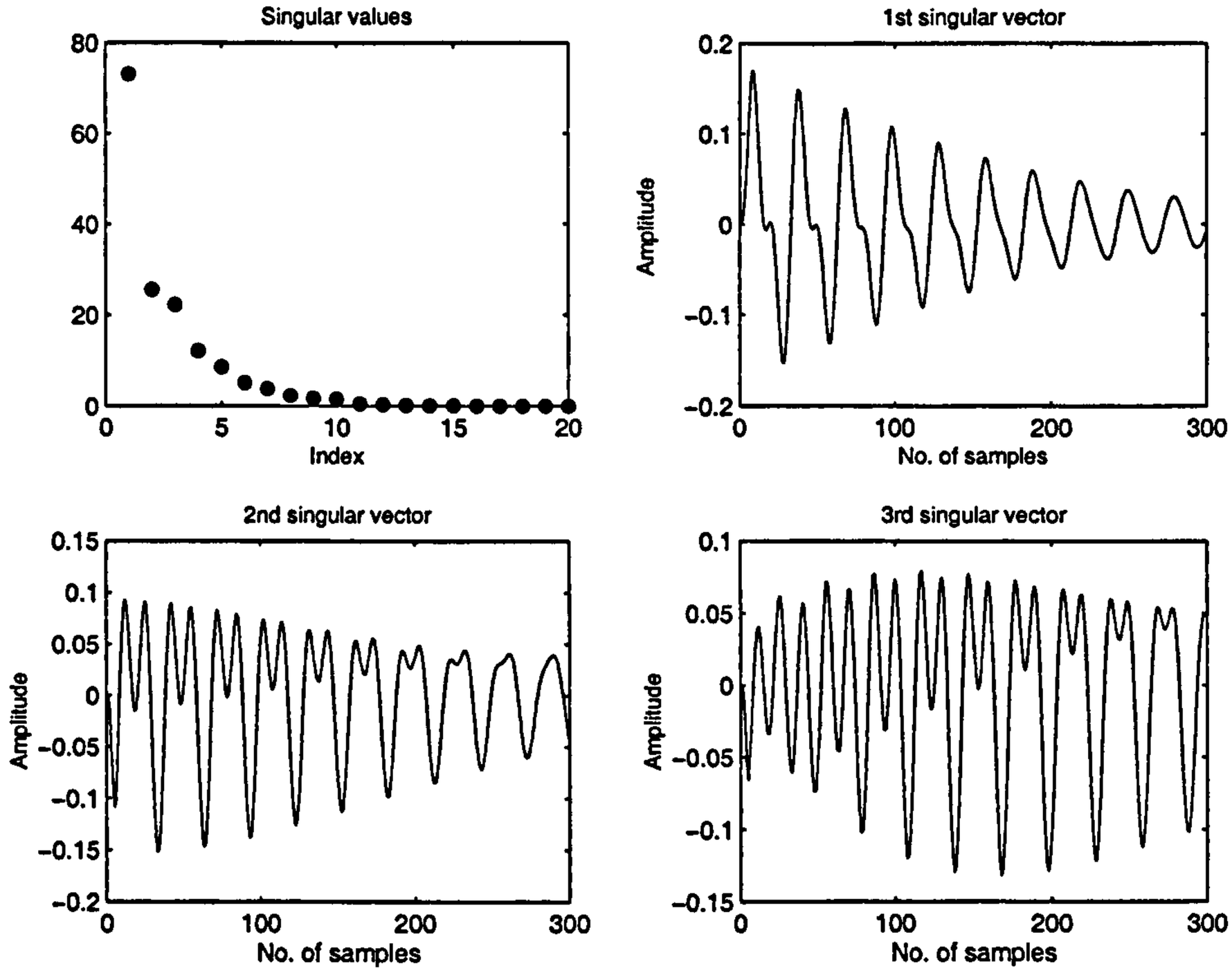


Fig 7.9 Singular values and the first three singular vectors of the signal s_1 formed by two frequency components (with delays between possible realisations).

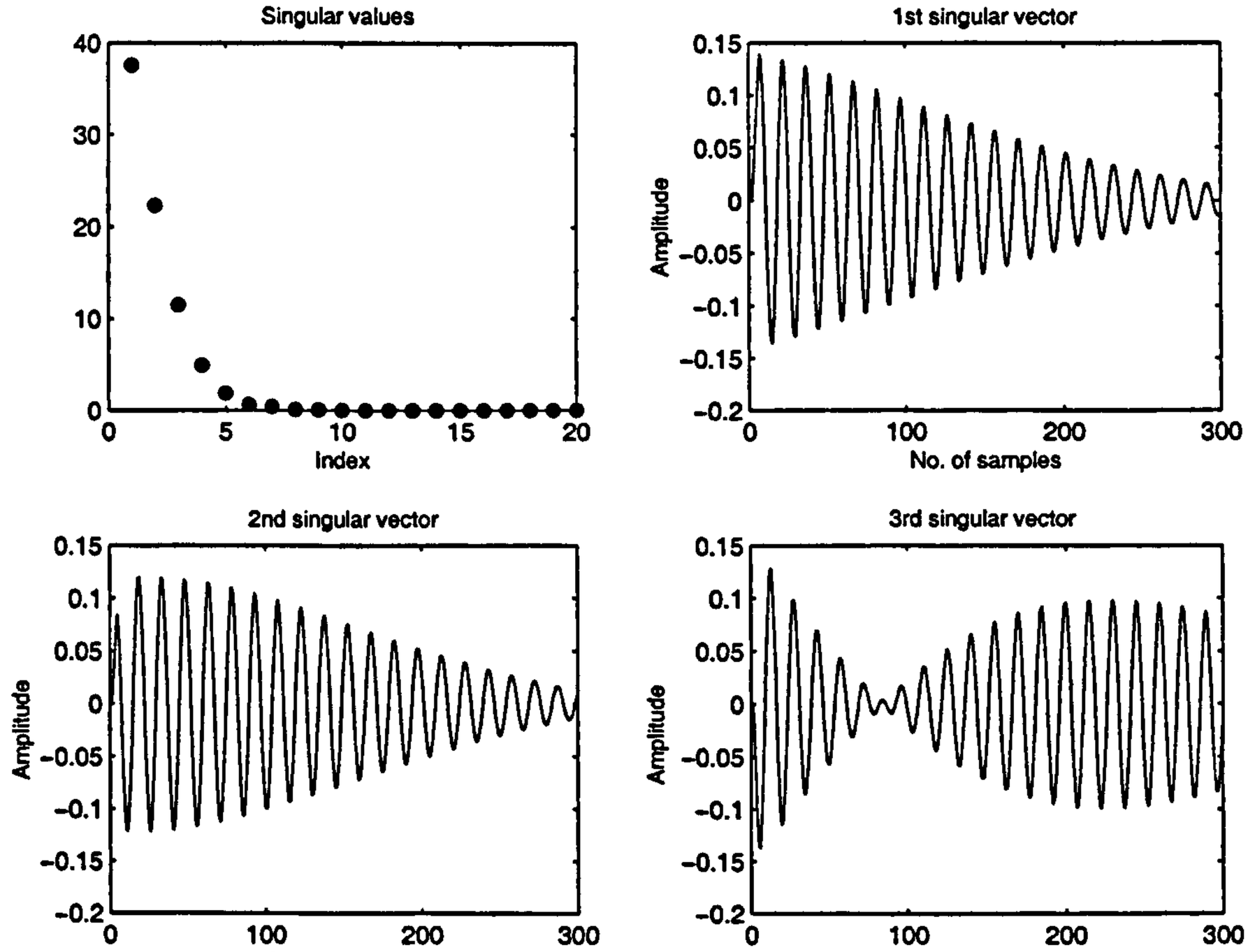


Fig 7.10 Singular values and the first three singular vectors of the signal s_2 formed by one frequency component (with delays between possible realisations).

The three signals were projected onto the basis sets Bs_1 and Bs_2 (figs 7.11 and 7.12.)

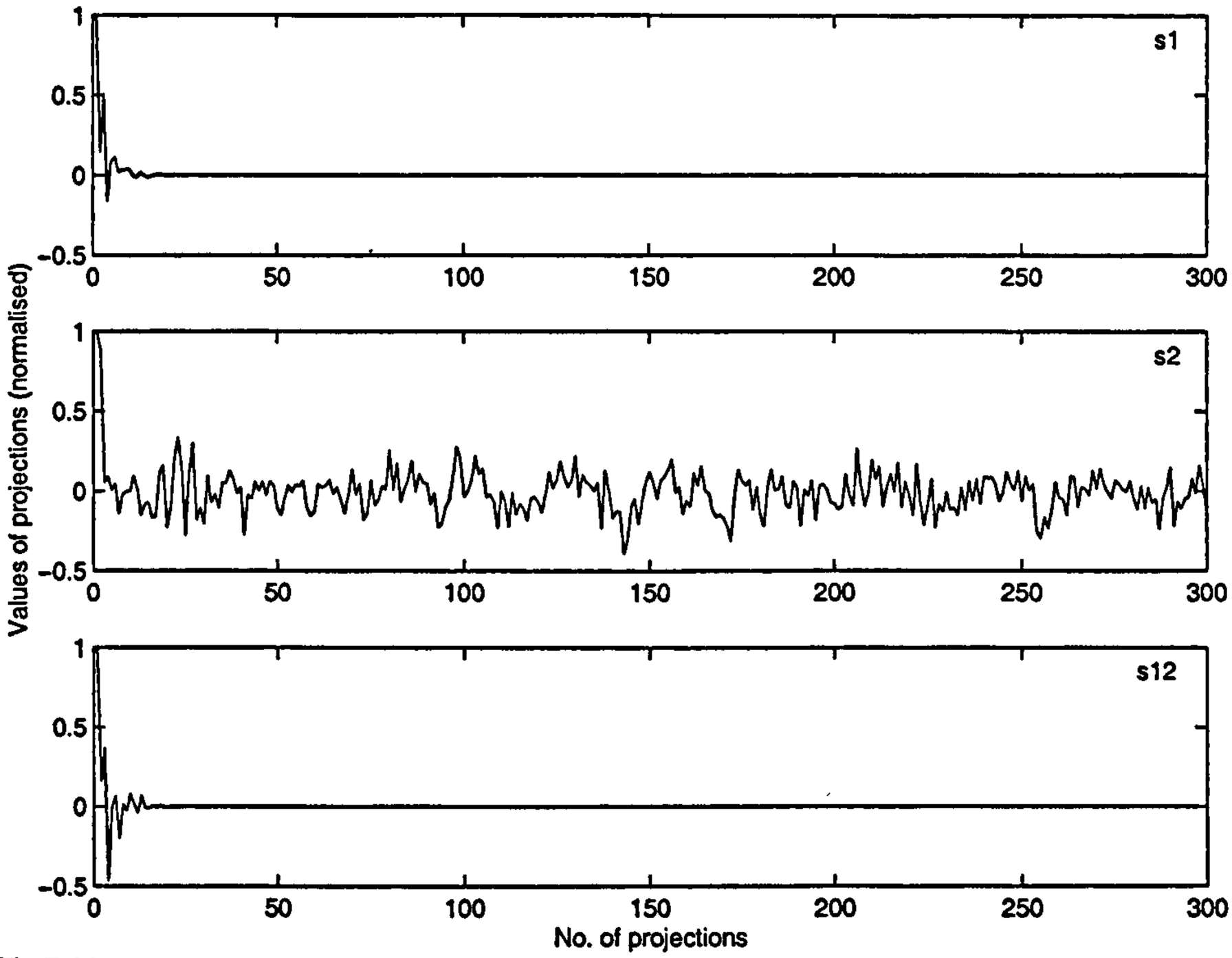


Fig 7.11 Projections of the signals s_1, s_2 and s_{12} on the basis set Bs_1.

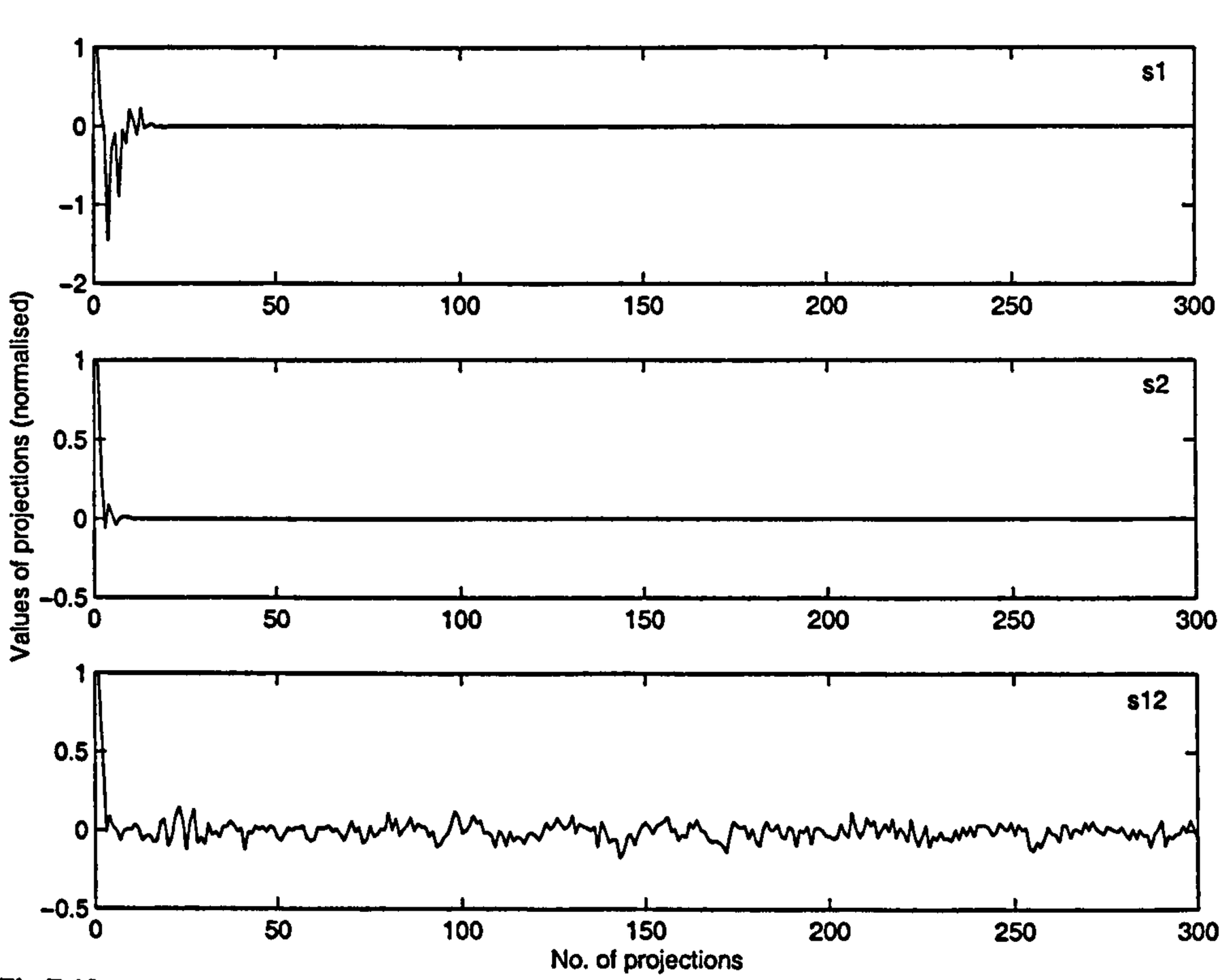


Fig 7.12 Projections of the signals s_1, s_2 and s_{12} on the basis set Bs_2.

As in the previous examples when a signal is projected in its own basis set it has a maximum value on the first projection and after a few projections with small values they become zero (fig.7.11 and 7.12). The signal s_1 when projected on the basis of s_2 does not present the random values after the first few value as happens for the projection of s_2 on Bs_1. The reason for this difference is because s_2 is a frequency component of s_1. The octave projected on Bs_1 has a similar pattern to the projection of the s_1 on the same base and that does not distinguish between those two signals. This simple analysis of the values of the projections is not enough to detect any difference in the projected values. The projection of the octave on s_2 gives different results (the octave has small values random distributed after the first ones) and it is possible to separate the octave s_{12} from s_2. This two results together of the projections of the octave on the basis Bs_1 and Bs_2 give enough information about the octave.

The Fourier transform of these three signals (fig.7.13) shows that it is not possible to make a distinction between the octave and the signal s_1.

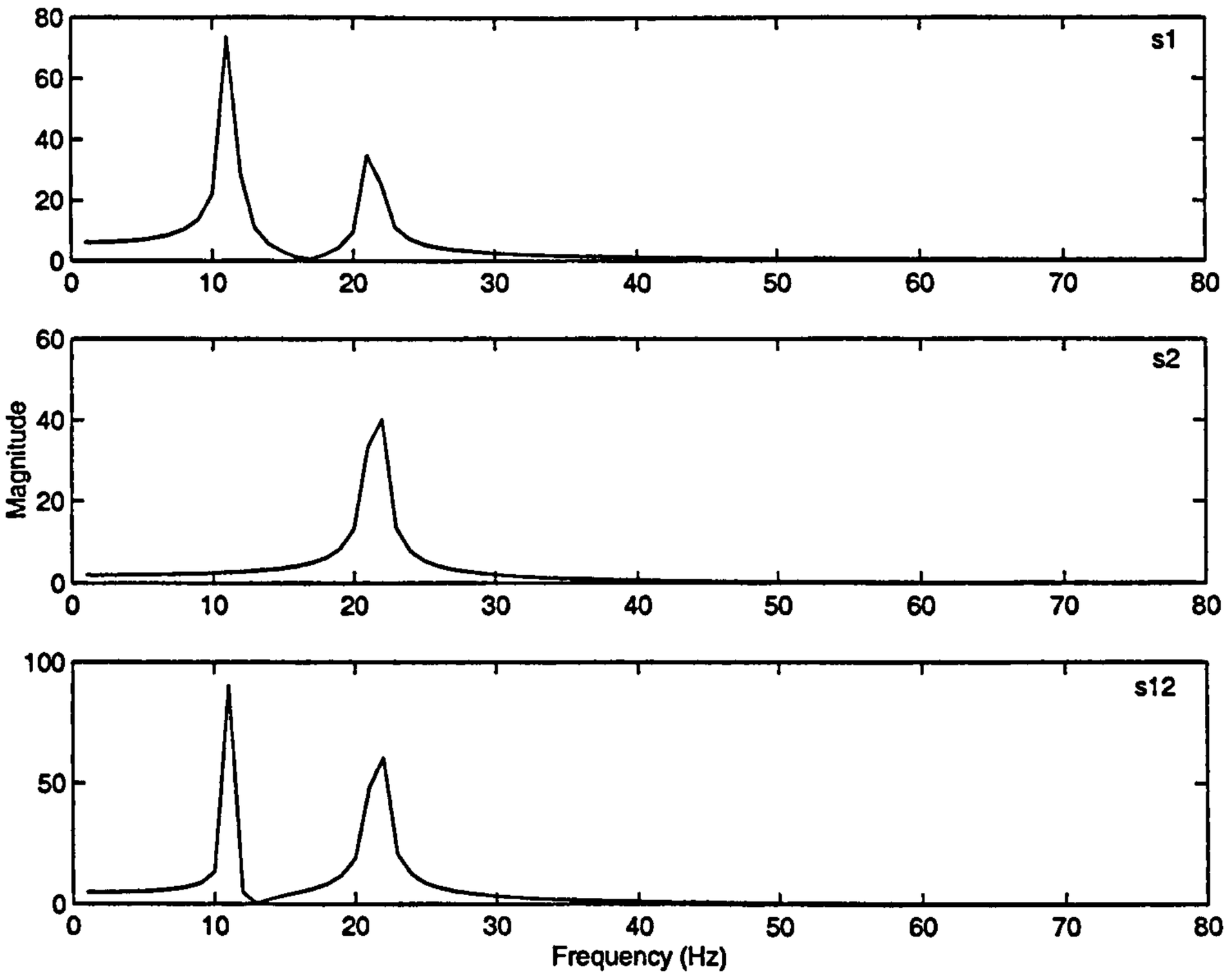


Fig 7.13 The Fourier transform of the signals s_1, s_2 and s_{12} (FT 300 points).

These resulting projections of the signals s_1, s_2 and s_{12} do not always have patterns similar to the ones presented. Some of the possible realisations when projected on the basis do not have the same results as we can see in the next figures.

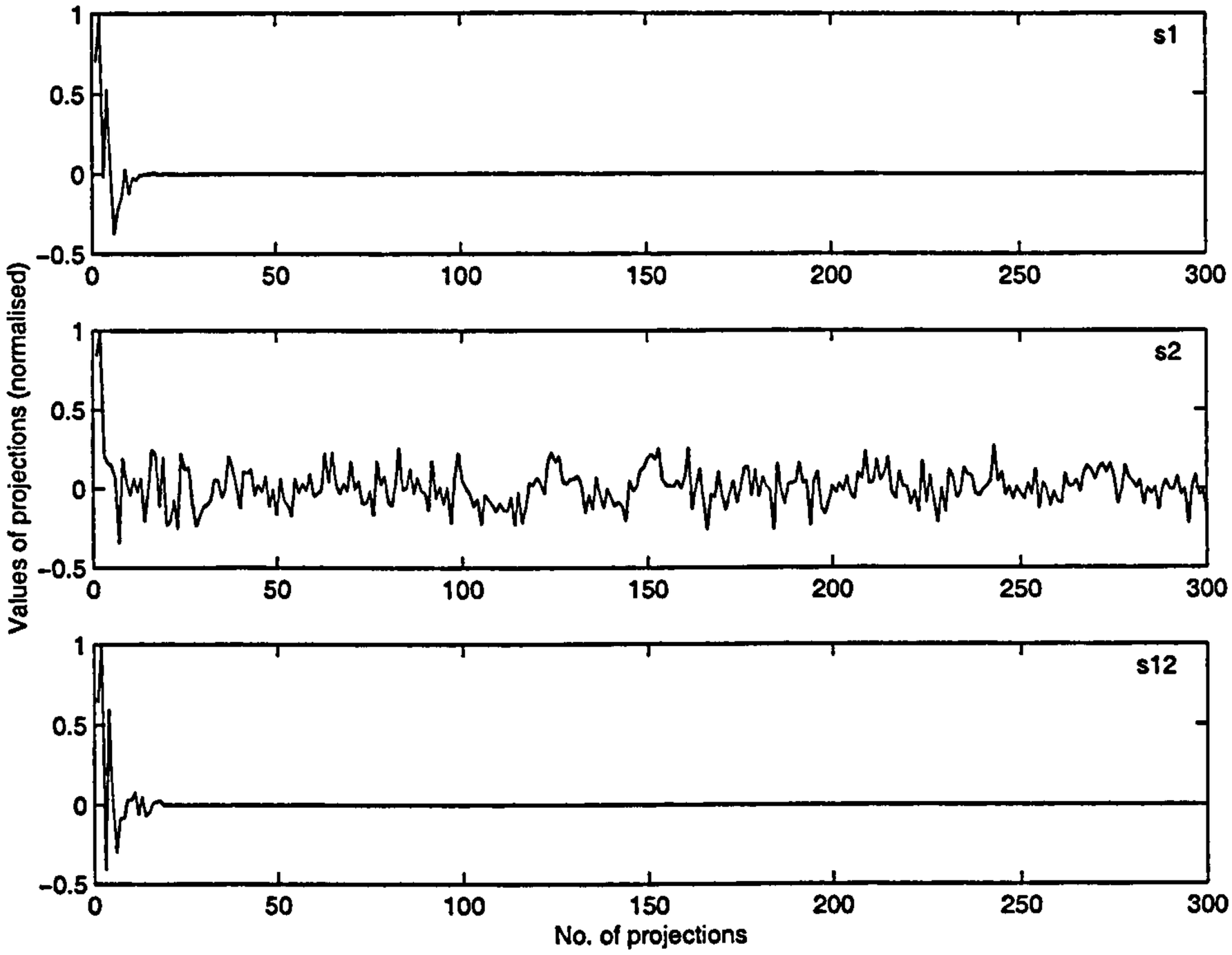


Fig 7.14 Projections of the signals s_1, s_2 and s_{12} on the basis set Bs_1.

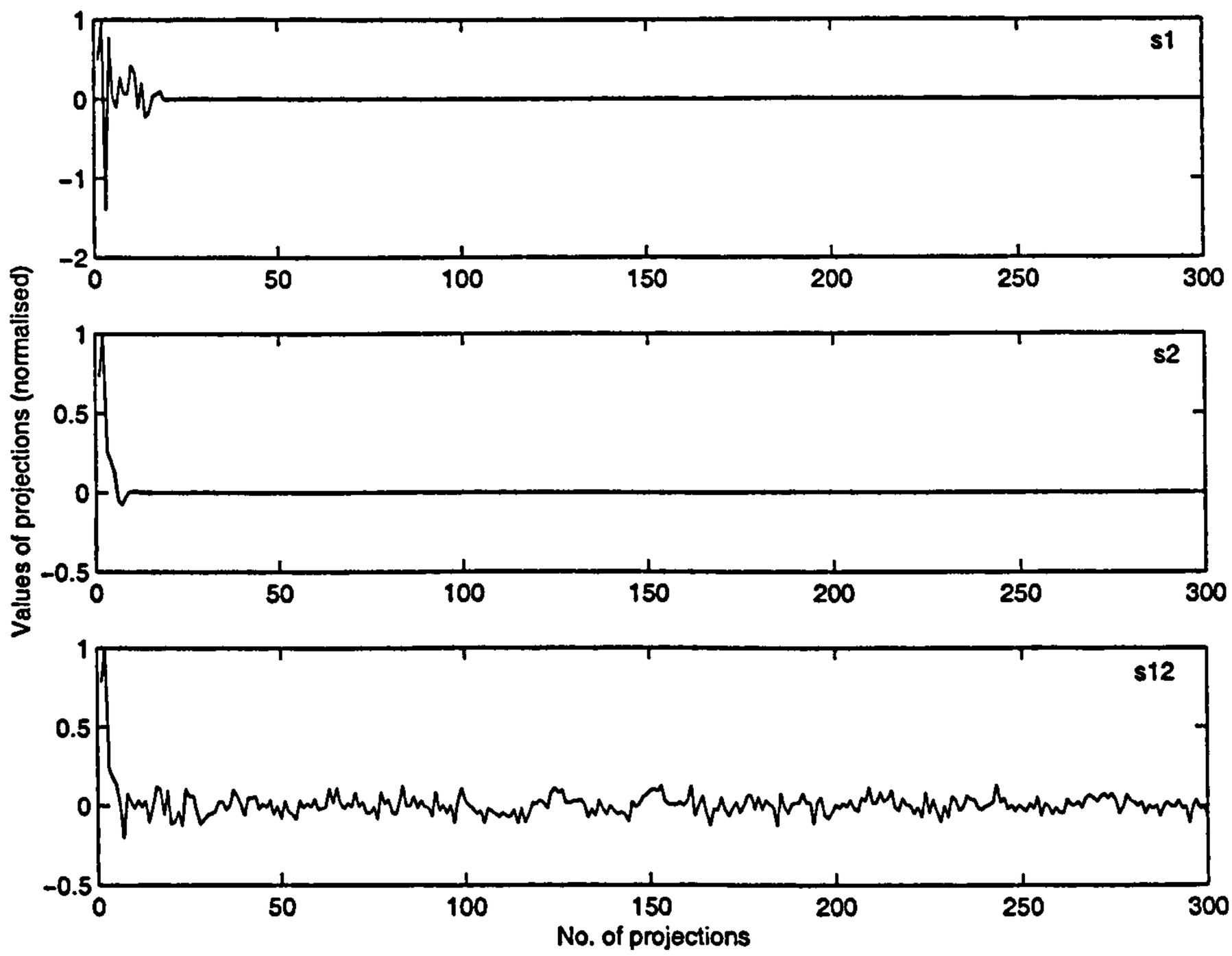


Fig 7.15 Projections of other realisations of the signals s_1, s_2 and s_{12} on the basis set Bs_1.

No longer do we see the maximum values on the first projection when the signal is projected on its own basis set. These results show that straightforward use of the values of the projections does not give always a consistent pattern for similar projections. We need to analyse the values of the projections in a different way. For that we will use the norm of those projections.

7.5 The norm of the projections

Instead of those projections we will use the norm of those values. First we will define which norm we are using for the measure of the presence of a note in a simulated signal.

The measure used to identify the note using the projection values is the norm of those values. The Euclidean norm of those projections is used as a measure of the presence of those notes in the piano signals. The Euclidean norm is defined as:

$$\|X\|_2 = \left(\sum_i |x_i|^2 \right)^{1/2} \tag{7.2}$$

The Euclidean norm of a signal projected on a set of orthonormal singular vectors remains invariant. Considering s a signal that is projected on the orthonormal basis set V, we have:

$$\|s.V\|_2 = \|s\|_2 \tag{7.3}$$

Only the value of the norm of the projection of the signal s on the space with basis set V does not give information about the signal's presence since it remains invariant. But we are projecting the signal in a space with 300 vectors of the new space. We can analyse how the norm varies when we project the signal on the first vector, on the first and second vectors and so on until it is projected on all vectors. This way we can see how the energy of the signal is distributed along the vectors of the space. If a signal belongs to the space into which is projected we expect that the norm values reach practically the maximum value (the norm of the signal) in the first few projections. For a possible comparison between projections of different signals we divide all the norm values by the Euclidean norm of the signal. This means that the norm of a signal projected in the new space (using all the vectors of the base) will be one.

7.5.1The norm of the projections of a simulated signal in a SVD basis set

The norm of the projection of signals was calculated for the signals of section 7.4. We determined the norm for signals s_1, s_2, s_{12} projected on basis Bs_1 and Bs_2 (fig.7.16).

The values of the projection of the three signals on Bs_1 are very similar but the s_1 norm values reach the maximum value (unity) faster, next is signal s_{12} and finally s_2. The norms of the signals projected on Bs_2 show now curves with distinguishing values. The octave s_{12} has its values between the projections of s_1 and s_2. Other norm curves for other possible events of those signals projected on Bs_1 and Bs_2 are similar with no significant differences as occurs for the values of the projections itself. These two graphs of the norm give enough information to separate and recognise the presence of an octave in the simulated signal.

In next sections we will analyse signals with more frequency components.

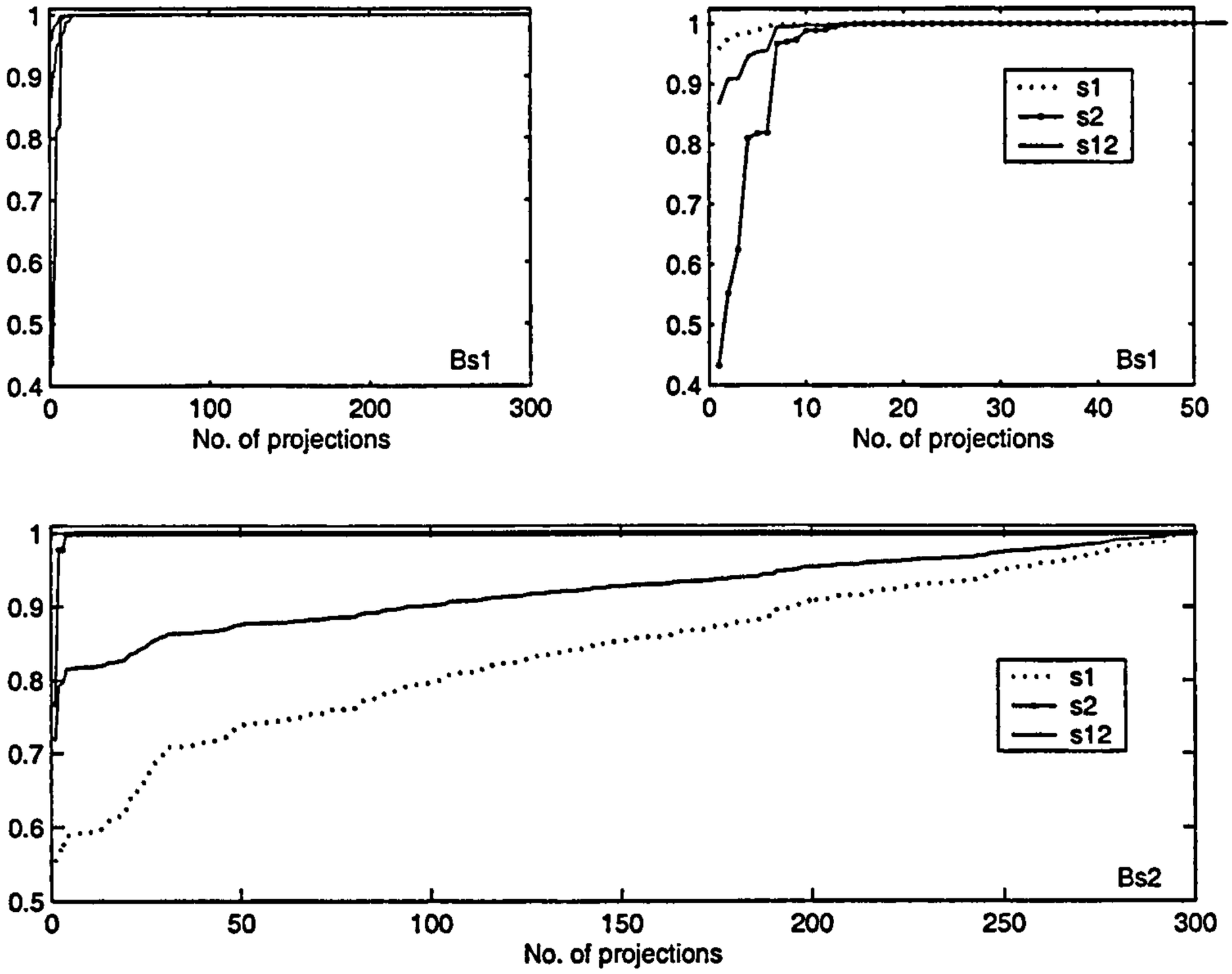


Fig 7.16 Norm values ('normalised') of the projections of signals s_1, s_2 and s_{12} on the basis sets Bs_1 (up left - all the projections, up right – a detail) and Bs_2.

7.6 Recognition of single notes with more than one frequency component

Real piano notes, has we have seen some examples in chapter 3, have several harmonics. We already simulated notes with two frequency components. In this example we simulate notes with five harmonics and consider whether of the recognition process used for single notes and octaves with one and two frequency components works for more complex signals.

We generate three signals s_1, s_2 and s_{12} again using equation (7.1). Signal s_1 will have 5 frequency components and s_2 has 2. Signal s_{12} is the sum of those two signals corresponding to a simulated octave. The parameters of the signals are represented in table 7.3.

	Frequency interval (Hz)	Centre frequency (Hz)	Amplitude interval	Damping interval
s_1 (5 frequency components)	[9.71,10.29]	10	[0.75,1]	[0.1,0.3]
	[19.71,20.29]	20	[0.5,0.75]	[0.12,0.32]
	[29.57,30.43]	30	[0.4,0.65]	[0.125,0.325]
	[39.43,40.58]	40	[0.3,0.55]	[0.32,0.12]
	[49.29 50.720]	50	[0.25,0.5]	[0.125,0.325]
s_2 (2 freq. comp.)	[19.43,21]	20	[0.75,1]	[0.1,0.3]
	[19.71,20.29]	40	[0.5,0.75]	[0.12,0.32]

Table 7.3 Intervals of variation of the components of the signals s_1 and s_2.

The two frequency components of the signals s_2 have the same centre frequency as the 1st and 3rd harmonics of the signal s_1. As in section 7.4 we introduce a delay between 0 and 5 samples for each signal. The amplitude of the each frequency components will decrease for higher order number of the simulated "harmonic". With these intervals of variation for each harmonic we generate signals whose components are slightly inharmonic.

The Fourier transform of those signals show that the signal s_{12} (simulated octave) has the same frequency components of the signal s_1. From those graphs it is not possible to distinguish from the signal s_1 and the octave.

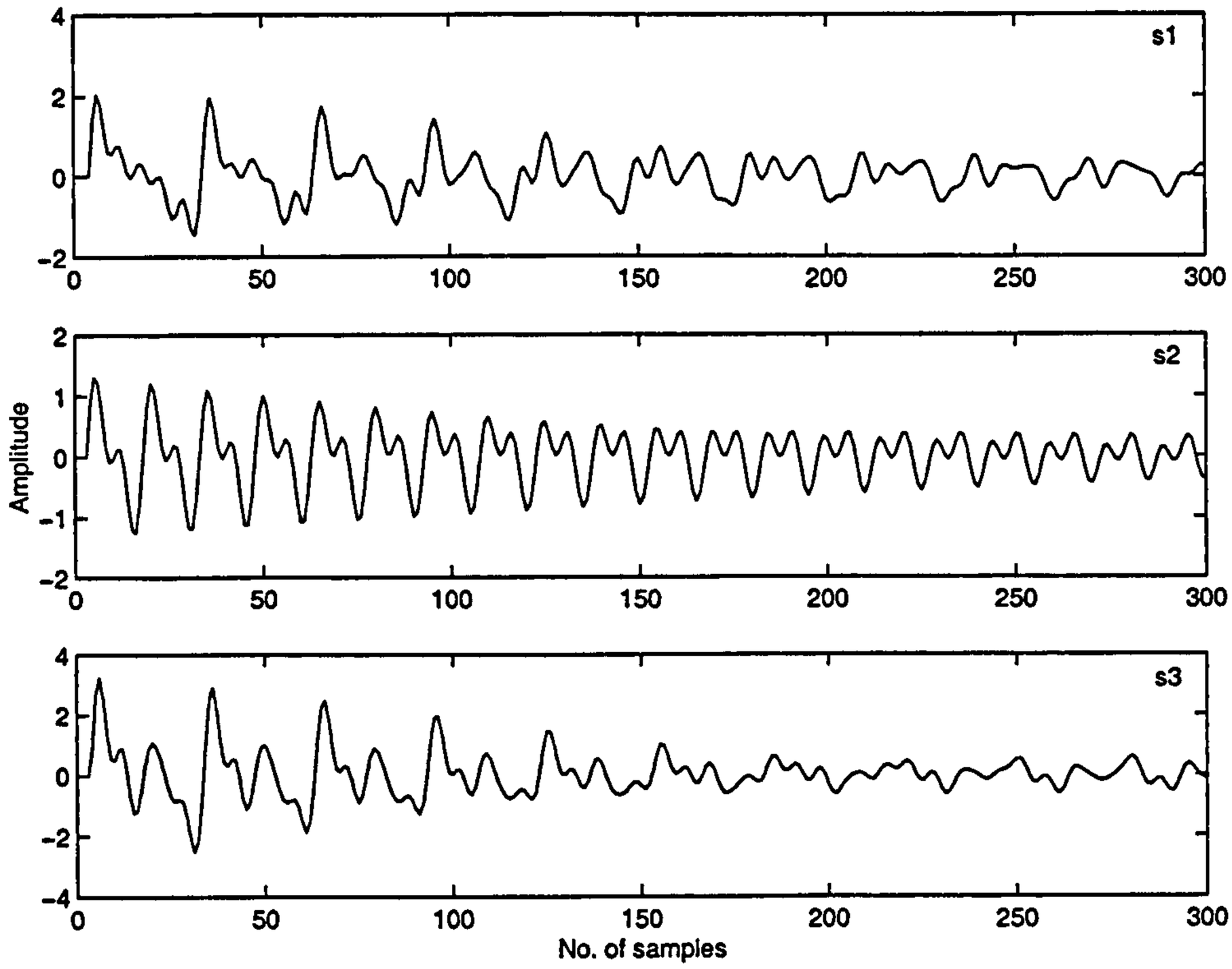


Fig 7.17 Time history of signals s_1 (5 frequency components), s_2 (2 frequency components), s_{12} ($s_1 + s_2$, 5 frequency components).

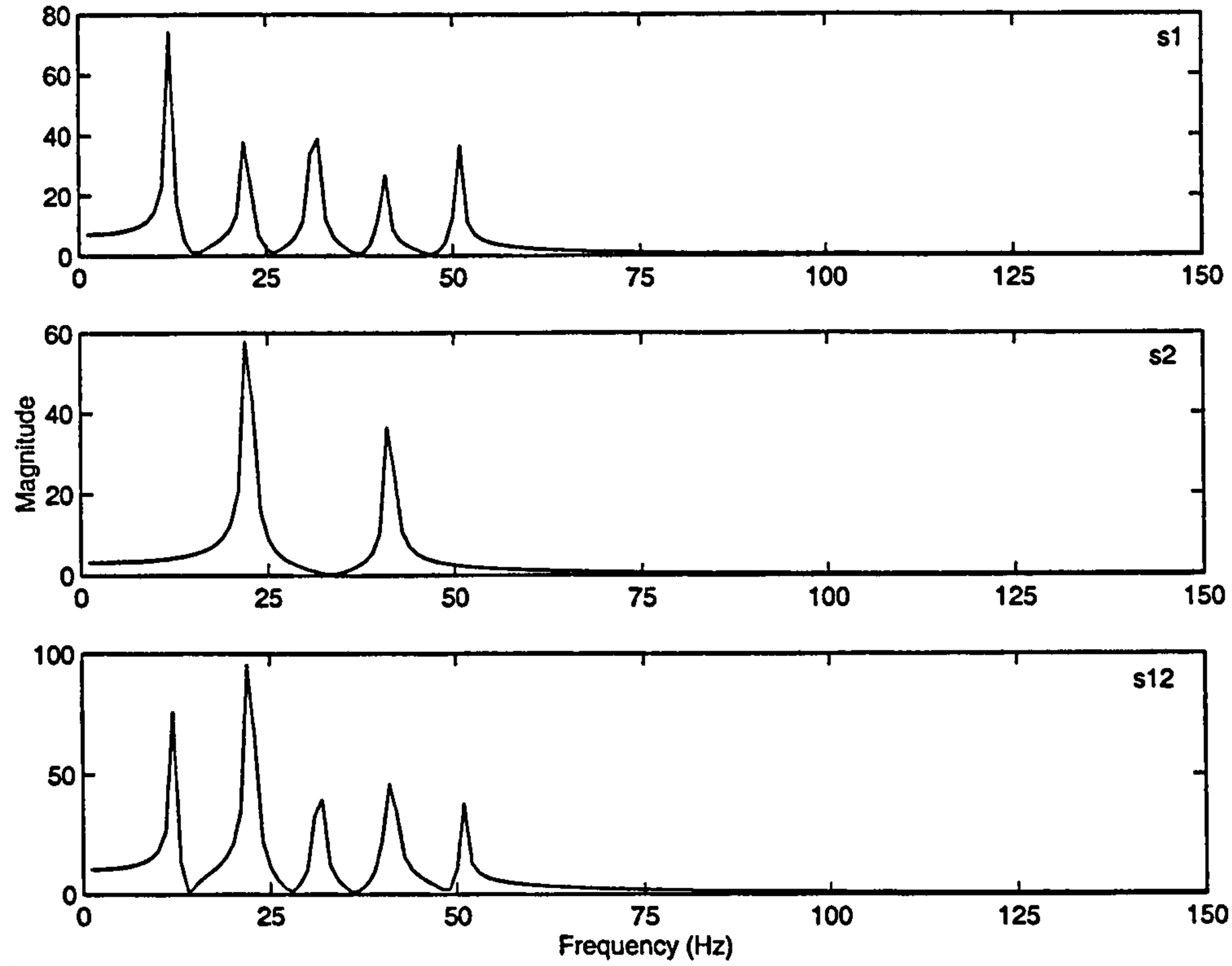


Fig 7.18 Fourier transform of signals s_1, s_2 and s_{12} (FT of 300 points)

We determined the SVD basis sets of signals s_1 and s_2. The number of nonzero singular values of s_1 is around 60 (we have 100 possible events and 60 of them are linearly independent – the number of harmonic components increases the number of independent events) and for s_2, 25 (fig. 7.20 and 7.21). The introduction of the harmonics on signals increases the number of positive singular values meaning that the energy of this new space is spread for a greater number of components.

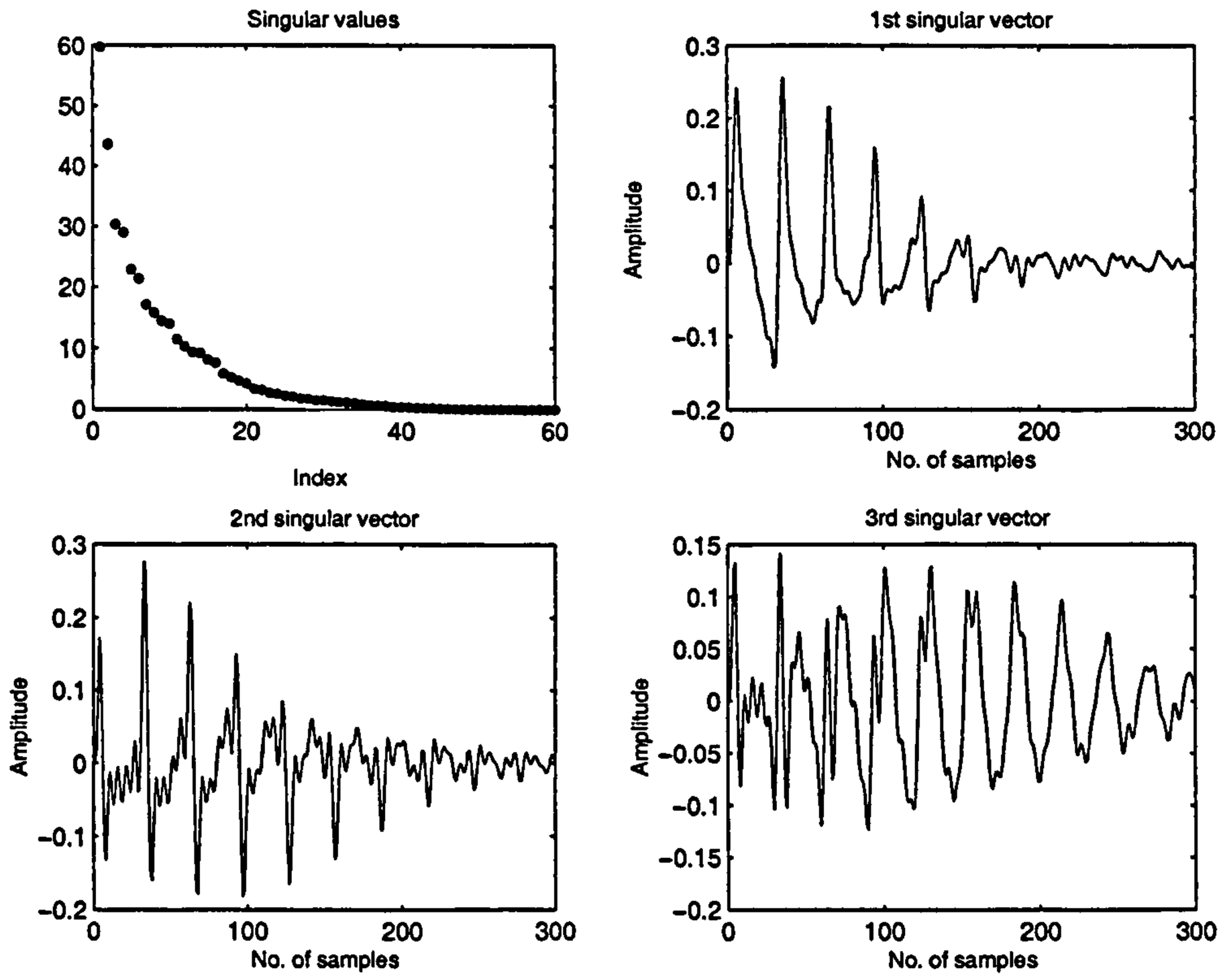


Fig 7.19 Singular values and vectors of signal s_1 (5 frequency components)

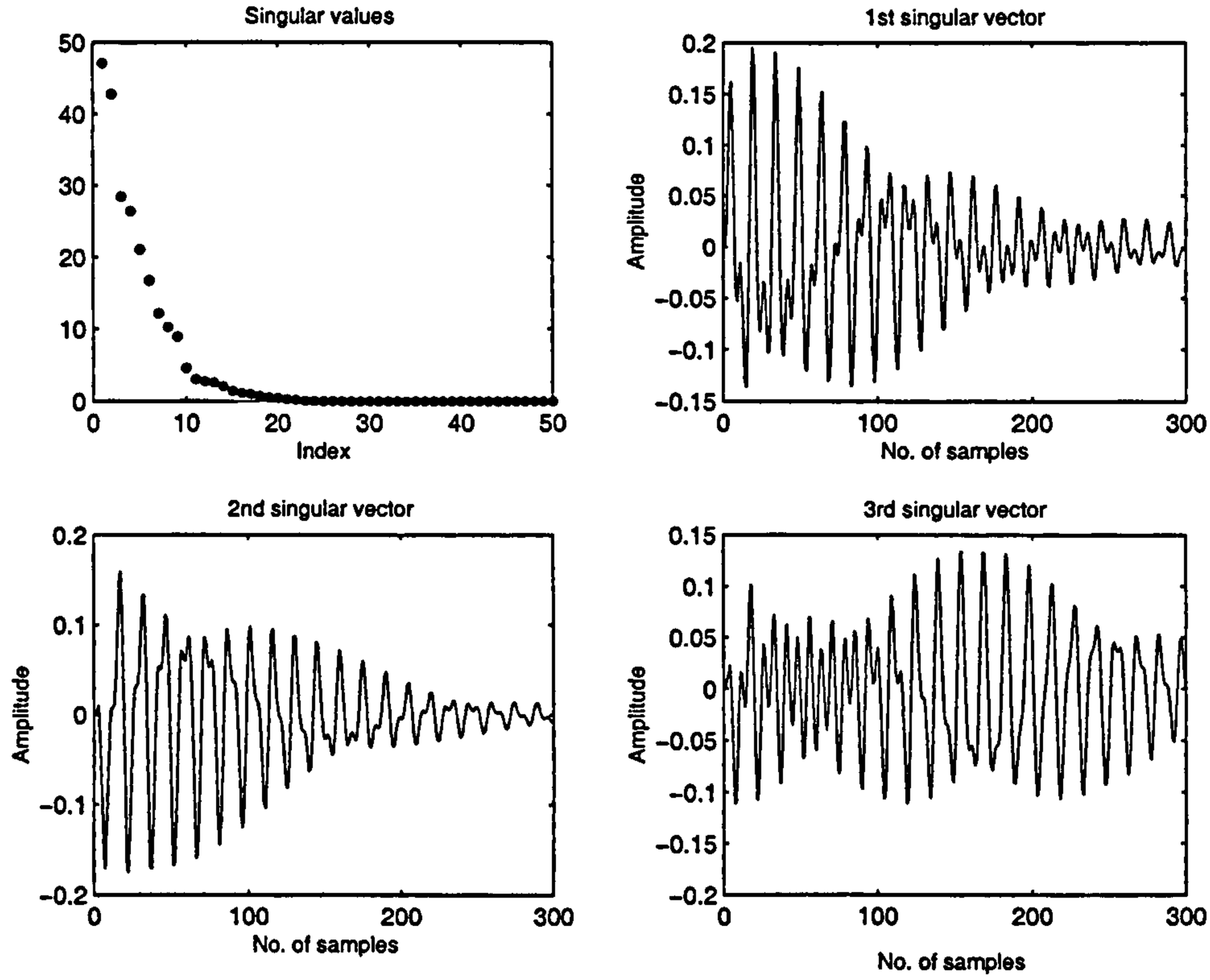


Fig 7.20 Singular values and vectors of signal s_2 (2 frequency components)

The first two singular vectors of the signals s_1 and s_2 have the same number of frequency components than the original signals s_1 and s_2 (5 components for s_1 and 2 for s_2 in fig.7.19 and 7.21).

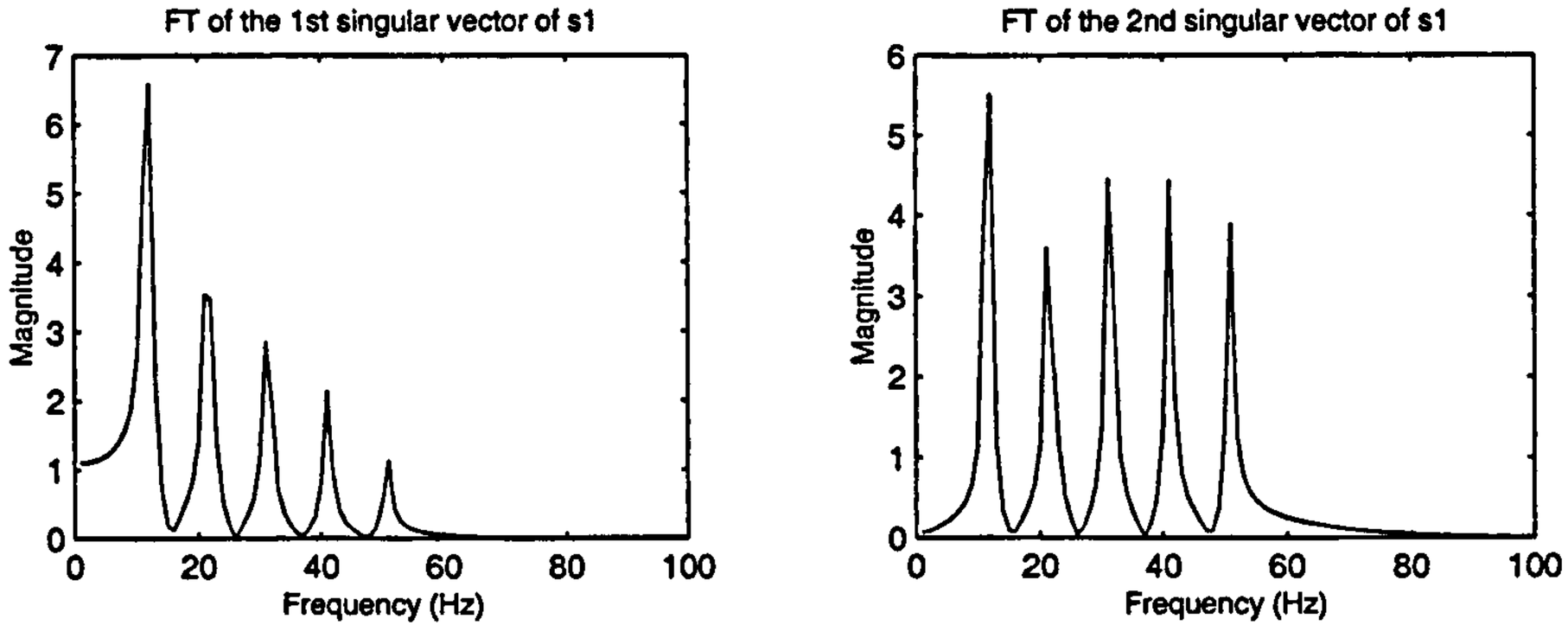


Fig 7.21 Fourier transform (300 points) of the singular vectors of signal s_1 and s_2.

The signals s_1, s_2 and s_{12} were projected on those basis sets: Bs_1 and Bs_2 (figs. 7.22 and 7.23). When the signal is projected on its own basis set we continue to have the more significant values on the first projections and after a certain number of projections (greater in the case of the signal with more frequency components s_1) the values become zero. The projections of the octave on Bs_1 are similar to the ones of s_1 and it is not possible to know exactly which one corresponds to which signal. For the case of the octave it seems that there are a few more projections with values different from zero (we remember that the two signals have both the same frequency components). When the octave is projected on Bs_2 the first projections have a similarity with the ones of s_2. After these values the projections seem to take a random pattern. In the projection of the octave on Bs_1 this method seem that is unable to detect the octave because the projections are quite similar. But when this two signals s_1 and s_{12} are projected on Bs_2 those signals are separated since the have very different projection pattern.

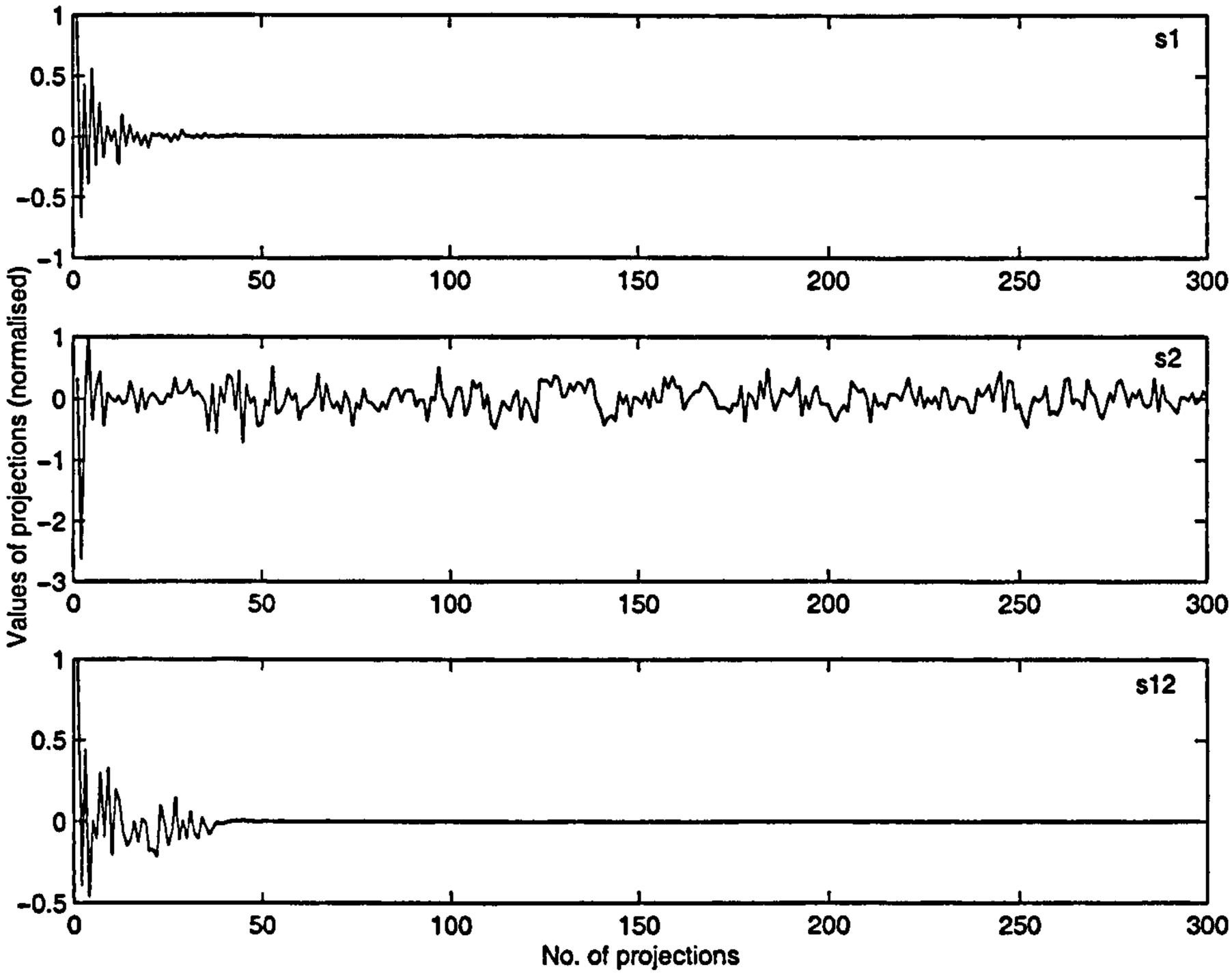


Fig 7.22 Projections of signals s_1, s_2, s_{12} on the basis set Bs_1.

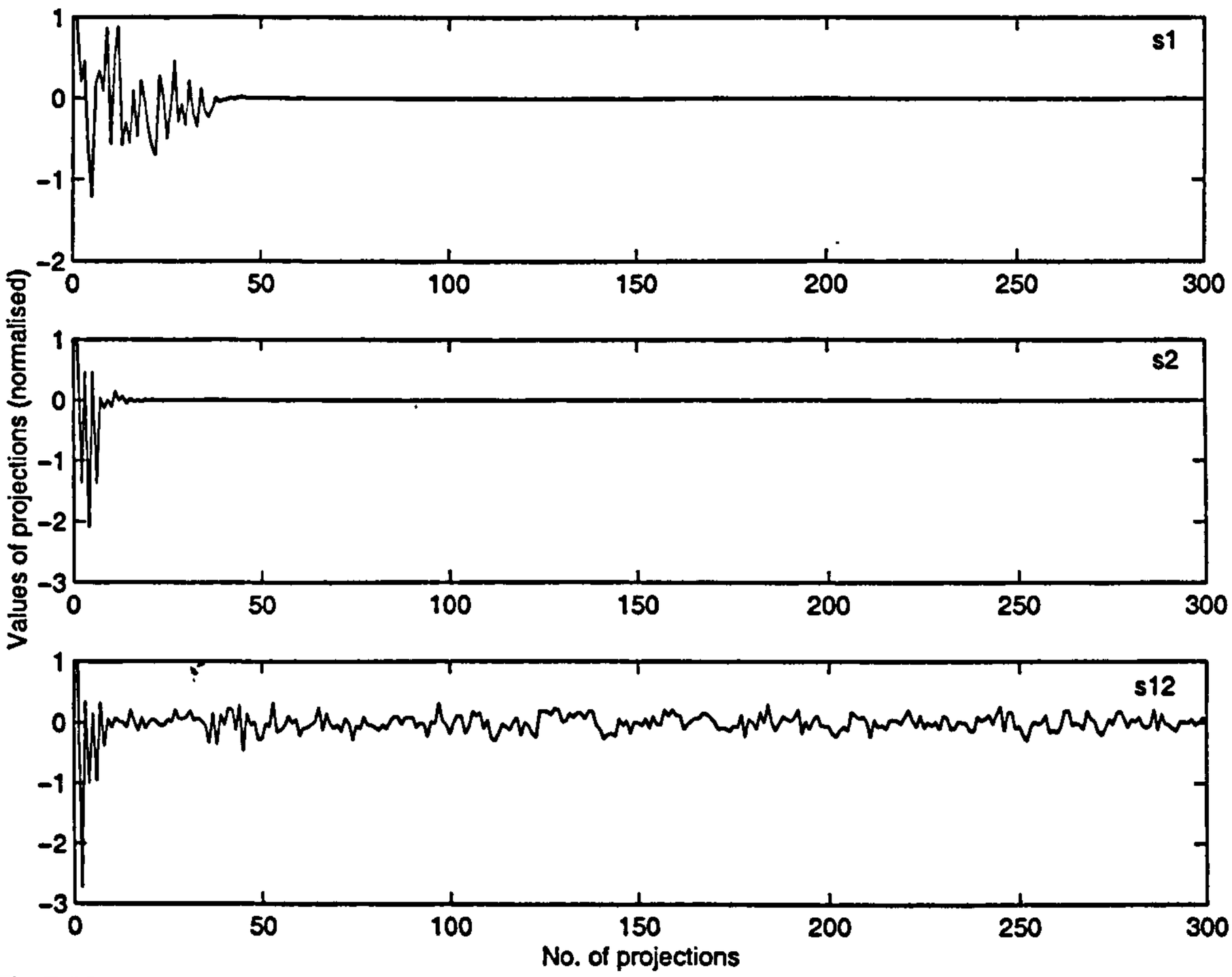


Fig 7.23 Projections of signals s_1, s_2, s_{12} on the basis set Bs_2.

The norm values for those projections can give more information about these signals.

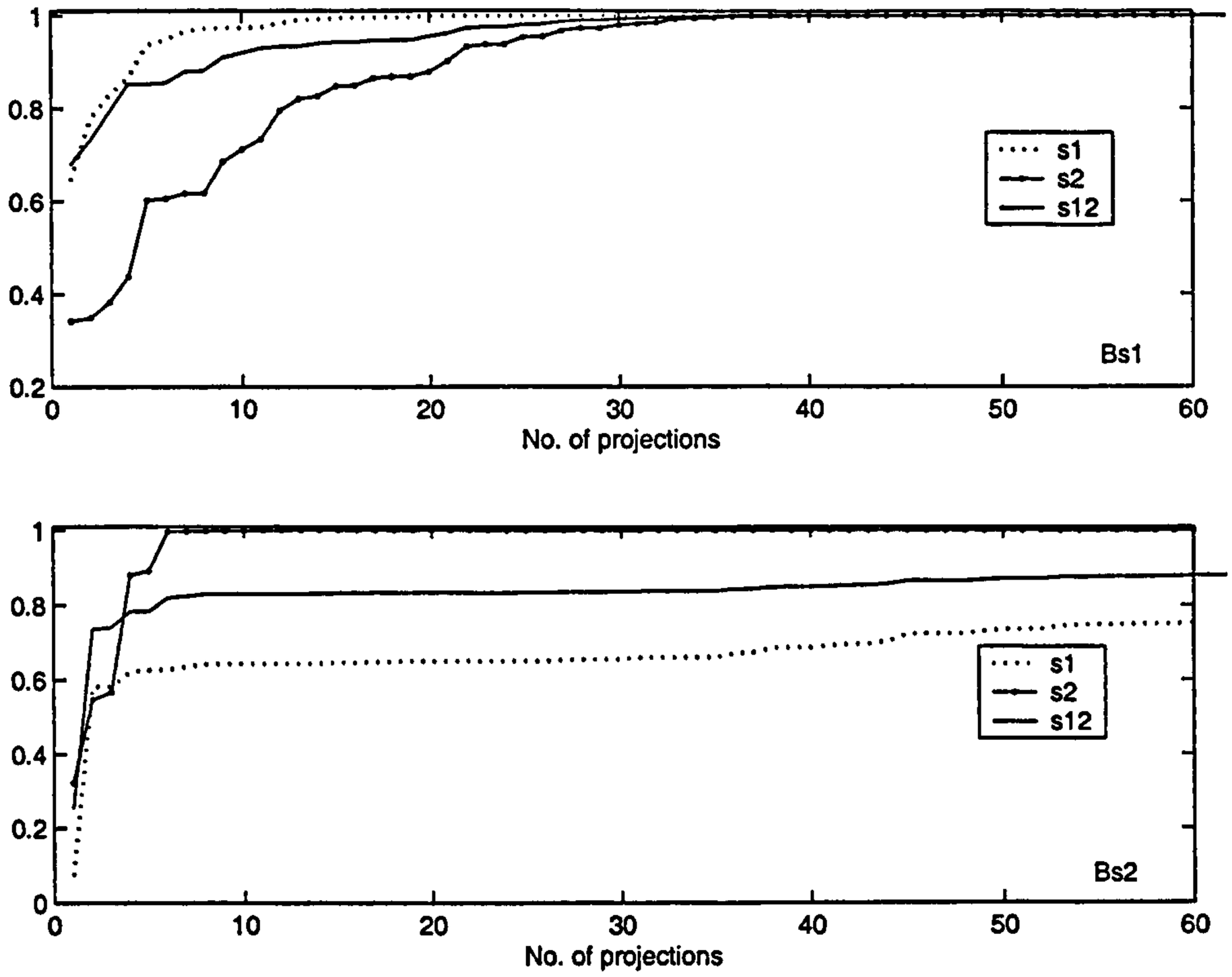


Fig 7.24 Norm of the projections of signal s_1, s_2 and s_{12} on the basis sets Bs_1 and Bs_2.

The norm of the signals s_1 when projected on basis Bs_1 reaches the maximum value (unity) before the other norms of s_2 and s_{12} (fig.7.24 Bs_1). After the 35th projection the norms are all in the maximum value. Both s_2 and s_{12} reach that value at the same projection but the octave always has higher values than s_2. These three curves indicate that there is a difference between the signals, although small, and it is possible to recognise the octave from the signal s_1. The projection of the same signals on Bs_2 shows three distinct curves. The norm of signal s_2 reaches the maximum value first on the 6th projection. The other norm curves only reach the maximum value on the last projections. The norm values of the octave are always above the ones of signal s_1. The values for the octave are always between the values for s_1 and s_2. This leads to the recognition of the octave and it is possible to distinguish between the octave, s_{12}, and signal s_1 that have the same frequency components.

This recognition process is able to recognise the presence of octaves even for more complex simulated signals with several frequency components. We analysed the performance of the method when a set of different single simulated notes are projected on the basis set of a signal. The basis set used to project the notes was the same Bs_1 (we remember this corresponds to a set of signals with the fundamental frequency drawn from an interval with centre frequency of 10Hz). A set of signals with 5 harmonics each were generated with centre frequencies related to the centre frequency of s_1 by musical half tones (from 10 to 20 Hz):

$$f_{c_i} = 2^{(i/12)}.10 \qquad i = 0,1,...,12 \tag{7.4}$$

The resulting norm values of the projections show a distinguishable curve that reaches the maximum value before any other curves (fig.7.25). That curve corresponds to the projection of a note similar to s_1: the fundamental frequency falls inside the interval of frequencies of the signal s_1 and it has the same number of harmonics.

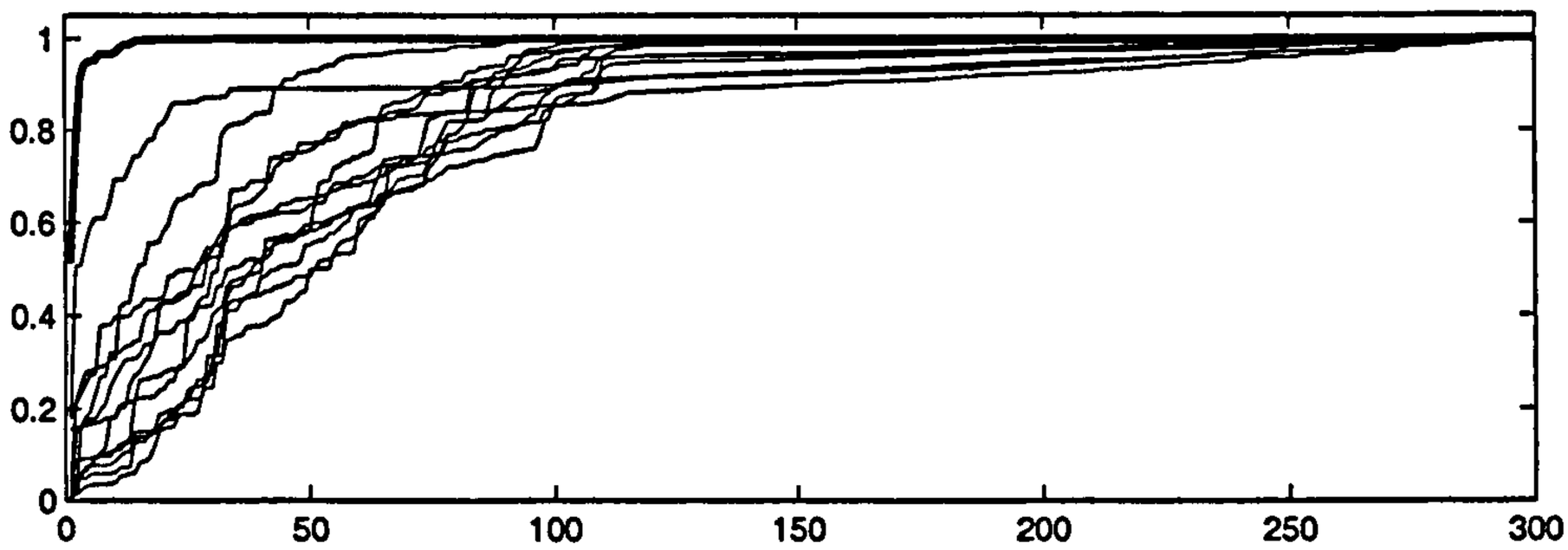


Fig 7.25 Norm of the projections of signals separated by musical half tones (from 10 to 20Hz) on the basis sets Bs_1.

7.7 Conclusions

The recognition process here presented is able to recognise simulated notes and to identify an octave. The SVD is applied to an ensemble of realisations of a signal (the same type of signal is defined by the centre frequency). The resulting basis sets are the "reference" of that note. To find if a certain note is present in a signal we project that signal on the different basis sets representing different simulated notes. The projections themselves self can give some information about the presence of the note on the signal when there is a maximum value on the first projection (on the most energetic singular vector). That does not always happen and we need to determine the norm of those projections to have more information. The norm of a signal when projected on its own basis set reaches the final value before any other signal. A very important result was that the identification of an octave was possible using this method.

PAGES

MISSING

IN

ORIGINAL

CHAPTER 8

APPLICATION OF SVD AND A RECOGNITION PROCESS TO REAL PIANO NOTES

Introduction

In chapter 7 we saw that it is possible to recognise simulated single piano notes (and octaves) when treated as stochastic transients. In this chapter we will analyse real music notes recorded from a piano. The recognition process for real piano notes and chords will be very similar to the one of the simulation case studied in the previous chapter. The performance of the method will be assembled in detail.

8.1 The recognition process

The recognition process for a real piano, based on the process for simulated notes, notes has the following steps:

1- Data acquisition: recording of the possible events of each piano note.

2- Determination of the singular vectors of each piano note.

3- Projection of the musical piano signals on the basis sets defined by the singular vectors of each note.
4- Detection of the presence of the single piano notes in the musical signal.

We will test this recognition process with single piano notes, two single notes played simultaneously (e.g. the octave) and more than two notes (chords). We seek to identify the notes presents in piano signals.

1- Data acquisition: recording of the possible events of each piano note.

The general idea for the recognition process applied to real piano notes (using SVD) is to have all the piano notes (88 notes corresponding to the 88 keys) represented by singular vectors. These singular vectors would be the reference of each piano note. All the piano signals for recognition would be projected on those sets of singular vectors to detect which notes are present in the signal.

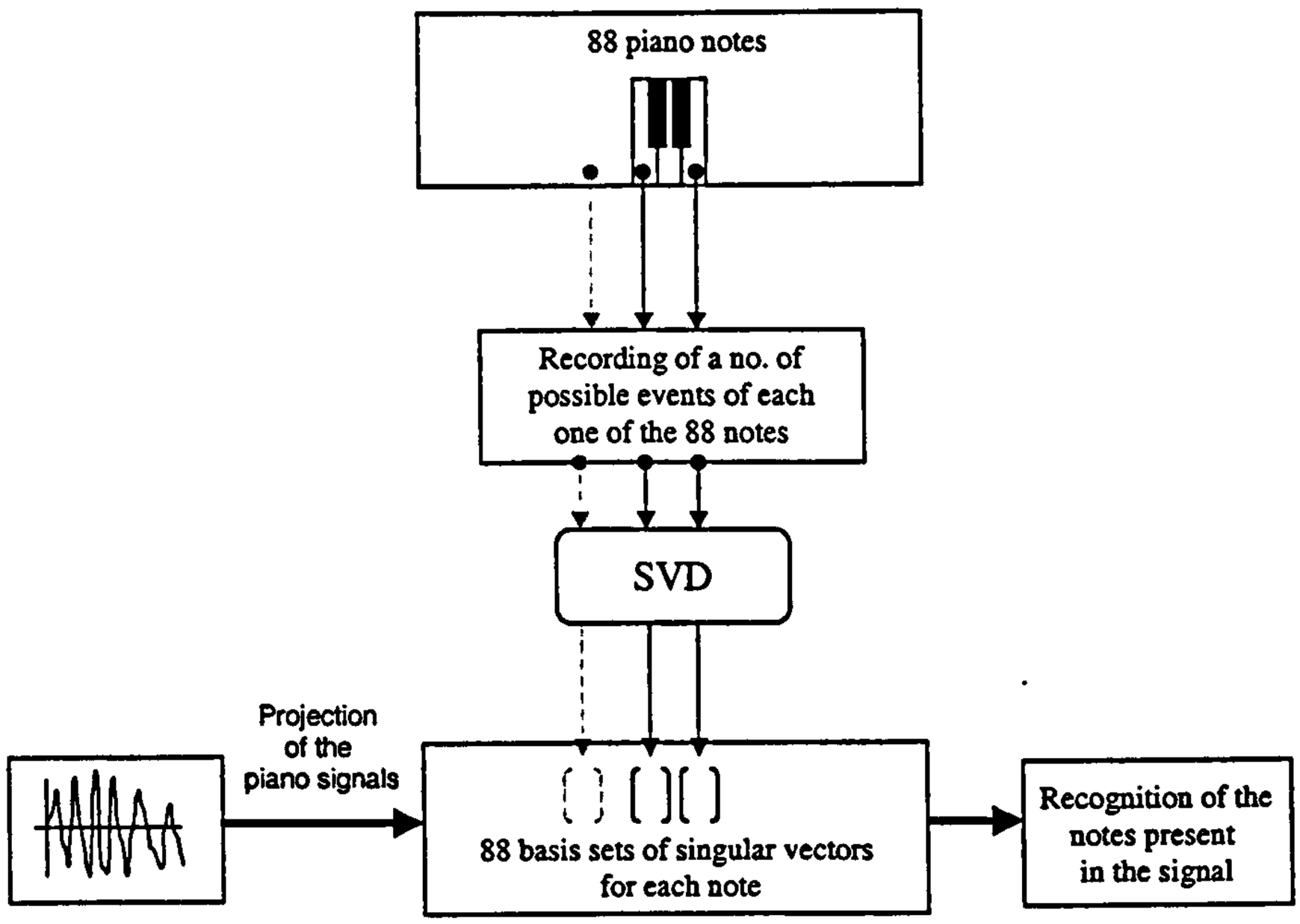


Fig 8.1 Structure of the recognition process for real piano notes

The recognition process, that includes SVD as a transform for data treatment of the musical signals, has to include a priori information about the single piano notes that will be the reference for the detection of those notes in any piano signal. For that we need to have an ensemble of possible realisations of each note that will form the time history matrix. SVD is then applied to those time histories and the resulting singular vectors of the decomposition will be the "singular notes" of each piano note that will be used as a reference for future piano signals in the recognition process.

The acquisition of real piano notes is described in annex 7. In order to apply SVD to real piano notes we recorded several events of different notes. These recorded notes give the a priori information about the signals that is needed for the recognition process using SVD.

For each note we recorded a set of realisations played with different dynamics. To cover a larger number of possible future events each note was played from *piano* (*p*) to *fortissimo* (*ff*), and with both pedals, *una corda* and *sostenuto*. We established 60 events for each note to cover all the range of different dynamics and the use of both pedals for each note. The number of samples taken from each event took into consideration the number of periods for the recorded notes and the computation time. The number of 2000 samples per event is a compromise those two aspects. These 2000 samples correspond approximately to 0.045s of the signal (the sampling frequency used in the recordings was $f_s = 44100\,\text{Hz}$). The 0.045s corresponds to a frequency of 22.05Hz. The piano note with a lower fundamental frequency analysed was G98Hz which has approximately 4 periods in 2000 samples. This value of 0.045 milliseconds of a musical signal is enough for the human ear to perceive the pitch of a note played (Gelfand, 1981).

The time histories of each note have 60 possible realisations with 2000 samples each. The analysis of this data will indicate if these dimensions are suitable for the recognition process.

2- Determination of the singular vectors of each piano note

With all the events recorded and established (the number of events per note and the number of samples per event) we build the time history matrices. The initial

conditions of the events are very important for the final result of the SVD. Previously it was verified that that the delay in the simulation is a very important variable that should be introduced in each event of a note, to avoid wrong identification of notes due to small errors in the detection of the beginning of the musical signal. The delays of the event must fall inside an interval that must not exceed a certain value. The size of the interval should not be greater than the time that the ear starts to discriminate in time between two notes (approximately 2ms (Ronken, 1970) which corresponds approximately to 88 samples for $f_s = 44100\,\text{Hz}$). The size of the interval that we choose is 60 samples that correspond to approximately 0.0016 ms. This value corresponds in some cases to 1/3 of the attack part of the signals.

SVD was applied to these time histories with delays between events for the recognition process. We also applied SVD to the time history with aligned possible realisations to determine the variability of the time history matrix of a piano note when played with different dynamics and with the two pedals.

3- Projection of the musical piano signals on the basis sets defined by the singular vectors of each note

Suppose now that we have a piano musical signal that we want to identify the notes that compose it. We will use the same process of the projections as was used in simulation. The signal is projected on the basis sets of the singular vectors of the different notes. Since the realisations in the time history matrix of a note have delays between them there is no need to align the signal to be recognised with the resulting singular vectors. We only need to ensure that the possible first sample of the signal (first sample of the attack) is inside the interval of delays (60 samples). The signal is projected initially onto all 2000 vectors of each basis set. As we will see later there is no need to use all the vectors to evaluate its presence or not in a signal.

4- Detection of the presence of the single piano notes in the musical signal.

As in chapter 7 we will not only use the direct values of the projections as for the simulation case since they do not give sufficient information. Real piano signals are complex with many harmonics that make the recognition difficult. We will use the Euclidean norm of the projections of the signal as defined in eq.(7.1). We analyse the norm values using different numbers of projections. We calculate the norm for the projection of the signal on the first vector, then on the first and second and so on until the 2000 vectors. All the values of the norm will be "normalised" by their division by the Euclidean norm of the signal. The norm of a signal projected on all the vectors of a basis set will be equal to the unity. In order to detect a note in a piano signal, according to the results for simulated signals, we expect to achieve a determined maximum value of the norm with a fewer number of projections when the note is present on the signal.

8.2 The SVD of single piano notes

We will see how the piano notes are decomposed using SVD. We will apply the SVD to the time histories of single piano notes. The decomposition will give information about the variability and the dimension of the system defined by events in the time history matrix. For this analysis we have to consider a time history where all the events are aligned at the starting point of the attack. Then we will see the SVD of a time history matrix where the events are not aligned. The resulting decomposition will be used on piano note recognition.

8.2.1The SVD of a piano note with all the events aligned

One of the important variables that can be introduced in the time history matrix is the delay between events as we saw in chapter 6. The presence of the delay between the events of the time history influences the results of the SVD.

We will see, in this section, the variability of a system when all the events of the time history of a real piano note are aligned at the starting point. This can give some information on how the variables in real piano notes vary for the same note played in different possible ways.

For this example we are going use the piano note C262Hz (see the sample time history and spectrum in fig.8.2). All the events were aligned. We used the correlation between events to align the events. In the case of real measured piano notes there is a DC offset component in the signals. As we saw in chapter 5 the first singular vector of the SVD of a signal represented by an ensemble is similar to the average of all the events. If there is a DC component on the signal that component will be present on the first vector as in the other vectors. This component is removed from the signals by extracting the mean. All the signals analysed had the mean subtracted.

To extract the mean we do an average of the set. For each event x_i we determine the mean value:

$$\overline{x_i} = \frac{1}{N}\sum_{n=1}^{N} x_i(n) \tag{8.1}$$

where N is the number of samples in each realisation. Then we subtract the mean values from each corresponding time series.

The singular values of the signal C262Hz have a curious feature when we think about the several ways that piano notes can be played. Although we played the notes with different dynamics and using both pedals, and used these events in the time history matrix, the dimension of the system or the number of linear independent rows is very small. We saw in chapter 3 that the spectra of the signals change with different dynamics: the number of harmonic components and the respective relations between the magnitudes change. Some nonharmonic components also appear for the *fortissimo* dynamics. The spectra also change with the use of the pedals *sostenuto* and *una corda*. The number of harmonics and the respective amplitude relations change for different dynamics.

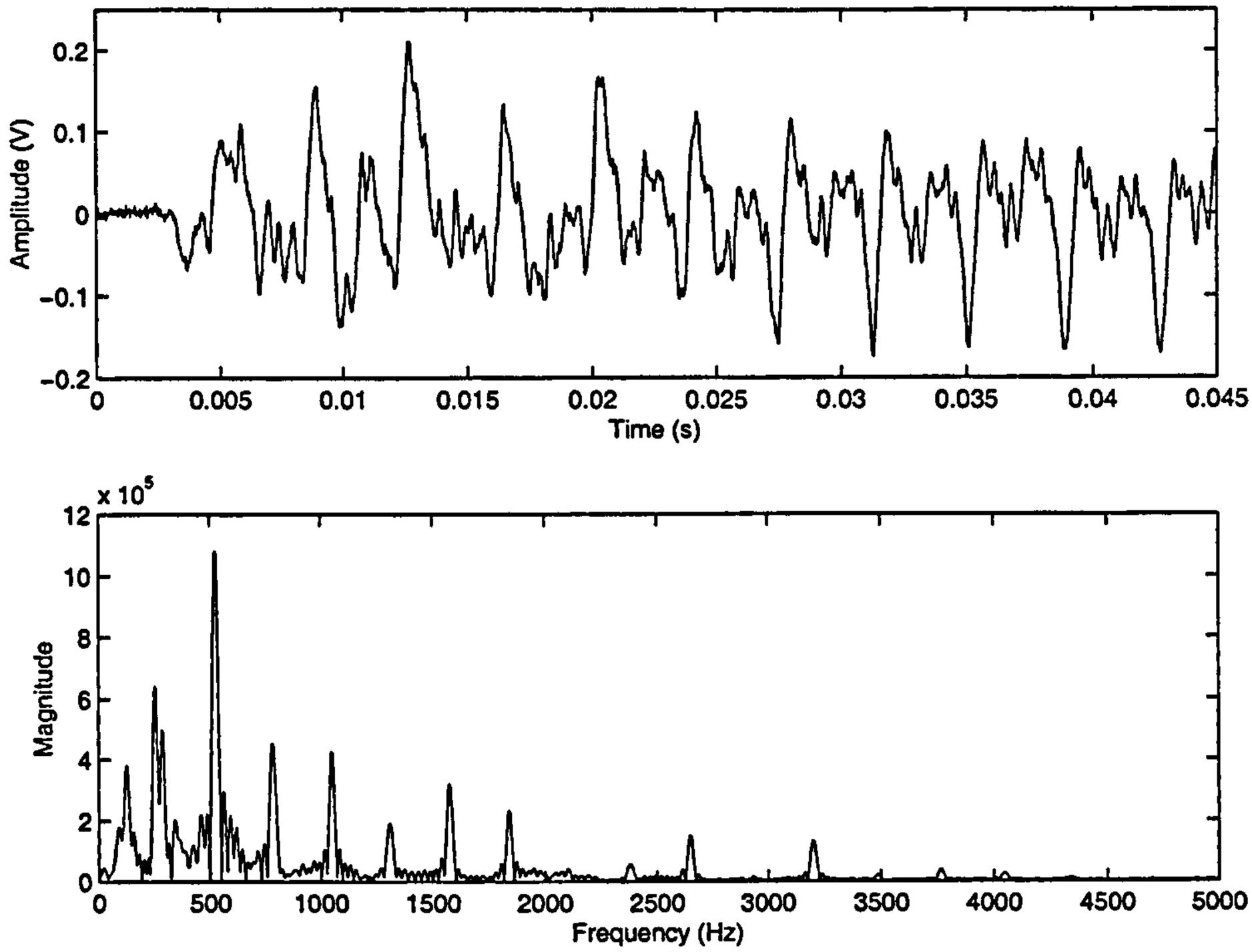


Fig 8.2 Time history and spectrum of piano note C262Hz.

Although we have this apparent variability on the piano notes, we can say that, practically, the dimension of the system is one since the first singular value has a very significant value (fig. 8.3) when compared with all the other ones (it is 7 times greater that the second value). This means that the variability of the piano note is small and although many parameters are present in the note production the time histories are more or less amplified versions of each other.

The first singular vector (fig.8.4) has a very similar shape to a realisation of the signal since the corresponding value has a very high value when compared with the others. In our case of the decomposition of nonstationary signals the greater the first value, the more similar is the shape to the realisation of the signal. In our modelled piano notes the same occurred. But the shape of second and next vectors seems to correspond to noisy signals. This is different from our model. While in the simulation cases we did not have noise in the signals in real case of the piano note we have the presence of noise. The shape of the second and next vectors and the correspondent values link them with the presence of noise. The spectra of those first four singular vectors (fig.8.5) show that the first vector is similar to a realisation. The note C262Hz has a strong frequency component at 131Hz. That same component has a very small magnitude on the first singular vector. On the second vector this component is dominant.

The information that SVD gives about piano signals, no matter how they are played, is that the time histories are quite similar resulting in a very significant singular value with correspondent vectors with a shape very similar to a realisation.

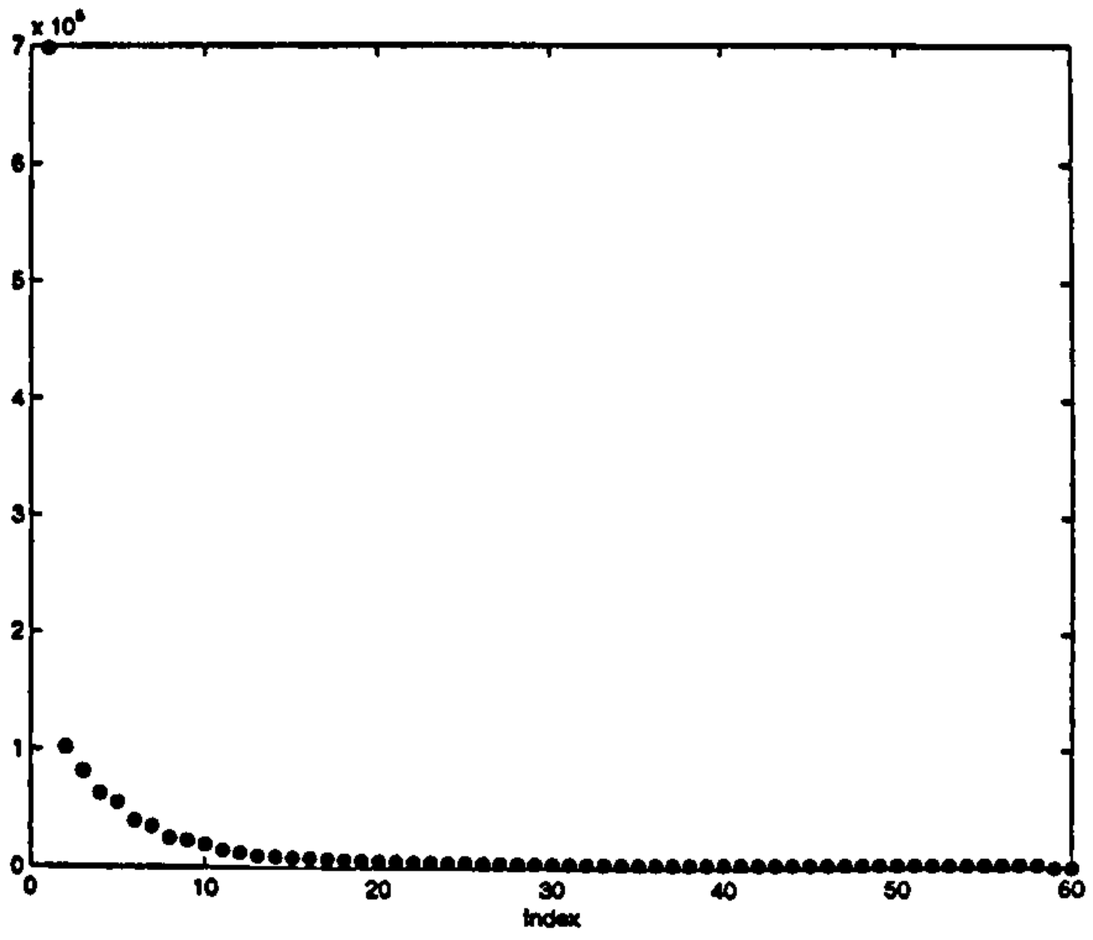


Fig 8.3 Singular values of a piano signal with all events aligned at the initial part of the attack.

The singular values are very small in relation to that first one and can be related to the presence of noise.

If we think in terms of the reconstruction of the signal the first vector is sufficient to define the original signal since most of the energy is concentrated on the correpondent singular value.

We can conclude for the chosen number of realisations and the number of samples that although the shape of the time history of the events of a real piano note changes with different dynamics applied to the signal and with the pedals *sostenuto* and *una corda*, the SVD of the matrix shows that the variability of the system is very small and can be reduced to only one dimension. Relating this with the results of the model we can conclude that when a piano note is played in several possible ways the shape of the time history continues to be very similar and the most important change is in the amplitude.

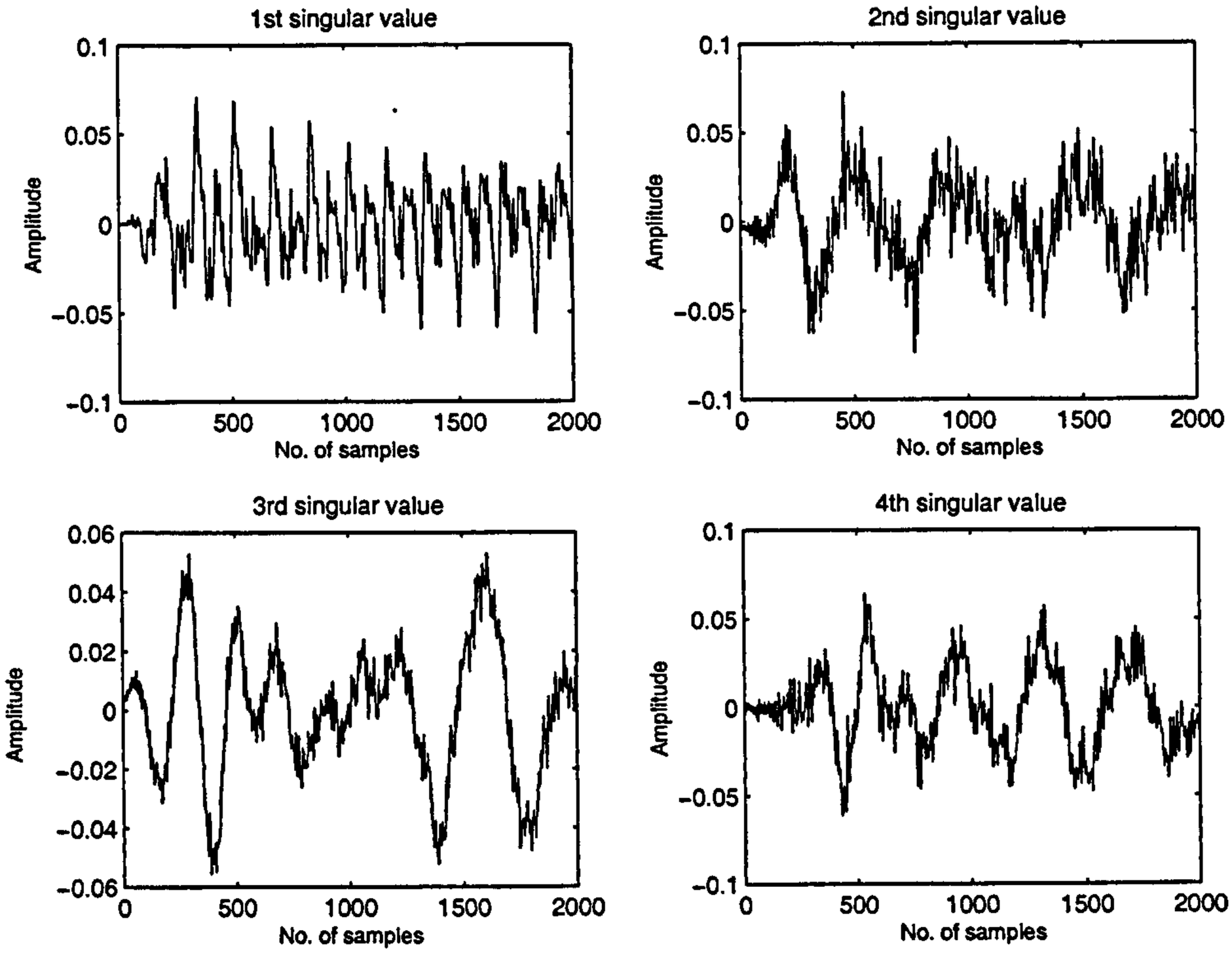


Fig 8.4 Singular vectors of a piano signal with all the events aligned at the initial part of the attack.

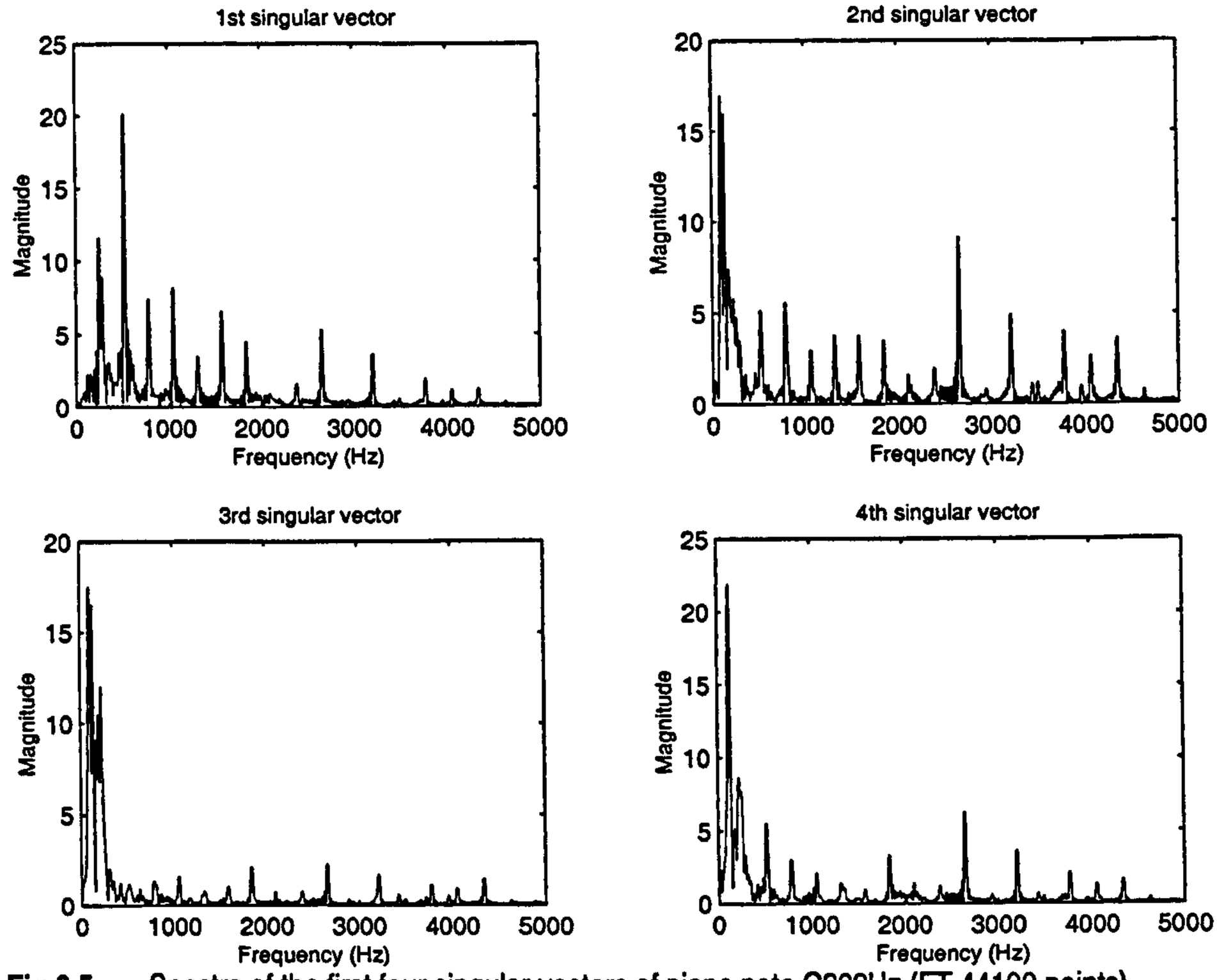


Fig 8.5 Spectra of the first four singular vectors of piano note C262Hz (FT 44100 points).

8.2.2 The SVD of a piano note with delays between events

The time history matrix, of the same piano note C262Hz was filled with the same events as in the previous section but this time they were not aligned at the starting point of the attack. This introduces a new variable in the system. Those delays could vary between 0 and 60 samples which corresponds to approximately 1 ms. If the interval of the delays is greater we can compromise the shape of the attack part of the signal in a way that the recognition can become inaccurate.

SVD shows that the dimension of the space defined by the delayed signals is now greater than for the example in the previous section. This was expected since a new variable, the delay, was included in the time history matrix. This was already shown in the analysis of the model piano note, in chapter 6. The increase of the dimension of the space, defined by the singular vectors of the signal, is reflected in the increase of the number of the singular values with significant magnitude. This means in terms of energy that there are more preferred directions by the signals in the new space due to the presence of more variables.

The first two values are detached from the rest (fig.8.6). When there is only one significant value the shape of the corresponding vector is similar to the signal itself. When there are more significant values the shape of the first vector becomes different from the signal. But those differences are not very great as we can see in fig.8.7 and 8.8. The first vector seems a smoothed version of the signal. The next vectors have a more significant contribution to the signal than the previous example without delay. The second singular value is near the first one and the shape of those vectors is quite similar These vectors have opposite orientation in relation to the zero line that corresponds to 90°. We had the same result in for the simulation when the delay was introduced. This suggests that the introduction of the delay changes the structure of the time history matrix and the resulting SVD becomes similar to a sine and cosine decomposition. The next vectors have a more significant contribution to the signal than the previous example without delay and are not noise-like.

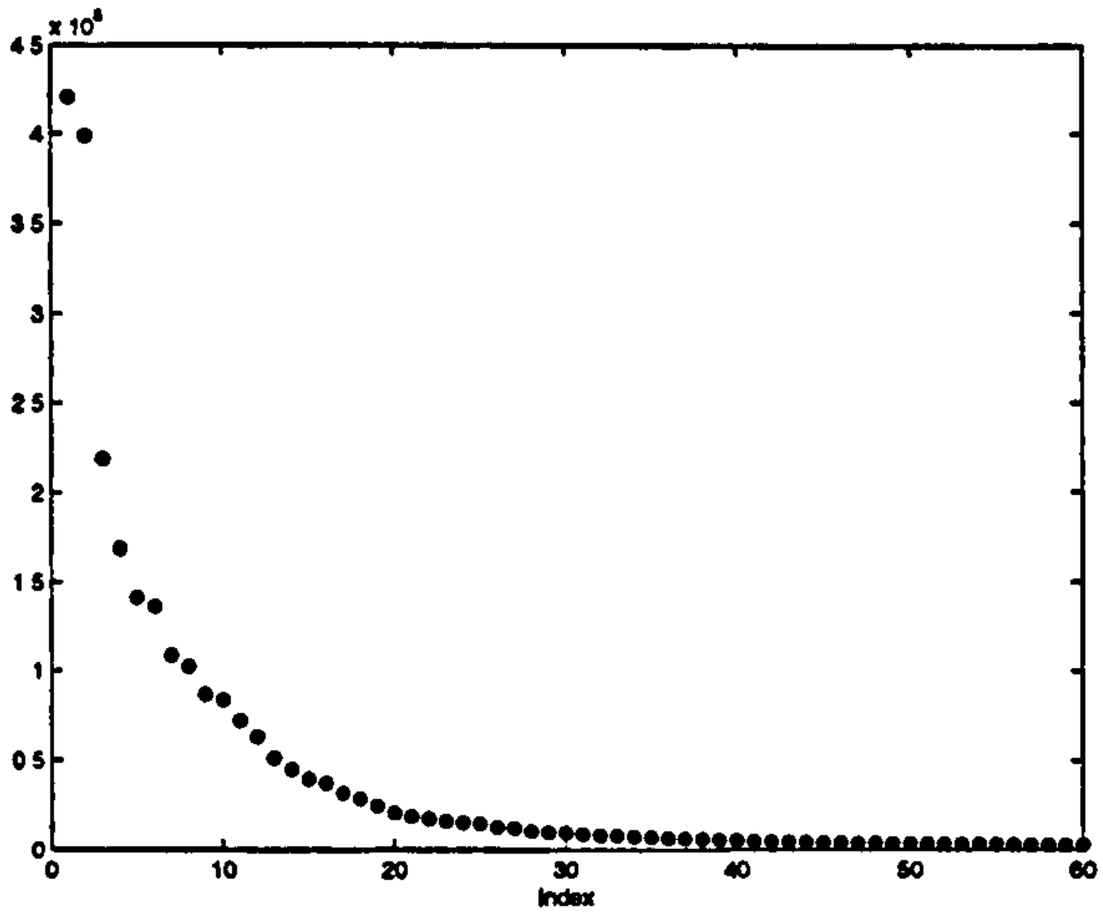


Fig 8.6 Singular values of the time history matrix of piano note C262Hz with delays between events.

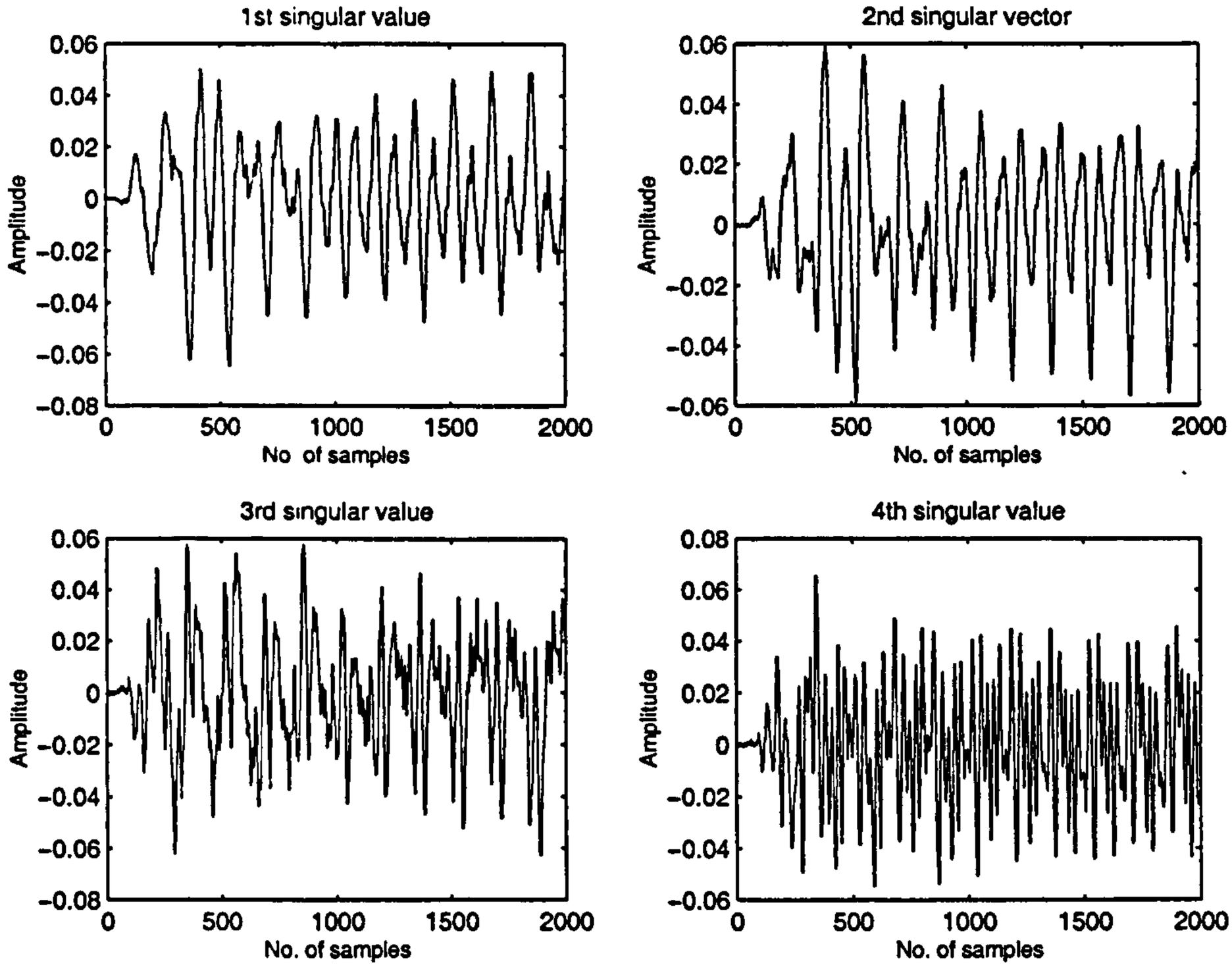


Fig 8.7 Singular vectors of the piano note C262Hz with delays between the events.

The spectrum of the first vector has the same frequency components as a realisation but the magnitude of 131Hz component is very small compared to that seen in the measurements. The frequency components are practically the same for these four vectors and what changes are their relative magnitudes. The number of higher frequency components of the first four singular vectors in this example with delays is smaller than for the previous case where all the events were aligned.

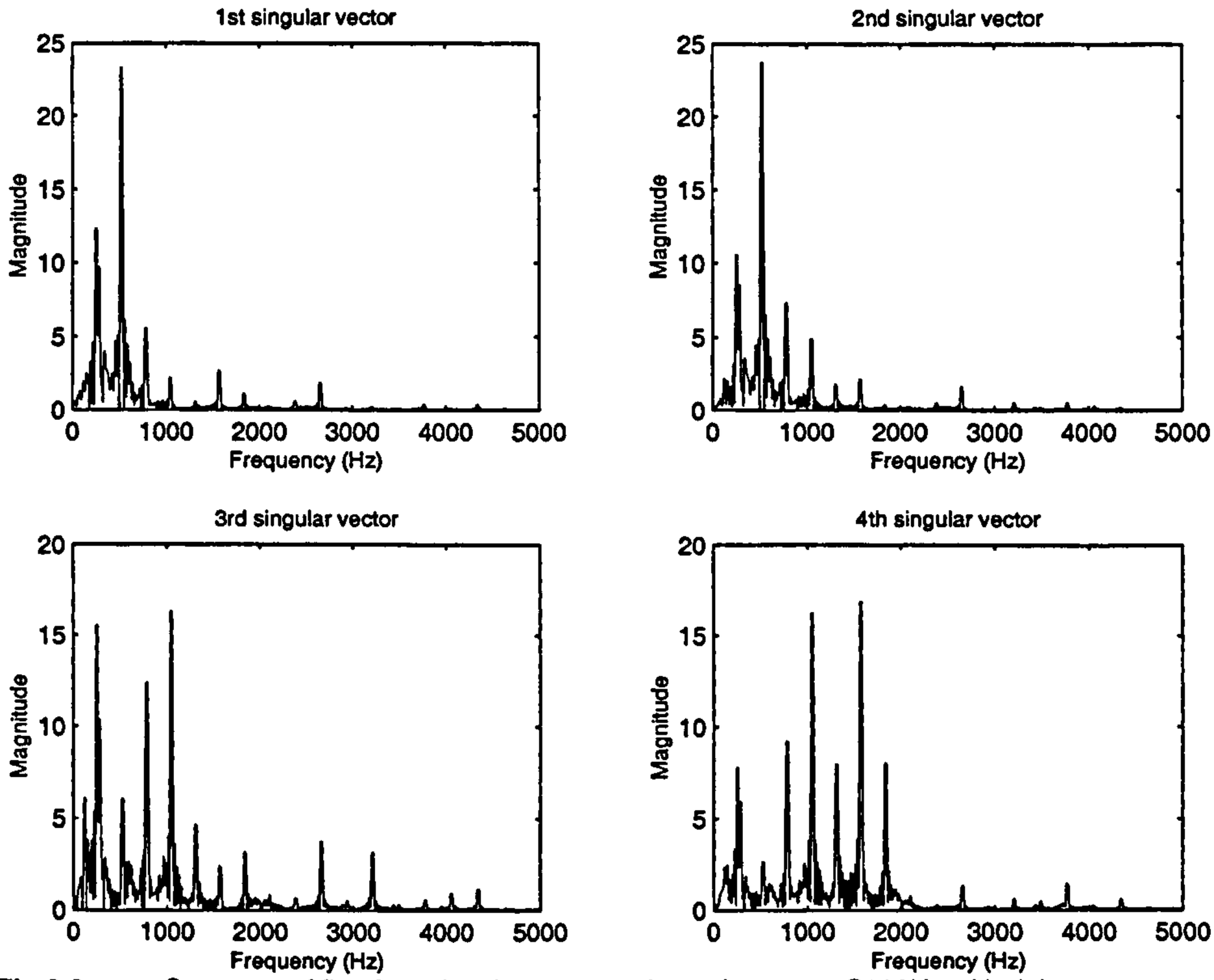


Fig 8.8 Spectrum of first four singular vectors of the piano note C262Hz with delays between the events. (FT 44200 points)

As for the simulation case we will use the SVD of delayed events of a single piano note for the recognition process. This way the detection of the exact starting point of a note for recognition and the exact alignement with the singular vectors for projection is not a concern. It is enough to have a value for the starting point that can have some error that fits inside of the delay interval of the events on the time history matrix.

8.3 Recognition of a single note

We saw previously the SVD of a piano note C262Hz. We will use the same note for the recognition process. The procedures are the same used in chapter 7 for simulated notes.

The piano notes for which we determined the SVD and used as the reference notes for identification are represented in the next figure. For each we took 60

realisations and determined the right singular vectors. For an easiest visualisation of the notes we represent them as the corresponding keys of the piano.

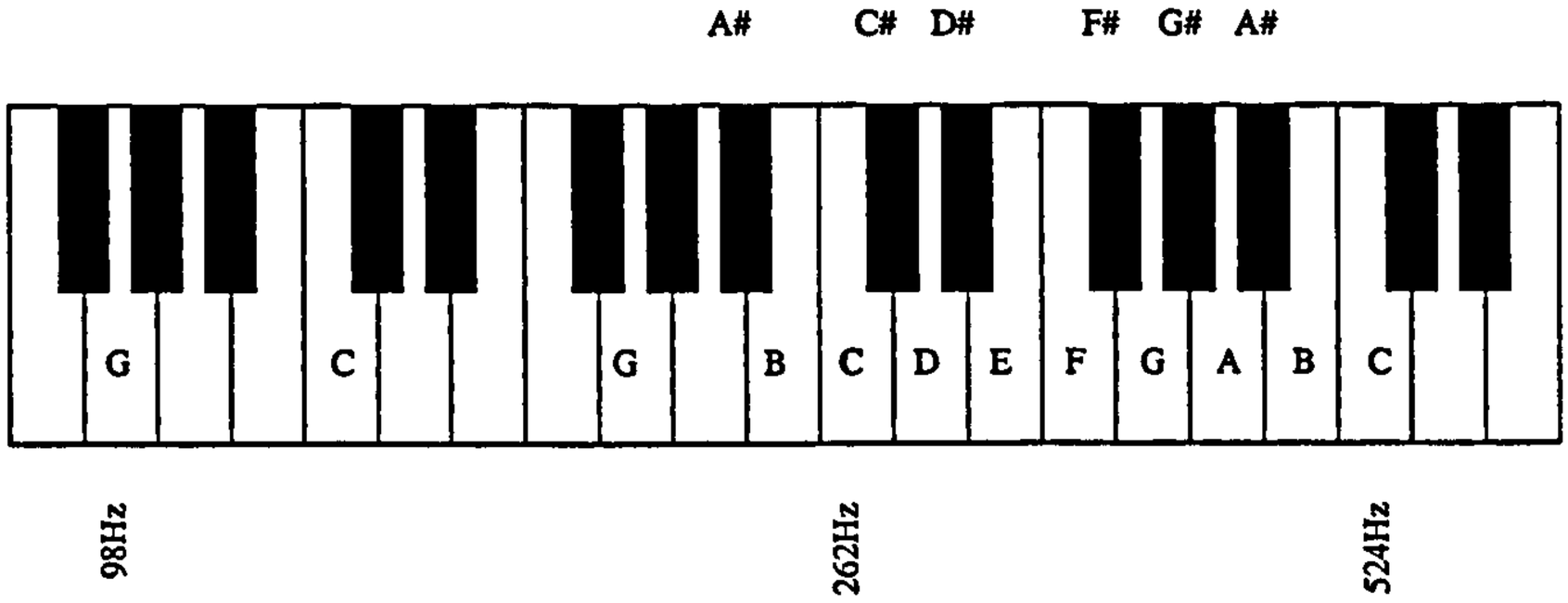


Fig 8.9 The set of piano notes used to calculate the SVD.

In this first example we will to see if it is possible to recognise the note C262Hz from the other recorded notes. A realisation of the note C262Hz will be projected in a set of singular vectors of different notes including the basis set corresponding to this note. (We continue to use the same notation used in previous chapters to address the basis sets of a signal: the basis set of the new space will referred with a B before the note name and frequency, e.g. for this note C the base is BC262Hz). Then we take the norm for the projections of the note on each basis set to recognise the presence of the note in the signal.

First we will see some examples of the projections of the note C262Hz on the base B131Hz, B262Hz and B524Hz. When C262HZ is projected on its own basis set after the first projections the values are the smallest of all the three examples (fig.8.10 a, b and c). The details of those projections (fig.8.10 A, B and C) show that they have a maximum value on the first or second projection, and about the 60^{th} projection the values become smaller and seem random (the values are normalised to allow comparison of the results). According to the results for the simulated piano notes the first projections of C262Hz on B524Hz (fig 8.10 C) are most likely to correspond to the note on its basis set

since it has the maximum value on the first projection. When we analyse all the projections of C262Hz on B262Hz the smallest values after the 60th projection indicate that the signal has the energy mostly concentrated in the first vectors. When the space where C262Hz is represented is not its own space, there will be more energy spread on those vectors after the 60th, especially when C262Hz is represented on B131Hz.

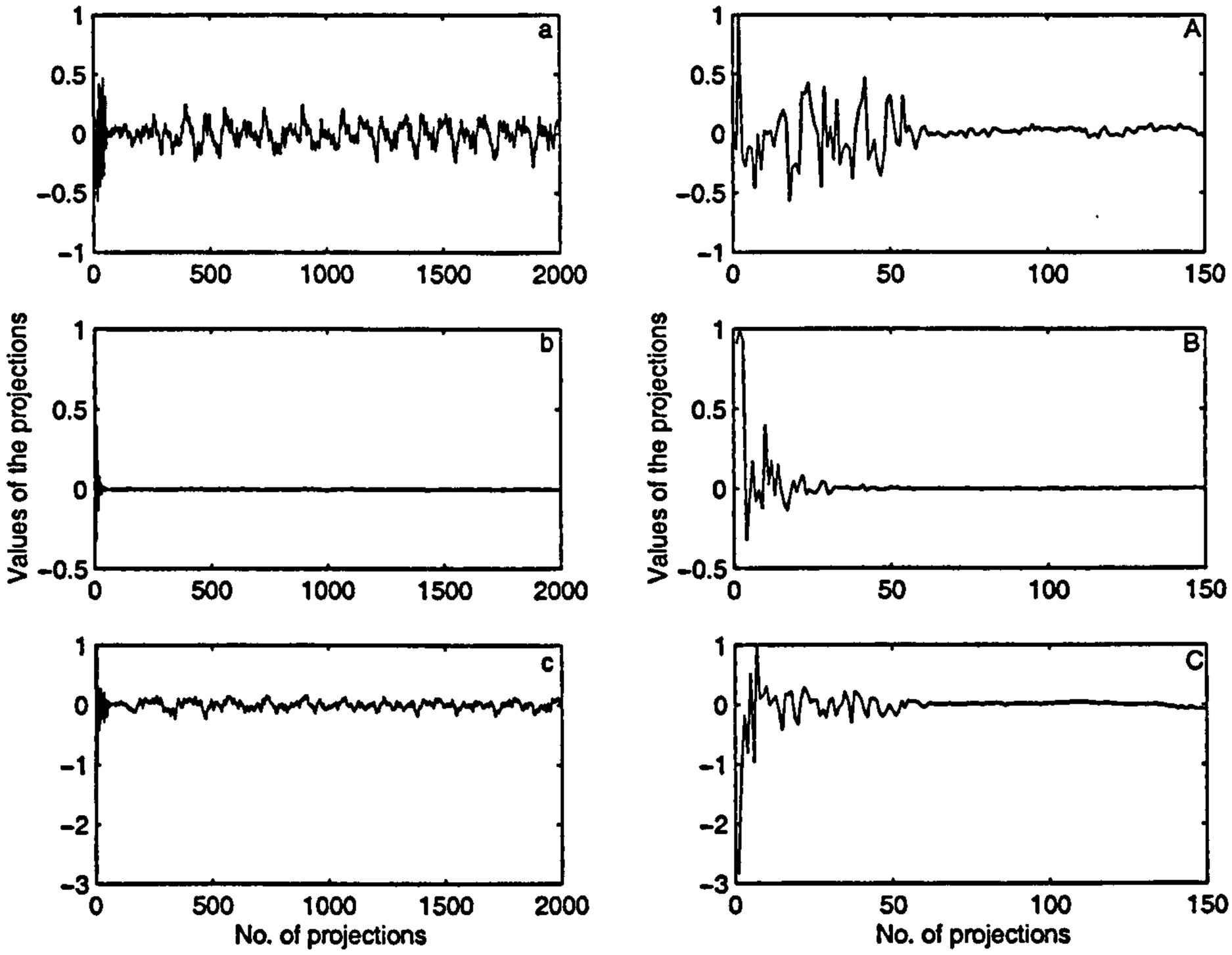


Fig 8.10 Projection values of the projection of note C262Hz on the singular vectors of C131Hz (a), C262Hz (b), C524Hz (c). Detail with the first 150 values C131Hz (A), C262Hz (B), C524Hz (C).

To quantify this in a way that we can detect the note in is own space we take the norm values as we previously did for the simulation case. We will now see the values of the norm of those projections of the note C262Hz in different basis sets of piano notes (see fig.8.11). The norm curve, that very rapidly reaches the final value of one, corresponds to the note projected on its own basis set. The others correspond to the signal represented in the SVD spaces of other signals. This confirms that when the signal is represented in its own basis set the energy is concentrated in a fewer number of vectors. These norm values show that it is possible to recognise the piano note C262Hz.

We present the norm values of the projection of this same note on all the basis sets of the recorded notes (represented on fig 8.9), C131Hz, G196Hz, A#233Hz, B247Hz,

C262Hz, C#277Hz, D294Hz, D#311Hz, E330Hz, F349Hz, F#370Hz, G392Hz, A#466, B494Hz and C524Hz. We want to be assured that the norm curve of the note C262Hz projected on its basis set is different from all others. In the graph of figure 8.12 we see that the norm values of C262Hz projected on its own basis set reaches the maximum value faster than other projections. For all the recorded notes there is no doubt about the identification of the note.

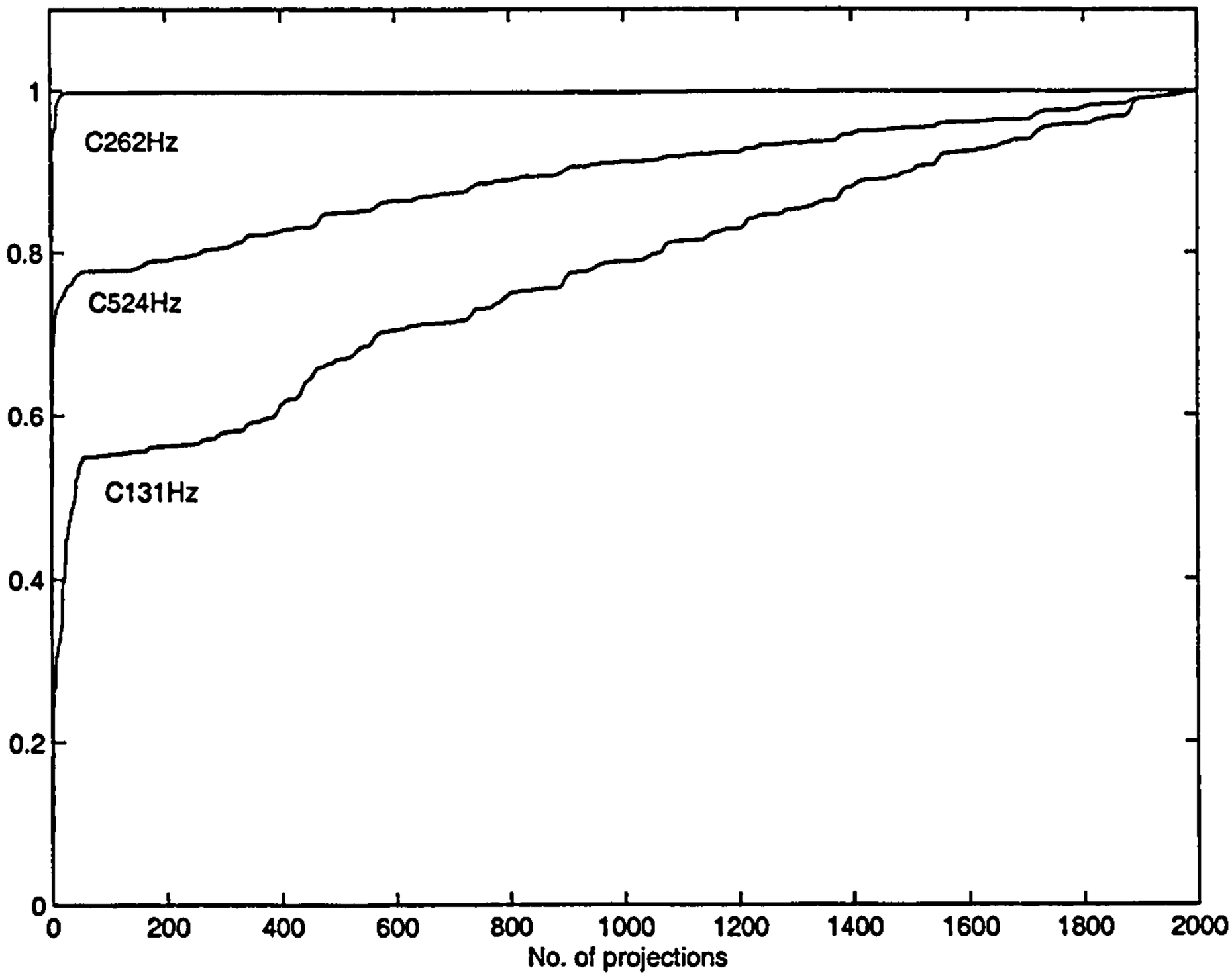


Fig 8.11 Norm of the projections of note C262Hz on the basis sets of C131Hz, C262Hz and C524Hz.

Analysing the results in more detail there is a common feature in all the values. At around the value norm corresponding to 60^{th} projection, there is a small step in all curves. They reach a certain value at that 60^{th} projection quickly. The number 60 is the number of events used in the time history matrix of each note. Since the events are independent (all the singular vectors different from zero) the dimension of the system defined by the events is 60. The dimension of the new space defined by the singular vectors is 2000. The initial space of the signal defined by the time history is embedded in a new space with 2000 linearly independent vectors as the basis set. When projecting a signal in its own space after the 60 projections we achieve what is called the null space.

When the number of events is changed the step on the norm values appears around that value.

The presence of this step on all the projections indicates that we do not need all the values of the norm (the different norms do not cross after the 60^{th} value) and the best value or values to identify a note are situated around the 60^{th} projection. We do not need to use all the singular vectors to project the signals for identification. That decreases the number of data stored for reference and the computing time for the calculations. (When the number of realisations of the time history matrix changes for example to 50, the step on the starts at the 50^{th} projection).

The results for the note C524Hz projected on all the basis sets are presented on fig.8.13. As before it is possible to recognise the norm curve corresponding to the projection of C524Hz on its own basis set. We note that for both cases of C262Hz and C524Hz the norm values of the projections on C524Hz and C262Hz respectively have higher values than the rest of the curves. If the values of the norm are high for the projection of the note C524Hz on the basis of C262Hz we would expect to have also high values for the projection of the note C262Hz on the basis of C131Hz. But that does not occur and those values are smaller.

For the recognition of single note this recognition process is very efficient. All the recorded notes were tested and the results were all similar: it is possible to recognise a single note when projected on a set of basis sets represented by the singular values of single piano notes.

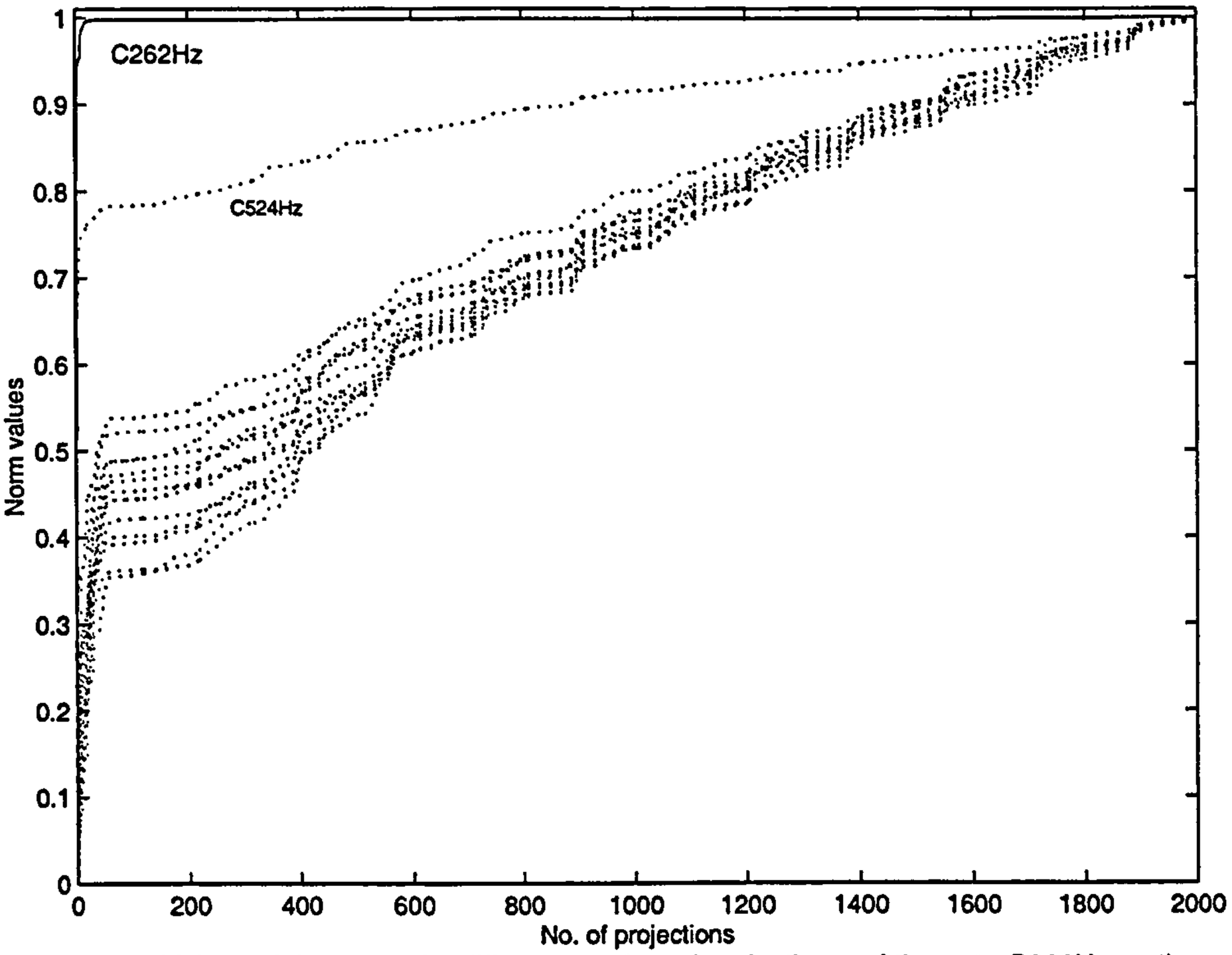


Fig 8.12 Values of the norm, for different number of projections, of the note C262Hz on the basis sets of notes C131Hz, G196Hz, A#233Hz, B247Hz, C262Hz, C#277Hz, D294Hz, D#311Hz and E330Hz, F349Hz, F#370Hz, A#466Hz, B494Hz, C524Hz.

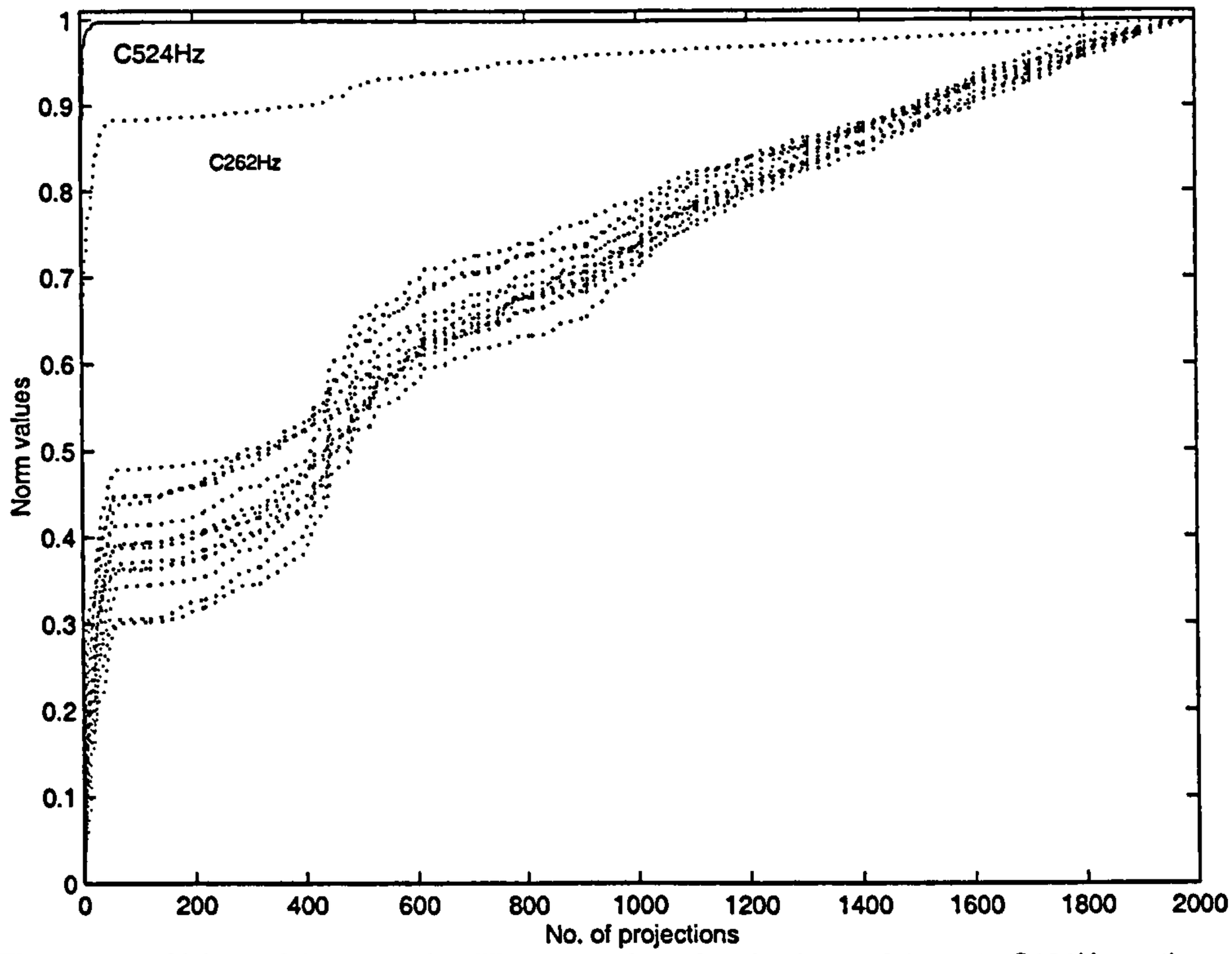


Fig 8.13 Values of the norm, for different number of projections, of the note C131Hz on the basis sets of notes C131Hz, G196Hz, A#233Hz, B247Hz, C262Hz, C#277Hz, D294Hz, D#311Hz and E330Hz, F349Hz, F#370Hz, A#466Hz, B494Hz, C524Hz.

8.4 Recognition of octaves

Now we will see if the recognition process works for octaves played in the piano. We will proceed the same way for the single notes.

In the simulation case the octaves were the linear addition of two signals with fundamental frequencies of f_1 and $f_2 = 2f_1$, and they could have some delay between them. In the real case the octave is not simply the sum of two notes with possible delays between them. In chapter three we mentioned that the strings are all connected to a soundboard. When two or more notes are played together there is coupling between the strings due to the connection to the soundboard. There is some interaction between the two sounds that will affect the time history of the resulting sound. We will see if it is possible with the SVD to recognise the components of a signal that is not the linear sum of two notes with coincident harmonics.

The first octave analysed is composed by the notes C262Hz and C524Hz. We will refer to this octave as 8C262Hz (the value 262Hz corresponds to the frequency of the fundamental of the note with lower frequency). The octaves were played without any special monitoring of the simultaneity of the notes and equal pressure applied on the keys. They were played just as any pianist does because this will correspond to the piano sound production in the real case. This implies that when we have more than one note played simultaneously there will generally be small delays between the notes that will change the time history characteristics of the signals. This introduces a new factor that will make it difficult to identify the notes as the slight different dynamics apply to both notes.

The Fourier transforms of the realisations of the signals C262Hz, C524Hz and 8C262Hz show that the octave has a very small frequency component at C262Hz but the frequency components are practically the same as seen in the note C262Hz (fig.8.14). The identification of the octave in the spectrum is not obvious from the other two spectra and it is not possible to determine the presence of two notes.

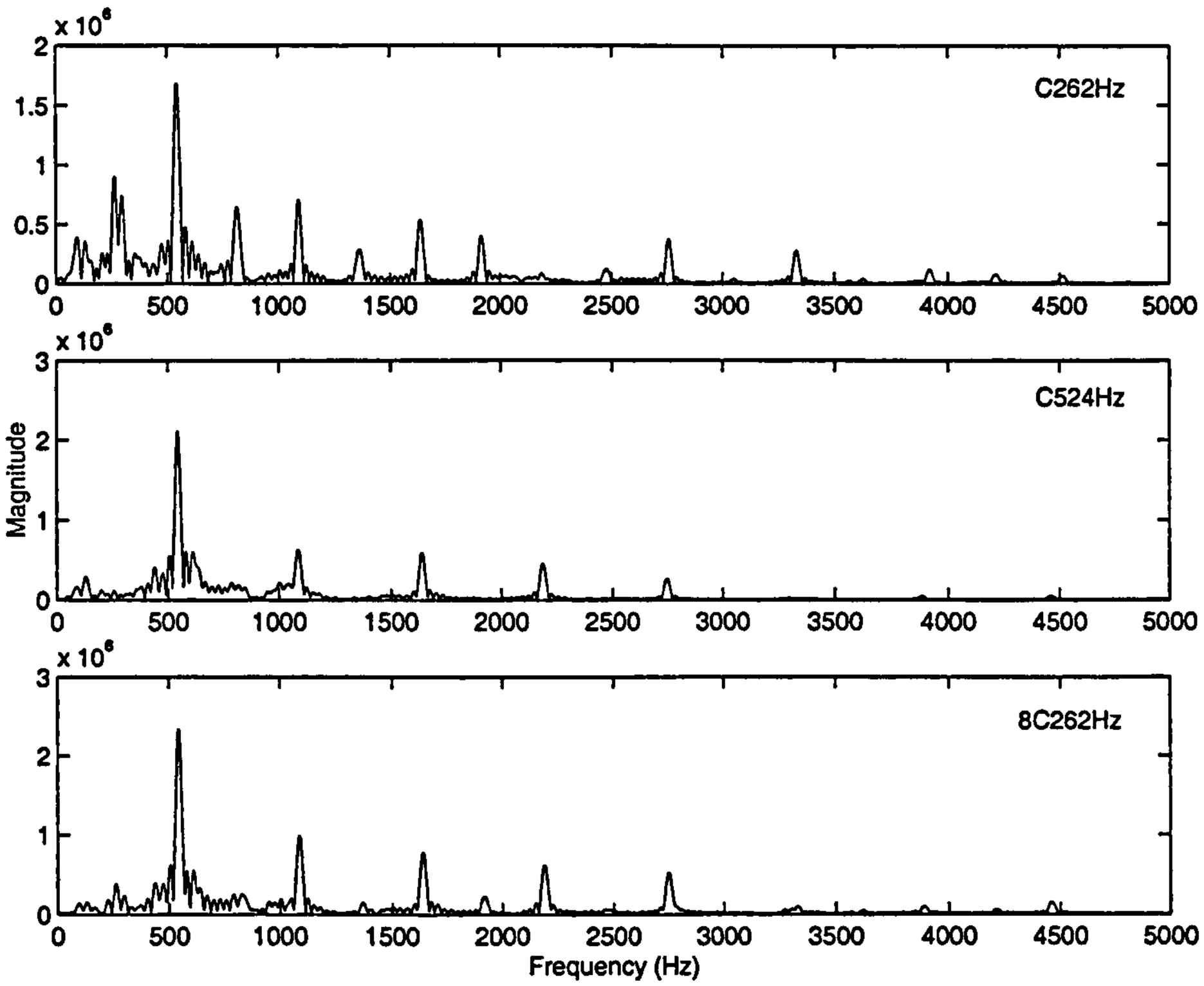


Fig 8.14 Fourier transform of the piano signals C262Hz, C524Hz and octave 8C262Hz (FT 44100 points)

For the recognition of the notes composing the octave we first projected the octave 8C262Hz on the basis BC262Hz and BC524Hz. The norm results were compared with the norm value of the projection of events of the notes C262Hz and C524Hz on their own basis and in the basis of the other note.

In the projection of the three signals C262Hz, C524Hz and octave on the basis of C262Hz (fig.8.15) we see that around the 60^{th} projection the values of the octave are in the middle of the other two. So there is a distinction between all the three signals when projecting on C262Hz.

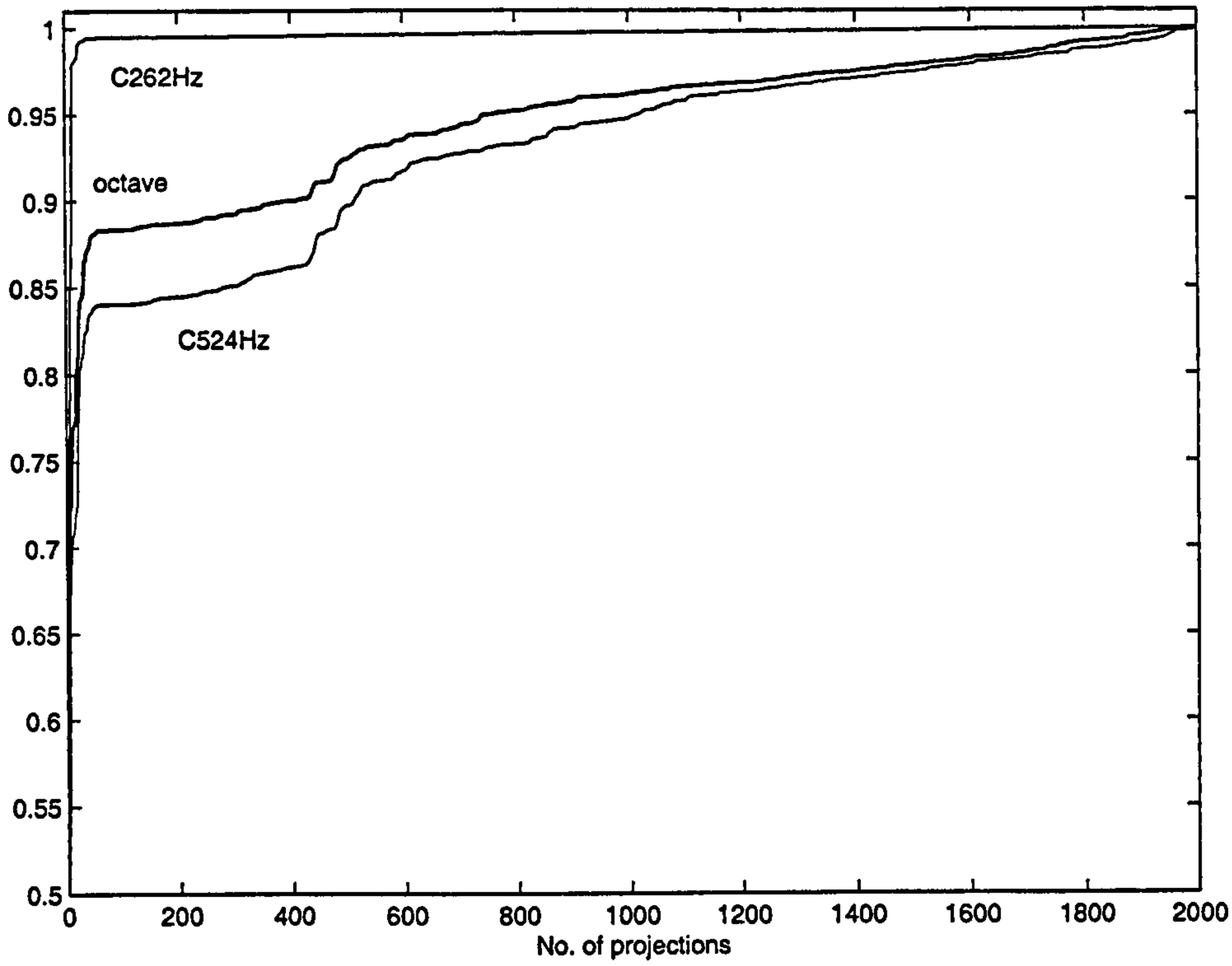


Fig 8.15 Values of the norm of the notes C262Hz, C524Hz and octave 8C262Hz on the vector basis set of C262Hz

In the projections of the same notes on BC524Hz again the values of the octave stay in the middle of the other two (fig.8.16). This way we can see that the norm values of the octave are different from the values of the note projected on its own base (C262Hz on BC262Hz and C524Hz on BC524Hz) but have higher values for the case when we project C262Hz on BC524Hz and vice versa. So the octave when projected on the notes that compose it has a value distinct from other projections.

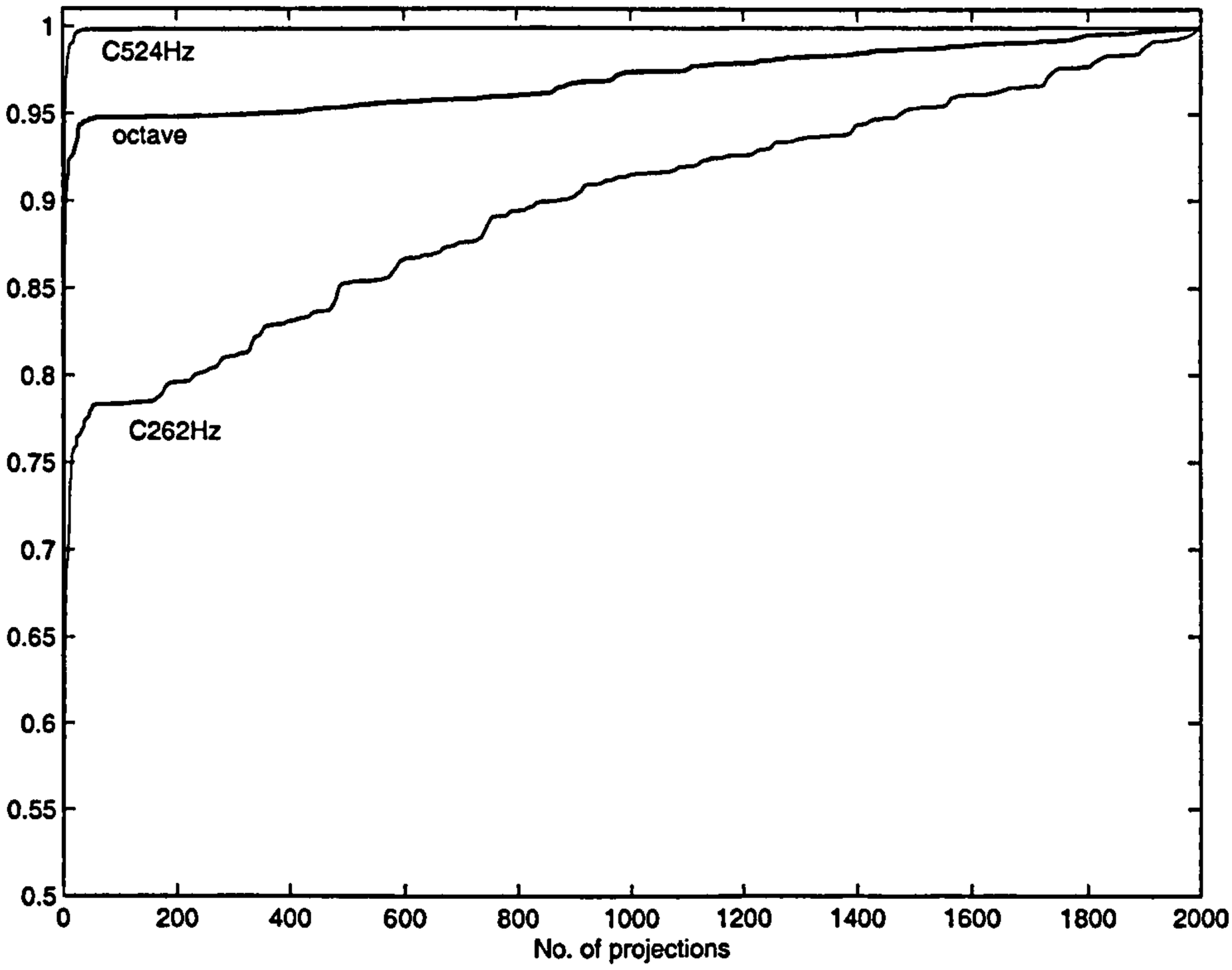


Fig 8.16 Values of the norm of the notes C262Hz, C524Hz and octave 8C262Hz on the vector basis set of C524Hz

The most probable mistake in note detection is to erroneously detect a subharmonic or a supra harmonic note that is not present in the signal. In the case of the octave we will see how the recognition process works when it is projected on the basis of notes that have half the frequency and four times the frequency of the lowest note in the octave. For the first case the note with half of the frequency value of the lowest note of the octave contains all the frequency components of the two notes of the octave.

The projection of the octave 8C262Hz on BC131Hz (fig.8.17), whose signal C131Hz contains all the frequency components of the two notes of the octave shows that it is unlikely to do a wrong detection of this note. So there is a little chance of confusing this subharmonic with a note present in the signal.

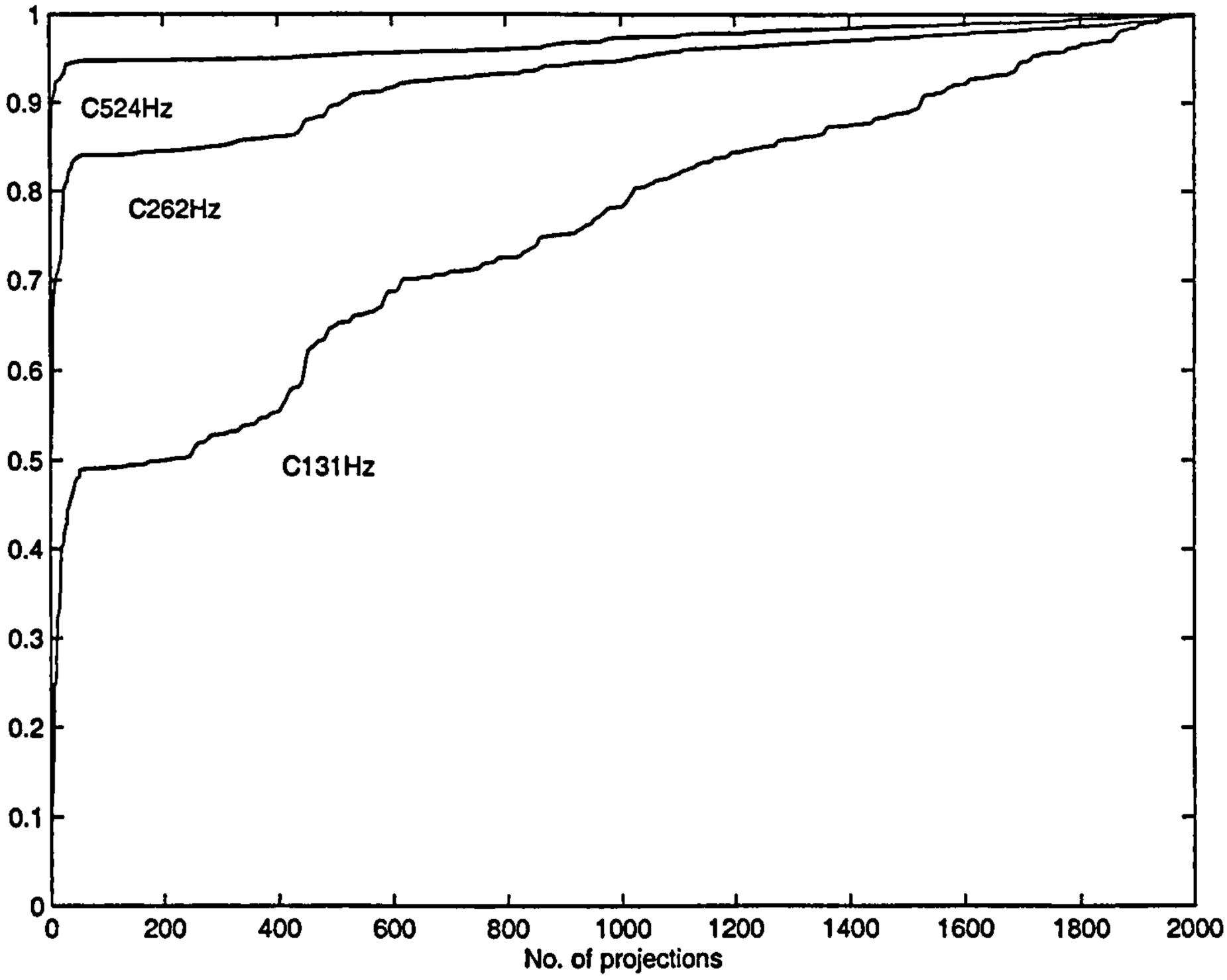


Fig 8.17 Values of the norm of the octave 8C262Hz projected on the vector basis set of C131Hz, C262Hz and C524Hz.

The results indicate that the projection of the octave on their component notes has a different result from the projection of the single note in its own basis set. But for an identification system we need to see how the norms behave when the octave is projected on all other notes. In a real detection case we do not know a priori the notes and the number of notes present in a musical signal. We expect that when projecting a signal in all basis sets the values of the norm around the 60^{th} projection will be higher than any other values and over a certain limit for that we can recognise the note as belonging to the signal. If the values of the norm stay above a certain value n the note is part of the piano signal.

The projection of the octave on all the basis sets show two norm values detached from the others (fig.8.18) corresponding to the two components of the octave piano signal, C262Hz and C524Hz. The graph of fig.8.13 corresponding to the projections of the note C524Hz is very similar to this graph for the octave 8C262Hz, but as we have seen in graphs of fig.8.16 the values of a note in the octave never reach unity until the last projection and are always smaller than those of the single note. For all the recorded

realisations of this octave the results were similar. The identification of the notes C262Hz and C524Hz was always possible.

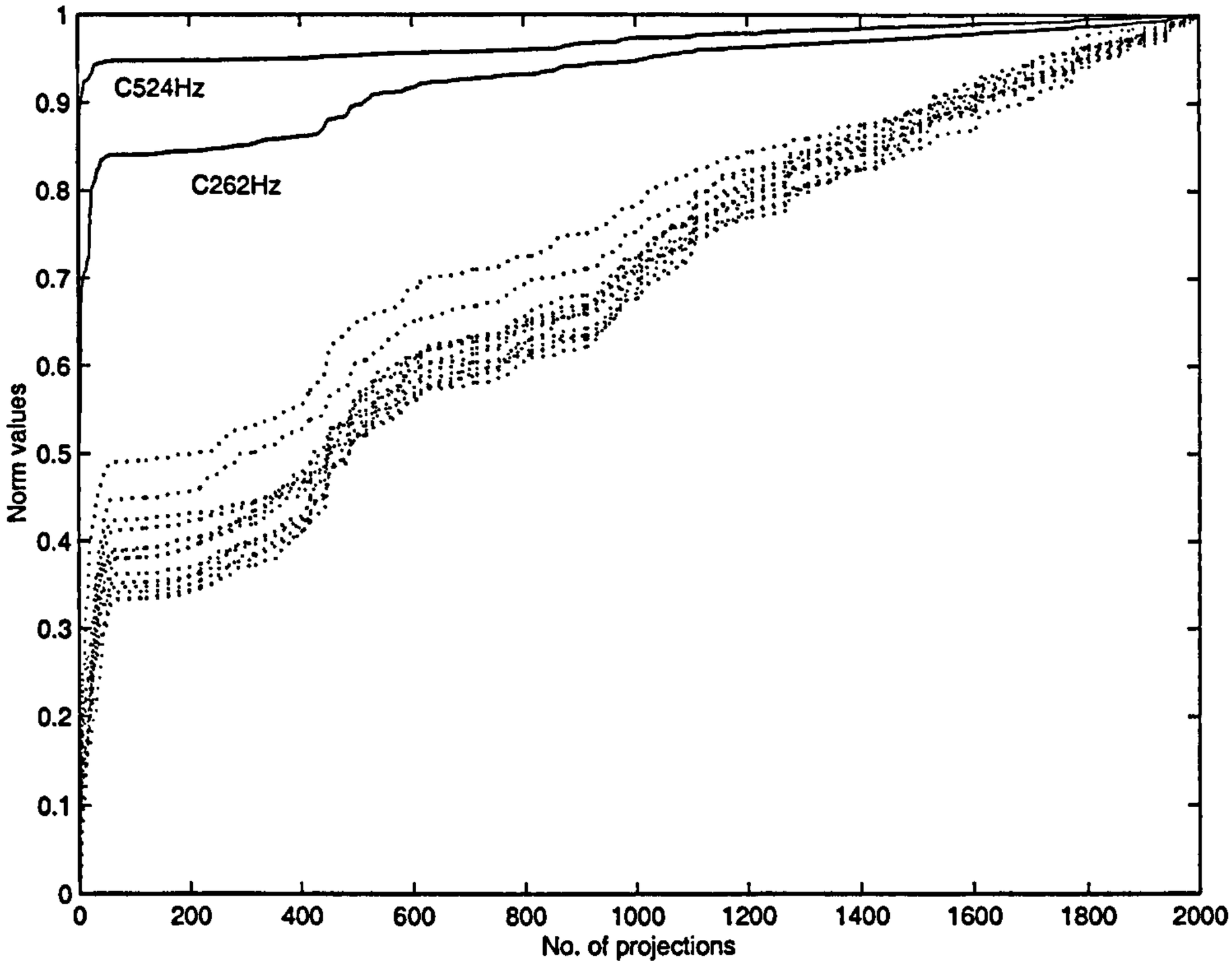


Fig 8.18 Values of the norm of the octave 8C262Hz projected on the vector basis set of all the other notes.

For other octaves 8A#233Hz, 8G196Hz it was also possible to identify the notes of the octave in all the recorded events. The results were similar to the ones presented for the octave 8C262Hz.

This identification was not possible for two of the octaves analysed, 8C131Hz and 8G98Hz, since the two norm values of the notes in the octave mix with the rest of the curves. These two octaves had the lowest fundamental frequency components of all the octaves analysed. We present the results of the octave 8C131Hz (fig.8.19) and one can see that the identification of both notes is not possible. Only the note C262Hz has its values significantly above the others. The C131Hz curve is mixed with the others corresponding to notes not present in the signal. Since the recognition does not work for the lowest values of the fundamental it is a problem of resolution. The 2000 points of each of the events (at sampling frequency of $f_s = 44100\,\text{Hz}$) could be insufficient to detect the presence of those notes on the signal. The spectra of that octave (fig.8.20), with the 2000 points and 3000 points show that the relative magnitude of the first

frequency components changes. We apply the recognition process to the signal with 3000 samples to analyse if there are any changes in the norm values that allow the identification of the octave components.

To determine the projections of signals with 3000 samples we had to again determine the SVD of the recorded single notes with the possible realisations of the signals with 3000 samples. The resulting basis sets have 3000 singular vectors with 3000 samples each.

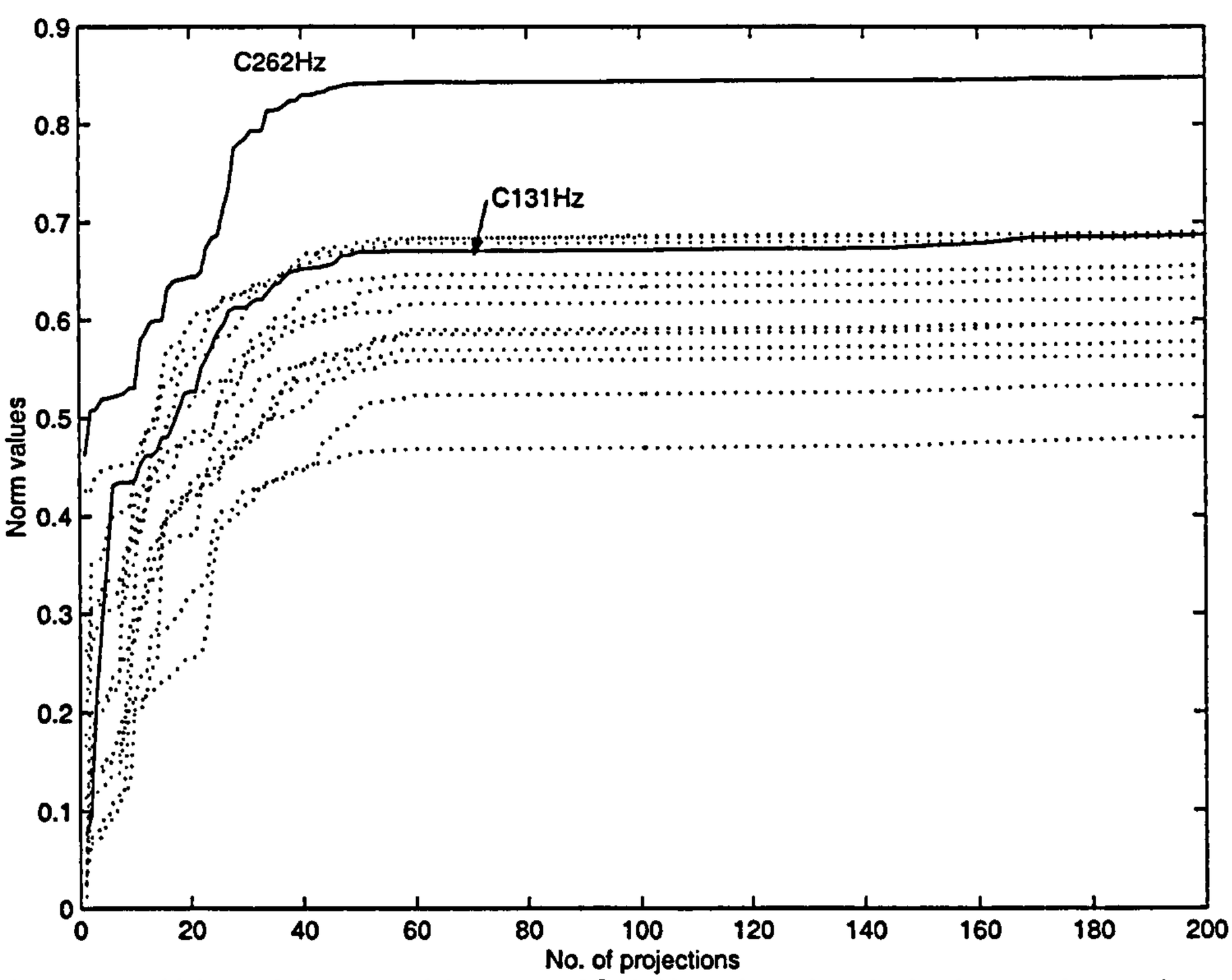


Fig 8.19 Values of the norm of the octave 8C131Hz projected on the basis sets of several single notes.

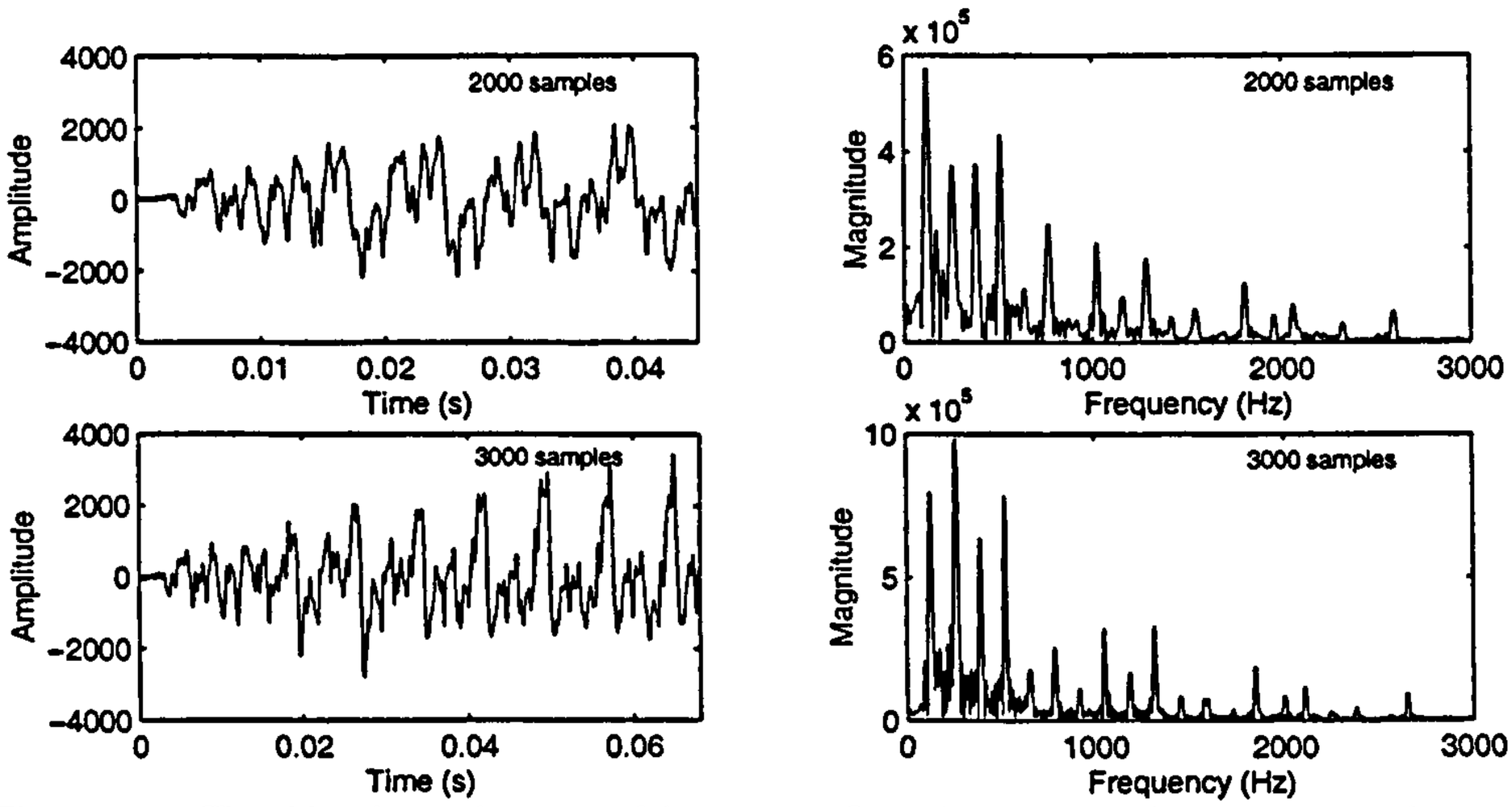


Fig 8.20 Time histories and spectra of the octave 8C131Hz for 2000 points and 3000 points (FT of 44100 points).

The norm values of the projection of the octave 8C131Hz with 3000 samples on the singular values with 3000 samples each show (fig.8.21) that now the norm curves of the notes belonging o the octave (C131Hz and C262Hz) have the highest values although the curve of the note C131Hz is still very near the others. The introduction of more samples improves the performance of the process for octaves with lower frequency component of 131Hz.

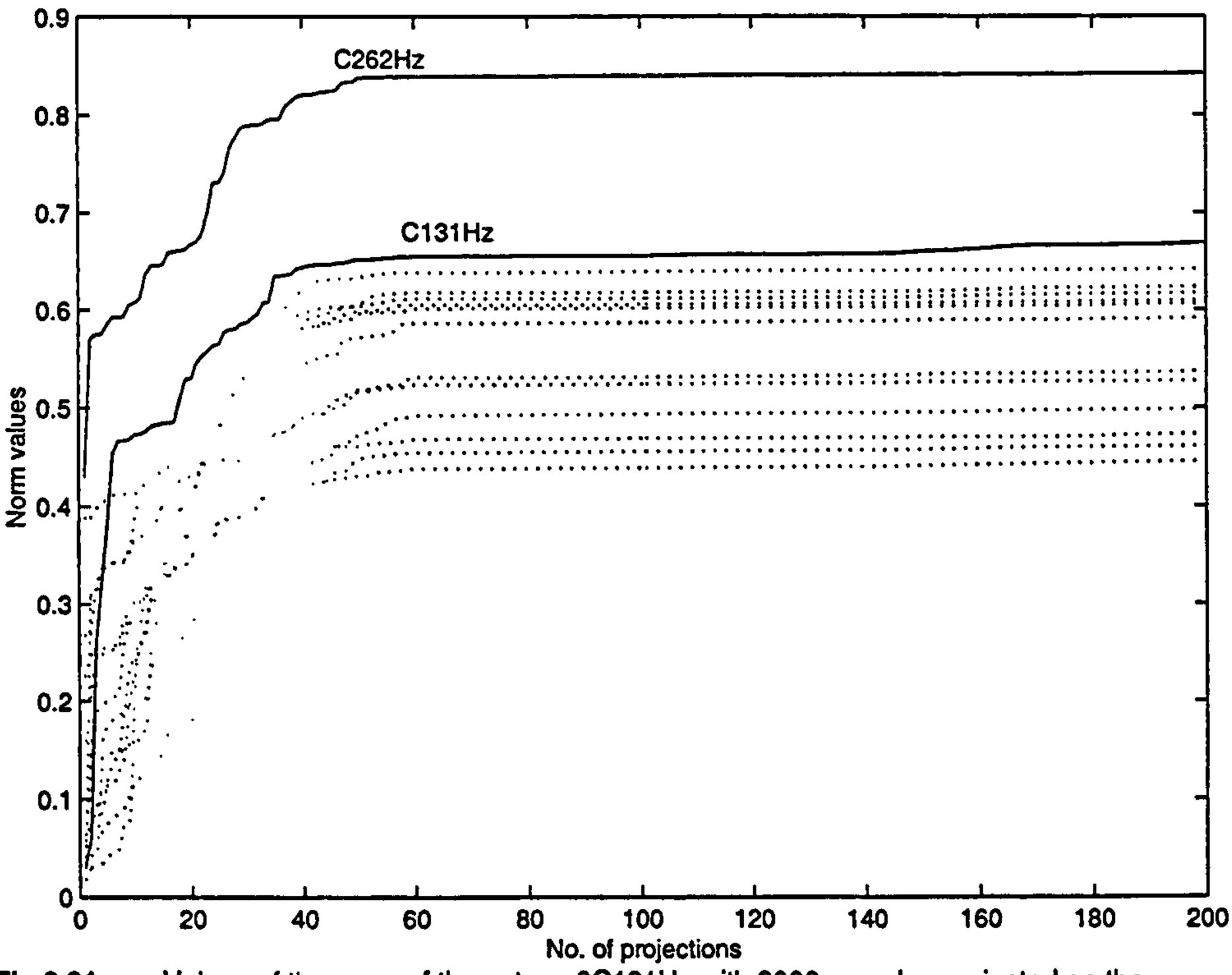


Fig 8.21 Values of the norm of the octave 8C131Hz with 3000 samples projected on the basis sets of several single notes.

The recognition of the octave is possible with SVD. The results indicate another way to analyse the results. We can search for the presence of the notes in the signal through the highest norm values and by the particular values of one note. In the example of the octave where one of the notes has its values mixed with others or above but very near the other note norm (always correspond to the note with lowest frequency) the other note has very high norm values but not as high when it is a single note. So only the value of that note alone in the octave indicates that we are in the presence of an octave.

8.5 Recognition of notes separated by an interval of a tone

We performed the recognition of signals with two notes played simultaneously separated by a tone interval: C262Hz and D293Hz. We seek to find the results for notes separated by small intervals. We show two examples (fig.8.22) of the norm values of these two notes and the respective Fourier transform. The note D293Hz has the highest values and note C is very near the other curves or is even mixed (on graph b there is a curve between C and D and corresponds to the noteG98Hz). When the relation between the frequency components of the two notes changes the norm values also change and they are more separated when there is a difference between the magnitudes of the fundamental. The curious note is that in example *a* there is a component at 98Hz and the curve of G98hz is mixed with the others while in example *b* there are no significant component at hat frequency and the curve is between D and C. The note G98Hz is in the harmonic series of both notes D and C. Other notes as C131Hz are also on the same series but it does not appear mixed with the curves of D and C.

The spectra show clearly two frequency components corresponding to the fundamental of each note. With the Fourier transform is possible to detect these two notes in the signal while with SVD we detect the note D293Hz but the results for the presence of note C are not clear. (The identification of the notes of this signal by the human ear is not an easy task. Generally it takes more time to identify the notes in this signal than others with notes forming greater musical intervals between them).

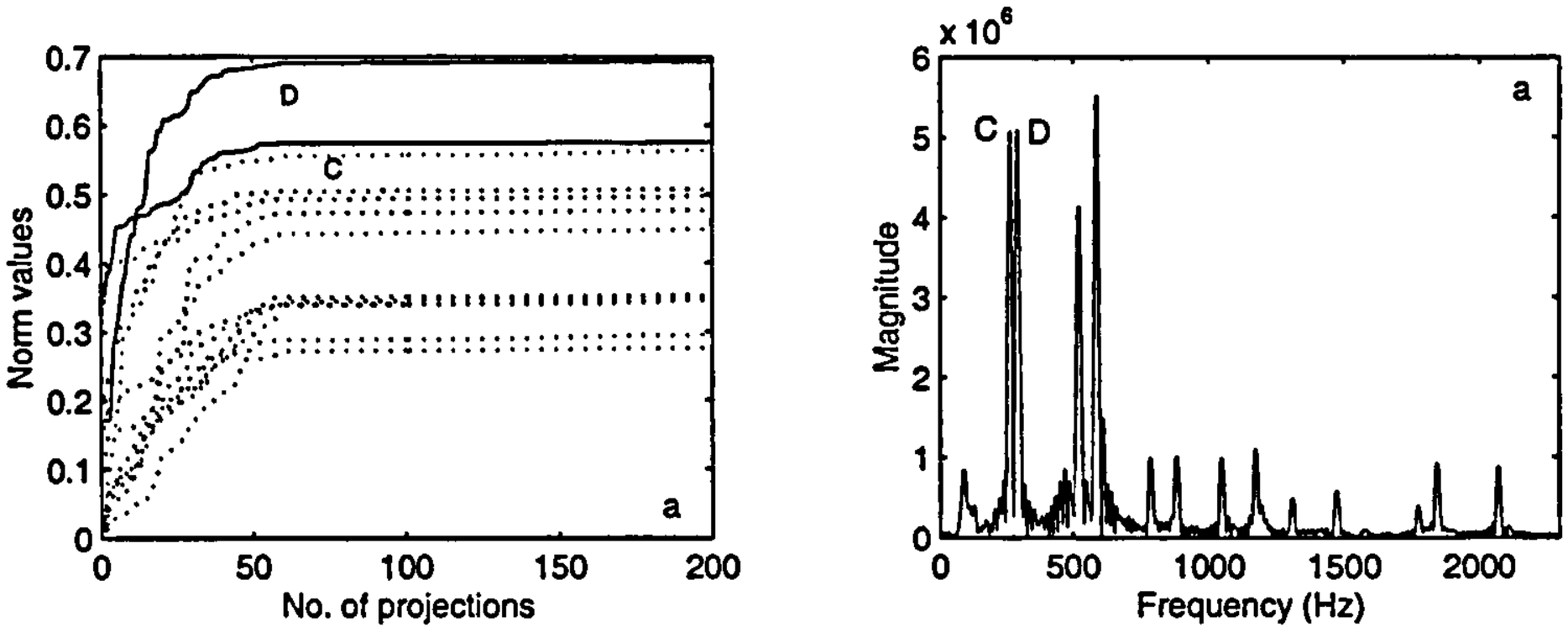


Fig 8.22 Values of the norm and spectra (FT of 44100 points) of the signal composed of notes C262Hz and D293Hz.

8.6 Recognition of notes in chords

Chords in music, by definition, are composed with more than three notes played simultaneously. There are a vast number of possible chords with three notes that can be played on the piano. The most common ones are the major and minor chords. As an example the major C chord as the notes, C, E and G, and the minor the notes C, E flat and G. All the chords recorded were played without special concern about the force applied to each key. As with the octaves they were played as a pianist would play them. Generally there is always a slight difference in the starting point of each note (not noticeable to the human ear) introducing very small delays between the notes played that influence the time history of the signal. The time histories are different, as we mentioned in the octaves section, and the resulting projections on the basis vectors of the single notes can reflect that (we should remember that we are working in the time domain when using the projections). The signal, in the case of a chord, is now defined by the combination of

three single piano notes that can be played with different dynamics and can have different delays. Another problem that we have to deal with now is whether the process is able to give information about the number of notes played. The curves for a note present in a signal may have maximum values in relation to others but they may not be sufficiently detached to determine the number of notes in the signal. For the case of octaves, for the example shown, it was possible to determine the number of notes since there were two norm curves detached form the rest when the signal was projected on the basis sets.

When a single note is projected on its own basis the norm achieves a value close to one at the 60^{th} projection. For the octaves, where two notes are played together, at the 60^{th} projection the values for the same note are smaller than for the case of the single note. For the chord with three notes we expect those values, corresponding to the of the notes present in the signal, to be even smaller than for the other mentioned cases. The energy of the signal is distributed through a greater number of frequency components and so it will be distributed through a greater number of singular vectors of any base.

8.6.1 Major chords with three notes

The first chord that we will analyse is the major C chord. The notes of this chord have the frequency values of C262Hz, E330Hz and G392Hz. The signal of this piano chord was projected on all the basis sets of singular vectors. We use first the 2000 samples of the signal and then 3000 samples to verify if the use of more samples improves the performance of the recognition process.

Analysing the results in graph 8.22 (for the 2000 samples) we see that there is no gap between the norm of the notes in the signal and out of the signal as it happened on the octave case. The notes C131Hz and G196Hz are not present in the chord but they have norm values higher than the notes C262Hz, E330Hz and G392Hz present in the chord. These three notes of the have the magnitudes very close to each other. Although the norm values are very close to each other after the 60^{th} projection this is not a general feature for this chord C major chord since for other similar chords this did not happen.

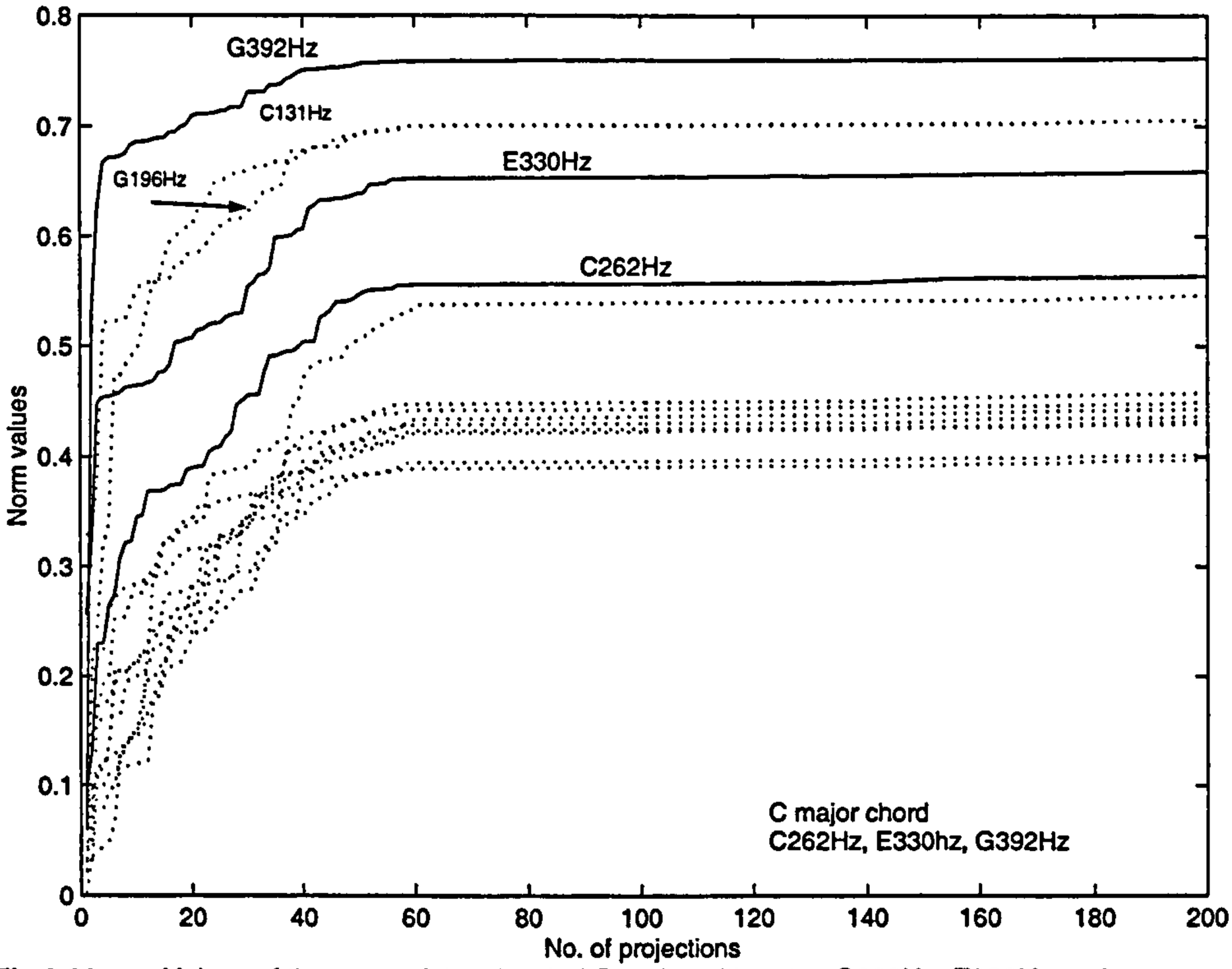


Fig 8.23 Values of the norm of the chord of C major with notes C262Hz, E330Hz and G392Hz projected on basis BC262Hz, BE330Hz and BG392Hz and other notes not present in the chord including BC524Hz (the arrow indicates the curve correspondent to the legend).

There could be an explanation for the higher values of the norm for those two notes. The notes C131Hz and G196Hz are subharmonics of the notes C and G present in the chord. The frequencies of the mode of vibration of the piano strings are related to the frequency of other notes on the piano. The modes of vibration corresponding to a certain musical note are called, in music, the harmonic series. All the notes C131Hz, G196Hz, C262Hz, E330Hz and G392Hz are in the harmonic series of the note C65.5Hz which is called the fundamental. The note G98Hz is a subharmonic of G392Hz in the chord but it does not belong to the harmonic series of the note C65.5Hz. The norm values of this note are mixed with the notes that do not belong to the chord. This seems to indicate that the projections detect the harmonic series common to the notes in the signal.

The fundamental frequencies of the notes in the chord have different magnitudes (fig.8.25 - 2000 samples). The relation between them seems to have a reflection on the values of the norm. The note G392Hz is the note with the highest magnitude for the fundamental have followed by E and C. The norm values of G (fig.8.23) are the highest followed by the note E and C.

For the same signal with 3000 samples the results are similar but there is a better separation between the notes in the chord and the other notes out of the chord (fig.8.24).

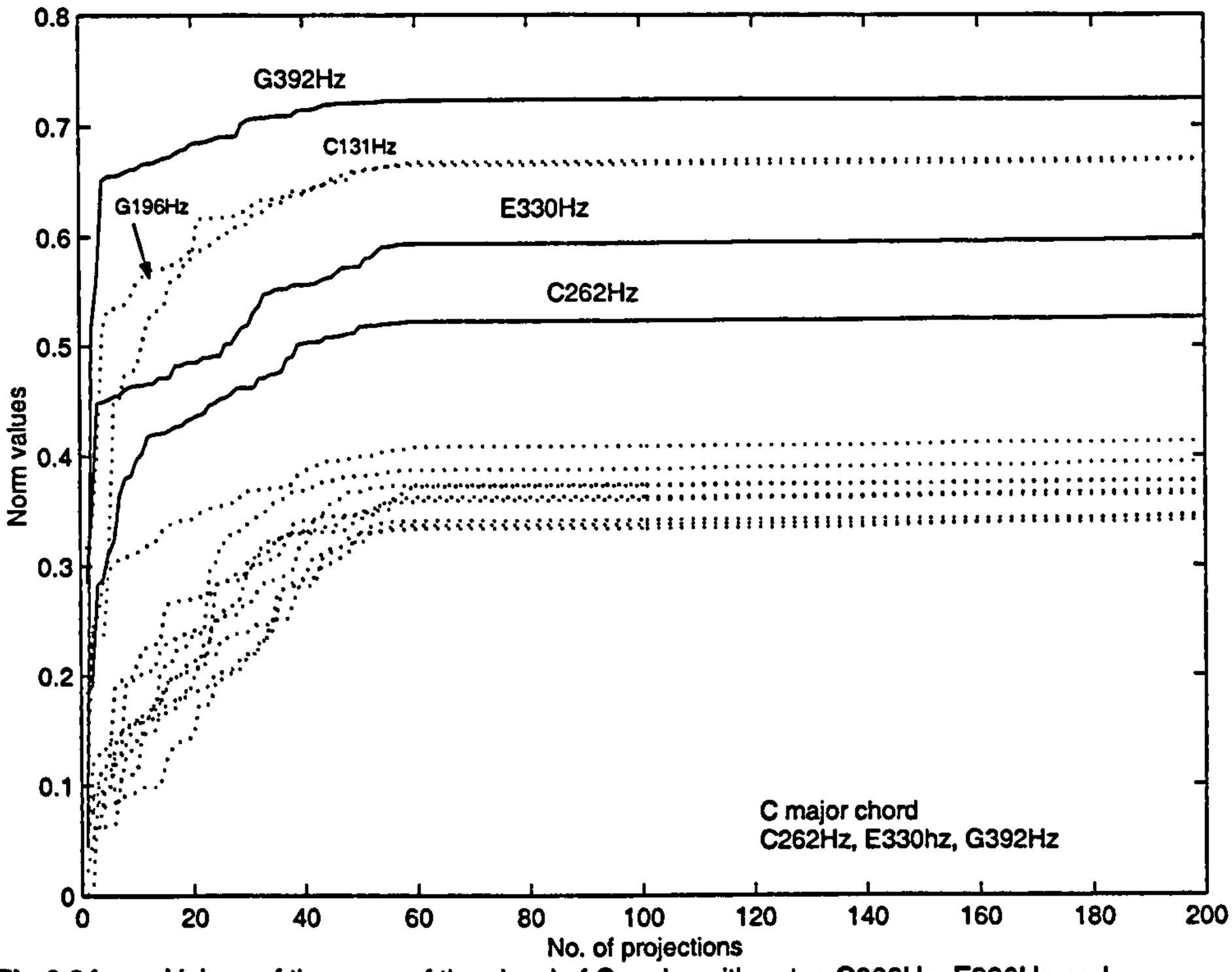


Fig 8.24 Values of the norm of the chord of C major with notes C262Hz, E330Hz and G392Hz with 3000 samples projected on basis BC262Hz, BE330Hz and BG392Hz and other notes not present in the chord including BC524Hz.

The Fourier transform of the chord with 2000 samples and 3000 samples show a better definition of the frequency components for the latter case (fig.8.25).

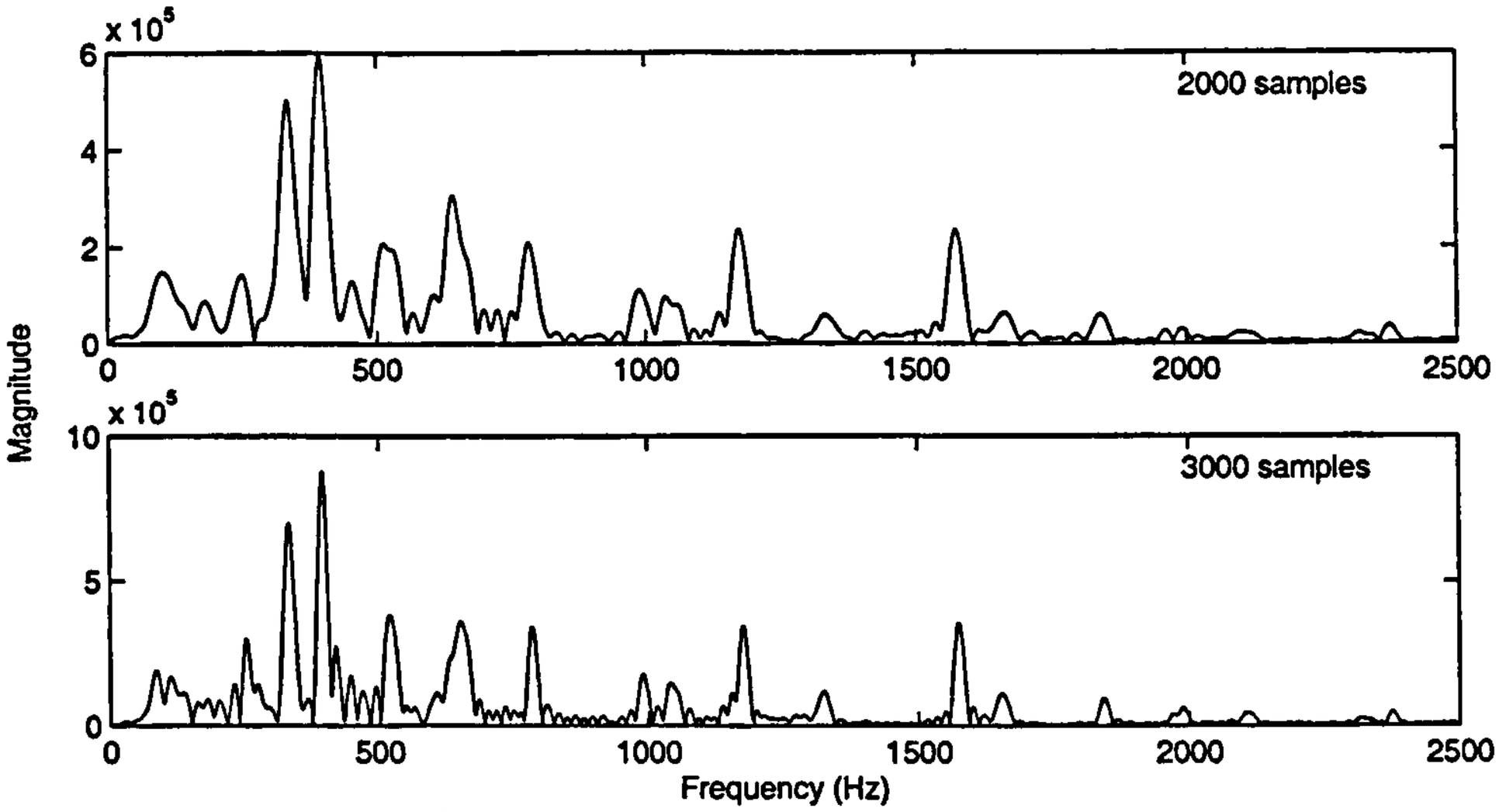


Fig 8.25 Spectra of the C major chord with 2000 samples and 3000 samples (FT of 44100 points).

We analysed more realisations of this C major chord. We also determined the Fourier transform of those chords to compare the results of the projections with the frequency components of the chords.

In the spectra (fig.8.26) we can see that there are frequency components with significant magnitude at lower frequencies than the lowest note in the chord, C262Hz. Those components have frequencies very near the subharmonics of the notes in the chord (95Hz and 121Hz). The norm values (fig.8.27) show that in all cases the notes C131Hz and G196Hz have high norm values and at least always higher than the C262Hz note of the chord. The note C262Hz has its norm values smaller than then projections on other notes out of the chord when its magnitude is noticeable smaller than the component at 121Hz (example *c* of figure 8.26). This frequency does not correspond exactly to the fundamental of any piano note. The nearest piano note is B124Hz.

The norm values of the notes in the chord (C262Hz, E330Hz, G392Hz) are related with the their magnitude in the spectra. In all cases the C262Hz note has the smallest magnitude of all the notes in the chord and the norm curve is the one that always has the smallest value of all three notes. In example *b* (figs.8.26 and 8.27) the components E and G have very near values and the norms have very near values. In example *d* the note E has greater magnitude than G and the norm values of E are greater than G. This indicates that the values of the norm of the notes in the chord are related with the magnitudes of the frequency components of the fundamental of those notes. But this only happens for the notes in the chord because, as we saw before, it also gives high values for absent notes in the chord, although there are frequency components in the spectra for those frequencies. Although the norm values are related to the magnitudes of the frequency components of the fundamental of each note the relation between the values at the 60^{th} projection is not the same as that between the magnitudes.

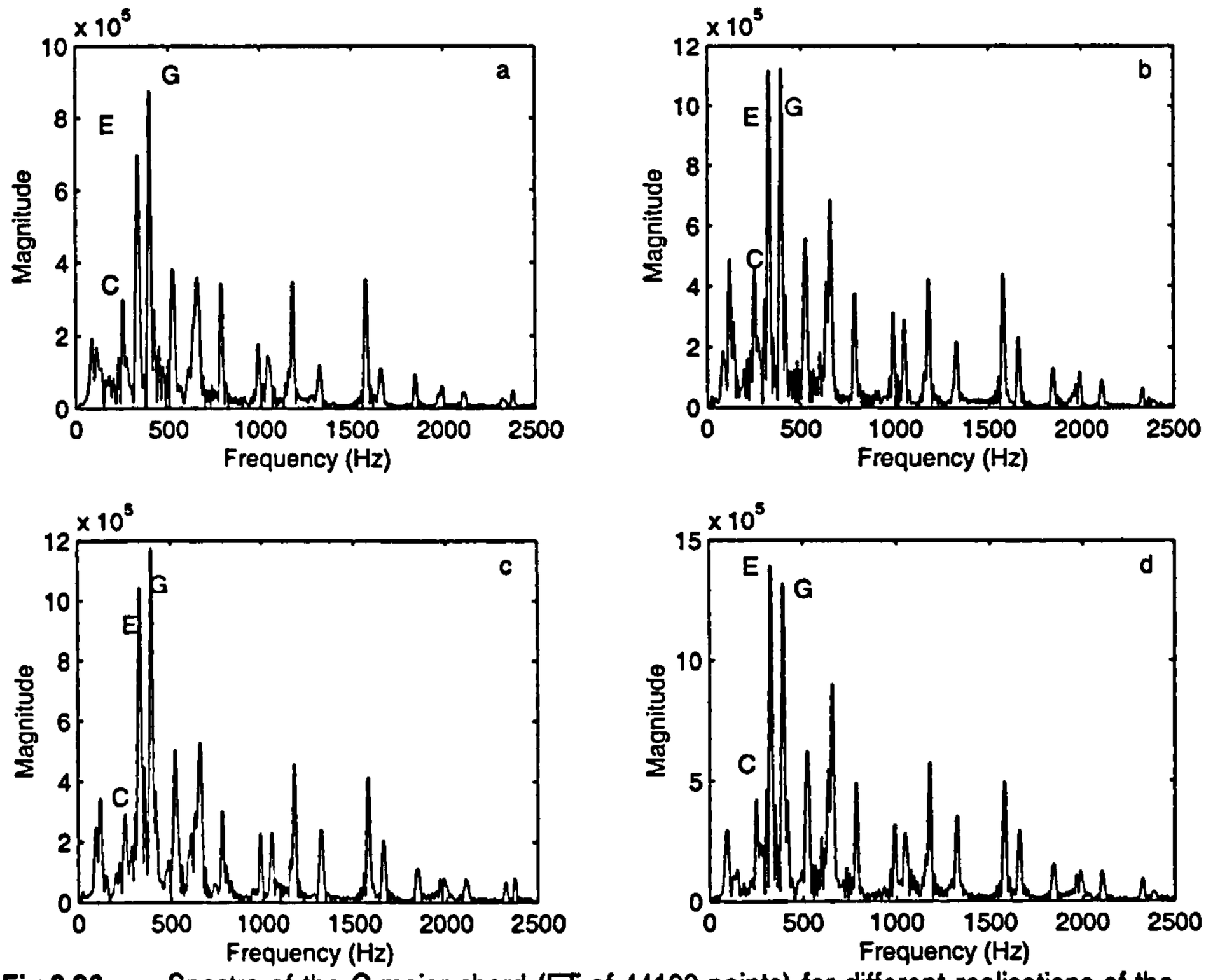


Fig 8.26 Spectra of the C major chord (FT of 44100 points) for different realisations of the same chord.

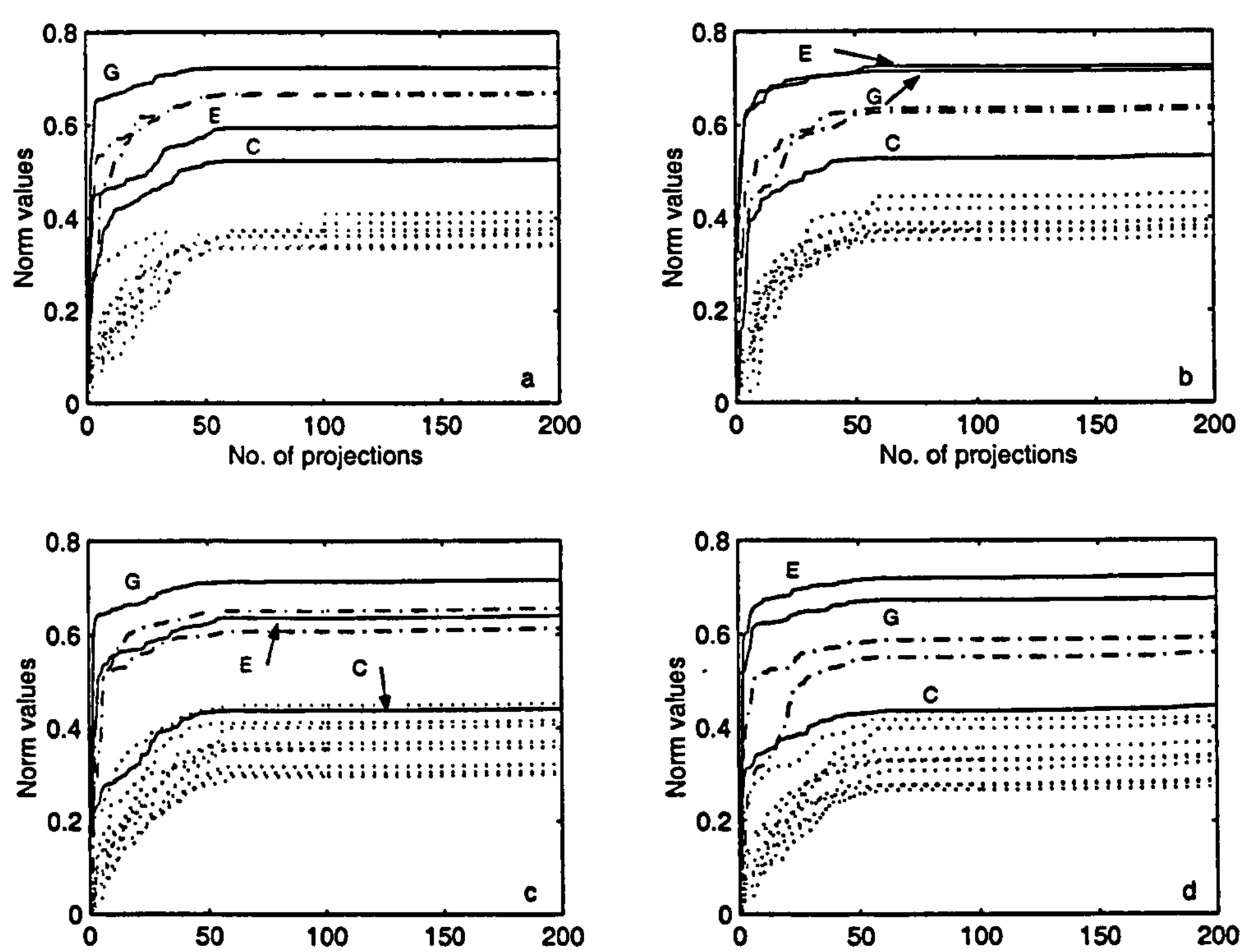


Fig 8.27 Norm values of the projection of the different realisations of the C major chord on different basis sets of singular vectors.

We will see the same C major chord (fig.8.28) but with other notes that belong also to the same harmonic series of C65.5Hz (in music this chord is called the 2nd

inversion of the C major chord). The chord is composed of the notes: G196Hz, C262Hz and E330Hz.

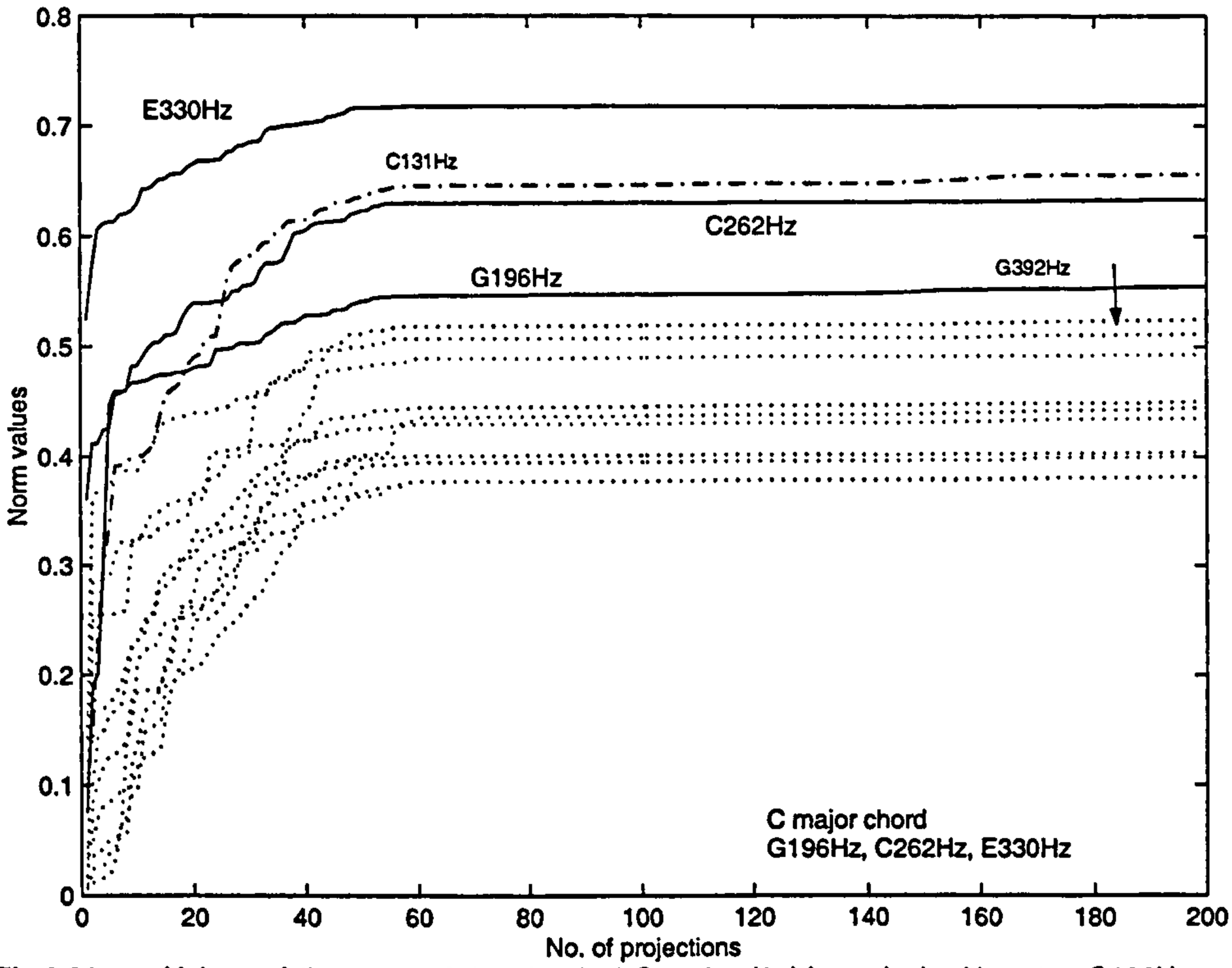


Fig 8.28 Values of the norm of the chord of C major (2nd inversion) with notes G196Hz, C262Hz and E330Hz projected on basis BG392Hz BC262Hz, BG392Hz and other notes not present in the chord.

In this chord although G196Hz is in the signal the norm values are lower than for the previous chords where the note was not present and the norm values corresponding to that note were the second highest. The note G392Hz now it is not on the signal and the value is very near the notes that do not belong to the chord and the division between the notes belonging to the chord and those out of the chord seems difficult. The results for this chord were similar with small variations and the results seem not to give a significant difference between this chord and the previous chord which makes the task of knowing which chord are the present difficult. The note G98Hz is a subharmonic of the note with the lowest frequency in the chord: G196Hz. But this note it is not in the harmonic series of C65.5Hz that contains all the notes of the chord (by that order G196Hz, C262Hz and E330Hz) and its norm values continue to be mixed with the other notes out of the chord.

The identification of the notes present in the analysed chords show that the generally the notes in the chord have one of the highest norm values at the 60th

projection. Notes that are not present on the chord but belonging to the harmonic series where those notes of the chord are included have one of the highest values. One of the problems in the note identification with this process is to determine the separation value at the 60^{th} projection between the norms that are indicates the notes in and out of the chord. Another problem is to know if the norm values appearing with the highest values are present in the signal or are only subharmonics. The spectra of those signals do not present significant peaks at those frequencies that justify their high values in the norm. We note that generally that the notes wrongly detected are always subharmonics. There was never the wrong detection of the notes with frequencies that are multiples of the fundamental of notes in the chords as C524Hz and G392Hz.

The notes of these major chords can be identified using the Fourier transform while with this method we can only identify some notes. With the Fourier transform we can also have difficulty in determining the lowest frequency note in the chord in the example of fig.8.25 *b* since there are subharmonics with higher magnitudes than the first note of the chord (the first note of a chord is the one with lowest frequency fundamental). The increase from 2000 samples to 3000 samples significantly improved the performance of the method in the case of the octaves. For the 3 note major chords the improvement was not so significant but the results were better. The results indicate that the use of more samples improves the performance (with the price of computation time and memory).

8.6.2 Major chords with four notes

We continue to analyse the same major chords of the previous section but now they will have one more note that is the octave of the note with the lowest frequency. As an example the C major chord will have the notes C262Hz, E330H, G392Hz and C524HZ and this last note is one octave above the note C262Hz. In Fourier based analysis the detection of this note C524Hz is very difficult. Generally those recognition processes fail to detect these notes since it has its frequency components coincident with the harmonics of note C262Hz. We will see if SVD gives information about the presence of this new note on the major chord.

The results are different from those obtained from the chord C262Hz, E330Hz, G392Hz and we can see that the note C524Hz has the highest norm values and it is detected although the harmonics are coincident with the C262Hz harmonics (fig.8.29). The graphs refer to four events of the C major chord with four notes. In the examples *a*, *b*, *c*, the norm value corresponding to C524Hz has the highest values. In the spectra (fig.8.30) these components also have the greater magnitude. In example *d* the magnitude is the second highest but its norm is the smaller of the four notes of the chord, including C262Hz that has a very small magnitude. In all the examples the magnitude of C262Hz is very small and the presence of the note C524Hz increases the norm values of C262Hz. The two norms of notes and C131Hz and G196Hz continue to be present with high values.

In the examples *b* and *d* the gap between the notes of the chord and the other notes (excluding the subharmonics) is noticeable and it is possible to determine the notes of the chord from the rest.

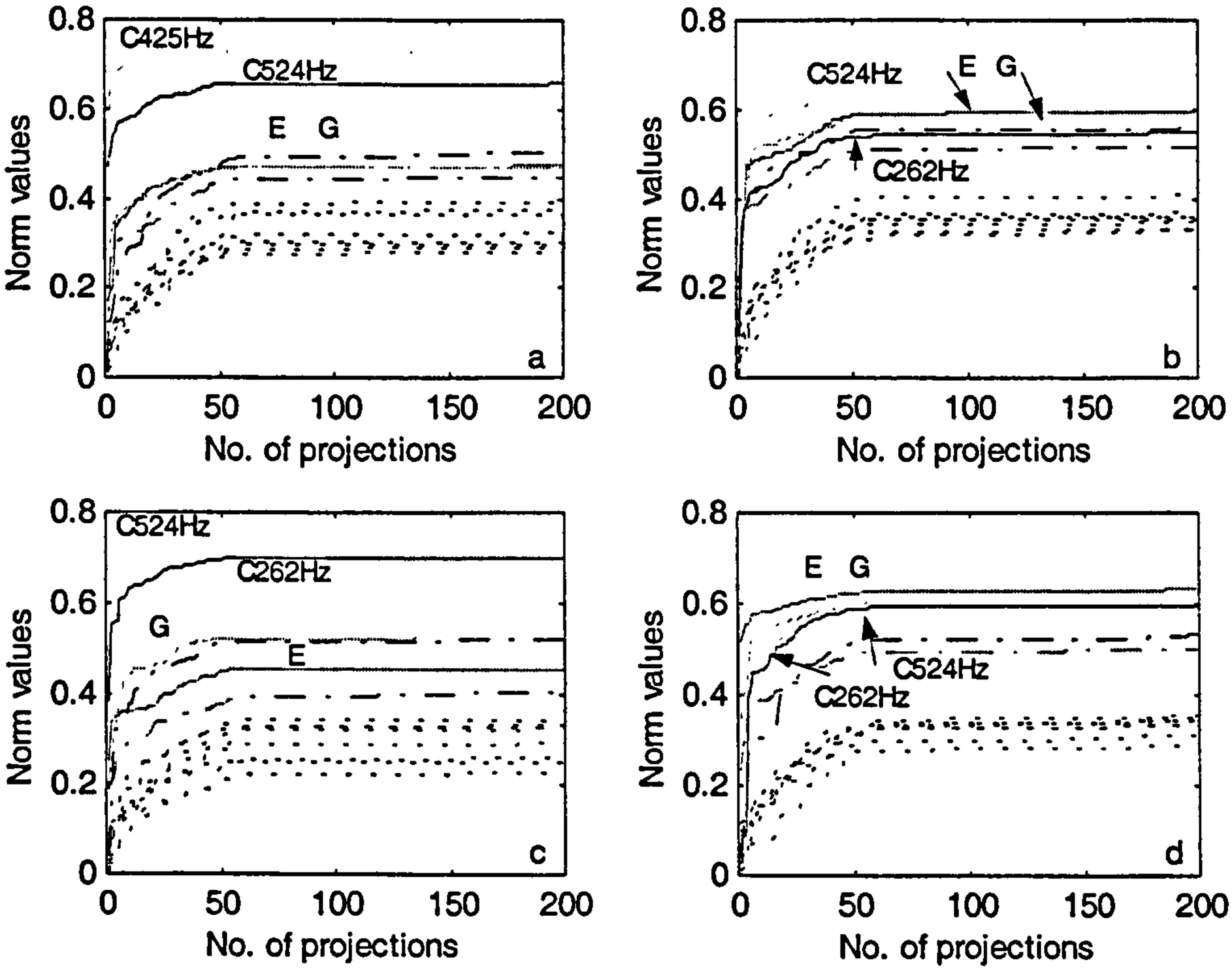


Fig 8.29 Values of the norm of the chord of C major with notes G196Hz, C262Hz, E330Hz and C524Hz.

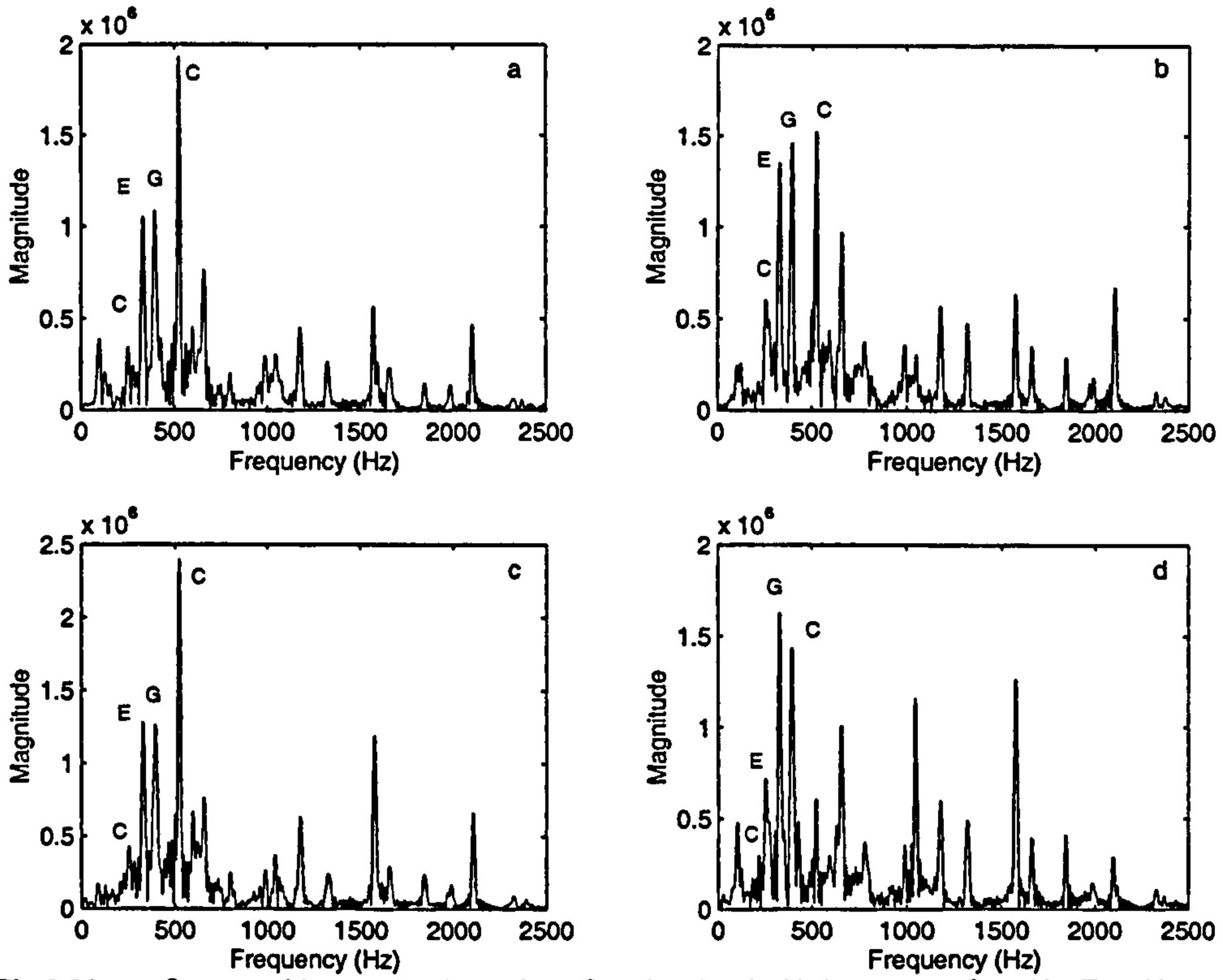


Fig 8.30 Spectra of four realisations of the C major chord with four notes: C262Hz, E330Hz, G392Hz and C524Hz.

The gap between the two notes C524Hz and C262Hz is enough to determine the presence of those notes in the signal. The other two notes are very near the rest of the norm values and we cannot conclude their presence in the signal. But the results are different for the three notes chord and also from the octave C262Hz and C524Hz that indicates that is possible to separate these signals from each other. In this example the SVD is able to indicate the presence of the octave in the signal.

The other C major chord with four notes analysed has the components G196Hz, C262Hz, E330Hz and G392Hz. It has the lower frequency note G196Hz doubled with the octave G392Hz (we show the example of a single realisation since the recorded events of the chord have very similar spectra and norm values). Again those two notes that are separated by an interval of an octave have the highest norm values. The subharmonic C131Hz is also present with a high norm value.

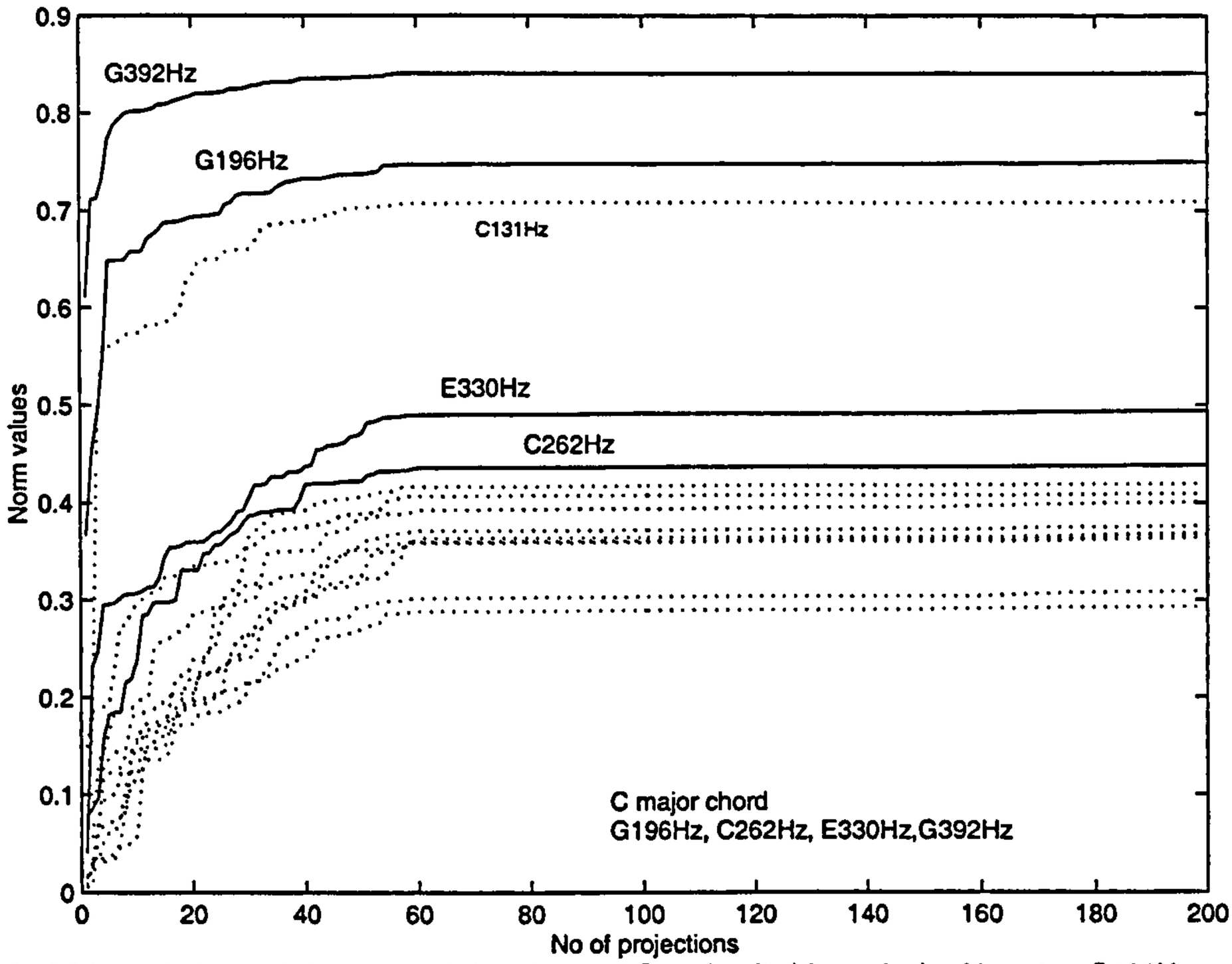


Fig 8.31 Values of the norm of the chord of C major (2nd inversion) with notes G196Hz, C262Hz, E330Hz and 524Hz projected on all basis sets.

The norm values for these major chords with four notes do not seem to give any information about the number of notes present on the signal. For the case of the octave it was possible to know that we had two notes for most of the cases. In the case of the three notes chord that was not possible anymore since there were notes with higher norm values corresponding to C131Hz and G196Hz. If the gap between the notes in the chord and the other notes (excluding the subharmonics) were greater enough we could determine the number of notes in the piano signal. We saw that there is a relation between the magnitudes of the fundamental frequency of the notes and the norm values. That relation fails for the subharmonics since they some times have very small magnitudes and the norm values are very high. This leads to a possible wrong detection of those notes.

The presence of the octave of the first note in the major chords is detected. With the Fourier transform the detection of that note would not be possible since the harmonics of those notes are coincident with the harmonics of the lowest note of the chord. The detection of other notes in the chord is more problematic since in some cases the norms are very near to the notes that do not belong to the chord. However, the results of the octave alone (8C262Hz see fig.8.18) and this chord, with two more notes E330Hz and

G392Hz, are different which allows to know that there are more notes present in the signal than the two notes of the octave.

Again as for the case of the three notes there seems to exist a pattern for this chord and the identification of the notes present in the chord could be performed based on the recognition of a similar pattern.

8.6.3 The seventh chords with four notes

The chords we analysed with four notes had the first note doubled with the octave. We saw that the presence of notes separated by an octave reinforce the detection of those notes in a chord. We will analyse now chords with four notes where none has an interval of an octave and see if it is possible to achieve satisfactory results.

The chord we are going to analyse is composed of single piano notes: C262Hz, E330Hz, G392Hz and Bs466Hz (the s is an abbreviation of sharp and Bs corresponds to the note A#466Hz).

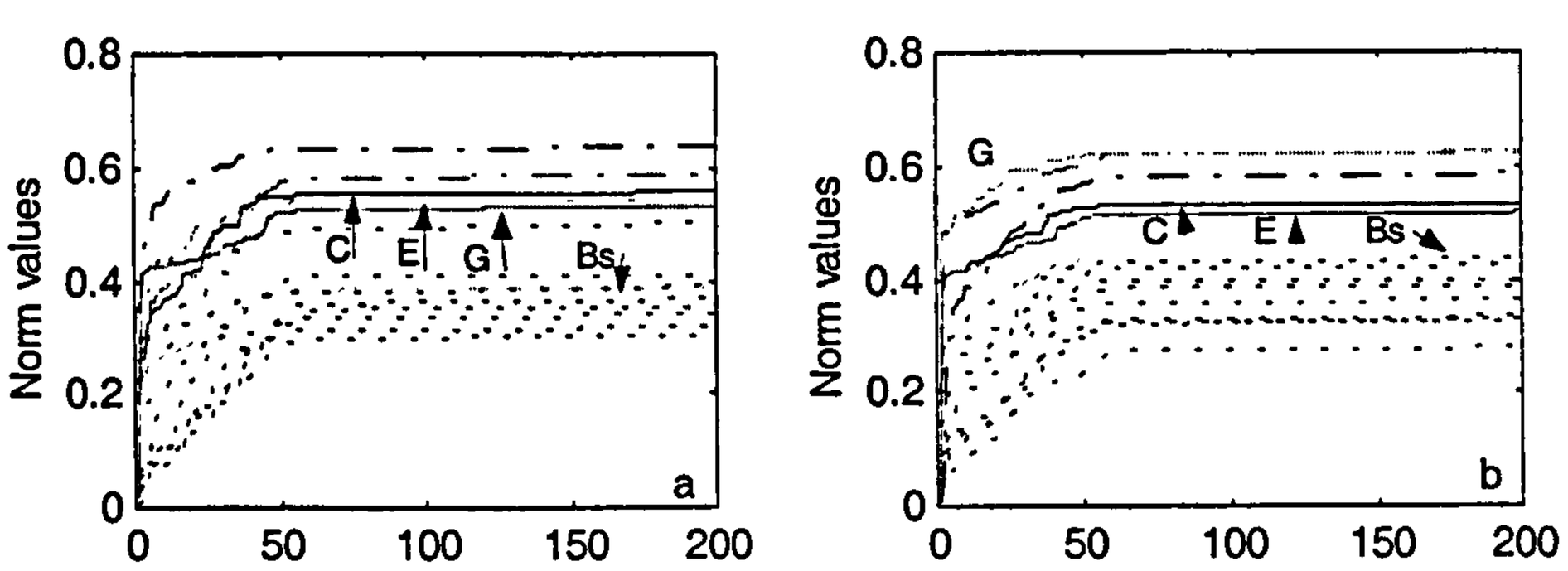


Fig 8.32 Values of the norm of the chord of C7th chord with notes G196Hz, C262Hz, E330Hz and Bs466Hz projected on different basis sets of single piano notes.

Another seventh chord analysed has exactly the same first three notes as the previous chords (C262Hz, E330Hz and G392Hz) and the fourth note is B. We seek to see if the difference on the fourth note will affect significantly the norm values.

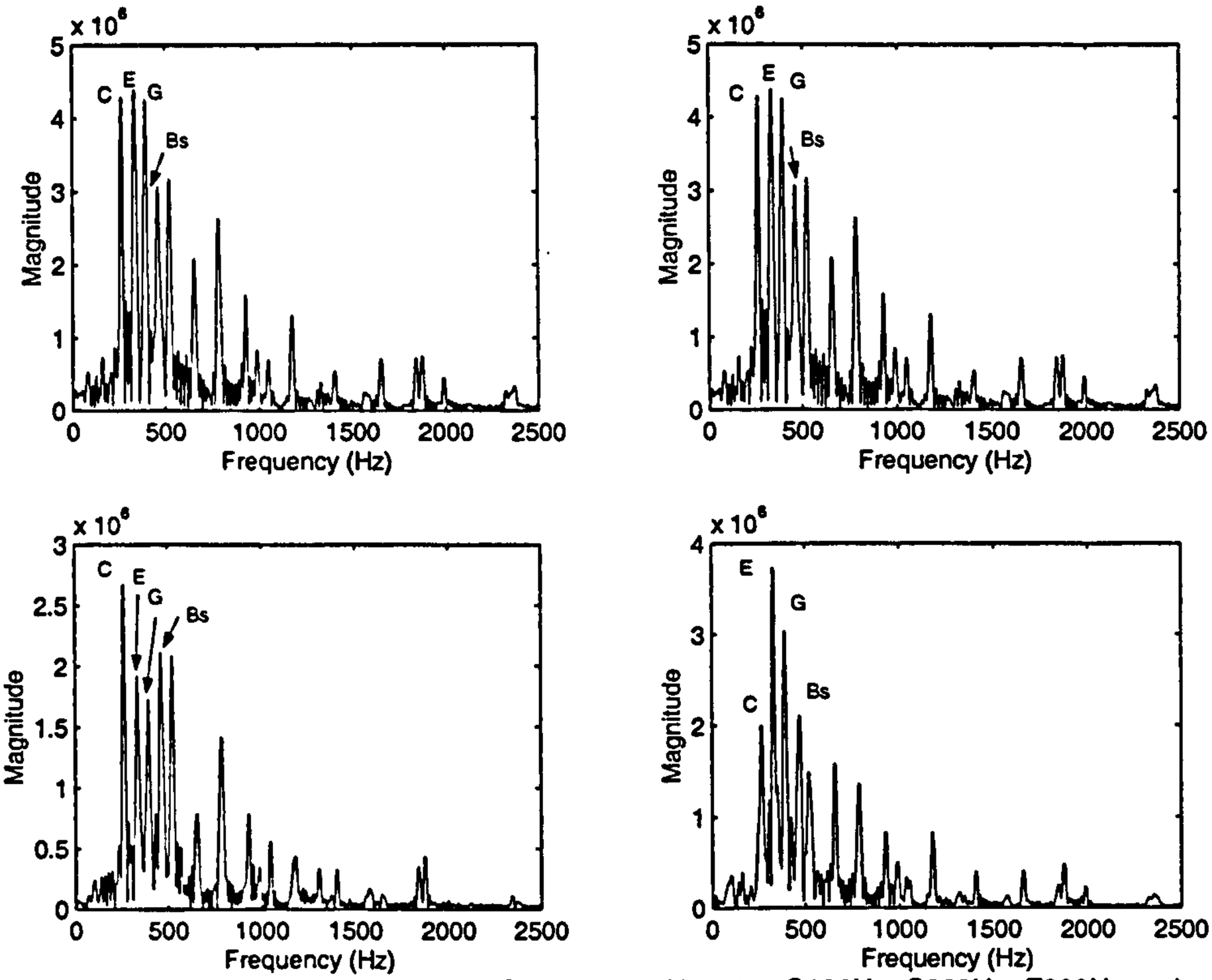


Fig 8.33 Fourier transform chord of C7th chord with notes G196Hz, C262Hz, E330Hz and Bs466Hz.

The norm values in example *a*, that are above all the norm values, correspond to notes G98Hz and C131Hz (the note Bs466Hz also belongs to the harmonic series of C65.5Hz). These two curves continue to be present with high norm values although they do not correspond to single piano notes present in the chord. In the same example the curve immediately under the three notes of the chord (C, E and G) corresponds to Bs233Hz that is a subharmonic of Bs466Hz. The curves of Bs233Hz have greater or very similar values to Bs466Hz. There is a small frequency component very near to the 233Hz.

The results for the different realisations of the same event are different and it does not seem possible to extract the exact information about the presence of notes in the chord or even a pattern for the norm values. What seems to be more constant in all the examples is the presence of the norm values of notes C131Hz and G98Hz. We do not know how the norms would behave if those notes were present in the chord. Their

presence in these examples of the seventh chord could indicate that we have a chord with notes from the C major chord.

The other seventh chord analysed has exactly the same first three notes, C262Hz, E330Hz and G392Hz, and the fourth note is B494Hz (half a tone higher than the note Bs of the previous example). The results present again two curves corresponding to the notes C131Hz and G98Hz. The highest value of curve at the 60th projection is slightly smaller than the values of the previous chord. While before the values where around 0.61,0.64, now they are around 0.57 and 0.6. There is not a special pattern from this chord and not the previous C seventh chord. The only difference is that B now has greater values than Bs. The curves that have higher values than B and are not in the chord correspond to the note B247Hz as we saw before with the note Bs233Hz.

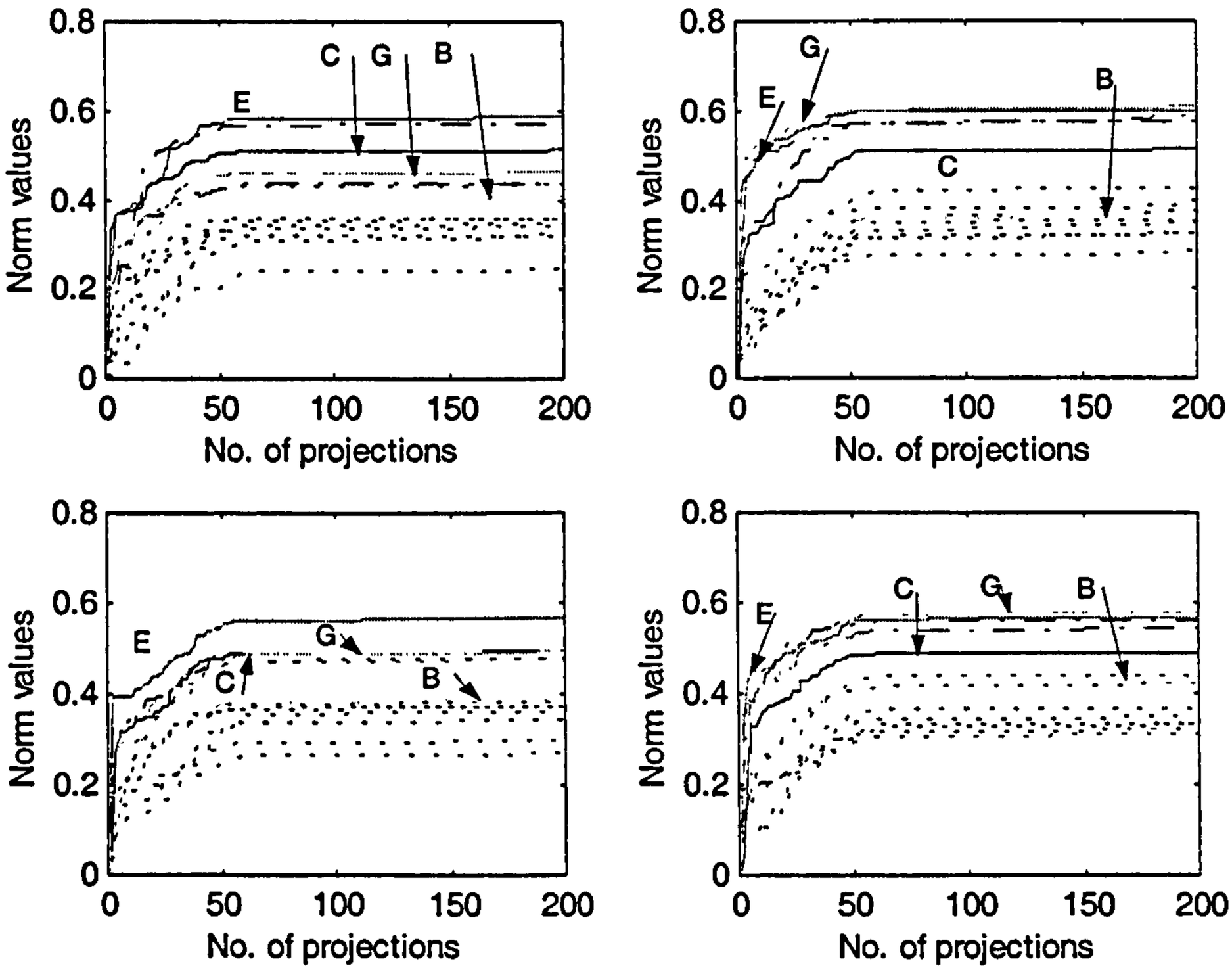


Fig 8.34 Value of the norm of C7th major chord with notes G196Hz, C262Hz, E330Hz and B494Hz projected on different basis sets of single piano notes.

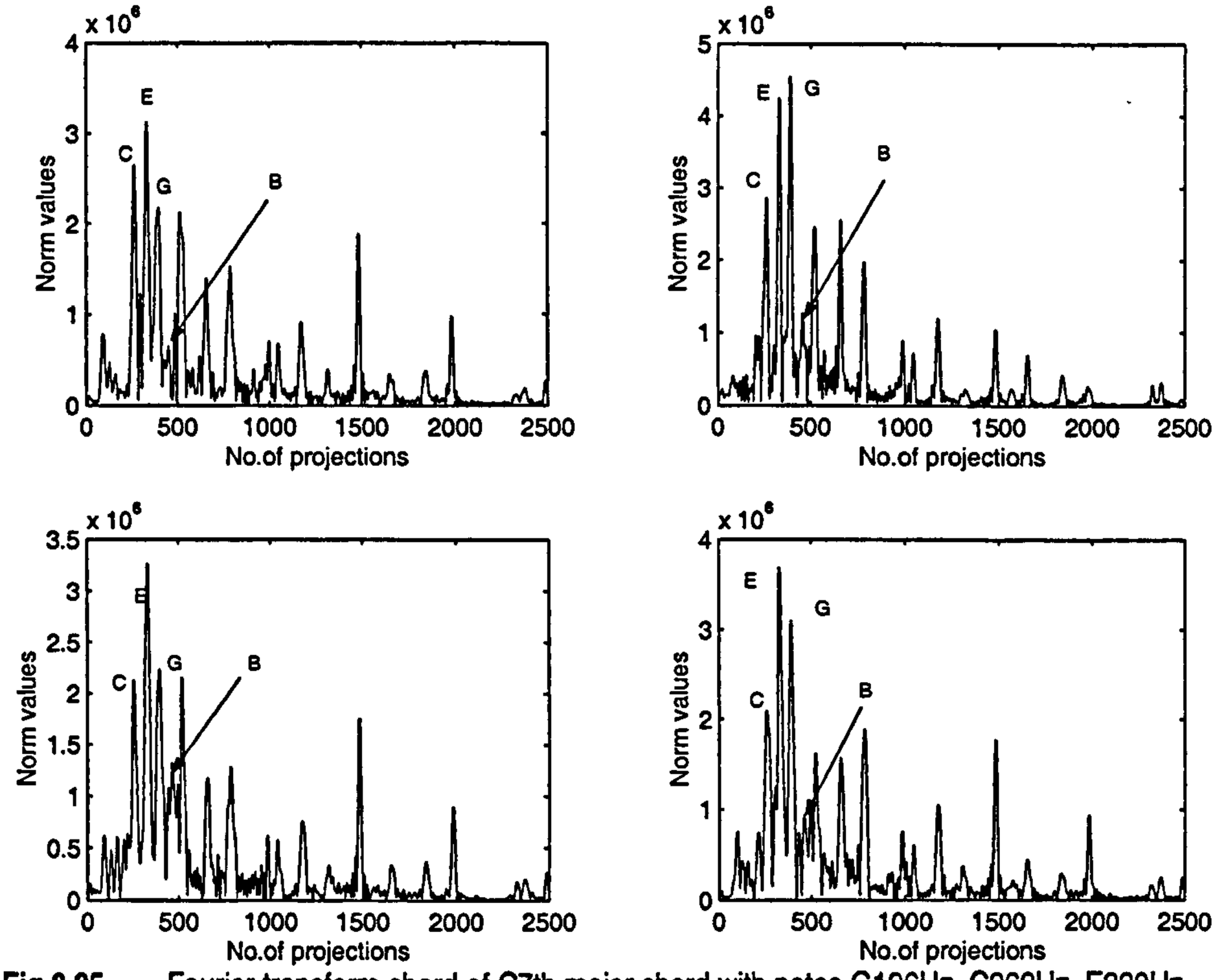


Fig 8.35 Fourier transform chord of C7th major chord with notes G196Hz, C262Hz, E330Hz and B494Hz (FT of 44100 points).

In both chords we have in common the presence of C131Hz G196Hz with the highest values and the notes C262Hz, E330Hz and G392Hz. The same happens for the C major chord and the difference between these chords and the C major is that the Bs and B have the norm curves after the notes mentioned above or at least the subharmonics of those notes that are one octave below. But this is not enough since we cannot establish a line that separates the notes in and out of the chord. Another indication that we do not have a C major chord, but a seventh chord, is the value of the highest norms at the 60^{th} projection is higher for the major chord.

8.7 Conclusion: The performance of the method

The recognition system implemented for piano note recognition has as the fundamental data analysis technique the SVD. The structure of the process is a set of all

the singular vectors representing each one of the piano notes. Those singular vectors are the reference of the original notes where the new signals are projected for recognition. The measure of the presence of a note on a signal is given by the norm of the projections of a signal in different number of vectors of each basis set. The curve of the norms have a shape, for real piano notes, that reaches a step at a certain number of projections that is directly related with the number of events in the time history matrix of the single note. The values of the norm at that step are the most important for our analysis since it will be used as the reference for the presence of the note on the signal.

The process has a good performance in the recognition of single notes. All the single notes recorded were perfectly recognised. The norm values of signals composed by a single note reach the maximum value of the norm practically at the 60^{th} projection. All the other notes have smaller values at that projection. The results for the single notes not only allow the information about which note is present on the signal as also indicate that there is only one note on the signal.

The identification of the octaves, one of the most difficult problems in automatic music transcription, is solved for some of the cases analysed (8A#233Hz, 8C262Hz), using 2000 samples of the signal (and singular vectors with 2000 samples each). The results of the norm compared with the Fourier transform of the signal (2000 samples of the signal) indicate that the number of samples used did not contain enough information for octaves composed by notes with frequencies of 131Hz and 98Hz. The number of samples used was increased to 3000. The results for the octaves where the recognition of both notes was not always possible improved and the norm values, of the note of the octave that are mixed with the other notes, are now above those values. Although the recognition of both notes of the octave was not always very clear for all the possible realisations of the octaves 8G98Hz and 8C131Hz we could be assured that the signal was an octave since there is only one detached value of the norm with a smaller value than for a single note. So we could conclude that we were in the presence of an octave. The note that is detached from others corresponds to the higher frequency note of the octave. The octave cannot be taken as a chord since the values of the norm are higher than the values of norm of the chords. While the detection of octaves is possible with SVD the detection of notes separated by a tone interval is not very clear. With the Fourier transform we have the opposite results: is very difficult to detect an octave but the results are very good for notes separated by small intervals.

The performance of the process on chords gives interesting results. We analysed in detail some chords with three and four notes. The chords chosen were the C major chord, in different inversions, and the seventh chords starting in the same C262Hz note. The identification of all the notes in the major chords with three notes is in some cases possible but there is the problem of the wrong identification of notes that have high norm values and are not in the chord. These notes are subharmonics of the notes of the chord and they are in the harmonics series of the note C65.5Hz. Although the Fourier transform presents in some cases components at those frequencies with small magnitude (smaller than the components at frequencies of the fundamentals of the notes played) the norm values are very high. The presence of this notes difficult the recognition of the notes really present on the signal. However they can indicate that we are in the presence of a C major chord.

When the chords have four notes and the fourth note has an interval of octave with the first one it is possible to detect the presence of that note since it has always the higher norm values. While with Fourier transform it is not possible to detect that note with SVD it is possible to do so. Again the two curves of G98Hz and C131Hz have high values. We remember that the three notes C262Hz, E330Hz and G392Hz were present in all chords (except for the three notes chord C major in the 2^{nd} inversion where G392Hz was absent) and they belong to the harmonic series of C65.Hz. Probably all the chords that have those three notes will present high values for the notes of the series.

For the case of the seventh chords again the three notes C, E, and G are present and the fourth note is Bs or B. Both Bs and B are in the harmonic series of the note C65.5Hz but the chord with the notes C262Hz, E330Hz, G392Hz and Bs466Hz are in the harmonic series exactly by that order (they are the 4^{th}, 5^{th}, 6^{th} and 7^{th} notes of the series) the note B is in a high order (although the notes are in the harmonic series the fundamental frequency of the single notes is equally tempered and their frequency is not exactly the same of the notes in the harmonic series). The results of both chords present similar characteristics in the sense that the notes that have highest values are C262Hz, E330 Hz, G 392Hz and the two subharmonics. We have the same for the C major chord. This makes the recognition of these seventh chords difficult. Some leads for the exact recognition of the notes of the chord is the presence of the 7^{th} notes Bs or B (or the subharmonics) on the high values and the value of the highest norm. When there are more notes in a chord the highest values of the norms are smaller. We saw that we have an exception when the four notes chord correspond to a major chord wit the highest

frequency note at a distance of an octave of the lowest one. The highest values can sometimes fall in the same interval as the same chord with three notes (around 0.7). For the seventh chords the high values can fall around 0.6.

Summarising the method has very good results for the recognition of single notes. All the recorded notes were perfectly recognised. For two notes signals it has a good performance in the recognition of octaves, where generally the Fourier based techniques fail. For chords where the notes are separated by small intervals the recognition of both notes is difficult while with the Fourier analysis is easy. When the piano signals are composed with more notes, 3 and 4, the identification becomes more complex. Some of the highest values of the norm correspond to absent notes on the chords: C131Hz and G196Hz. These notes are subharmonics of the notes (or of some of the notes in the case of the seventh chords) in the chord and they are in the same harmonic series of the noteC65.5Hz. Since the criteria of the notes detection is high norm values they could be wrongly detected. The same C major chord (3 notes) played in different inversions (G196Hz, C262Hz, E330Hz and C262Hz, E330Hz, G392Hz) does not present much difference for both cases. The note G196Hz is in the chord for one case and out for the other and the norms for both cases do not present distinct features.

CHAPTER 9

CONCLUSION

9.1 Summary of results

In this research we have presented a new approach to the automatic transcription of music. Previous work in this area, in general used Fourier based analysis. All the approaches considered present the same problem of the identification of notes with coincident harmonics. Early methods fail to detect the presence of such notes or can lead to errors in detection. Following a review of the existing methods we decided to use the KLE. In KLE the basis sets in wich the signal is represented are determined by the signal itself instead of the "imposed" sines and cosines of the Fourier transform. Another important characteristic is that they can be applied to nonstationary signals.

We demonstrated theoretically that the KLE is similar to SVD and presented practical results verifying this. While for the KLE we need to determine the autocorrelation function of the signal and built an autocorrelation matrix in order to determine the eigen values, for the SVD we only need the time histories. For reasons of simplification it was decided to replace the KLE with the SVD.

In this thesis we analysed simulated and real piano signals. The piano notes were considered as nonstationary stochastic processes. The KLE and SVD were applied to nonstationary simulated signals and SVD to real piano signals. The KLE of

nonstationary signals is applied to the nonstationary autocorrelation matrix determined by ensemble averaging and this is equivalent to the product of the transpose of the matrix by itself. We show that SVD has similar results for the most significant vectors when applied directly over the time history matrix of the possible events of the nonstationary process. In this way is we were able to replace the KLE by the SVD.

Based on the SVD technique we proposed a recognition process for piano signals. This method uses a priori data for each piano note that is decomposed using SVD. The singular vectors, of the resulting decomposition of each note forming a set of orthonormal vectors, are the reference for the identification of future piano events. So the methodology is that we have single notes as the reference, represented by singular vectors, and they are used to recognise any piano signal composed of one or more individual notes. Any signal to be recognised is projected on the sets of singular vectors representing each note.

A first attempt with the process was performed with the simulated notes. The results of this simple and highly controlled situation helped to establish an approach to the real piano notes. We performed the recognition of simulated single notes and octaves and showed that it is possible to distinguish between the octave and the notes in the octave (the note with the lowest frequency).

In the case of real piano notes the recognition method works with a high accuracy for single notes. All the single notes analysed were perfectly recognised. It is possible to perform the recognition of the notes present in an octave with this method. Some factors have to be taken into account for the recognition process to work. The first concern is the right construction of the time histories of the single notes. The 60 events we used covered a great variation on the way a note is played and worked well. The SVD of the aligned possible events of a single note show that the variability of the system is small. The energy is mostly concentrated on the first vectors and the other vectors have very small associated values. Although there are many variables influencing the piano note production the final results seem to indicate that generally the notes are almost amplified versions of each other. Another important factor was to allow the presence of delays between events. Those delays must not exceed a certain value when the human ear starts to distinguish between two notes. With the introduction of the delays the identification of the notes becomes simpler since there is no need to determine the exact starting point with great accuracy.

The recognition of the notes present in a signal composed of two notes as in the case of the octave was resolved, with this method, for the cases presented. For the octaves 8C262Hz, 8A#233Hz and 8G196Hz the recognition was possible with 2000 samples representing approximately 0.045s of the signal. For other octaves such as 8C131Hz and 8G98Hz the identification is more difficult. The increase of the number of samples for these two cases improved the performance of the method but demanding more computation time. The data length is still not enough for octaves with lower frequency components. However in all the cases the note of the octave with high frequency is always detected and the values of the norm allow the identification of the octave with only that note.

For signals composed of two notes separated by tone the method it is not always able to detect the two notes. In one of the cases presented the two notes have the maximum values but we cannot infer the number of notes present.

For more complex sounds composed of three and four notes the results are ambiguous. One of the most significant results is the high norm values of some notes out of the chords that belong to the harmonic series that contains the chord. So we detect the notes of the chord and other notes in the same harmonic series. It seems that with this process we manage to detect the notes of belonging to the same harmonic series even when they are not present in the signal.

In summary, the proposed recognition method is able to detect the notes present in an octave. The performance of the method was proven to improve with the increase of the number of samples (especially for signal with low frequency components). The application to chord recognition shows considerable promise, but it is not always successful.

9.2 Possible future work

The recognition process using SVD as the fundamental technique of data analysis works with very good results in the recognition of single notes played on a piano. In this present work we only recognised notes played on the same piano. The recognition of notes played on other pianos would be another step forward in the development of the

process by the inclusion of those notes in the time histories of the single notes. This inclusion of notes from different pianos would transform the process to a more generalised system of recognition and probably any piano note from any piano could be recognised.

We saw that when the number of samples of the signal increased from 2000 to 3000 (the size of the events in the time history matrix also increased from 2000 to 3000 samples) the performance of the recognition process improved. The performance of the method could be verified for higher number of samples to determine if this improvement continues to exist. This would imply large computation time in the determination of SVD and the projections. However the number of projections can be reduced beyond that corresponding to the number of possible events in the ensemble thereby the computation.

Pitch Identification with an Auditory Model

In the auditory model proposed by Meddis and Hewitt one of the most important stages is the basilar membrane simulation. That is achieved with a filter bank with 128 digital critical-band filters (gammatone filters). The centre frequencies of these filters are equally spaced between 80Hz to 8KHz in an equivalent rectangular bandwidth scale (ERB). The use of 128 filters will provide a good resolution for the harmonics while details in the temporal domain will be lost. This model was revisited again (Meddis & O'Mard 1997) and it was used now 60 linear fourth order gammatone filters between 100 and 8KHz.

Another model proposed by Van Immerseel and Martens (1992) has a filter bank with only 20 band pass filters. The design of the filters of the bank is based in the psychophysical data on simultaneous masking (Zwicker and Feldtkeller, 1967) and it is achieved in two stages. The first consists in a decimation unit and the second in a filter bank of fourth-order Butterworth band pass filters. This model was the one followed to simulate the basilar membrane filtering.

The decimation unit

The decimation unit has three equal decimation filters (FIR low-pass filters) whose coefficients are in table A.1. The decimation unit produces four decimation products that are sampled at rates $fs/2^K$ (k=0,1,2,3). Supposing that the sampling frequency of the input signal is 1024Hz, the first output has a sampling frequency of 1024Hz. The second one has a sampling frequency of 512, the third 256Hz and the fourth 128Hz.

m	h_m
0	0.365
1	0.2833
2	0.1078
3	-0.0260
4	-0.0558
5	-0.0237
6	0.0036
7	0.0064

Table A1.1 The FIR low pass filter coefficients $H(z) = \sum_{-7}^{7} h_m z^{-(m+7)}$

A similar decimation unit was designed using the same data for the low pass filters in table A.1. The input signal is composed by a set of equal amplitude harmonics in sinus phase with frequencies of 50, 150 and 250Hz sampled at 1024Hz. The output in time domain and spectra for this decimation unit is in figures A1.1 and A1.2.

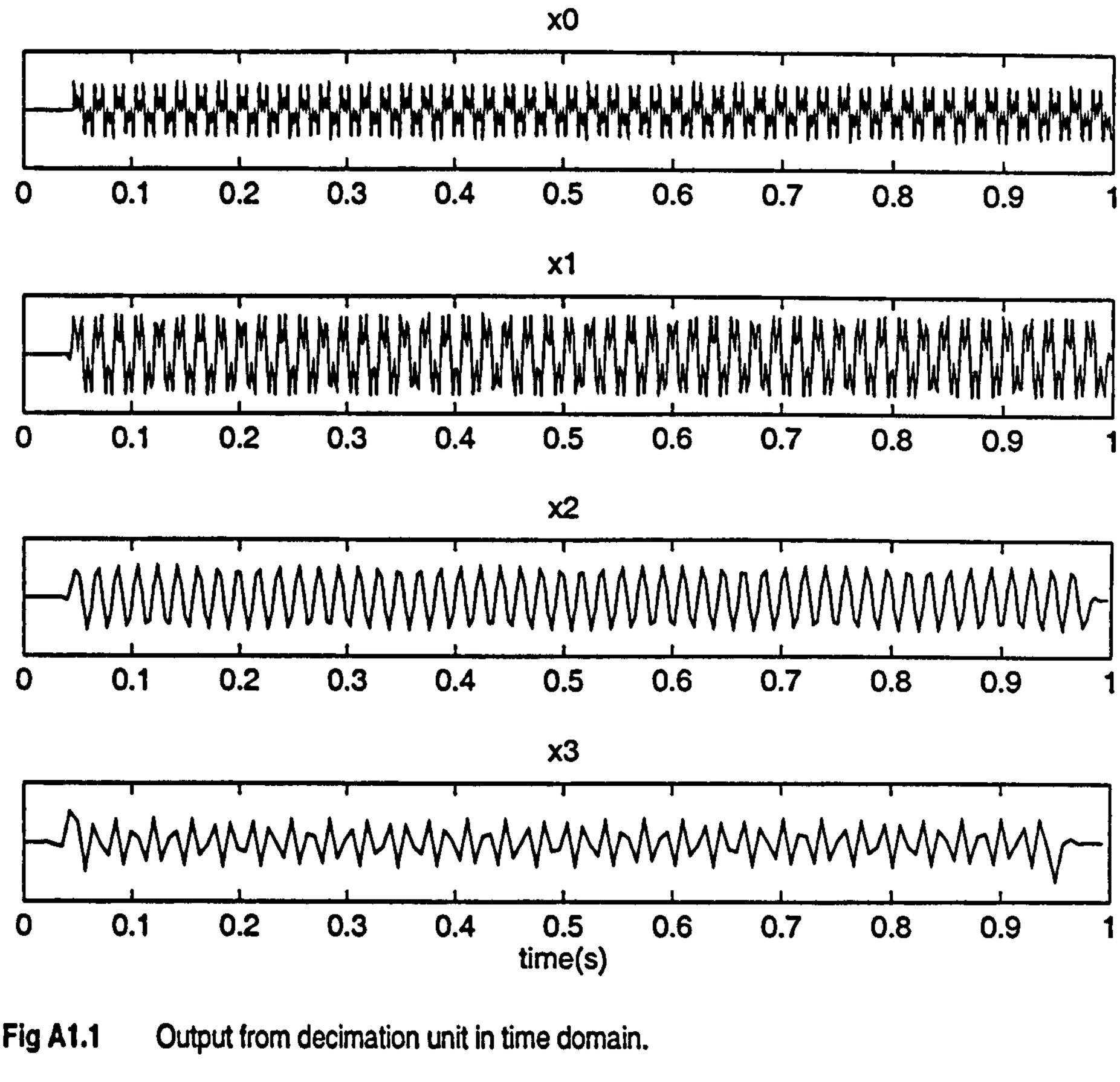


Fig A1.1 Output from decimation unit in time domain.

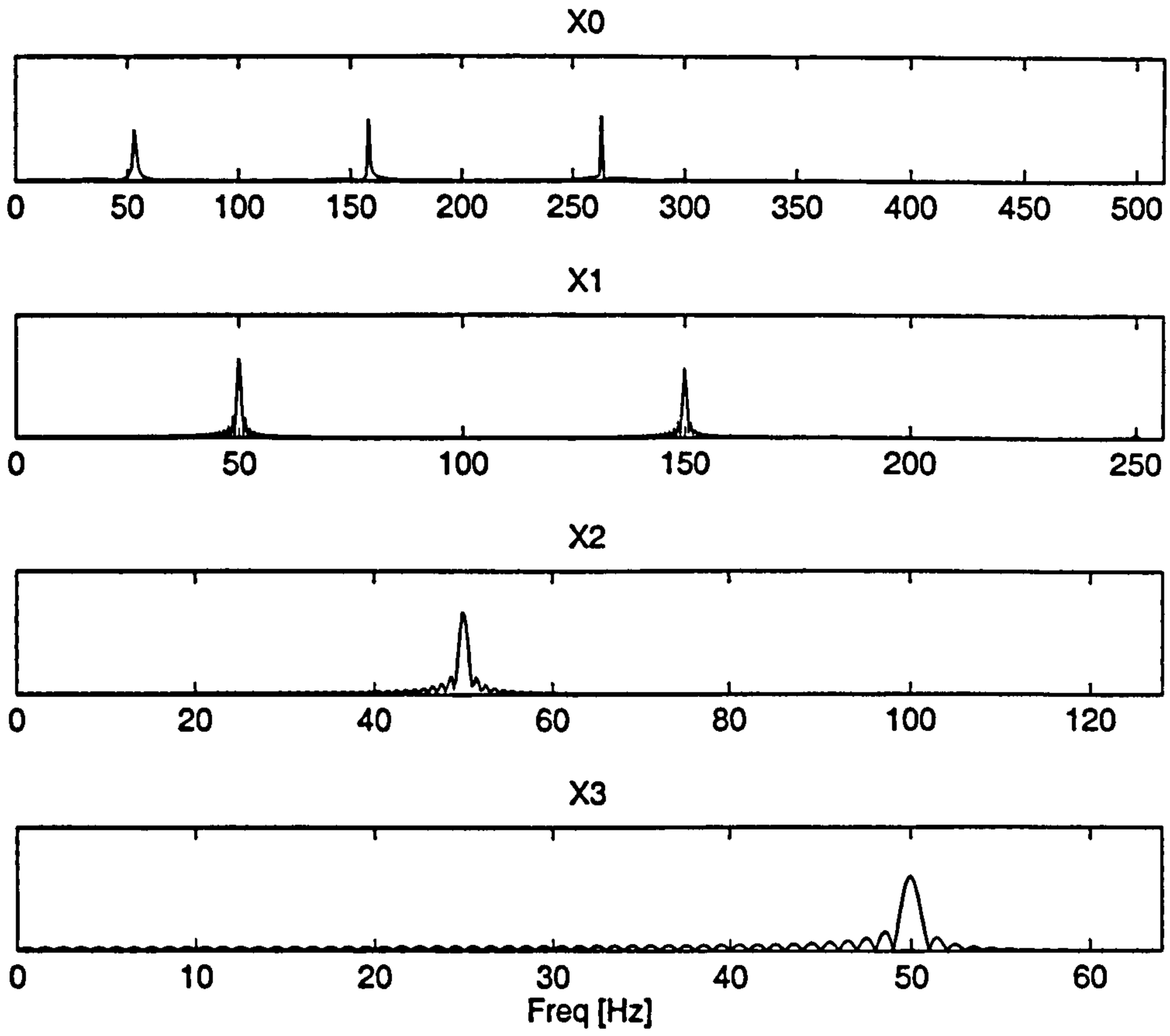


Fig A1.2 Output from decimation unit - frequency domain.

For each output X0, X1, X2, X3 the sampling frequency is respectively 1024, 512, 256 and 128Hz, and all the harmonics that have frequencies above half of the sampling frequency are previously filtered this way avoiding aliasing. As an example the sampling frequency in X2 output is 256Hz and it can be seen in the graph of fig.A1.8 that all the harmonics of the input signal with frequency above 128Hz are not present (150 and 125Hz). The only harmonic present is the 50Hz components.

The outputs from the decimation unit will be connected to input of the filter bank.

The filter bank

The next step is concerned with the filter bank. Fig.A1.3 shows the frequency responses of each of the band pass filters designed following roughly the steps indicated by Van Immerseel and Martens (1992). The filters are digital bandpass Butterworth filters of fourth order with centre frequencies mapped according to a critical band units scale (Zwicker 1990) showed in table A1.2 (it was used 18 filters). The total frequency response is in fig.A1.4 and the total impulse response in fig.A1.5. The Nyquist frequency rate of the filters is 10 KHz.

No. of the filter	Centre freq.(Hz)	Critical BW (Hz)
1	150	100
2	250	100
3	350	100
4	450	110
5	570	120
6	700	140
7	840	150
8	1000	160
9	1170	190
10	1370	210
11	1600	240
12	1850	280
13	2150	320
14	2500	380
15	2900	450
16	3400	550
17	4000	700
18	4800	900

Table A1.2 Centre Frequency mapping and the number of the correspondent filter in the bank.

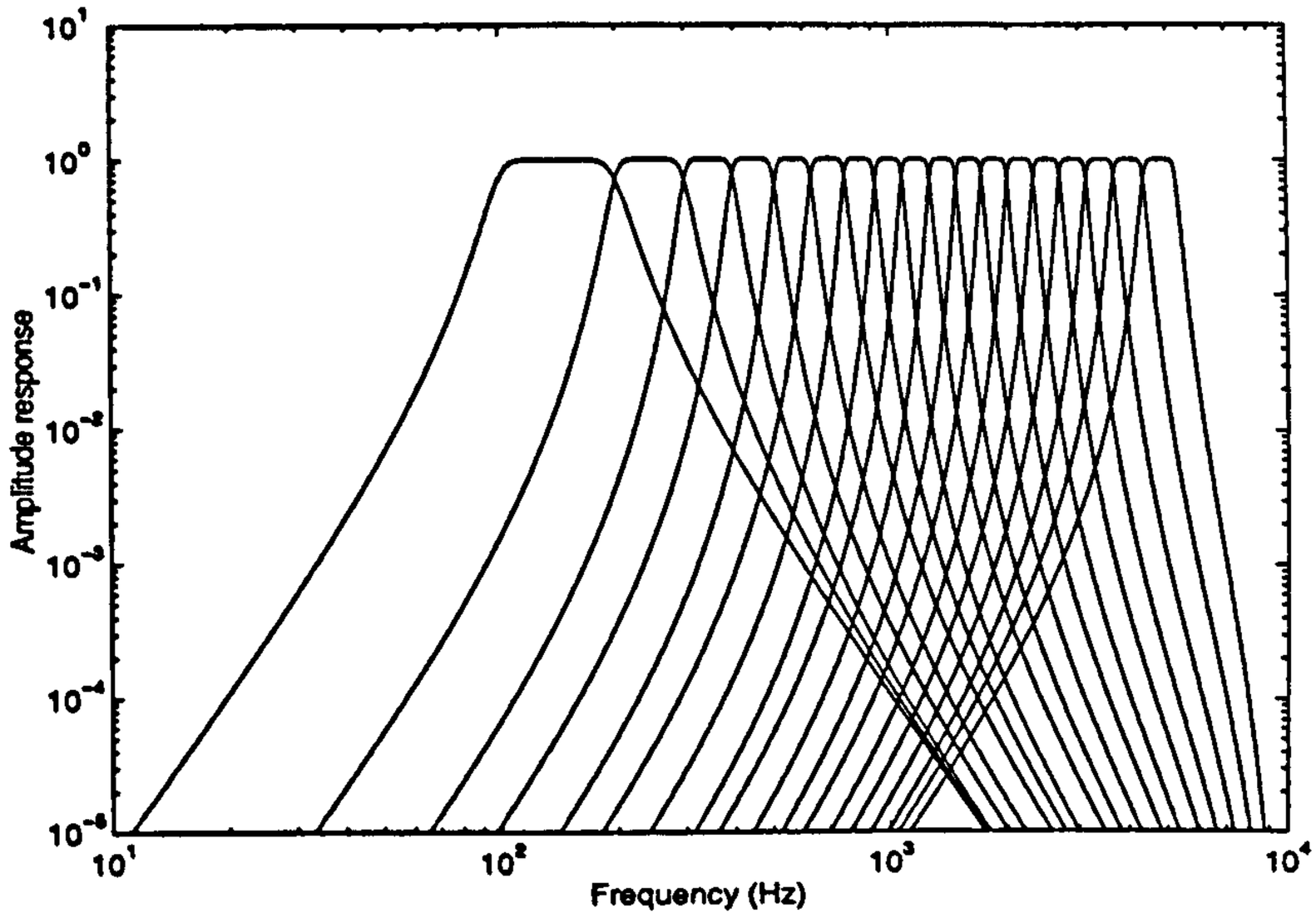


Fig A1.3 Frequency response of each filter of the bank.

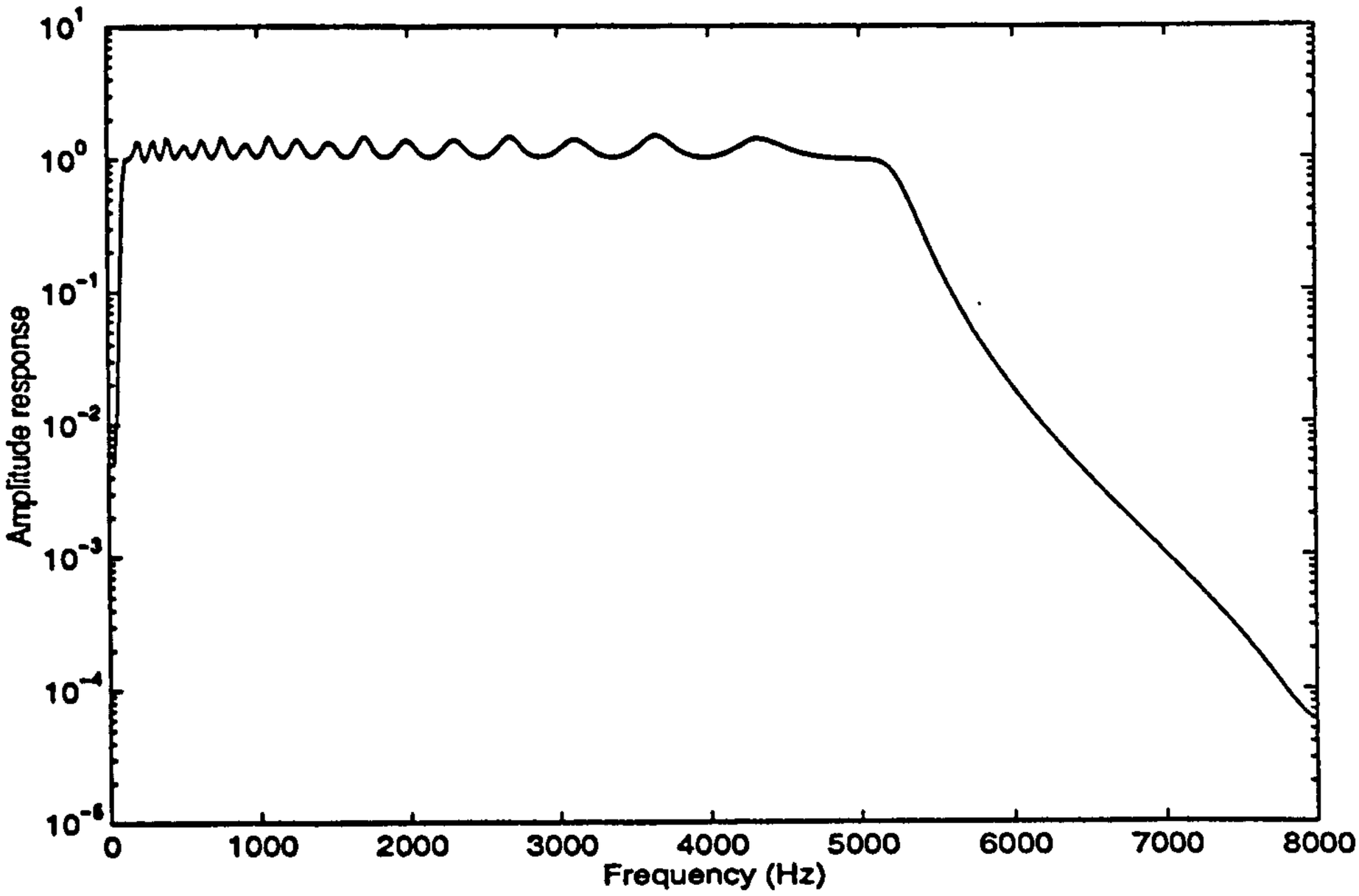


Fig A1.4 Total frequency response of the filter bank.

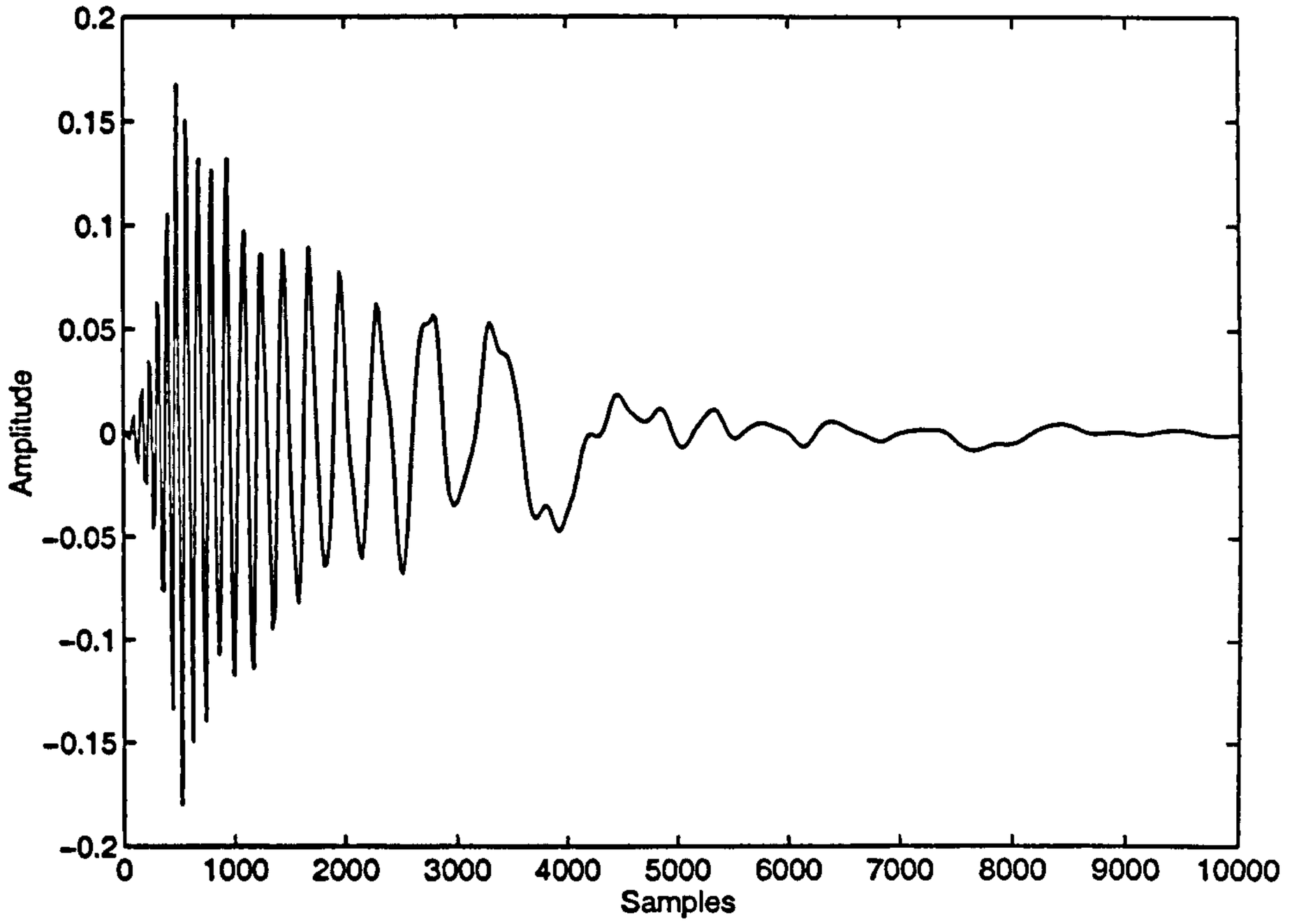


Fig A1.5 Total impulse response of the filter bank.

The impulse response of the filter bank for a signal with components of pure sine waves with zero phase, equal amplitude and frequencies of 450 and 1000Hz sampled at 20KHz (these frequencies correspond to the centre frequencies of two of the filters of the bank nos. 4 and 8 - see table A1.2) is in fig.A1.6.

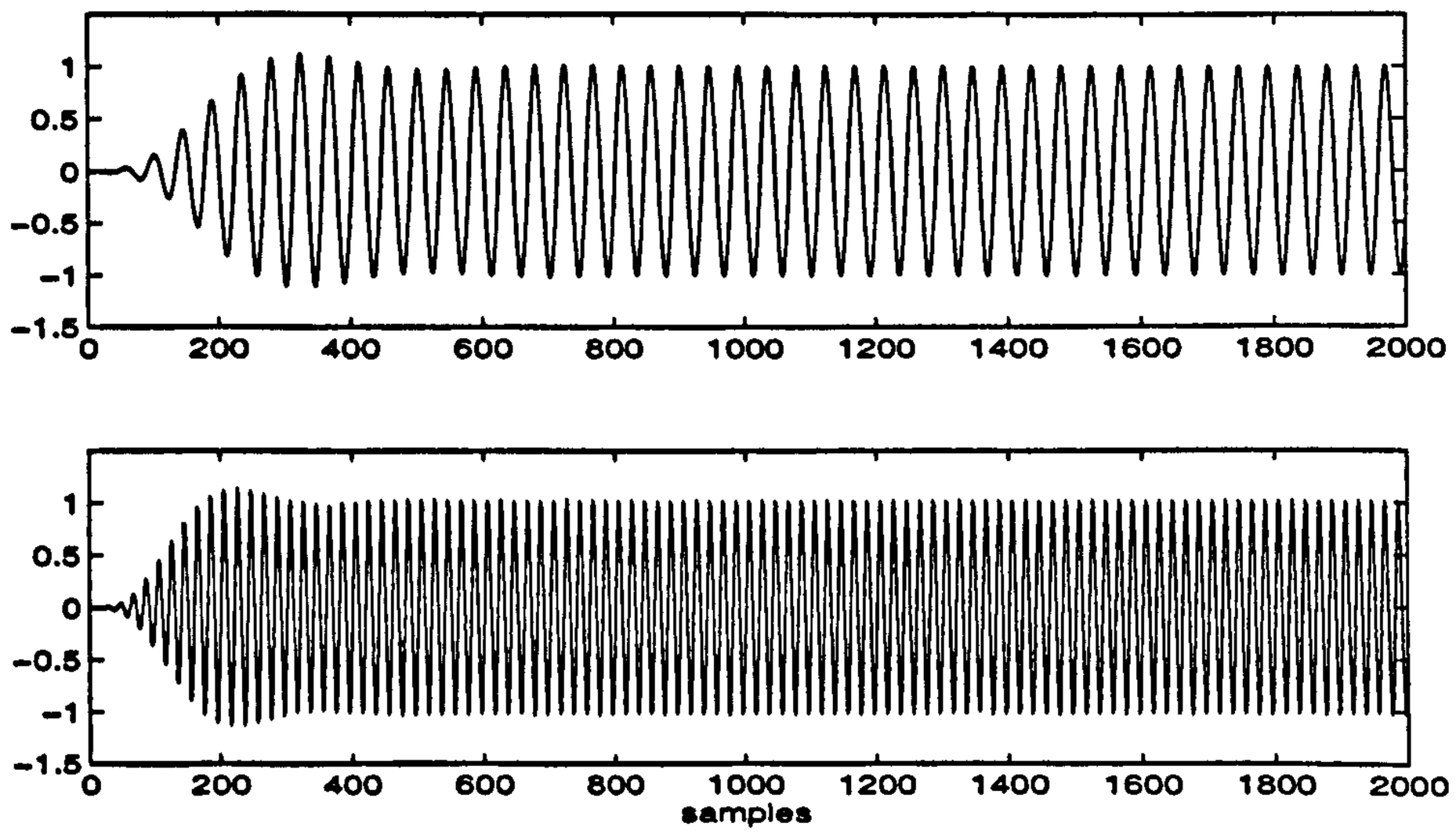


Fig A1.6 Output of filters 4 and 8 of the filter bank for a signal with harmonic components of 450 Hz and 1 KHz.

ANNEX 2

Pitch Identification by Summary Autocorrelation Function

The Meddis and Hewitt (1991) model has a processing sequence where the first stages simulate the behaviour of the outer and middle ear effects (fig.A2.1). The next stage performs the basilar membrane filtering. The last stages produce the autocorrelograms for the different channels that represent the nerve fibres of the ear. The autocorrelograms are aggregated to produce a summary autocorrelation function that will give information for the pitch detection. According to Meddis and Hewitt (1991) the autocorrelation function contains the necessary information for the purpose of simulation of the human listener.

The input signal is low pass filtered (outer and middle ear) and then is filtered in a filter bank. Each of the outputs from the filter bank passes through a hair cell simulator. The output of this simulator is, for channels carrying a high amplitude filter bank output, approximately a compressed half-wave rectified version of the input. For low amplitude filter bank output The output of the hair cell simulator follows the input with little rectification or compression (Meddis and Hewitt 1991). The correlation between these output signals from the filter bank and hair cell simulator gives the information for pitch identification. The summary autocorrelation function is the sum across all these channels.

Autocorrelation is the method that extracts the periodicity and the summary autocorrelation function represents the pitch information.

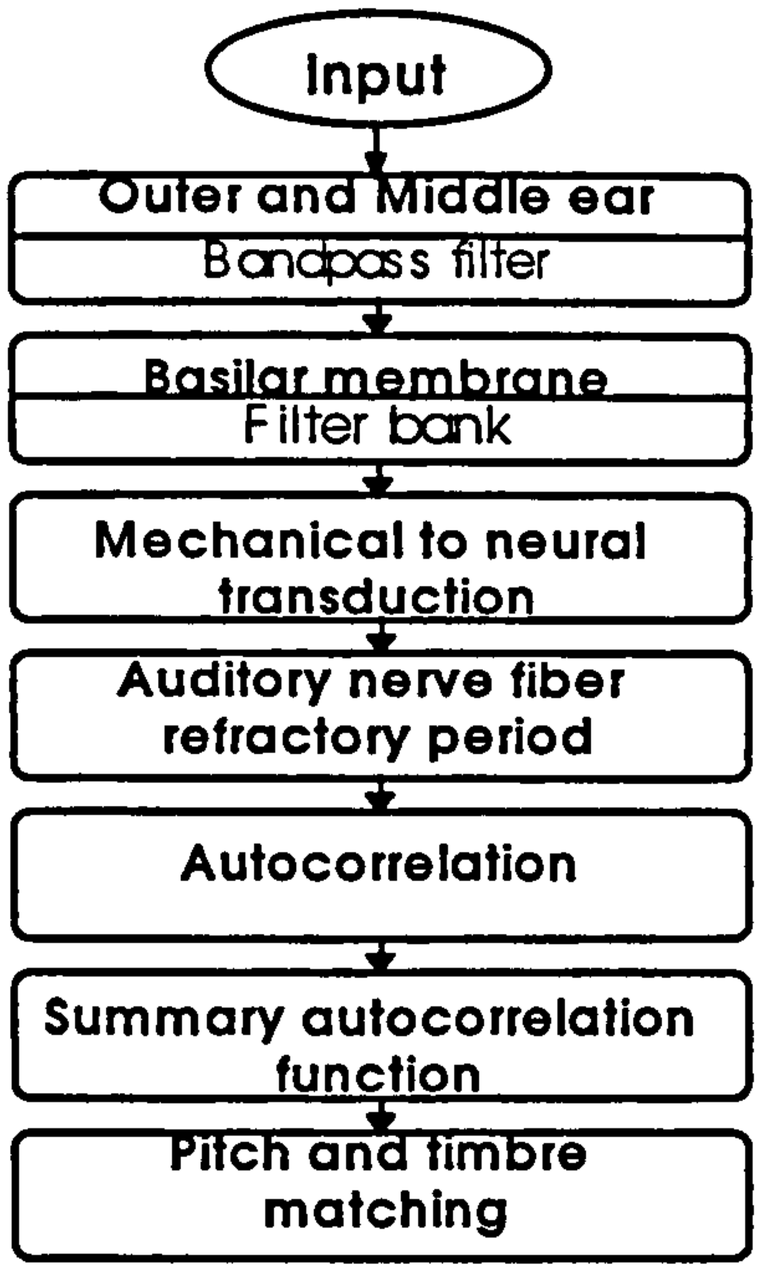


Fig A2.1 Meddis' Model Sequence.

A simulated autocorrelogram was performed using as an output from the filter bank a sine wave. The correlation gives the relation between the sine wave and its half wave rectified version. The expression for the autocorrelation function is in this case:

$$R = \int_{a}^{b} \sin(\omega_0 t).\sin\omega_0(t+\tau)dt \qquad \text{(A2.1)}$$

Since the rectified wave has half of the period equal to zero the limits of integration will be:

$$R = \int_{0}^{(\pi/\omega_0-\tau)} \sin(\omega_0 t).\sin\omega_0(t+\tau)dt \qquad \text{(A2.2)}$$

The final expression for the autocorrelation function is:

$$R = \left[(\pi / 2\omega_0) - \tau / 2\right]\cos(\omega_0\tau) + (1/2\omega_0)\sin(\omega_0\tau) \qquad \text{(A2.3)}$$

This equation (A2.3) can be rewritten as follows:

$$R = \frac{1}{2\omega_0}\left[(\pi - \omega_0\tau)\cos(\omega_0\tau) + \sin(\omega_0\tau)\right] \qquad \text{(A2.4)}$$

with

$$0 \le \tau \le \frac{\pi}{\omega_0} \qquad \text{(A2.5)}$$

The next figures A2.2 and A2.3 show the autocorrelograms for different values of ω and the summary autocorrelogram function. The summary autocorrelogram function (fig.A2.3) is the sum of the autocorrelograms of fig.A2.2.

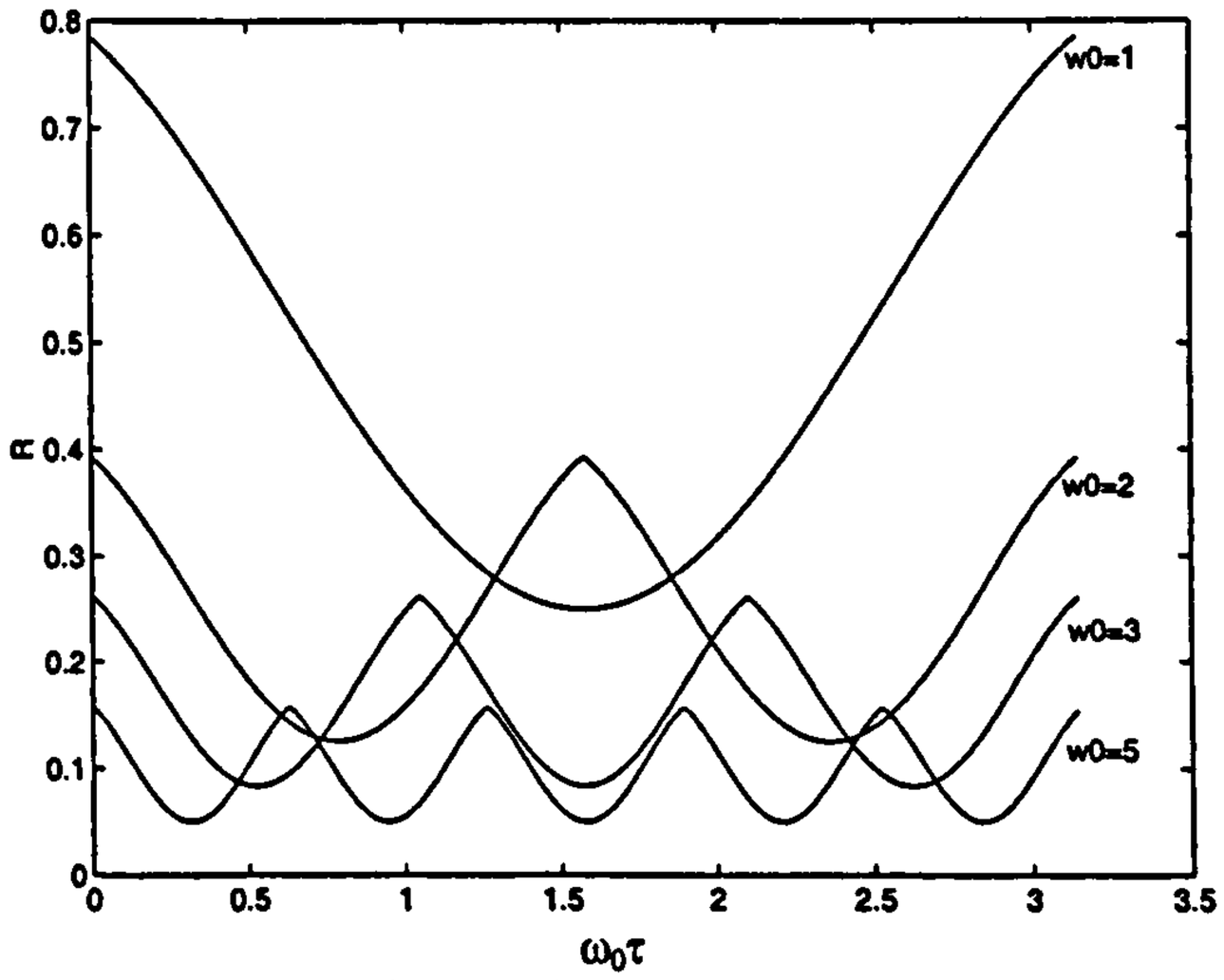


Fig A2.2 Autocorrelation function

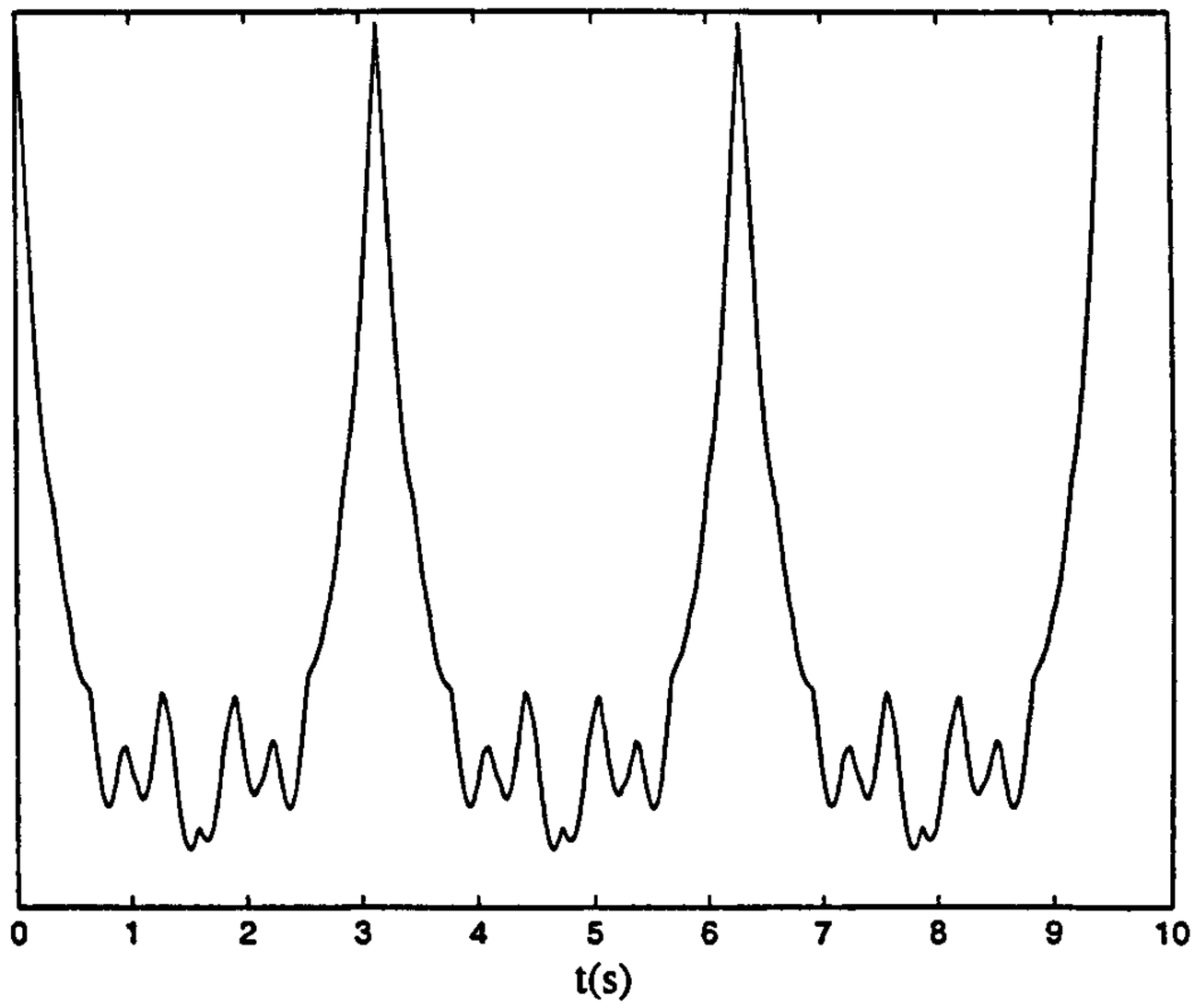


Fig A2.3 Summary autocorrelogram function (ACF)

The first peak in the summary autocorrelogram function graph is at t=3.14s. This peak corresponds to the fundamental frequency. As we have seen in (A2.5) the values of $w_0\tau$ stay between zero and π. For $w_0 = 1$ the value of τ is equal to π which corresponds to a frequency of $1/\pi$ which is the same value of the fundamental frequency represented in the graph.

ANNEX 3

Proof of the Karhunen-Loève Expansion

To prove the KLE we have to verify if equation 4.8 is satisfied when a stochastic signal is represented by the KLE.

Consider a stochastic process $X\mu(t)$ defined on the interval $a \le t_1 \le b$ and $a \le t_2 \le b$. The autocorrelation function $R(t_1, t_2)$ is expressed as:

$$R(t_1, t_2) = E\left[X_\mu(t_1) X_\mu(t_2)\right] \tag{A3.1}$$

Replacing the stochastic process $X(t)$ by its KLE expansion (equation 4.6) on equation A3.1, the correlation becomes:

$$R(t_1, t_2) = E\left[X_\mu(t_1) X_\mu(t_2)\right] = E\left[\sum_{n=1}^{\infty} \sqrt{\lambda_n} c_n(\mu) \Phi_n(t_1) \sum_{m=1}^{\infty} \sqrt{\lambda_m} c_m(\mu) \Phi_m(t_2)\right] \tag{A3.2}$$

Since the coefficients of the KLE expansion, $a_n(\mu)$ and $a_m(\mu)$ are random, and the basis functions $\Phi_n(t_1)$ and $\Phi_m(t_2)$ are deterministic we can write:

$$R(t_1,t_2)=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sqrt{\lambda_n}\sqrt{\lambda_m}\Phi_n(t_1)\Phi_m(t_2)E\left[c_n(\mu)c_m(\mu)\right] \qquad \text{(A3.3)}$$

Using the fact that the random coefficients of the KLE are orthogonal (equation 4.7):

$$R(t_1,t_2)=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sqrt{\lambda_n}\sqrt{\lambda_m}\Phi_n(t_1)\Phi_m(t_2)\delta_{nm}$$

$$=\sum_{m=1}^{\infty}\lambda_m\Phi_m(t_1)\Phi_m(t_2) \qquad \text{(A3.4)}$$

Replacing the expression of the autocorrelation in the left side of equation 4.8:

$$\int_a^b R(t_1,t_2)\Phi_n(t_1)dt_1=\int_a^b\sum_{m=1}^{\infty}\lambda_m\Phi_m(t_1)\Phi_m(t_2)\Phi_n(t_1)dt_1 \qquad \text{(A3.5)}$$

For the orthonormal eigen functions of the KLE expansion we have:

$$\int_a^b\Phi_m(t)\Phi_n(t)dt=\delta_{nm} \qquad \text{(A3.6)}$$

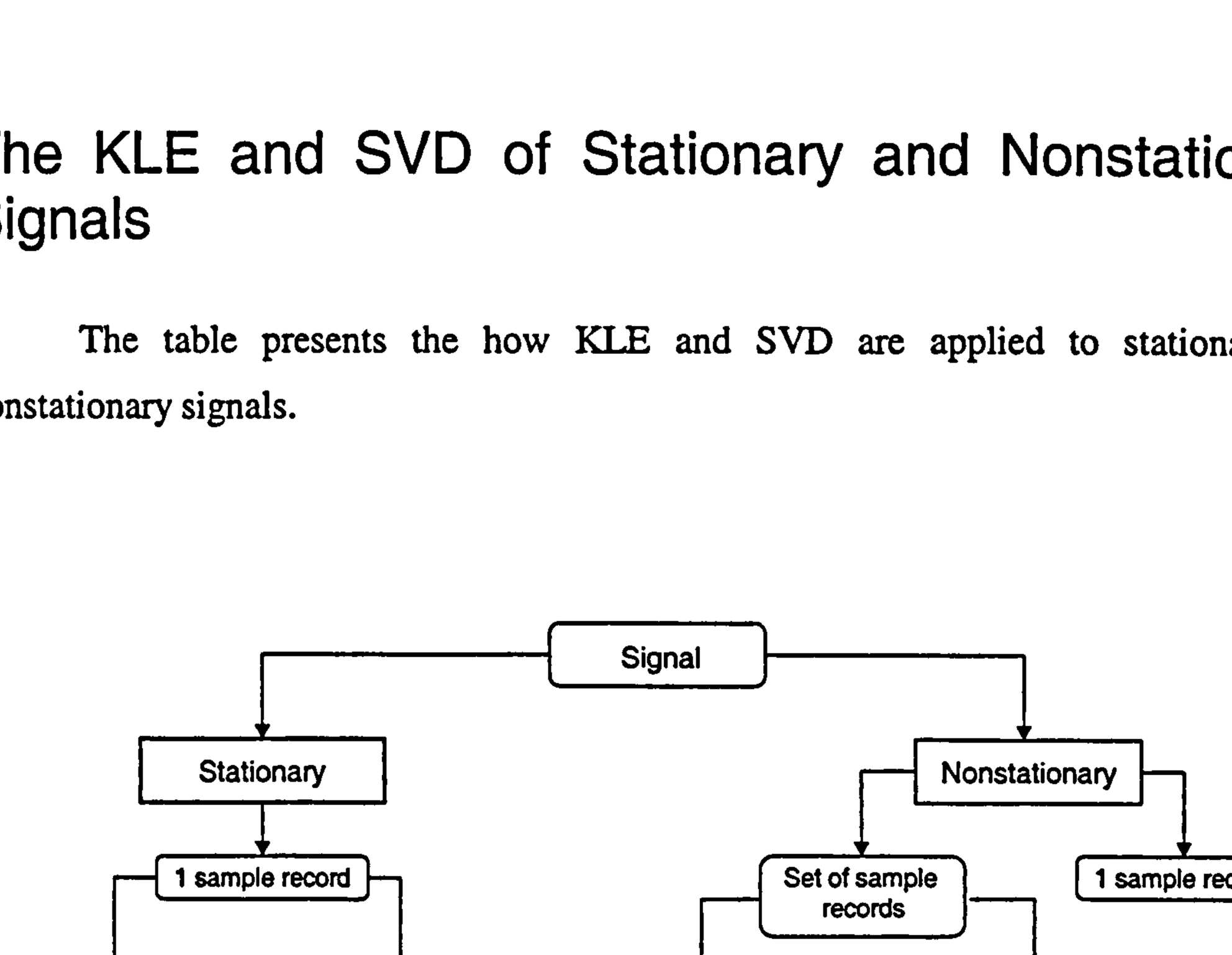
Then the expression A3.3 becomes:

$$\int_a^b R(t_1,t_2)\Phi_n(t_1)dt_1 = \sum_{m=1}^{\infty}\lambda_m\Phi_m(t_2)\delta_{nm} = \lambda_n\Phi_n(t_2) \qquad \text{(A3.7)}$$

and corresponds to equation 4.8. Equation 4.8 is satisfied when a stochastic signal is represented by the KLE.

The KLE and SVD of Stationary and Nonstationary Signals

The table presents the how KLE and SVD are applied to stationary and nonstationary signals.



Tab. A4.1 The KLE and SVD of stationary and nonstationary signals.

When the signal is nonstationary the KLE and SVD, for an ensemble of possible realisations of the signal, have similar eigen and singular vectors. When there is only one realisation of the signal KLE for the nonstationary case the KLE and SVD are generally not applied. Marmarelis proposed a time-average estimator that remains unbiased for the autocorrelation function using a single sample record. According to Marmarelis the nonstationary autocorrelation function can be expressed as the product of two uncorrelated functions: the single record estimator and an orthonormal basis. No further work, on this subject, was developed by Marmarelis. If the estimation of the autocorrelation function could be performed using only one possible realisation than the KLE could be applied to that estimation.

The Modulation of the Singular Vectors of a Nonstationary Sine Wave

The 3rd and 4th singular values of the SVD of a nonstationary sine wave defined as:

$$x_i(t) = \sin(2\pi f_i t) \qquad \text{(A5.1)}$$

where f_i represents the frequency values draw form an interval with uniform distribution, [9.75,10.25]Hz and centre frequency of 10Hz are amplitude modulated. They have the shape of amplitude modulated sine waves. The modulation has a shape similar to a sine (or cosine) functions. On the 3rd vector that modulation has a frequency about 0.5Hz and for the 4th vector it is about 0.75Hz. The Fourier transform of those vectors presents two frequency components approximately at 9.75Hz and 10.75Hz.

The amplitude modulation of a sine wave (the carrier signal) is:

$$A_c\left[1+m(t)\right]\sin\left(2\pi f_c t\right) \tag{A5.2}$$

where A is the amplitude of the carrier signal and $m(t)$ is the modulation function. We consider for the case of the vectors that the modulation function:

$$m(t)=A_m\sin\left(2\pi f_m t\right) \tag{A5.3}$$

The modulated signal can be written as:

$$x(t)=A_cA_m\left[\sin\left(2\pi f_c t\right)+\sin\left(2\pi f_m t\right)\sin\left(2\pi f_c t\right)\right]=$$

$$=A_cA_m\left[\sin\left(2\pi f_c t\right)+\frac{1}{2}\left(\cos\left(\left(f_c+f_m\right)t\right)-\cos\left(\left(f_c-f_m\right)t\right)\right)\right] \tag{A5.4}$$

The Fourier transform of the modulated signal is then:

$$X(f)=\frac{A_cA_m}{2i}\left[\left(\delta(f+f_c)+\delta(f-f_c)\right)\right]+$$

$$+\frac{A_cA_m}{4}\left[\delta(f+f_c+f_m)+\delta(f-f_c-f_m)-\delta(f+f_c-f_m)-\delta(f-f_c+f_m)\right] \tag{A5.5}$$

For the vectors the carrier frequency is 10Hz which is the centre frequency of the interval where the frequency takes its values. At 10 Hz there is no frequency component for the 3rd and 4th vectors. This indicates that the amplitude modulation is a double side band modulation. Then the signal can be written as:

$$x(t) = \frac{A_c A_m}{2}\left[\cos\left((f_c + f_m)t\right) - \cos\left((f_c - f_m)t\right)\right] \qquad \text{(A5.6)}$$

The Fourier transform is:

$$X(f) = \frac{A_c A_m}{4}[\delta(f + f_c + f_m) + \delta(f - f_c - f_m) -$$

$$-\delta(f + f_c - f_m) - \delta(f - f_c + f_m)] \qquad \text{(A5.7)}$$

According to the equation we should have frequency components at $10 \pm 0.5\,\text{Hz}$ for the 3rd vector and $10 \pm 0.75\,\text{Hz}$ for the 4th vector. Since the frequency resolution of the spectrum is 1/3Hz the frequency components of the 4th vector could fall in the same values of the 3rd vector.

ANNEX 6

Equal Temperament

The western music since the 18^{th} century uses the equal temperament to tune the instruments. An interval of octave, as the interval of C262Hz and C524Hz is divided in twelve equal parts (intervals). The fundamental frequencies of the notes in the interval of octave are related as 2:1. This value is divided in twelve equal parts each one corresponds to:

$$h = \sqrt[12]{2} \approx 1.5092 \qquad \text{(A6.1)}$$

The value of h corresponds to the smallest musical interval between two notes in the temperament tuning and gives the frequency relation between two consecutive notes in the piano. This interval is known has a half tone.

The reference frequency f_r for tuning the musical instruments is 440Hz that correspond to the note A. From there the fundamental frequency of the other notes is given by:

$$f_n = (h)^n . f_r \qquad n = \ldots -2,-1,0,1,2\ldots \qquad \text{(A6.2)}$$

The value n is an integer that can take positive and negative values. For the negative values we will have the frequency of notes below A and for positive n we will have the note above A.

For our simulation we will consider that the frequency variation for a note will be smaller than that half tone (to separate one note from another). The central frequency of the interval, f_c, will be any convenient chosen frequency. The lower and upper limits of the interval will be determined by the division and multiplication of f_c, by the value 1.029 that corresponds to a quarter of tone. That for sure will be large enough to include the variation s of the fundamental frequency mentioned in chapters 3 and 7.

ANNEX 7

Data Acquisition

The piano data was digitally recorded with a DAT Sony 60ES. The sampling rate used for the recordings was 44100Hz. The recordings were performed with a B&K (Brüel & Kjaer) 4165 (mono) microphone, amplified by the B&K 2609 amplifier. The microphone was placed over the piano case, 60 cm up from the strings.

The piano used was a Yamaha baby grand. The data was live recorded with ambience noise of the room. Each single piano note were recorded several times using different dynamics and using also both pedals.

REFERENCES

Askenfeld,A.& Jansson,E.(1990). "From touch to string vibrations.I: The motion of the key and hammer", J. Acoust.Soc.Am. 88(1),52-63, July 1990

Askenfeld,A.& Jansson,E.(1991). "From touch to string vibrations.II: Timing in the grand piano action", J. Acoust.Soc.Am. 90(5),2383-2393, November 1991

Askenfeld,A.& Jansson,E.(1993). "From touch to string vibrations.III: String motion and spectra", J. Acoust.Soc.Am. 93 ,2181-2196, 1993

Gelfand,S.A. (1981)."Hearing An Introduction to psychological and physiological acoustics", Marcel Dekker, NewYork (1981).

Golub,G.H., Van Loan,C.F. (1996). "Matrix Computations", 3rd Ed., John Hopkins, London

Gray,R.M.(2000). "Toeplitz and Circulant Matrices: A review", http://www-isl.stanford.edu/~gray/toeplitz.pdf

Hall,D.E.(1986). "Piano string excitation in the case of small hammer mass", J.Acoust.soc.Am., 79 141-147 February 1986

Hall,D.E.(1992). "Piano string excitation VI: Nonlinear modelin", J.Acous.Soc.Am. 92(1),95-105, July 1992

Jeans,J.(1937). "Science and Music", Cambridge, University Press, 1937

Kashino,K.,Nakadai,K.,Kinoshita,T.,Tanaka,H.(1995). "Aplication of the Bayesian probability network to music scene analysis". Proc. Int. Joint Conf. on AI Workshop on Computational Auditory Scene Analysis, Montreal

Loève,M.M. (1955). "Probability theory", Princeton,N.J.:VanNostrand

Lutkepohl,H.(1996). "Handbook of Matrices", Wiley& Sons, England 1996

MartinK.D.(1996)."A blackboard System for Automatic Transcription of Simple polyphonic Music", M.I.T Media Laboratory Perceptual Technical Report No. 399, 1996

Meddis,R.&Hewitt,M.(1991). "Virtual pitch and phase sensivity of a computer model of the auditory PeripherY.I: Pitch Identification," J.Acoust.Soc.Am.89(6) 2866-2882, 1991

Moorer,J.(1977)."On the transcription of musical sound by computer", Computer Music Journal 1(4): 32-38

Newman,A.J.(1996). "Model Reduction via the Karhunen-Loeve Expansion Part I: An Exposition" http://www.isr.ud.edu

Ronken,D.A. (1970). "Monaural detection of a phase difference between click", JASA 47, 1091-1099.

Sjostrom,E. (1996). "Circulant Matrices" http://www.mai.liu.se/~evlun/pub/lic/node9.html

Stapleton,J. & Bass,S.(1988). "Synthesis of musical tones based on the Karhunen-Loeve Transform", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 36 No.3 March 1988

Walmsley,P. P.J.& Goodsill,S. & Rayner,P(1999)."Polyphonic Pitch Tracking Using Joint Bayesan Estimation of Multiple Frame Parameters", Proc.1999 IEEE Workshop on Applications of Signal Processing to Audio Acoustics, Oct 1999

BIBLIOGRAPHY

1. Algazi,V.R. & Sakrison,D.J.(1969), On the Optimality of the Karhunen-Loeve Expansion,IEEE Transactions on Information Theory, 319-320, March 1969
2. Anderson,E.J.(1997)."Limitations of Short Time Fourier Transforms in Polyphonic Pitch Recognition,"Project Report Recognition, May,1997
3. Askenfeld,A.& Jansson,E.(1990)."From touch to string vibrations.I: The motion of the key and hammer", J. Acoust.Soc.Am. 88(1),52-63, July 1990
4. Askenfeld,A.& Jansson,E.(1991)."From touch to string vibrations.II: Timing in the grand piano action", J. Acoust.Soc.Am. 90(5),2383-2393, November 1991
5. Askenfeld,A.& Jansson,E.(1993)."From touch to string vibrations.III: String motion and spectra", J. Acoust.Soc.Am. 93 ,2181-2196, 1993
6. Backus,J.(1969)."The acoustical foundations of music", W.W.Norton & Company,Inc., New York (1969)
7. Benade,A.(1976)."Fundamentals of Musical Acoustics",Oxford University Press, New York, 1976
8. Bendat,J.S.& Piersol,A.G.(1971). "Random data: analysis and measurements procedures", Wiley-Interscience,New York,1971
9. Brown,J.(1989).''Calculation of a ''narrowed''autocorrelation function'',J.Acoust.Soc.Am. 85(4), April 1989
10. Brown,J.(1999)."Computer identification of musical instruments using pattern recognition with cepstral coefficients as features," J.Acoust.Soc.Am.105(3)1933-1941,1999
11. Chaigne,A.(1994)."Numerical simulations of piano strings.I.A physical model for a struck string using finite difference methods", J. Acoust.Soc.Am.95(2),1112-1118, February 1994
12. Cheveigne,A .(1993)."Separation of the concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J.Acoust.Soc.Am. 93(6) 3271-3290,1993
13. Cheveigne,A .(1998)."Cancellation model of pitch perception," J.Acoust.Soc.Am.103(3)1261-1271
14. De Poli,G Piccialli,A. & Roads,C.(1991),"Representations of Musical Signals
15. Eronen,A.&Klapuri,A.(2000)."Musical instruments recognition using cepstral coefficients and temporal features", Signal Processing Laboratory, Tampere, Finland
16. Everson,R. & Sirovich,L.(1995).''Karhunen-Loeve procedure for gappy data", J.Opt.Soc.Am A/Vol.12,No.8 1657-1664, August1995
17. Fancourt,C.L.&Principe,J.C.(?)."Competitive Principal Component Analysis for Locally Stationary Time Series",Department of Electrical Engeneering, University of Florida

18. Galembo,A.& Askenfelt,A.(1999), "Signal Representationand Estimation of Spectral Paramenters by Harmonic Comb Filters with Application to the Piano'', IEEE Trans.on Speech and Audio Processing,Vol.7,No.2,197-203,March 1999
19. Gelfand,S.A. (1981)."Hearing An Introduction to psychological and physiological acoustics", Marcel Dekker, NewYork (1981).
20. Gray,R.M.(1972). "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices", IEEE Trans.on Information Theory,Vol.IT-18,No.6,Nov 1972
21. Gray,R.M.(2000). "Toeplitz and Circulant Matrices: A review", http://www-isl.stanford.edu/~gray/toeplitz.pdf
22. Golub,G. & Van Loan,C.(1989) "Matrix Computations", The John Hopkins University Press, 2nd Ed. London
23. Hall,D.E.(1986). "Piano string excitation in the case of small hammer mass", J.Acoust.soc.Am., 79 141-147 February 1986
24. Hall,D.E.(1987). "Piano string excitation III: General solution for a soft narrow hammer", J.Acous.Soc.Am. 81(2),547-555, February 1987
25. Hall,D.E.(1992)."Piano string excitation VI: Nonlinear modeling", J.Acous.Soc.Am. 92(1),95-105, July 1992
26. Hammond,J."Applied digital signal processing," Vol.1 Notes
27. Hartmann,W.M.(1996)."Pitch, periodicity, and auditory organization", J.Acoust.Soc.Am. 100(6) 3491-3502, (1996)
28. Hermes,D.J.(1987)."Measurement of pitch subharmonic summation", J.Acoust.Soc.Am.83(1)257-264
29. Hilai,R. & Rubinstein,J. (1994). "Recognition of Rotated images by invariant Karhunen-Loeve Expansion", J.Opt.Soc.Am A/Vol.11,No.5 1610-1618 May 1994
30. Houstma,A.J.M. & Smurzynsky,J.(1990)."Pitch Identification and discrimination for complex tones with many harmonics", J.Acoust.Soc.Am.87(1)304-310
31. Irino,T. & Patterson,R.D.(1996)."Temporal asymmetry in the auditory system", J.Acoust.Soc.Am. 99(4) Pt.1,2316-2331, 1996
32. Kashino,K.,Nakadai,K.,Kinoshita,T.,Tanaka,H.(1995). "Aplication of the Bayesian probability network to music scene analysis",Proc. Int. Joint Conf. on AI Workshop on Computational Auditory Scene Analysis, Montreal
33. Kirby,M. & Sirovich,L. (1990). "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12 No.1,103-108, Jan 1990
34. Klapuri,A (2001)
35. Leonard,F.(1998). "Object transformand time-frequency worms"
36. Loève,M.M. (1955) "Probability theory", Princeton,N.J.:VanNostrand
37. Lutkepohl,H.(1996). "Handbook of Matrices", Wiley& Sons, England 1996
38. Maher,R.C.(1990)."Evaluation of a Method for separating digitized duet signals," JAES,Vol.38,No.12, 956-979,1990
39. Maher,R.C.(1994). "Fundamental frequency estimation of musical signals using a two way mismatch procedure", J.Acoust.Soc.Am.95(4),April 1994
40. Marmarelis,V.(1981), "A single-record estimator for correlation functions of nonstationary random processes", Proceed.IEEE,Vol.69,No.7,July 1981
41. Marmarelis,V.(1983). "Practical Estimation of Correlation Functions of Nonstationary Gaussian Process", IEEE Trans.on Information Theory,Vol.IT-29,No.6,Nov 1983
42. Martin,K.D.(1996)."Automatic transcription of the simple polyphonic music," M.I.T Media Laboratory Perceptual Technical Report No. 399,1996
43. MartinK.D.(1996)"A blackboard System for Automatic Transcription of Simple polyphonic Music," M.I.T Media Laboratory Perceptual Technical Report No. 399, 1996
44. Martin,K.D.& Scheirer,E.D. & Vercoe,L.(1998). "Music Content analysis through Models of Audition", Presentes at ACM Multimedia 98,Sep t 1998
45. McNab,R.J., Smith,L.A. & Witten,I.H.(1996). "Signal processing for melody transcription", Proceedings of the 19th Australasian Computer Science Conference, Melbourn, Australia, 1996

46. Meddis,R.&Hewitt,M.(1991)."Virtual pitch and phase sensivity of a computer model of the auditory PeripherY.I: Pitch Identification", J.Acoust.Soc.Am.89(6)2866-2882, 1991
47. Meddis,R.&Hewitt,M.(1991)."Virtual pitch and phase sensivity of a computer model of the auditory periphery.II: Phase Sensivity", J.Acoust.Soc.Am.89(6)2883-2894, 1991
48. Meddis,R.&O'Mard,L.(1997),"A unitary model of pitch perception", J.Acoust.Soc.Am. 102(3) 1811-1820, 1997
49. Meddis,R.(1986)."Simulation of mechanical to neural transduction in the auditory receptor",J.Acoust.Soc.Am.79(3)702-711, 1986
50. Meddis,R.(1987)."Simulation of auditory-neural transduction: further studies", J.Acoust.Soc.Am.83(3)1056-1063, 1987
51. Moorer,J.(1977)."On the transcription of musical sound by computer", Computer Music Journal 1(4): 32-38
52. Newman,A.J.(1996). "Model Reduction via the Karhunen-Loeve Expansion Part I:An Exposition" http://www.isr.ud.edu
53. Newman,A.J.(1996). "Model Reduction via the Karhunen-Loeve Expansion Part II: Some Elementary Examples", http://www.isr.ud.edu
54. Patterson,R.(1994),''The Sound of a sinusoid:Spectral models'', J.Acoust.Soc.Am. 96(3), Sep.1994
55. Patterson,R.D.&Allerhand,M.H.(1995)."Time-domain modelling of peripheral auditory processing: A modular architecture and software platform," J.Acoust.Soc.Am.98(4)1890-1894, 1995
56. Pentland,A. Moghaddam,B. &Starner,T.(1994)"View-based and Modular Eigenspaces for face recognition",MIT Media Lab. Perceptual Computing Section Technical Report No. 245
57. Press,W. et all (1986)."Numerical Recipes in Fortran The art of scientific computing", 2nd Ed. Cambridge University Press
58. Ronken,D.A. (1970) 'Monaural detection of a phase difference between click', JASA 47, 1091-1099.
59. Rossi,L. & Girolami,G.(1997)."Identification of polyphonic piano signals," Acustica-Acta acustica vol.38 (1997) 1077-1084), 1997
60. Rossing,T.D.(1990),"The Science of Sound"Addison -Wesley Publishing Company, Inc. Second Edition
61. Russel,D.A.(1997)."The piano hammer as a nonlinear spring", http://gmi.edu/~drussell/Piano/NonlinearHammer.html
62. Sayegh,S.et all(1997)."A neural network approach to timbre discrimination of identical pitch signals," J.of Intelligent Systems Vol.7,3-4,1997
63. Schreirer,E. (1998), PhD. Proposal, Http://sound media.mit.edu/~eds/papers/phdprop
64. Schroeder,M.&Atal,B(1962)."Generalized Short-Time Power Spectra and Autocorrelation Functions,JASA,Vol.34,No.11 1962
65. Schroeder,M.(1968)."Period histogram and product spectrum: new methods for fundamental-frequency measurement,"J.Acoust.Soc.Am.(43)829-834
66. Sirovich,L.(1987). "Turbulence and the dynamics of the coherent structures Part I: Coherent Structures", Quarterly of Apllied Mathematics,Vol.XLV,No.3, 516-571, Oct 1987
67. Sjostrom,E. (1996). "Circulant Matrices" http://www.mai.liu.se/~evlun/pub/lic/node9.html
68. Stapleton,J. & Bass,S.(1988)."Synthesis of musical tones based on the Karhunen-Loeve Transform",IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 36 No.3 March 1988
69. Stewart,G.W.(1993), "On the early history of the singular value decomposition", SIAM review, Vol.35, No.4, 551-566 ,1993
70. Suzuki,H.(1987)."Model analysis of a hammer-string interaction"J.Acoust,Soc.Am. 82(4) October 1987
71. Taylor,J.(1983)." Digital Filter Design Handbook,"New York, M.Dekker,Inc
72. Terada,T.&Nakajima,H.&Tohyama,M.& Hirata,Y.(1994), "Nonstationary Waveform Analysis and Synthesis Using Generalized Harmonic Analysis", Proceed.IEEE-SP, 429-432,Oct 1994

73. Uenohara,M.&Kanade,T.(1997)."Use of Fourier and Karhunen-Loeve Decomposition for Fast Pattern Matching With a Large Set of Templates",IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.8, August 1997
74. Van der Horst,R.&Leeuw,A.R.&Dreschler,W.A.(1999),"Importance of temporal-envelope cues in consonant recognition,"J.Acoust.Soc.Am.105(3)1801-1809, 1999
75. Van Immerseel,L.M.&Martens,J.(1992)."Pitch and voice/unvoiced determination with an auditory model",J.Acoust.Soc.Am.91(6)3511-3526, 1992
76. Vu,H.L.&Lois,L.(2000),"Efficient Distance Measure for Quantisation of LSF and its KL Transformed Parameters",IEEE Trans.Speech and Audio Processing, Vol.8, No.6, 2000
77. Walmsley,P.J.& Goodsill,S. & Rayner,P.(1999), "Multidimensional Optimisation of Harmonic Signals", http://www.stat.umn.edu/PAPERS/tech.reports/tr560.ps
78. Walmsley,P. (1997), "Signal Separation of Musical Instruments", First year report
79. Walmsley,P. P.J.& Goodsill,S. & Rayner,P(1999), "Polyphonic Pitch Tracking Using Joint Bayesan Estimation of Multiple Frame Parameters", Proc.1999 IEEE Workshop on Applications of Signal Processing to Audio Acoustics, Oct 1999
80. Will,T. (1999), "Singular value decomposition", http://davidson.edu/math/will/svd/index.html
81. Yamashita,Y. & Ogawa,H (1996),"Relative Karhunen-Loeve Transfrom", IEEE Transactions on Signal Processing, Vol.44,No.2, 371-378, February 1996
82. Zwicker ,E & Fast,H (1990)."Psychoacoustics Facts and Models," Ed.Springer-Verlag, Berlim