

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

A Study of Automatic
Transcription of Music Using a
Standard PC

Includes CD(s)

Victor Poon

2001

MASSEY UNIVERSITY



1061533714

Abstract

This thesis describes using a Personal Computer to identify notes that are played by a musical instrument. Several groups have been doing this work with more sophisticated laboratories and equipment with only moderate success.

We have found the waves created by musical instruments vary, between instruments, a great deal in their stability and inherent vibration. It was more difficult to identify notes with very low frequencies than those with more central frequencies. We found it was very important to choose the correct starting point for the analysis with Fourier Transform otherwise we would not be analysing the stable stage of the wave.

We tried simple strategies to initially reduce the number of computer operations, and memory requirements, with marginal success. We followed with more complex subtraction strategies which were much more successful. The most useful technique involved creating a “calculated percentage multiple” which was almost 100% successful in identifying single notes. For multiple notes we were surprised to find that a group of five different instruments from MIDI were a better source of “known” notes to compare with the “unknown” notes than the MIDI equivalent of the real instrument playing the music.

These methods were developed using midi instruments but were verified using a real grand piano.

We suggest some further lines of enquiry that may make this technique more successful.

Acknowledgement

I would like to express my thanks to my supervisor, Dr. Peter Kay, for his help and advice during the completion of this thesis.

I am also very grateful for the help and the time given by Dr. Murray Sampson and Mrs. Lorraine Sampson who patiently proof read this thesis.

Finally I would like to thank my parents Yu Lin Poon and Tit Wing Poon for their financial and family support during my university study.

Contents

Contents	i
List of Figures	iv
List of Tables	ix
Introduction	1
Literature Review	3
Chapter One: Background	10
1.1 Music Notes	10
1.2 Digital Audio.....	12
1.3 Wave File Format.....	12
1.4 MIDI.....	15
1.5 MATLAB.....	17
1.6 Fourier Transforms.....	18
1.6.1 The differences between DFT and FFT	19
1.6.1.1 Time efficiencies.....	19
1.6.2 Symmetry	20
Chapter Two: Using MIDI To Analyse Sound	23
2.1 Sample Rate, Sample Size, Frequency and Wavelength.....	25
2.1.1 The sample rate	25
2.1.2 The frequency, the wavelength and the sample size	26
2.2 The Effect of Using Different Sample Sizes.....	28
2.3 The Effect of Using Different Starting Points	31
2.4 Recording in MIDI.....	36
2.5 Playing Two Notes Simultaneously on a Single Instrument.....	41
2.6 Playing Three Notes at Once.....	45

2.7	Why is the Highest Peak of the Single Note not the Answer in Some Cases?	46
2.8	What is the Relationship between a Single Note and Two Notes or Three Notes?.....	48
2.9	The Reason for Different Instruments having Different Results from the Fourier Transform	50
2.10	Using the Fourier Transform of Recorded Notes to Identify Unknown Notes	58
2.10.1	The file size of the recording format.....	58
2.10.2	The choice of representative value.....	58
2.11	Possible Filtering Methods to Identify Unknown Notes.....	62
2.11.1	The first strategy	63
2.11.2	The second strategy	64
2.11.3	The third strategy	66
2.11.4	The fourth strategy	67
2.11.5	The fifth strategy	69
2.11.6	Summary	71
2.12	The Conditional or Tolerance Values	72
2.12.1	The first strategy	72
2.12.2	The second strategy	73
2.12.3	The third strategy	74
2.12.4	The fourth strategy	75
2.12.5	The fifth strategy	76
2.13	The Accuracy Rate for Different Strategies when Playing a Simple Single Note	77
2.13.1	Application of using strategies 1 to 4.....	77
2.13.2	Application of strategy 5	79
2.14	Application of Strategies 3-2-1 when Playing Two Notes.....	81
2.15	Application of Strategy 5 on Two Notes	83

2.15.1	Finding the first solution	83
2.15.2	Subtraction of notes to find the second solution	84
2.15.2.1	Subtraction excluding the first solution	85
2.15.2.1.1	Simple subtraction.....	85
2.15.2.1.2	Complex subtraction method 1	87
2.15.2.1.3	Complex subtraction method 2	87
2.15.2.2	Summary of subtraction excluding the first solution	90
2.15.2.3	Preliminary conclusions about strategy 5	91
2.16	Application of Strategy 5 on Three Notes Together	91
2.17	Sources of Error within Calculations	92
2.18	Finding the Number of Notes that are Playing Simultaneously?	93
Chapter Three: Real Instruments		98
3.1	Application of Strategy 5 on a Single Note	105
3.2	Application of Strategy 5 on Two Notes Played Together	105
3.2.1	Experimented results.....	106
3.2.2	Estimating the number of solutions.....	107
3.3	Application of Single Note Using MIDI as the Unknown Note	108
3.4	Application on Two Notes Using MIDI as the Unknown Note.....	109
3.4.1	Known notes from a single instrument	109
3.4.2	Known notes from 5 different instruments	110
3.5	Practical Experiment	111
3.6	Result of Method B	111
Chapter Four: Conclusion		116
Reference		119

List of Figures

Chapter One: Background

Figure 1.1	Example of a sound wave (Acoustic Grand Piano from MIDI, A4)	14
Figure 1.2	The number of operations of using DFT and FFT	20
Figure 1.3	Results of FT using MATLAB	20
Figure 1.4	Results of FT after rearrangement.....	21

Chapter Two: Using MIDI To Analyse Sound

Figure 2.1	The concept of translating text files to <i>.wav</i> files.....	25
Figure 2.2	Wave format for A0 (Samples from 0 to 80,000)	26
Figure 2.3	Wave format for C8 (Samples from 0 to 3,150)	27
Figure 2.4	The results of Fourier Transform in 2D when using different sample sizes on an acoustic grand piano when playing the note C2 (65.41 Hz)	30
Figure 2.5	The results of Fourier Transform in 3D when using different sample sizes on an acoustic grand piano when playing the note C2 (65.41 Hz)	30
Figure 2.6	Simplified amplitude evolution of a music note	32
Figure 2.7	The results of Fourier Transform in 2D when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz).....	33
Figure 2.8	The results of Fourier Transform in 3D when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz).....	33
Figure 2.9	The results of Fourier Transform in 2D when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz) (in percentages)	34
Figure 2.10	The results of Fourier Transform in 3D when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz) (in percentages)	34
Figure 2.11	The results of Fourier Transform when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz) after totalling all the spectra	35

Figure 2.12	The results of Fourier Transform in 2D when using different starting points, on an acoustic grand piano when playing the note C2 (65.41 Hz) in percentages after totalling all the spectra.....	35
Figure 2.13	Example of recorder from MIDI (75) when playing the note A4 (440 Hz) (from Sample '0' to Sample 1500).....	36
Figure 2.14	The sound wave of two different instruments when playing C2 (65.41 Hz).....	47
Figure 2.15	The results of Fourier Transform on an acoustic grand piano when playing C2 (65.41 Hz).....	47
Figure 2.16	The results of Fourier Transform on a recorder when playing C2 (65.41 Hz)	47
Figure 2.17	The sound wave of an acoustic grand piano when playing C4	48
Figure 2.18	The sound wave of an acoustic grand piano when playing A4.....	48
Figure 2.19	The sound wave adding up the value of C4 and A4 on an acoustic grand piano.....	49
Figure 2.20	The sound wave for an acoustic grand piano when playing C4 and A4 together.....	49
Figure 2.21	The sound wave of the note A4 on an acoustic grand piano (MIDI 1)	50
Figure 2.22	The sound wave of the note A4 on an acoustic guitar (steel) (MIDI 26)	50
Figure 2.23	The sound wave of the note A4 on an acoustic bass (MIDI 33)	50
Figure 2.24	The sound wave of the note A4 on a violin (MIDI 41).....	51
Figure 2.25	The sound wave of the note A4 on a recorder (MIDI 75).....	51
Figure 2.26	The sound wave of the note A4 when using different instruments..	52
Figure 2.27	The results of the Fourier Transform of the note A4 on an acoustic grand piano (MIDI 1)	53
Figure 2.28	The results of the Fourier Transform of the note A4 on an acoustic guitar (steel) (MIDI 26).....	53
Figure 2.29	The results of the Fourier Transform of the note A4 on an acoustic bass (MIDI 33)	53
Figure 2.30	The results of the Fourier Transform of the note A4 on a violin (MIDI 41)	54

Figure 2.31	The results of the Fourier Transform of the note A4 on a recorder (MIDI 75).....	54
Figure 2.32	Example of tone spectra in FT for A4 (440 Hz)	57
Figure 2.33	The results of Fourier Transform when playing C2 (65.41Hz) on an acoustic grand piano.....	59
Figure 2.34	Figure 2.33 with note boundaries added	60
Figure 2.35	The “lowest” results of Fourier Transform from within each frequency range when playing C2 (acoustic grand piano).....	61
Figure 2.36	The “highest” results of Fourier Transform from within each frequency range when playing C2 (acoustic grand piano).....	61
Figure 2.37	The “average” results of Fourier Transform from within each frequency range when playing C2 (acoustic grand piano).....	61
Figure 2.38	Summary of different methods of subtraction when using average frequency starting point.....	90
Figure 2.39	Summary of different methods of subtraction when using lower frequency starting point.....	90

Chapter Three: Real Instruments

Figure 3.1	An example of the process of detecting the starting point when playing with real instruments	102
Figure 3.2	A close up of the earlier parts of the FTs in Figure 3.1	103
Figure 3.3	The sound wave of C2.....	104
Figure 3.4	The maximum difference in each FT	104
Figure 3.5	The sum of difference in each FT	104

Appendix B

Figure B.1	The sample of acoustic grand piano when playing the note C2 (65.41 Hz).
Figure B.2	The results of Four Transform in 2D when using different sample sizes on an acoustic grand piano when playing the note C2 (65.41 Hz).
Figure B.3	The results of Four Transform in 3D when using different sample sizes on an acoustic grand piano when playing the note C2 (65.41 Hz).

- Figure B.4** The results of Four Transform in 2D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz).
- Figure B.5** The results of Four Transform in 3D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz).
- Figure B.6** The results of Four Transform in 2D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.7** The results of Four Transform in 3D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.8** The results of Four Transform in 2D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz) after totalling all the spectra.
- Figure B.9** The results of Four Transform in 2D when using different starting point on an acoustic grand piano when playing the note C2 (65.41 Hz) in percentages after totalling all the spectra.
- Figure B.10** The sample of bright acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.11** The results of Four Transform in 2D when using different sample sizes on a bright acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.12** The results of Four Transform in 3D when using different sample sizes on a bright acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.13** The results of Four Transform in 2D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.14** The results of Four Transform in 3D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.15** The results of Four Transform in 2D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.16** The results of Four Transform in 3D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.17** The results of Four Transform in 2D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz) after totalling all the spectra.

- Figure B.18** The results of Four Transform in 2D when using different starting point on a bright acoustic piano when playing the note C2 (65.41 Hz) in percentages after totalling all the spectra.
- Figure B.19** The sample of an electronic acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.20** The results of Four Transform in 2D when using different sample sizes on an electronic acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.21** The results of Four Transform in 3D when using different sample sizes on an electronic acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.22** The results of Four Transform in 2D when using different starting point on an electronic acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.23** The results of Four Transform in 3D when using different starting point on an electronic acoustic piano when playing the note C2 (65.41 Hz).
- Figure B.24** The results of Four Transform in 2D when using different starting point on an electronic acoustic piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.25** The results of Four Transform in 3D when using different starting point an electronic acoustic piano when playing the note C2 (65.41 Hz) (in percentages).
- Figure B.26** The results of Four Transform in 2D when using different starting point on an electronic t acoustic piano when playing the note C2 (65.41 Hz) after totalling all the spectra.
- Figure B.27** The results of Four Transform in 2D when using different starting point on an electronic acoustic piano when playing the note C2 (65.41 Hz) in percentages after totalling all the spectra.

List of Tables

Chapter One: Background

Table 1.1	The process of finding frequencies	11
Table 1.2	Table of frequencies displayed in octaves	11
Table 1.3	Table of the instruments in MIDI and the numbers to denote each instrument	16
Table 1.4	The number of operations using DFT and FFT	19

Chapter Two: Using MIDI To Analyse Sound

Table 2.1	The estimated duration of notes	27
Table 2.2	The relationship between frequency, the results for the Fourier Transform and the sample size.....	29
Table 2.3	The results of the highest peak of using Fourier Transform on recorder (MIDI 75).....	37
Table 2.4	The accuracy rate when comparing the expected frequencies and the highest peak of the Fourier Transform grouped into types of instrument	40
Table 2.5	The results of playing two notes on MIDI by using recorder	41
Table 2.6	The results of playing two notes using recorder on MIDI	43
Table 2.7	The accuracy when using different ranges of expected note on the recorder	44
Table 2.8	The accuracy rate when using the two highest peaks to predict the results on different instruments.....	45
Table 2.9	The results of the Fourier Transform of the note A4 on different instruments	55
Table 2.10	The results of the Fourier Transform of the note C4 on different instruments	55
Table 2.11	A theoretical example of the highest peaks.....	63
Table 2.12	Applying the first strategy to Table 2.11	64
Table 2.13	Establishing the values accepted.....	65

Table 2.14	Applying the second strategy to Table 2.13.....	66
Table 2.15	A theoretical example applying strategy 3 using the information from Tables 2.13 and 2.14	67
Table 2.16	Table 2.13 with showing the difference between successive peaks....	68
Table 2.17	Applying strategy 4 to Table 2.16.....	69
Table 2.18	Applying strategy 5 using the information from Tables 2.13, 2.14 and 2.15	70
Table 2.19	Summary of different minimum values using various starting points on 5 different instruments (in percentages).....	73
Table 2.20	Minimum differences for 5 different instruments.....	74
Table 2.21	Minimum matches for 5 different instruments.....	75
Table 2.22	The number of zeros when using different conditional percentages...	75
Table 2.23	The minimum results of “known” minus “unknown” using different conditional percentages.....	76
Table 2.24	Results with notes from the same instrument	77
Table 2.25	Results with notes from 5 different instruments	78
Table 2.26	The summary of the number of rejected known notes using strategy 3 followed by strategies 1 and 2 in alternative order	78
Table 2.27	The summary of applying three different strategies in series showing the number of the correctly located notes	79
Table 2.28	The summary of the rank of the correct notes using strategy 5	80
Table 2.29	The results of choosing the highest rank in strategy 5 as the solution	81
Table 2.30	The average number of known notes in different instruments that have been filtered out when using average frequency as the starting point	82
Table 2.31	The average number of known notes in different instruments that have been filtered out when using lower frequency as the starting point....	82
Table 2.32	The number of known notes that have been filtered out as we apply the strategies.....	83
Table 2.33	The number of cases which the highest rank is equal to the expected note.....	83
Table 2.34	The number of results is correct using strategy 5 (with repeated answer)	85

Table 2.35	The number of cases where the second solution is correct when preventing the first solution from being repeated	86
Table 2.36	The number of correct second solution using strategy 5 with complete FT results.....	87
Table 2.37	The number of correct second solution using strategy 5 using different subtraction methods	89
Table 2.38	Summary of the accuracy rate using different methods to find the solution when three notes are played together	91
Table 2.39	Summary of the averages of the highest peak of the FT results using different methods when two notes are played together.....	94
Table 2.40	Summary of the averages of the highest peak of the FT results using different methods when three notes are played together.....	95

Chapter Three: Real Instruments

Table 3.1	Summary of the accuracy rate when playing single notes	105
Table 3.2	Summary of the accuracy rate in different experiments when two notes are played together	106
Table 3.3	Summary of the averages of the maximum height of FT in different calculation method in the three experiments in section 3.2.1	107
Table 3.4	Summary of the accuracy rate when using different sets of known notes including results from the previous section	108
Table 3.5	Summary of the accuracy rate using single MIDI instrument (AGP) as the known note	109
Table 3.6	Summary of accuracy rate of using 5 different instruments as the known note.....	110
Table 3.7	Summary of the accuracy rate playing a simple tune	111
Table 3.8	The accuracy rate when playing two notes using Method A (same as Table 3.2)	112
Table 3.9	The accuracy rate when playing two notes using the lowest of the maximum difference between the previous FT and the current FT as the starting point.....	112
Table 3.10	The difference between the results in Tables 3.8 and 3.9	113
Table 3.11	The accuracy rate when playing two notes using the lowest of the average difference between the previous FT and the current FT as the starting point	113

Table 3.12	The difference between the results in Tables 3.8 and 3.11	114
Table 3.13	A summary of the differences in the accuracy rate between method A and the two versions of method B.....	114

Appendix B

Table B.1	The comparison between the highest peak from Fourier Transform and the expected results when using different instrument on MIDI.
Table B.2	The results of playing three notes together.
Table B.3	The relationship on an acoustic grand piano using different starting point and compared to the original curve (15 Periods of wavelength) in percentage.
Table B.4	The relationship on an acoustic guitar (steel) using different starting point and compared to the original curve (15 Periods of wavelength) in percentage.
Table B.5	The relationship on an acoustic bass using different starting point and compared to the original curve (15 Periods of wavelength) in percentage.
Table B.6	The relationship on a violin using different starting point and compared to the original curve (15 Periods of wavelength) in percentage.
Table B.7	The relationship on a recorder using different starting point and compared to the original curve (15 Periods of wavelength) in percentage.
Table B.8	Summary of the number of correct results when three notes are played together.
Table B.9	Number of Errors occurs using the absolute value method with complete FT results when two notes are played together.

Appendix C

Table C.1	The results using different types of known notes when playing a song on a real grand piano.
------------------	---

Here we provide a CD which contains the programs we wrote to enable us to perform the experiments and analyses.

Introduction

This thesis describes the work we have done to study the automatic transcription of music to printed notation using a standard PC. After the subject of this thesis was accepted we began a literature review. This review showed we would need to reduce the scope of the work to enable us to make a valid contribution to the field. We have therefore limited the study to identifying notes by comparing them to MIDI and also notes from the same instrument. Later in our work we realised that the selection of the starting point for the comparison was very important. However the time available did not allow us to more than touch on the issue. This complexity has also created difficulties for other work (Hamer 2001 [16]) done by larger groups using more sophisticated equipment.

We quote from another worker “Our scope and treatise is limited by several factors, but especially by the limited amount of resources compared to the wide range of topics that are related with music transcription. Moreover, engaging in a research that is quite new to our laboratory, analysis of musical signals, called for paying the required attention to just finding the right points of emphasis and avoiding wrong assumptions in an early phase.” (Klapuri 2002 [21]) The quoted author is one member of a group six. We quote further discussion on commercial products “Polyphonic transcribers – sad to say – work very poorly. Monophonic systems are more robust, of course.”

Our research has shown similar difficulties to those of other workers and taking into account the limitations, our accuracy has been acceptable.

Chapter 1 gives the background to the basic techniques used in this thesis. In sections one to four, we will explain the music and file systems we used. In sections five and six, we will look at the advantages of using MATLAB and the use of Fourier Transform.

Chapter 2 explains and reports on the analysis of MIDI generated files. It covers the analysis on single and a small number of notes played together.

Chapter 3 will perform the similar analysis on a real piano.

We then report our conclusions.

Literature Review

When one examines the literature, it becomes clear that the mathematical study of music is a difficult field studied over many years and from many different aspects. Such studies are the basis for the ability to transcribe music using a PC. In this review we give a general overview with reference examples rather than extensive lists. Our other approach is to cite reviews, textbooks and homepages to enable the reader to home in on any particular interest they may have.

When the unaware see sheet music they would be inclined to believe that what they see tells them what they will hear, if it is played by a competent musician. Unfortunately the simple indication of the note to be played does not reveal the way the music is created by instrument or voice. The sound produced varies a great deal depending on the instrument. The pitch is the same but the timbre is different.

Pitched instruments may make their sound through mechanisms that depend on creating a vibration in a physical object and modifying it before it transmits through air to the audience. A "string" maybe struck as in a piano or "rubbed" as in a violin. The length of the string, either in the instrument, or modified by "fingering", will set the frequency of the vibration, hence the sound. There will usually be more than one frequency created as part of one "note". Further modification of sound is via the construction material of the instrument, through hollow spaces such as in the body of a violin, or sounding boards in a piano.

Wind instruments make sound through the use of air forced through by either special oral techniques or mechanisms such as reeds to cause vibration. Their sound is affected by the length of the tube the air passes through. Most Brass instruments shorten the length of the tube by using valves to close off part of the tube. Other wind instruments such as the flute, depend on shortening the amount of tube that can allow air vibration, by covering the holes in the tube. Shortening the tube length alters the frequency of the note. see Suggs (1966) [35]. The physics of musical instruments is

very complicated and is explained in Fletcher and Rossing (1998) [10]. This reference may be the prior edition of Rossing, Wheeler and Moore (2002) [34] but is sufficiently different to be worth referring to separately.

These differences, and others, are the source of the complications found in the automatic transcription of music. Notes may "overlap" or be played together.

The science of sound is explained in Rossing, Wheeler, Moore (2002) [34].

A recent review of mathematics and music is Benson (2002) [3]. This is an expanded and augmented version of notes given for an undergraduate course at the University of Georgia. Water waves are waves where the local movement is at right angles to the direction of propagation (transverse waves) whereas sound waves are longitudinal waves with the local movements in the same direction as the propagation. Sound waves have four main attributes:

1. Amplitude is the size of the vibration and is perceived as loudness. The amplitude of typical everyday sound is only a small fraction of a millimetre.
2. Pitch corresponds to the frequency of the vibration.
3. Timbre corresponds to the shape of the frequency spectrum of the sound, and
4. Duration which is the length of time for which the note lasts.

However most vibrations do not consist of a single frequency and naming the "defining" frequency can be difficult. Music should be defined more in terms of the perception of sound rather than in terms of the sound itself. The perceived pitch of a sound may represent a frequency not actually present in the waveform -- a "missing fundamental". This phenomenon is a part of "Psychoacoustics".

Benson describes the functioning of the human ear and its limitations. It "hears" within a range of about 20Hz (Hertz) to 20,000 Hz although it may "feel" the sound below 20Hz.

The relevance of sine waves in relation to sound is explained as lying in the differential equation for simple harmonic motion which can be taken as a close

approximation to the equation of motion of a particular point on the basilar membrane of the human ear or anywhere else on the chain from the outside air to the cochlea. The limitations of the approximation are discussed. Harmonic motion is explained in mathematical terms. This leads to damped harmonic motion and resonance.

Refer to Benson (2002) [3] where he then discusses Fourier theory and a useful mathematical system, which is often used for the analysis of sound, being Fourier Transform.

The article usually regarded as the original, announcing the fast Fourier Transform as a practical algorithm, is Cooley and Tukey (1965) [6]. An early algorithm is given by Gold and Rabiner (1969) [12]. This algorithm has been used much later in New Zealand to develop an initial system for the identification of melodies. McNab, Smith, Bainbridge and Witten (1997) [25]. An earlier description of signal processing, which involves measurement of spectra using Fourier transform is given in the text, Rabiner and Gold (1975) [31].

The creation of music by instrument is one way of starting the chain. The other is the voice. Some studies are involved in relating sound produced to the characteristics of the larynx. An example is Bagshaw, Hiller and Jack (1993) [1]. The next link is the transmission of sound through air and its subsequent interaction with objects within the space or enclosing the space. These matters of acoustics are of importance in the transcription of music because they indicate the possibility of the sound being changed before it is identified. (Kinsler, L.E, Frey, Coppens, A.B. , and Sanders 1982 [19]). The shape of the room in which it is played, and the nature of the floor and its foundations, are two examples of the factors which affect the acoustics.

The human perception of music is another aspect which is also very complicated (Roederer 1995 [33]). The functioning of the human ear is an integral part of this perception. (Hudspeth 1985 [17], 1989 [18]) is another reference for this aspect. Further references are continually updated on the homepage (CSTR Publications 2001 [7]). Neuropsychology is the discipline that studies the neural system processes and functions linking the input received from the environment and body with the full behavioural and mental output. A text covering this is McAdams and Bigand (1993) [24]. A homepage that covers this field is Neuropsychology Central (2001) [29].

The transcription of music involves artificial perception and music recognition rather than human perception. Tanguiane (1993) [36] gives us a good overview of this topic. Work dates back to the early 1970's with the first experiments on automatic notation of monophonic music and a program for transcribing polyphonic music performed on a computer-wired keyboard. The limitations experienced at that time included only admitting pitched sounds, avoiding some pitch combinations, and requiring simple rhythmic structure and a constant tempo. Artificial intelligence methods began to be used in the early to the mid 1980's.

Artificial perception is the usual approach to the pattern recognition which is necessary to develop a computer system for the automatic notation of the performed music. It attempts to follow nature. A difficulty is that if a system of artificial perception is developed for one type of music, it may not be as successful with different styles or musical cultures. Another problem is that compared to human perception, a much longer passage of music is required for artificial perception to establish a reliable result. A singing voice, being less stable than an instrumental sound, is more difficult to analyse. By 1985 the studies branch into either chord/note recognition or rhythm/tempo recognition.

An early method for rhythm recognition was to develop an hypothesis concerning the rhythmic structure from the very first events which is then continuously confronted with current data and being modified if necessary.

We paraphrase a passage that states that a particular difficulty is that there are no explicit definitions of notes, chords, rhythm, and tempo, and so their recognition is complicated. Goto and Muraoka (1997) [14] discuss the issue of the evaluation of beat tracking systems and the need to rely on human intervention as an evaluation tool since the beat is a perceptual concept a person feels in music which is difficult to define in an objective way. In designing their measure for evaluation they considered subjective hand-labeled beat positions to be the correct beat times. The positions can be finely adjusted by playing back the audio with click tones at beat times and the user also defines a hierarchical rhythmic structure. They later (Goto and Muraoka, 1998 [15]) report on a real-time beat tracking system that recognises a rhythmic structure in real-

world audio signals sampled from popular-music compact discs. A homepage as an entry into the literature on rhythm/tempo recognition is Goto (2002) [13].

A recent method (Klapuri, Virtanen and Holm 2000 [22]) for the estimation of the multiple pitches of concurrent musical sounds comprised of sung vowels and the whole pitch range of 26 musical instruments had error rates for mixtures ranging from one to six simultaneous sounds were 2.1%, 2.4%, 3.8%, 8.1%, 12% and 18% respectively. In musical interval and chord identification tasks, the algorithm outperformed the average of ten trained musicians. This gives an indication of the current accuracy being achieved.

It is interesting to realise that music recognition and voice recognition use the same methods and the studies overlap. A problem in common is the separation of several sounds being heard by the recognition system some of which are "interference" outside the voice or music being studied. A recent review is De Cheveigne (1993) [8]. Results from an implementation of this approach illustrated its ability to analyse complex, ambient sound scenes that would confound previous systems.

Homepages for groups working in this field include MIT Media Lab (2002) [27] and Klapuri (2002) [21]. Because these are updated continually, they are ongoing sources of recent work also giving references to other workers.

A system of artificial perception, which has some similarity to artificial intelligence is the use of neural network. Examples of this concept are given in Roberto (1993) [32]. Other work with neural networks can be accessed at CSTR publications (2001) [7].

The review to this stage showed that it was going to be very difficult for us to create a research project within the limit of time and facilities. Klapuri (2002) [21] mentions a similar problem. It is interesting to note that Klapuri made this statement after completing the thesis (Klapuri 1998 [20]) within Tampere University of Technology, Finland, which has an Audio Research group. Klapuri (1998) [20] gives a review of systems of artificial perception. The phrase onset time is defined as the instant of time when the sound starts playing.

An important distinction made by Tanguiane (1993) [36] is quoted.

"The difference between artificial perception and artificial intelligence in pattern recognition is understood as follows. Artificial perception is used for discovering structure in visual and audio images by self-organisation of data and segregation of patterns. Artificial intelligence is used for pattern identification by their matching to known concepts. Usually, the identification of already segregated patterns is much simpler than their recognition in data flows: thus artificial perception and artificial intelligence are complementary."

This distinction by Tanguiane (1993) [36] is not necessarily clear in the use of the phrase, artificial intelligence, by other workers, but it leads to another approach which is used in our studies.

In August 1983, music manufacturers agreed on a protocol that is called the "MIDI 1.0 Specification". General MIDI (Musical Instrument Digital Interface) is a standard that defines specific locations for different instrument sounds in the present memory of synthesisers (MIDI Manufacturers Association Incorporated 2001 [26]).

Therefore MIDI could be used as the known concept, referred to by Tanguiane (1993) [36] to develop a system of musical note identification. Benson (2002) [3] gives a recent review of concepts in digital music.

Work which has some similarities to ours has used matching methods to compare a hummed tune to a recorded database of songs. We refer to Ghias and Logan (1995) [11] and Blackburn and DeRoure (1998) [4]. They have used a pitch tracking method represented as a sequence of relative pitch changes (i.e a melodic pitch contour). We note that when a person hums a song it is monophonic.

The New Zealand Digital Library MELody inDEX (McNab, Smith, Bainbridge and Witten (1997) [25]) is a system that accepts acoustic input from the user created by a few notes sung into the microphone. It transcribes the melodies automatically from the microphone input then searches a database for tunes that contain the same or similar sung patterns. The tunes retrieved are ranked according to the closeness of the match. Different search criteria were used such as melodic contour, musical intervals and

rhythm, and tests were performed using both exact and approximate string matching. They also performed tests on how people remember tunes. They concluded from these experiments that people needed a choice of several matching procedures and should be able to explore the results interactively in their search for a particular melody. This recent effort shows the recurring need for human interaction in these processes. The voice range was limited to from F2 to G5. This reference is dealing with monophonic sound and uses the algorithm on Gold and Rabiner (1969) [12]. The system depends on the user separating each note by singing da or ta to create a note boundary. The solutions to the other problems and the techniques used are described. On going access to the work of this group can be made via the homepage of Bainbridge (2002) [2].