Whistle Recognition Applied in Searching Music Database

By

Nghiem Anh Tuan 3189

Dissertation submitted in partial fulfillment of
the requirement for the
Bachelor of Technology (Hons)
(Information Communication Technology)

NOVEMBER 2005

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak

CERTIFICATION OF APPROVAL

Whistle Recognition Applied in Searching Music Database

by

Nghiem Anh Tuan

Dissertation submitted in partial fulfillment of
the requirement for
Bachelor of Technology (Hons)
(Information Communication Technology)

Approved by,

(Dr. Abs Md. Said)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
November, 2005

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

NGHIEM ANH TUAN

ACKNOWLEDGEMENT

With the help and support from many people while working on this project, I would like to relish this opportunity to express my gratitude to all who has in one way or another helped me to complete this work.

First of all, I would like to thank to Dr. Abas M Said, the project supervisor who gave me idea of this project as well as a lot of supportive and constructive ideas. His suggestions have guided me through this project.

I also gratefully acknowledge all my friends who have given invaluable review and feedbacks on the project. Special thanks go to Nguyen Duy Vinh for his time testing on this system.

Last, and most of all, I am thankful for endless encouragement and care from my Dad, Mum, my brother and my love. They have given me the strength, motivation and determination in all my undertakings.

ABSTRACT

This project aims at constructing an innovative Music Search Engine. The system takes human whistle as input, and return full information of the song which user has just whistled. Due to the fact that human being remembers the melody of composition better than the name or lyric, the project purpose is to demonstrate a new searching method which is based on composition melody. There are two main phases in Music Searching Engine: 1) Whistle Recognition: converting whistle to music note and 2) Music Search: finding the similar songs with the user whistle. Music Search Engine mainly bases on the application of Digital Signal Processing and Dynamic Programming theory. The project was successful in constructing Music Search Engine. For demonstration purpose, author has built an online music search application which is derived from Music Search Engine above. The demonstration purpose is to show the applicability and potentially commercial value of the project. This search engine as well as the online application was constructed using JAVA technology.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPER 1: INTRODUCTION

1.1 Background of the study

In recent years, the online music industry has achieved a great deal of development. Nowadays we can listen to any songs in the world with just an internet connection. Internet has become a big pool of entertainment, including music. However, the rapid growth has resulted in the difficulty of searching music. The metadata such as title, author, and date of recording is not sufficient to find a composition online. The first difficulty belongs to human beings. Human tends to remember the melody of the music rather than those metadata which are called non-melody information. The second difficulty is that there are a lot of compositions which have similar metadata such as same title; same artist and so on .This results in time-consuming searches for a right composition.

This project is about the music searching by whistle which is a combination of whistle recognition and searching engine. Basically, this system is designed to assist users in finding composition. The system requires a database of compose in MIDI file types. The system is designed to be integrated in any future potential applications which require composition searching. The engine is made up of several modules: noise filter module, whistle recognition module or searching database module.

In the project, I would like to introduce an online music searching web application which is considered as an expanded instance of the system. The application will receive as human's whistle as input. The sound of whistle will be recorded and analyzed by system. The system will transform it into a searchable key before the key is used to match with pool of composition in database. The most matching songs will be listed and ordered in proper ranking in term of music melody similarity. By using this application, user can find a composition even though he or she does not remember the title of the composition or name of artist.

1.2 Problem statement

1.2.1 Problem Identification

Conventionally, in order to find a composition we need to remember some metadata of composition such as apart of title or artist name of the composition. However due to human characteristic, we tend to remember the melody rather than title of composition, hence we may not find that composition even though we may know the melody of it very well.

Moreover, due to the rapid growth in music industry, there are many compositions with same titles available on the internet for example. This makes difficult for us to find the right composition in the jungle of same title compositions. The only method we can use is trial and error, going through list of compositions one by one. This method is ineffective and time-consuming.

1.2.2 Significant of the Project

If the system is successfully constructed, it will assist us a lot in searching for compositions that we want. The integration of whistle recognition and powerful search engine will bring us a flexible option. We just need to whistle an arbitrary part of composition; system will find the most closed compositions to our whistle.

With the success of project, the problem of searching foreign songs could be solved nicely. The advantage of melody-based searching method is that human do not need to know the language of the songs but still they can perform searching. This advantage can not be found in text-based searching method.

1.3 Objective and scope of the study

The project concerns the comparison of key pattern between the original compositions inside database and the whistle of human after going though some preprocessing. The main purpose of this application is to demonstrate a new music searching method and to assess the potent of its commercial value.

The main feature of the application are that 1) it receives human whistle as input 2) it then transforms raw sound of human whistle into note of music. This process is known as whistle recognition. 3) It compares key feature of human whistle with key feature of compositions in database and then 4) it returns the information of the closet compositions against the human whistle in term of melody. The database is constructed by extracting raw MIDI music files. Only the most important information such as main theme of composition is saved.

1.3.1 Objectives of the Project

The objectives are as follows:

1. To transform whistle into music notes with digital signal processing methods.
2. To compare the similarity between two MIDI compositions and quantize it into a form of number in which the order of similarity can be ranked.
3. To provide an innovative Music Search Engine that can be integrated with other related applications.
4. To assess the potential value of music searching by whistle.

1.3.2 Scope of the Study

The scope of study is as follows:

1. Focus on the transformation of human whistle from raw wav data to understandable music notes representation.
2. Focus on a music search engine with capability of searching by human whistle.
3. Focus on construction of a web application which derives from that music search engine to assess the applicable and commercial value.

CHAPER 2: LITERATURE REVIEW

2.1 The Nature of whistle

Whistle is the action of blowing air through lips to make a sound. The flow of air generates a complicated wave which is called whistle sound. The pressure of the sound then travels through the air to the listener's ear or microphone of computer. Microphone is a special device which is used to capture the vibration of sound and convert it into electronic signal.

The frequency of a sound wave is called pitch. When human whistles a high music note, the energy of whistle sound concentrates in high frequency area. Vice versa, when human whistles low music note, the energy of whistle sound concentrates in low frequency area. This characteristic of whistle sound is very important in whistle recognition process later on.

Another important characteristic of human whistle is its less-dependence on the person who whistles. In another word, whistle sound is simple and varies little from person to another. This is one of main reasons why whistling is chosen to boost up accuracy of recognition processes. The figure below shows that from spectrogram view, whistle sound is less complicated than signing sound in comparison with sinus tone.

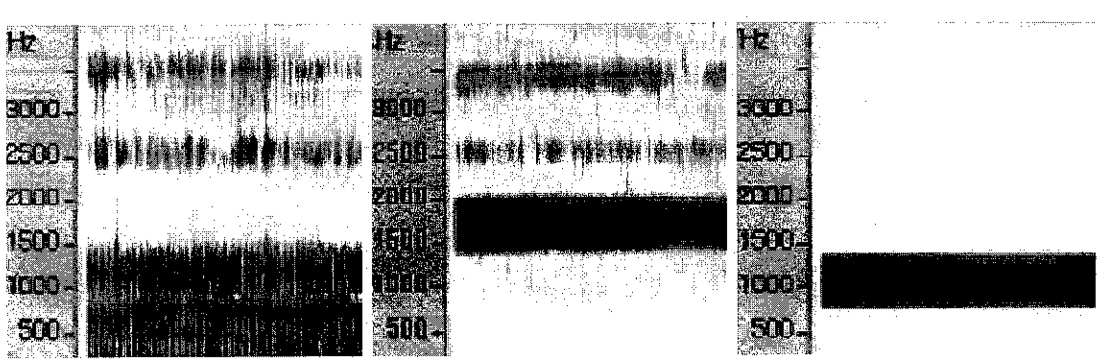

Figure 1: ***Spectrogram of singing, whistling and sinus tone***

2.2 Digital Signal Processing

Digital signal processing (DSP) is the study of signals in a digital representation and the processing methods of these signals. DSP and analog signal processing are subfields of signal processing. DSP has three major subfields: audio signal processing, digital image processing and speech processing. The DSP has applications such as noise filtering, voice recognition and so on. DSP provide a quite number of methods in which the signal data can be analyzed and extracted feature. Whistle is a sound which is represented in the form of signal in general. The DSP is very important in the project especially in whistle recognition parts:

- Whistle sound signal preprocessing such as filtering environment noise
- Fast Fourier Transform (FFT) is very famous algorithm in DSP to transform signal from time domain to frequency domain.

2.3 WAV and MIDI

WAV is the standard archive format of Windows® Operating System in which the sound is captured and recorded digitally. After being recorded by microphone, the human whistle will transform into digital signal through the digitization process. The digital signal of sound will be recorded into WAV directly without any data compression. The disadvantage of WAV file format is its large file size. However, it possesses an advantage that is the processing speed. The sound signal stored in WAV format can be loaded very fast to process.

MIDI stands for Musical Instrument Digital Interface. MIDI is a standard protocol that allows electronic instruments to communicate with each other and with computers, which today is found in various signal production and control devices in the studio and/or stage. It allows one musician to write, perform, and record and entire orchestral composition with only a computer, some instruments, and some software. The advantage of MIDI is its small file size. While the disadvantage is its dependence on instrument library in sound card, that results in the inflexibility of describing natural sounds.

2.4 Approximate String Matching algorithms

Approximate String Matching (ASM) is an algorithm which is used to evaluate the similarity between two given strings. ASM algorithm has much application in automatic grammar correction as we can notice in Microsoft Office Word for instance. ASM also expand its area of application to bioinformatics such as searching and matching DNA.

An approximate string or pattern matching algorithm compares two sequences of data and calculates their similarity by way of an *edit distance*. A set of rules specific to the data being compared defines the costs of operations which are used to transform one sequence to the other, and the total of these costs is given as the edit distance. Fundamental operations include inserting or deleting an element from one of the sequences, and transforming one individual element into another. As a simple example, in comparing two character sequences `abbc` and `abcd`, the former can be transformed to the latter by many sequences of steps; two such transformations are shown in figure below. The algorithm identifies the transformation sequence yielding the smallest edit distance, reporting that value as the matching score for the pair.

`Abbc → abc → abcd`   `abbc → abcc → abcd`

Deletion insertion   transform   transform

Some of the ASM child algorithms are as follows:

- *Edit Distance*: The algorithm calculates the minimum number of basic editing operations which can transform one string into the other. Typical basic editing operations are insertion, deletion, substitution and adjacent transposition. The figure below show how to determine the edit distance between the strings "AVERY" and "GARVEY"

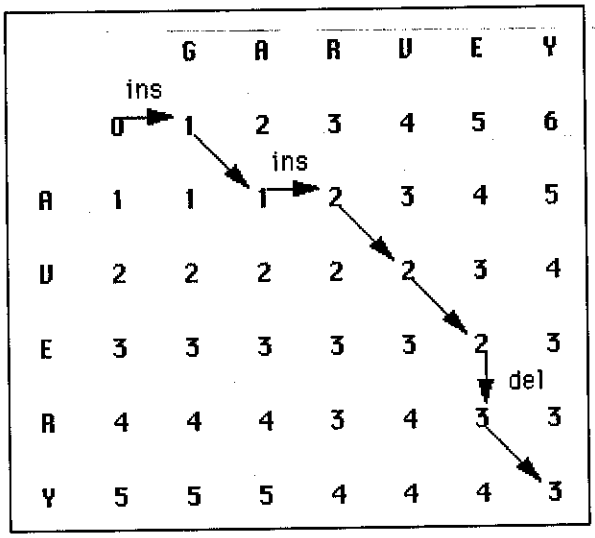

Figure 2: *Edit Distance algorithm in matching "AVERY" and "GARVEY"*

- *Smith-Waterman distance*: A variance of Edit Distance but it is more accurate. However, its disadvantage is the computation complexity.
- *Jaro distance metric*: The Jaro distance metric takes into account typical spelling deviations.
- *BLAST*: Derived from Smith-Waterman and has been recognized as the fastest approximate pattern matching which is used in bioinformatics.

2.5 Related works

2.5.1 Query by humming – Ghias

Hummed queries may be recorded in a variety of formats, depending upon the platform-specific audio input capabilities of Matlab. For example: 16-bit, 44Khz WAV format on a Pentium system, and 8-bit, 8Khz AU format on a Sun SPARCstation. Pitch tracking is performed in Matlab, chosen for its built-in audio processing capabilities and the ease of testing a number of algorithms within it. Autocorrelation, Maximum Likelihood, and Cepstrum Analysis are compared (for pitch tracking).

2.5.2 TuneServer:

A Query by humming system under development since 1997 with approximates matching, editing distance, with a vantage point index. The system used the contour- based in searching music. It used the Edit Distance algorithm to evaluate music similarity. Their publications could be found in Lutz Prechelt, Rainer Typke: An Interface for Melody Input(http://rainer.typke.org/publications.0.html). The newest version of TuneServer is called Musipedia which is whistle-based searching system. The database contains large number of songs About 10,000 classical themes, 1,700 popular themes, and 15,000 folk songs.

2.5.3 Meldex/Greenstone

Meldex/Greenston is The New Zealand digital library which stands for MELody index. It contains folk songs with content-based and metadata search possibilities. The system used Monophonic, Approximate String Matching and Partial Matching to perform searching.

CHAPER 3: METHODOLOGY AND PROJECT WORK

3.1 Methodology

To develop the music search engine as well as the web application, I have specified a number of necessary phases as a guideline in completing the development of this project.

3.1.1 Project Planning

Project planning phases took me a week to complete. In this phase, I did a rough feasibility study of the project problem proposed by my supervisor. The feasibility study provided me a guideline to estimate the time, scope and budget to guarantee the completion of project. During this phase, I determined the scope of the study as well as the objectives needed to achieve at the end of project lifetime.

The basic problem of project was read and analyzed carefully during this time. The purpose was to ensure that all the basic and foreseen obstacles are considered so that time and effort are allocated reasonably.

3.1.2 Project Analysis and Literature Review

Due to the time constraints, fourteen week project life time, my first two week was allocated for project analysis and literature review. The objectives of this two week were:

- To find out of the similar works which have been done previously
- To find out related research to solve project problem such as Fast Fourier Transform, Approximate String Matching, parson Code as well as natural whistle characteristic.
- To find out any open-source that could benefit the system development. The found open-source will significantly reduce the amount of time to resolve transitional problem and shift my valuable time into project further enhancement.

- To sketch out the application framework in term of techniques, technologies with respect to time constraint, referred resources, as well as my capability.

3.1.3 Project Design

Project design phase took me three weeks to complete. In this phase, I came up with a detail design document of system. The project was designed base on concept of OOP.

There were several experimentations being done on critical parts of project especially the parts which required special techniques or technology. The purpose was to make sure there was no difficulty in term of concept before actual implementation of project was launched. All foreseen factors should be considered and taken notes.

3.1.4 Project Implementation

In this phase, I put design into implementation. The core of music searching engine was constructed firstly and tested with a JAVA application interface. The recognition part was completed in the first week after this phase started. My strategy was to test system carefully part by part from the beginning because the inaccuracy in any part of system could result in the failure of system in the next part. The unit testing and module testing was carried on seriously.

One of the most time-consuming sections was Similarity Algorithm. I decided to use Edit Distance method as the similarity-evaluation function. However there were certain modifications being made to original Edit Distance algorithm in order to enhance the accuracy and sensitivity of melody-similarity evaluation.

3.1.5 Evaluate Result

In the last phases, the system testing was performed. The performance of system is accessed through number of testing cases. I did the alpha testing by myself as well as some of my friends.

3.2 Tools

I have decided to use JAVA as programming language to construct Music Searching Engine due to:

- The ease of Object-Oriented language
- The available support of open-sources community which could boost up development stage, make less burden of classic problem.
- The power of JAVA in web application as well as mobile devices.

Beside, Sound Forge 7.0 is the powerful tool which I used to study the whistle sound and perform sound recording, enhancing and analyzing.

CHAPER 4: RESULT AND DISCUSSION

4.1 Result

4.1.1 Whistle Recognition

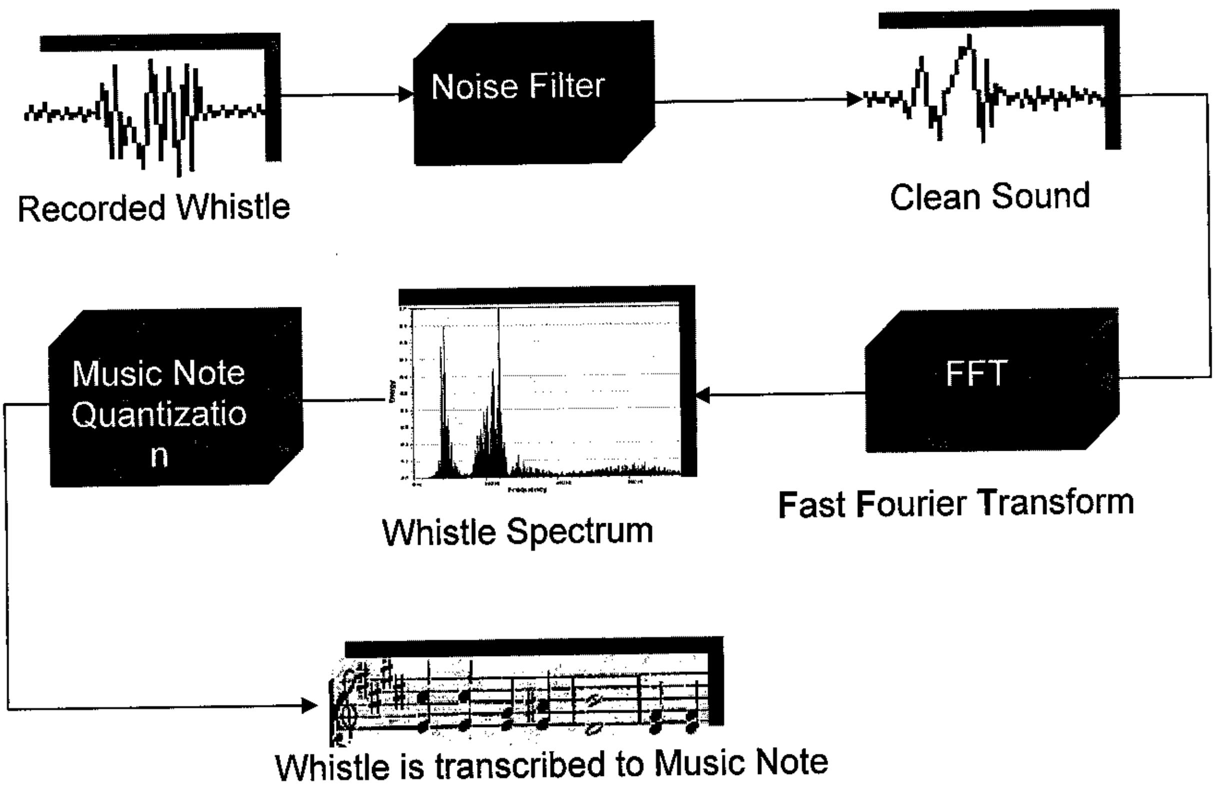

Figure 3: ***Whistle Recognition phase***

I have succeeded in constructing the Whistle Recognition module. Basically, this module will take care of human whistle from recording to building up the transcript of whistle. Whistle Recognition phase contains 3 main processes:

1. **Noise Filter process**: After recording through microphone, the whistle of human is stored in wave form. Because the recorded sound contains noise from environment, the Noise Filter process is needed to filter out the unwanted signal information. Specifically, in the first step, all the too low and too high frequency are canceled out, only the frequency portion which contains whistle is kept. This filter is called Band Pass filter. The second step, all the frequencies where energy

is too low are canceled out. As a result, only the strong signals in interested frequency area are kept for further processing.

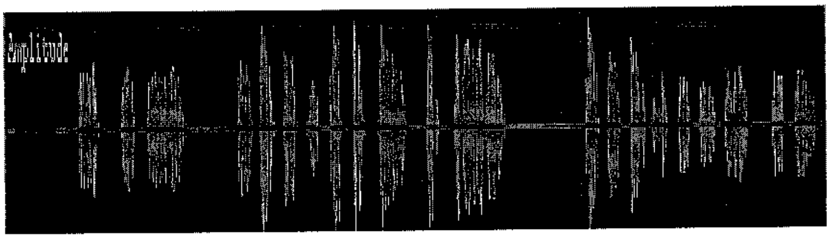

Figure 4: The time-domain wave form of whistle

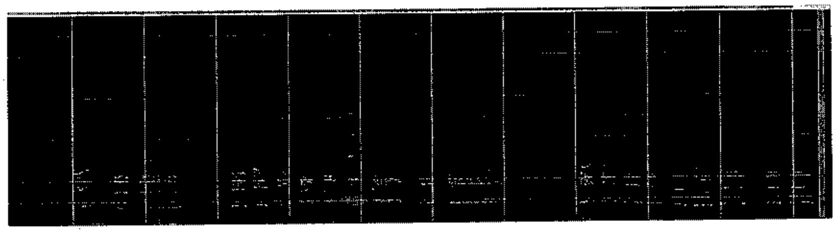

Figure 5: ***The spectrogram of whistle***

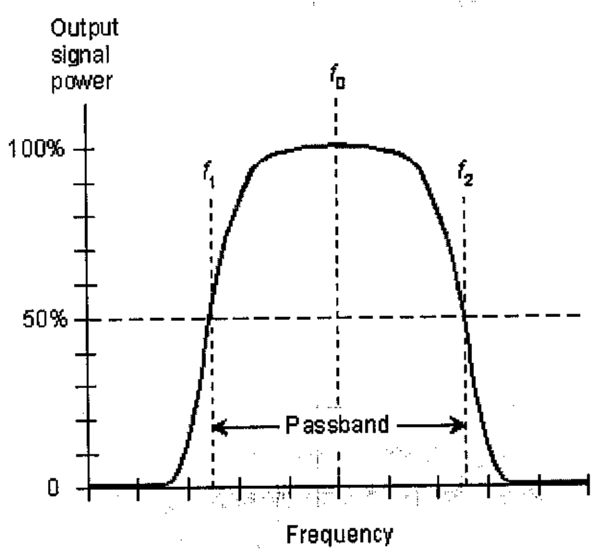

Figure 6: ***The band pass filter which is used in noise filtering process.***

2. **Fast Fourier Transform process**: The clean sound resulted from first process is transformed from Time Domain to Frequency Domain. This is done with Fast Fourier Transform method. FFT is a classically famous algorithm. I have adopted the FFT open source to integrate with my system.

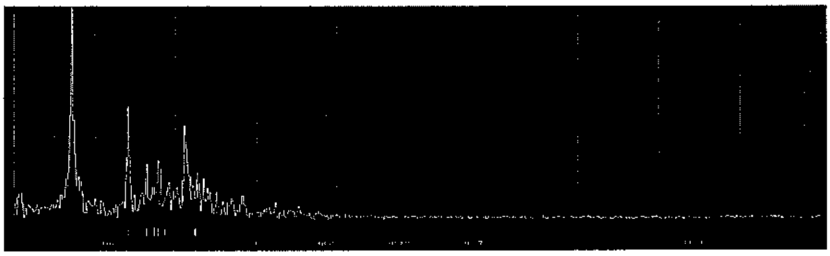

Figure 7: ***The whistle represented in frequency domain.***

3. **Music Note Quantization process**: With the Whistle Spectrum resulted from the FFT process; the system determines the frequency-area which has strongest energy. That frequency-area is used to decide the music note pitch because it has domination power over other frequencies in constructing whistle.

Based on testing result, Whistle Recognition could recognize and transcribe human whistle with confidence level 80%.

4.1.2 Music Searching

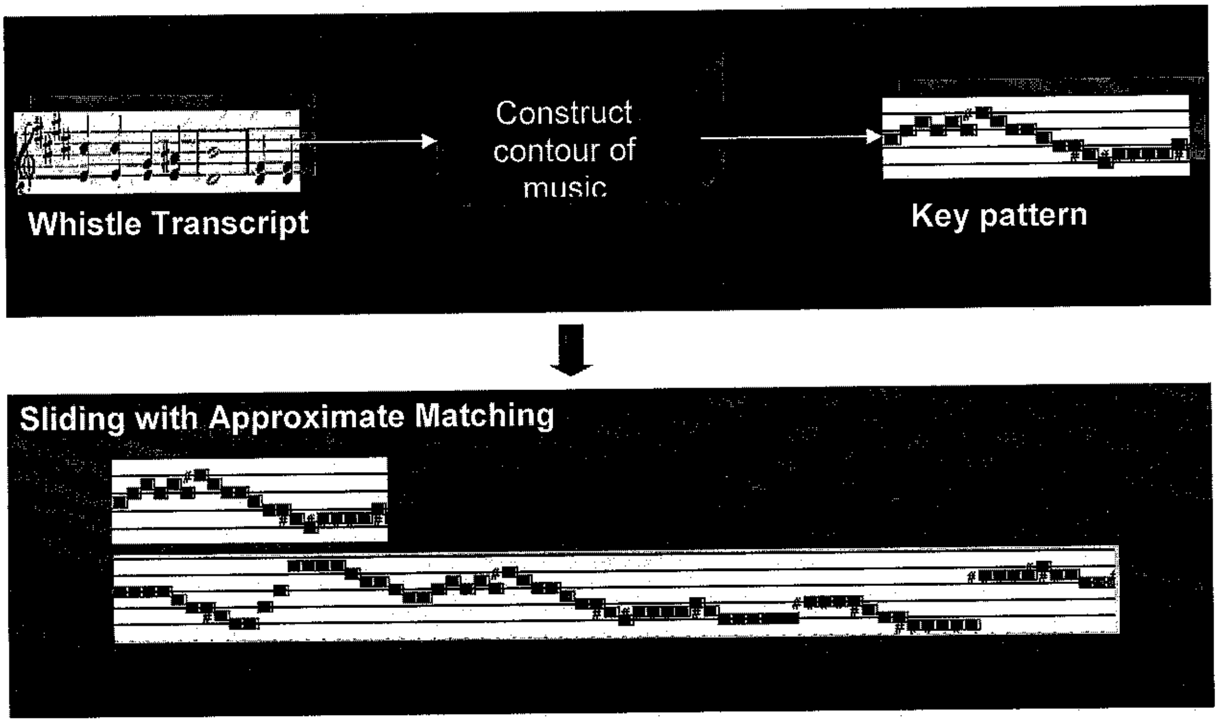

Figure 8: ***Music Searching phase***

This module function is to find the composition of which melody was whistled by user. There are two main processes in this module:

1. **Music's contour construction**: The whistle transcription resulted from previous module undergoes a process called "Music's contour construction". As the result of this process, the shape of music note pitch is depicted in the form of Parson Code.

Sign	**Meaning**
R	If the current note pitch is equal to previous note pitch (Remain)
U	If the current note pitch is higher than previous note pitch (Up)
D	If the current note pitch is lower than previous note pitch (Down)

Table 1: Parson Code table

The contour of music is described as a string contains R, D, U letter, for example "DRUUUUUU" (the parson code of beginning of Yesterday- The Beatles).

2. **Database Searching with Approximated String Matching technique**: The parson code resulted from precious process is used as the key pattern to search compositions in database. In here, compositions are loaded from database and compared with the key pattern one by one to find the best match. The key pattern is slid along the compositions; the Edit Distance algorithm will calculate the similarity between the key pattern and the corresponding portion of each composition. The system need to slide the key pattern because user could whistle at arbitrary point in the song. The maximum value of similarity between whistle key pattern and portions of a composition will become similarity value between whistle and the composition. As a result, system return a list of compositions which was already put in order based on similarity value.

4.1.3 Demo of online music search *fMusic* system:

In order to prove the applicable and potentially commercial value of Music Search Engine, I have built an online music search application called fMusic. fMusic is fully written in Java Server Page, utilized the functionality of Music Search Engine. fMusic allows user to record his/her whistle through internet, processes that whistle on the fly then gives back user information of the song which user has just whistled. The interface of fMusic is designed with simple interface, Google-like look. fMusic also provides Advance Search function in which user has more option of search such as: search by duration, search by title, author name, genre and so on. The result pages return results in order of similarity ranking; the closer similarity will be displayed first. User also can listen to each song to determine the song he/she wants.

4.1.3.1 The interface of fMusic

In the front page of fMusic, user can simply click on **Rec** button and start whistling to microphone. The **Rec** button now changes to **Stop** button. User clicks on this Stop button to finish recording. Immediately, the whistle of user will be analyzed to convert into key pattern by system. This key pattern is placed in the textbox on the screen. With understanding of

Parson Code, advance user can modify the key pattern before performing music search. At this point, the Music Search button will be enabled for user to start searching. By click on this Music Search button, the system will start searching through database to find the best match.

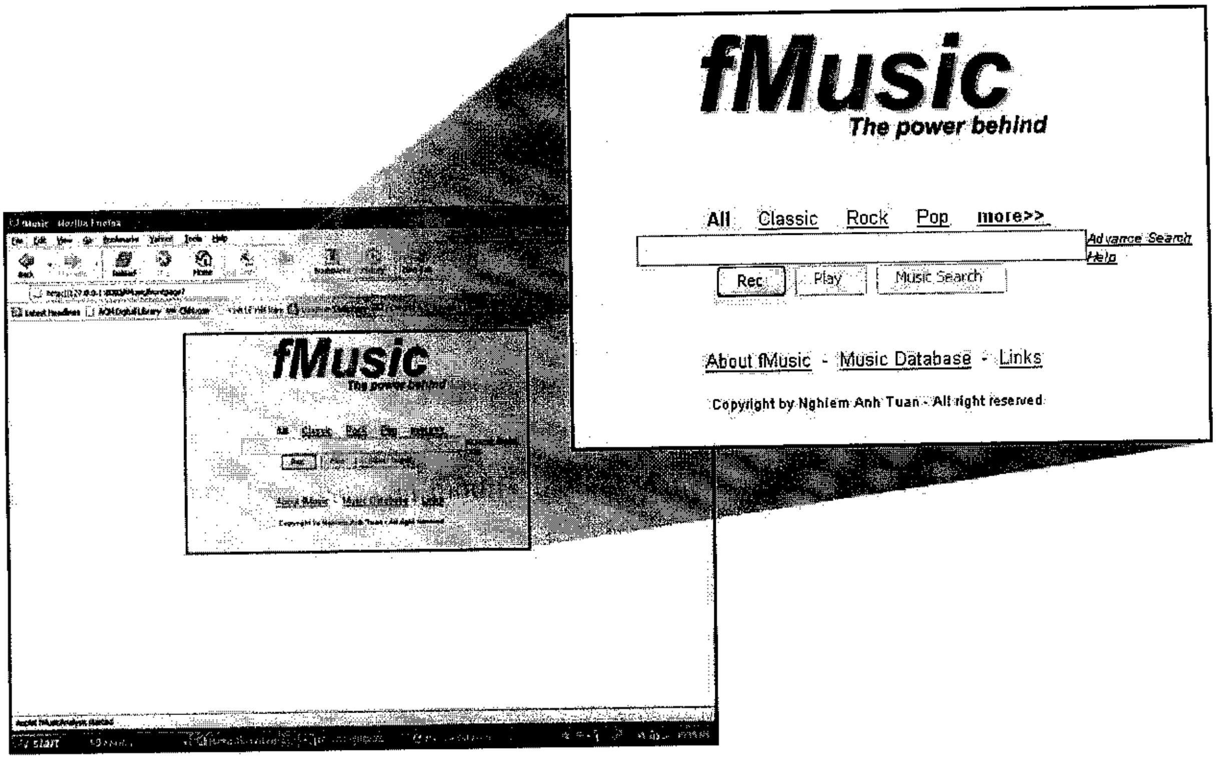

Figure 9: ***The front page of the fMusic web application***

After performing the music search, system display list of compositions as shown in figure 9. Basically, the information of each composition contains title, artist name, genre, comment and the percentage of its match with user whistle. Beside, fMusic gives user option to listen to each composition simply by clicking on QuickTime plug-in player on the page. However, it only plays the main theme of composition. In order to play full composition, user needs to click on the title of composition. A small window will pop up and play the requested composition as shown in figure. In this page, fMusic still provides recording portion on the top of the page so user still can continue his or her searching.

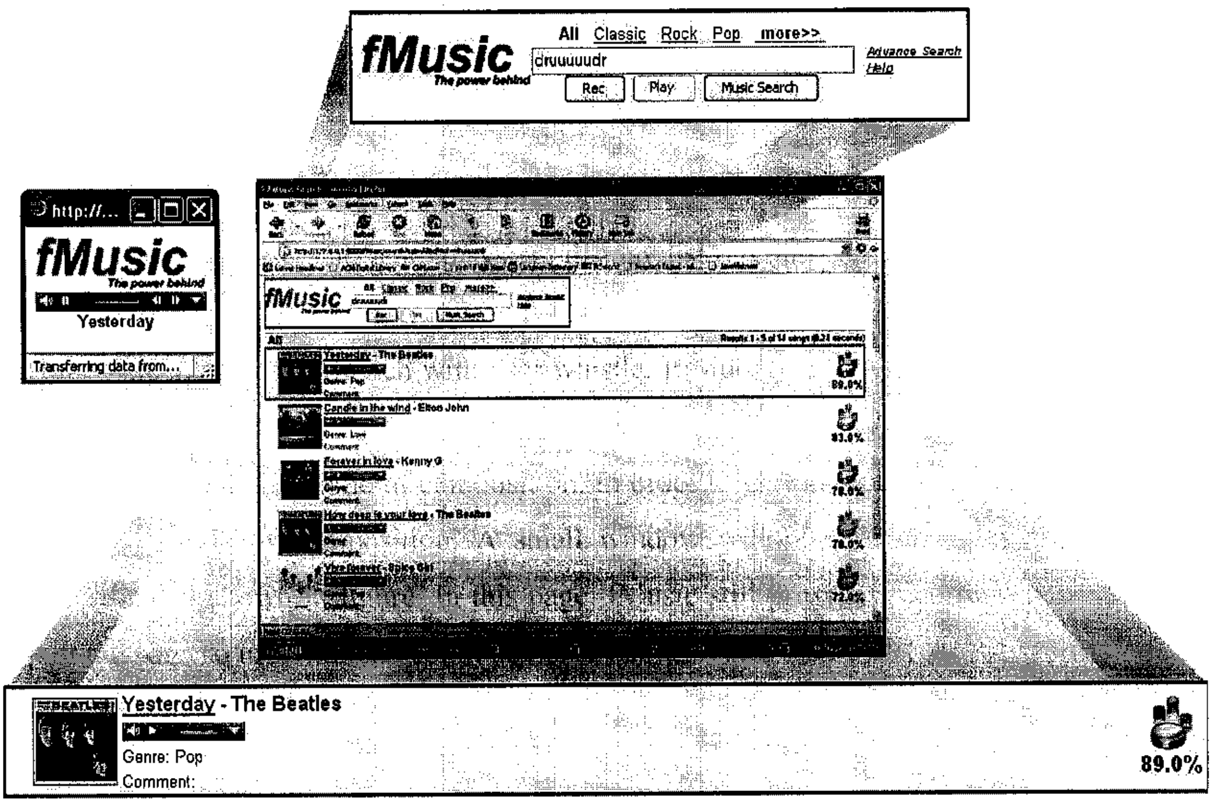

Figure 10: ***Result returned from the fMusic application after user whistles.***

In Advance Search page, user has more options to customize music search. As showing in figure below, user can choose searching by pitch, searching by duration or both pitch and duration. He or she also can customize the number of results returned as well as the minimum ranking. In the bottom of page there is place for metadata of compositions. User can fill in the title of composition, artist name or genre to help system to search more effectively and efficiently.

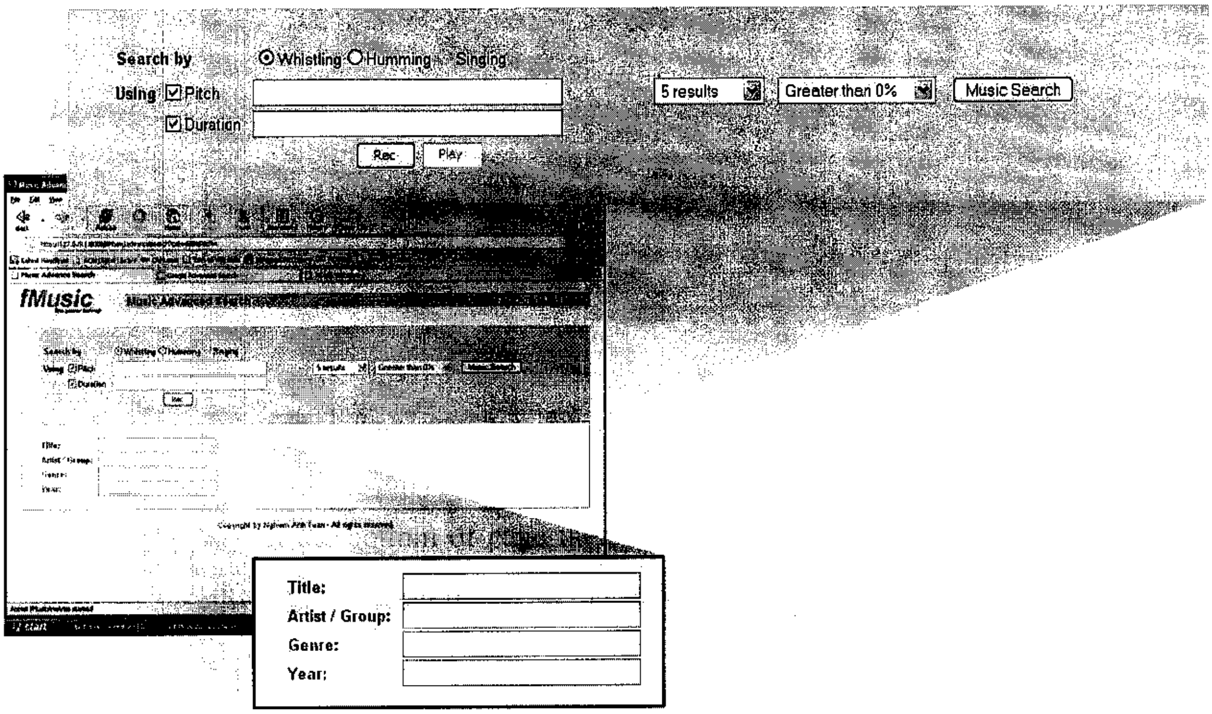

Figure 11: ***The Advance Music Search function with capability of performing search by pitch, duration, or both of them.***

4.1.3.2 System requirement of fMusic

In order to run fMusic on web browser, the client computer needs to have:

- Java Runtime Enviroment (JRE) version 1.5.0.4 or higher
- QuickTime 6.0 plug-in
- JAVASCRIPT is enabled in web browser
- Applet record permission is granted in JRE.

4.2 Discussion

In the system testing phase, I found out that the searching with both pitch and duration returns better ranking than searching with pitch or duration separately. The ranking is better when there are less or no compositions with same ranking value. Although duration is not good as pitch in acting as searching key, it can be used parallel with pitch to provide better result. In specific, an estimated good search based on pitch 85% and duration 15%. Below is the table and graph in one of my experiments on system using pitch + duration and pitch, duration separately.

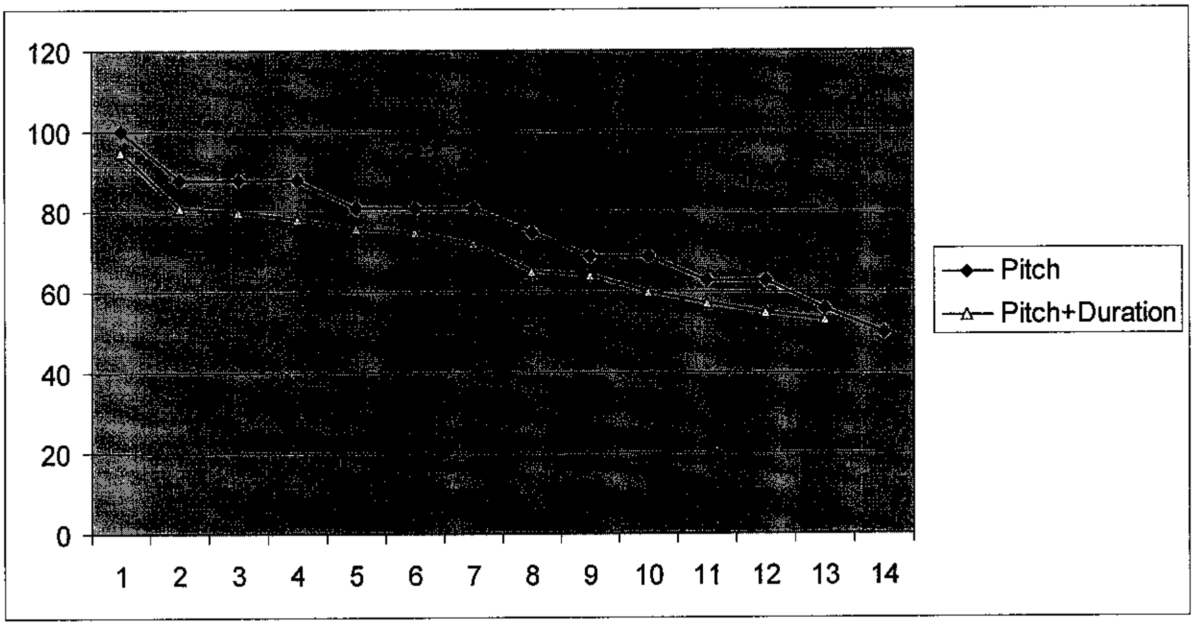

Figure 12: ***Comparison between the result from searching by Pitch only and searching with both Pitch+Duration. The diagram show that result from searching with both Pitch and Duration return less compositions with same value than result from searching with only Pitch. As we can see the effect on the diagram, Pitch results has staircase shape while Pitch+Duration results have values going down gradually.***

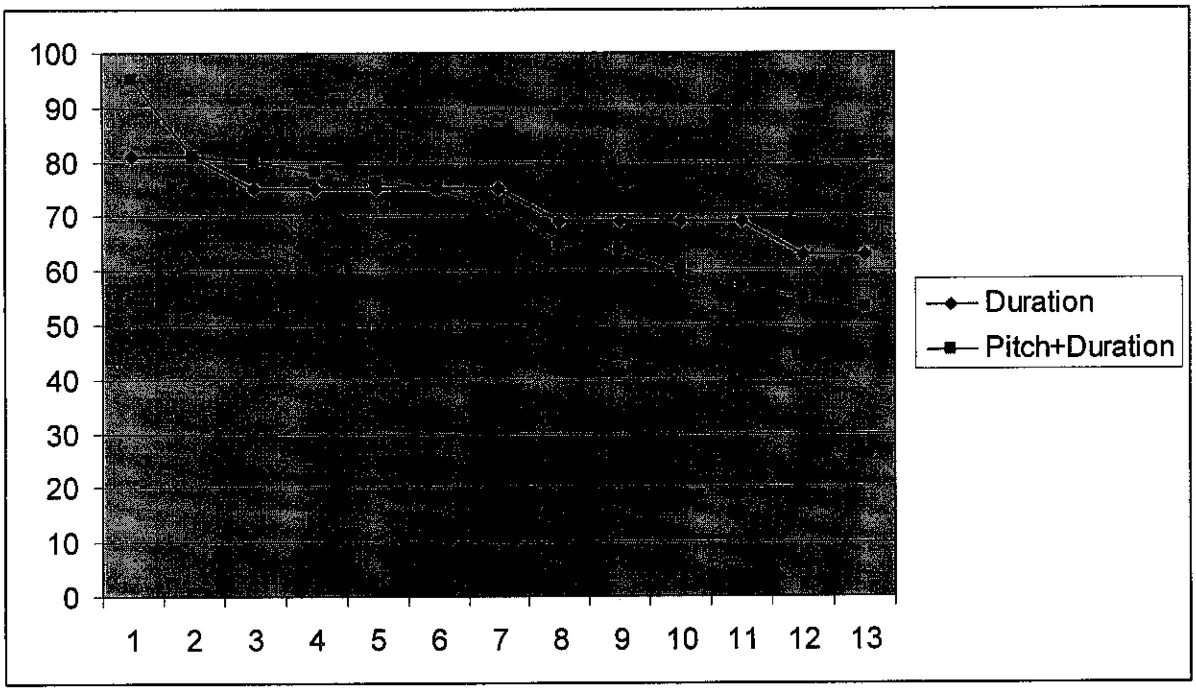

Figure 13: ***Comparison between the result from searching by Duration only and searching with both Pitch+Duration. The diagram show that result from searching with both Pitch and Duration return less compositions with same value than result from searching with only Duration. As we can see the effect on the diagram, Duraion results has staircase shape while Pitch+Duration results have values going down gradually.***

CHAPER 5: CONCLUSION AND RECOMMENDATION

5.1 Conclusion

With a rapid development of entertainment industry, the music searching gradually becomes heavy task for human being. Searching music by whistle has shown the potent of an alternative searching method for as well as a compliment to conventional method. I strongly believe in the applicable and commercial value of the Music Search Engine by whistling in future. The system has been completed in time. For the Final Year Project purpose, I have achieved the objectives. I completed to transform human's whistle to music note representation. I have successfully adopted Edit Distance and SmithWaterman algorithm to compare melody similarity. These algorithms are the heart of music search engine which I also completed in time. This search engine has been integrated into searching application called fMusic. The success of this integration has shown the potent of the engine in integrating with other related applications. Despite success in this first step, the system has not been fully assessed in term of performance in working with larger database and in term of user dependency. So this project still serves as the demonstration of new music searching method and it needs further intensive enhancement and testing in order to apply to reality.

5.2 Recommendation

Personally, due to my belief in system's value, I will continue to develop it further. There are still a lot works being saved for further development such as: 1) Upgrading to humming recognition, in which system can detect pitch from human humming due to the fact that not everyone knows whistling. 2) Enhancing Approximate String Matching which uses BLAST algorithm. BLAST is considered the most powerful algorithm in this moment. 3) Enhancing database searching method, currently linear searching is used which is not efficient. 4) Constructing automatic theme extraction tool which can automate manual database update process.

CHAPER 6: REFERENCES

[1] Prechelt and Typke, An interface for melody input, 2001
http://portal.acm.org/citation.cfm?id=376978

[2] Ghias et al.. Query By Humming -- Musical Information Retrieval in an Audio Database, 1995 http://www.cs.cornell.edu/Info/Faculty/bsmith/query-by-humming.html

[3] McNab et al. The New Zealand Digital Library MELody index, 1997
http://www.dlib.org/dlib/may97/meldex/05witten.html

[4] Deitel deitel : Advanced Java Programming 2002

[5] D. Greenhaus. About the digital tradition.http://www.deltablues.com/, 1992.

[6] 7. W.B. Hewlett and E. Selfridge-Field, editors. *MelodicSimilarity: concepts, procedures and applications*.MIT Press, 1998.

[7] W. Jay Dowfing. Scale and contour: T\vo components of a theory of memory for melodies. *Psychological Review, 85(4):341-354,* 1978.

[8] A. Ghias, J. Logan, D. Chamberlin, and B. Smith. Query by humming. In ACM Multimedia, 1995.

[9] B. Whitman and S. Lawrence. Inferring descriptions and similarity for music from community metadata. In Proceedings of the 2002 International Computer Music Conference. Sweden, 2002.

[10] A. Berenzweig1, B. Logan2, Daniel P.W. Ellis1, B.Whitman3 A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures,
http://www.ee.columbia.edu/%7Edpwe/pubs/ismir03-sim-draft.pdf , 2003

[11] Music Similarity Metrics
http://www.ee.columbia.edu/%7Edpwe/research/musicsim/metrics.html, 2003

[12] Code for the Earth Movers Distance (EMD)
http://ai.stanford.edu/%7Erubner/emd/default.htm, 1999

[13] P. Weijnitz : Monophonic Music Recognition,
http://stp.lingfil.uu.se/~perweij/gslt/kurs/speech/final_speech_2002_weijnitz.pdf 2003

[14] Liljencrants. Speech signal processing. Handbook of Phonetic Science. 1997

[15] Parsons,D. The directory of tunes and musical themes,1975