

# **Computer Assisted Music Instrument Tutoring Applied to Violin Practice**

**Lu Huanhuan**

A THESIS SUBMITTED  
FOR THE DEGREE OF MASTER OF SCIENCE  
DEPARTMENT OF COMPUTER SCIENCE  
SCHOOL OF COMPUTING

**NATIONAL UNIVERSITY OF SINGAPORE**

2009

©2009

Lu Huanhuan

All Rights Reserved

## **Abstract**

Computer Assisted Musical Instrument Tutoring Applied to Violin Practice

Lu Huanhuan

Lecture and practice are the two most important phases in the learning of musical instruments. In contrast to their comparable importance, while lecture is well studied in music education and Computer Assisted Musical Instrument Tutoring (CAMIT), practice is receiving far less attention especially when it is unsupervised.

This thesis focuses on the everyday practice of beginning musical instrument learners and propose a general framework for designing CAMIT systems focusing on unsupervised practice. The thesis also presents interactive Digital Violin Tutor (iDVT), a practical CAMIT system that follows the proposed framework and aims at assisting amateur violin players in unsupervised practice. iDVT provides accurate, informative and intuitive feedback that smooth the learning experience of beginners.

# Contents

<b>List of Figures</b>	<b>v</b>
<b>Chapter 1 INTRODUCTION</b>	<b>1</b>
1.1 Violin Is Difficult for Beginners . . . . .	2
1.2 The Predicament of Practice . . . . .	3
1.2.1 Unsupervised Practice Is Dominant and Crucial in Instru- ment Learning . . . . .	3
1.2.2 Unsupervised Practice Remains to Be Improved . . . . .	4
1.2.3 Unsupervised Practice Will Not Be Replaced in A Short Time	5
1.3 CAMIT Can Help in Unsupervised Practice . . . . .	6
1.4 Thesis Contribution . . . . .	7
<b>Chapter 2 LITERATURE REVIEW</b>	<b>8</b>
2.1 Overview of Current CAMIT Systems . . . . .	8
2.1.1 CAMIT Projects with General Goals . . . . .	8
2.1.2 CAMIT Projects with Specific Goals . . . . .	10
2.2 DVT: The Predecessor of iDVT . . . . .	11
2.3 Summary . . . . .	12
<b>Chapter 3 GENERAL FRAMEWORK</b>	<b>13</b>
3.1 What Is Needed in Unsupervised Practice . . . . .	13

3.1.1	Verification . . . . .	13
3.1.1.1	Self Verification . . . . .	14
3.1.1.2	External Verification . . . . .	15
3.1.2	Instructions . . . . .	15
3.1.2.1	Descriptive Instructions . . . . .	16
3.1.2.2	Demonstrations . . . . .	16
3.1.3	Motivation . . . . .	17
3.2	General Framework for CAMIT System in Unsupervised Practice .	18
3.2.1	Performance Evaluator . . . . .	19
3.2.1.1	Recorder . . . . .	19
3.2.1.2	Transcriber . . . . .	19
3.2.1.3	Evaluator . . . . .	20
3.2.2	Interactive Feedback Generator . . . . .	20
3.2.2.1	Reflector . . . . .	20
3.2.2.2	Instructor . . . . .	22
3.2.2.3	Motivator . . . . .	23
3.2.2.4	Attention Points for Interactive Feedback Generator	24
3.3	Additional Criteria for A Good Design . . . . .	25
3.3.1	Low Cost . . . . .	25
3.3.2	Simplicity . . . . .	25
<b>Chapter 4</b>	<b>iDVT: AN IMPLEMENTED EXAMPLE</b>	<b>27</b>
4.1	Overview . . . . .	27
4.2	Hardware Setting and System Work Flow . . . . .	28
4.3	Technical Details . . . . .	30
4.3.1	Recorder . . . . .	30
4.3.2	Transcriber . . . . .	31
4.3.2.1	Audio Processing . . . . .	33

4.3.2.2	Video Processing . . . . .	33
4.3.2.3	Audio-Visual Fusion . . . . .	34
4.3.3	Evaluator . . . . .	34
4.4	My Contribution . . . . .	35
<b>Chapter 5</b>	<b>USER INTERFACE DESIGN</b>	<b>36</b>
5.1	Overview . . . . .	36
5.2	Functionality . . . . .	38
5.2.1	Reference Panel . . . . .	38
5.2.2	Performance Analysis Panel . . . . .	39
5.2.3	Video Analysis Panel . . . . .	39
5.2.4	Embodiment of Interactive Feedback Generator . . . . .	41
5.3	Layout . . . . .	41
5.3.1	Five-line Staff . . . . .	42
5.3.2	Piano Roll . . . . .	43
5.4	Implementation . . . . .	46
5.4.1	Overview . . . . .	46
5.4.2	Five-line Staff . . . . .	46
5.4.3	Performance Analysis . . . . .	46
5.5	My Contribution . . . . .	48
<b>Chapter 6</b>	<b>ITERATIVE USABILITY EVALUATION</b>	<b>49</b>
6.1	Participant . . . . .	49
6.2	Evaluation Strategy . . . . .	50
6.3	Evaluation Sessions . . . . .	50
6.3.1	Teachers' Session . . . . .	50
6.3.2	Students' Session . . . . .	51
6.4	Summary of Evaluation . . . . .	52

<b>Chapter 7</b>	<b>CONCLUSION</b>	<b>54</b>
7.1	Future Work . . . . .	55
7.1.1	Performance Evaluator . . . . .	55
7.1.2	Interactive Feedback Generator . . . . .	56
7.1.2.1	Instructor . . . . .	56
7.1.2.2	Motivator . . . . .	56
7.2	Further Usability Evaluation . . . . .	57

# List of Figures

3.1	General framework for CAMIT system assisting unsupervised practice.	18
4.1	Hardware setting of the system. . . . .	29
4.2	iDVT fully implements the performance evaluator as the technical core. . . . .	31
4.3	Work flow of the transcriber. . . . .	32
5.1	iDVT incorporates the interactive feedback generator in the user interface. . . . .	37
5.2	User interface of iDVT. . . . .	37
5.3	Fingering and Bowing. . . . .	40
5.4	Five-line Staff. . . . .	42
5.5	Piano Roll. . . . .	43
5.6	Upper panel:Reference Piece; Lower panel:Piano Roll Comparison.(Blue for correctly played notes. Gray for wrongly played notes. Red for corresponding reference for wrongly played notes.) . . . . .	45



## Acknowledgments

I wish to express my deep and sincere gratitude to my supervisors, Associate Professor Leow Wee Kheng and Assistant Professor Wang Ye, for their invaluable support, encouragement, supervision and useful suggestions throughout this research work. Their guidance enabled me to complete my work successfully.

I would like to thank my partner, Zhang Bingjun, for his cooperation in this research project. I really appreciate his time and endeavors in the project. I am also grateful for his advice that gives me encouragement and enlightenment.

I also would like to acknowledge all those who give me advice, comments and evaluations regarding the iDVT system and the thesis.

Finally, I would like to thank my family for their love and support. They are always the light that guide me through hardships and doubts.

# Chapter 1

## INTRODUCTION

Lecture and practice are the two most important phases during the learning process of musical instruments. In contrast with their comparable importance, while lecture is well studied in music education and Computer Assisted Musical Instrument Tutoring (CAMIT), practice is receiving far less attention especially when it is unsupervised.

This thesis focuses on the everyday practice of beginning musical instrument learners and proposes a general framework for designing CAMIT systems focusing on unsupervised practice. The thesis also presents interactive Digital Violin Tutor (iDVT), a practical CAMIT system that follows the proposed framework and aims at assisting amateur violin players in unsupervised practice. iDVT provides accurate, informative and intuitive feedback that smooth the learning experience of beginners.

The thesis is divided into seven chapters:

**Chapter 1 Introduction** explains the motivation and goals of the thesis.

**Chapter 2 Literature Review** discusses the related CAMIT systems and literatures.

**Chapter 3 General Framework** describes the general framework for designing CAMIT systems in unsupervised instrument practice.

**Chapter 4 iDVT: An Implemented Example** describes iDVT, an implemented example of the proposed framework.

**Chapter 5 User Interface Design** explains the user interface design of iDVT.

**Chapter 6 Iterative Usability Evaluation** presents the usability evaluation of iDVT.

**Chapter 7 Conclusion** summarizes the work and plans for future work.

## 1.1 Violin Is Difficult for Beginners

Over the past hundreds of years, the glorious violin is so loved by people that it wins the fame of “the queen of instrument”. Even today, it is among the most popular instruments, which attracts millions of learners all over the world.

However, as the proverb goes, “it is the first step that is troublesome”. Learning to play a violin is not an easy task, especially for beginners.

Violin is difficult for beginners. Unlike piano or guitar, whose keys or frets offer an explicit references for the player to find the correct fingering position, violin has no specific markers for such correspondence. Moreover, due to the special vibration pattern of bowed string (compared to plucked string in the case of guitar), it is subtle for beginners to control the pressure, position and direction of bowing in order to make a conventionally acceptable sound. Demanding as the inherit characteristics of violin is, the learning curve for amateur violin players is rather steep, even frustrating.

Despite the challenges from technical points of view, beginners are also confronted with the predicament of practice, which adds more difficulty to their learn-

ing.

## 1.2 The Predicament of Practice

The learning cycle of beginners in violin and almost all musical instrument playing can be divided into two essential phases, lecture and practice.

Lecture is the phase when teachers take active role in the learner's learning process, who use their professional expertise to equip learners with the knowledge of musical instrument playing and supervise their practical playing.

Practice is the phase when learners take the sole charge of the learning, who consolidate what are taught in lectures, train themselves to control the instrument and sharpen their musical acumen to evaluate the performance through repetitive practicing all by themselves.

The predicament of practice plagues many learners in their early days of learning. Among the many practical reasons that lead to the predicament, the lack of supervision is the core issue.

### 1.2.1 Unsupervised Practice Is Dominant and Crucial in Instrument Learning

"Practice makes perfect" is one of the best-known mottoes among instrument learners. In fact, practice takes the majority of the time learners spent in learning an instrument. Take violin as an example, violin learners most commonly take one or two hours of lecture every week and spend one or two hours every day to practice at home. For hardworking students, the practice time may be even longer. As learners get more and more experienced and mature in playing, lecture will take less and less proportion in the learning cycle, while practice become more and more overwhelming in time consumption.

Practice can be categorized as either supervised or unsupervised, according to whether there are professionals supervising the learner's practice or not. For most of the learners and in most of the times, the practice is unsupervised. The reason is simply that professionals are usually expensive and not always readily available. This fact establishes the dominance of unsupervised practice not only in practice, but also in the instrument learning as a whole.

Thus, the efficiency and effectiveness of unsupervised practice is really crucial to the learners' progress.

### **1.2.2 Unsupervised Practice Remains to Be Improved**

With the absence of supervision from professionals like teachers, the efficiency and effectiveness of practice is weakened, especially among beginners.

Teachers are one of the most influential factors in the current music education, especially in beginning stages. Qualified music teachers are experts in music education acquainted with systematic pedagogy and methodology developed by generations of music educationist and practitioners. Take violin as an example, after more than three hundred years of development and refinement, violin tutoring is considered quite well-studied and mature. Moreover, teachers provide learners with an informative, interactive and supervised learning environment. They can interact with learners, supervise their performance and give valuable instructions and feedback in a timely manner.

But with the absence of teachers in practice, learners can enjoy none of the above benefits. They have no one to show them the right things to do, to tell them whether they are playing correctly or not and to solve problems they can not handle. Especially for beginners, who have very limited music knowledge, music sense and command of the instrument, the outcome of unsupervised practice may be far worse than expected. One common consequence is that learners spend lots of

time and effort in practicing, but show little progress. Moreover, it is also possible that all the time and effort end up consolidating the wrong thing. As the saying goes, “Practice makes permanent”, it will take additional time and efforts to bring the learner back onto the right track.

In the case of children, one of the largest groups of music beginners, the problems is more severe. Being immature both cognitively and psychologically, they will face more hardship and sense of frustration, which may result in their negative attitude towards instrument learning and wear out their interest and initiative.

In light of the above facts, current unsupervised practice is really in need of change and improvement to make the learning more efficient and effective.

### **1.2.3 Unsupervised Practice Will Not Be Replaced in A Short Time**

Seeing the downside of unsupervised practice, it is natural to think of replacing the unsupervised practice with the supervised one. It is reasonable at first glance; however, it is not widely applicable at least in current circumstance.

Practice is time-consuming, yet teaching resource is scarce. The limited teaching resource is definitely impossible to meet everyone’s “One on One” need in practice, when *everyone* here counts in millions. In addition, teaching resource is expensive. Attending lectures is already costly according to current market price, it will be absolutely luxury to think of having one teacher in companion every time one practices. Some learners may be lucky enough to have parents knowing the instrument and having the time to supervise the practice. However, most people simply don’t have this privilege.

Therefore, unless time and cost efficient methodologies are introduced to change the way learners practice, unsupervised practice will keep its dominant status in musical instrument learning.

### 1.3 CAMIT Can Help in Unsupervised Practice

The predicament of practice has been existent for quite a long time. However, people have always been trying to extend help in unsupervised practice using technology. From the tuning fork (which helps beginners tune the instrument themselves) to the recordings (which offer demonstrations for learners to refer to), new technologies keep bringing convenience to and boosting the effectiveness of unsupervised practice.

Now with the prevalence of personal computers and advancement of computer science, the potential of computer technology in promoting music education is catching the eyes of both music educationists and computer scientists. Computer Assisted Musical Instrument Tutoring (CAMIT) stands out as a hot research topic to answer the call for computer technologies in musical instrument tutoring and learning.

As described in [PWT07], many CAMIT projects have come into existence. They take the advantage of multimedia technology in helping learners to learn musical instrument and has won quite a lot of positive feedback. However, with CAMIT being a relatively developing research field, few researchers have paid specific attention to unsupervised practice. This motivates this thesis, which aims at clarifying the important issues and proposing general framework related to the application of CAMIT system in helping unsupervised musical instrument practice of amateur players. It also motivates interactive Digital Violin Tutor (iDVT), a practical CAMIT system developed following such guidelines, which offers learners with useful assistance during unsupervised violin practice.

## 1.4 Thesis Contribution

The contributions of this thesis are as follows. Firstly, it investigates the application of CAMIT systems in unsupervised practice, an important phase in musical instrument learning that are rarely addressed by CAMIT researchers. Secondly, it proposes a general framework for CAMIT systems that focuses on improving unsupervised practice. It is applicable but not limited to bowed string instrument like violin, viola and erhu. Thirdly, it describes the design and implementation of interactive Digital Violin Tutor (iDVT), a system developed following this framework and evaluates its performance in assisting violin learners' everyday practice.



# Chapter 2

## LITERATURE REVIEW

Over the past fifteen years, a number of CAMIT projects have come into existence to assist in musical instrument tutoring and learning.

### 2.1 Overview of Current CAMIT Systems

#### 2.1.1 CAMIT Projects with General Goals

There are some large CAMIT projects aiming at general music educational goals and attempting to provide a complete learning environment. They focus on proposing innovative approaches in both technological and pedagogical level. From technological point of view, they explore solutions for common CAMIT problems such as performance evaluation and feedback. From educational point of view, they leverage present computer multimedia and network technology to enhance self learning, group learning or distance learning.

Piano Tutor [DSJ<sup>+</sup>90][DSJ<sup>+</sup>93] is the pioneer in CAMIT which dated back to 1990. The aim of Piano Tutor is to teach beginners how to play the piano. The core of this project is an expert system that embodies knowledge about teaching the piano. The system keeps track of the user's profile, chooses suitable practice

materials and gives feedback based on the evaluation of the user's performance. Using MIDI piano keyboards instead of acoustical pianos, researchers bypassed the problem of music transcription and focused on designing the core knowledge system for enhanced interactive learning.

IMUTUS [FLO<sup>+</sup>04] [SAH05][SHA04] is an open platform for training students on non-MIDI musical instrument. It mainly focuses on recorder, a traditional wind instrument widely taught in European schools. The key components of IMUTUS are a virtual teacher and a score viewer. The virtual teacher focuses on performance evaluation, which transcribes the user's play into MIDI and does the evaluation. The score viewer is the graphical interactive user interface that reflects user's own performance, shows the evaluation results and gives comments or hints. It also explores and includes components like optical music recognition and distance learning, which may be helpful for teachers and students.

VEMUS [web07b] [FLO07] pushes the work of IMUTUS a step forward. Instead of focusing on recorder, VEMUS embraces more popular wind instruments, such as the flute, the saxophone and the clarinet. Besides self learning and distance learning, it also explores the possibility of enhancing music education using computer technology in the classroom. The score viewer in IMUTUS is further enriched by emoticons, hand written annotations, audio annotations and real-time audio processes.

i-Maestro [web07a] is an ongoing project having board coverage of CAMIT. It covers self learning, collaborative learning and distance learning. It also touches on various aspects like gestural interfaces, augmented instruments and symbolic music representation. The project is still in progress and we look forward to their further results.

Piano Tutor and IMUTUS show the interest of early CAMIT researchers in completely replacing real teachers with computers in teaching learners musical

instruments. With such a big goal, the two systems have to tackle the knowledge system, the performance evaluation and the user interface all at one time, all of which are difficult even in today's point of view. Due to technological constraints, they are implemented with compromises here and there (like using MIDI piano keyboards instead of acoustical pianos in the case of Piano Tutor and using simple instrument recorder, which is more general music educational than musical instrument oriented in the case of IMUTUS).

Although regarded as the successor of IMUTUS, VEMUS is not representing the same design goal of IMUTUS. It is shifting its focus to distance learning and collaborative learning, all of which aim to offer teachers and students with a virtual learning environment bringing in new learning experience.

All of the above systems touches CAMIT in unsupervised practice, but none puts it as the main focus and discuss it in detail.

### **2.1.2 CAMIT Projects with Specific Goals**

There are also quite a few small CAMIT projects with specific goals. They usually start with a particular need or problem in real application scenario and offer solution for a specific goal. A big proportion of them are researching on how to offer meaningful feedback to users, which is also relevant to our application scenario.

PianoFORTE [SWK95] is the early work for visualizing real piano performance. The aim of this work is to convert dynamics, tempo, articulation and synchronization of both hands into expressive symbols, which will facilitate the understanding of the evaluation results. In [FMC05], a visualization that integrates multiple feedback sources are provided in real time. In [HSD06], a review of real-time visual feedback in singing training is given. In [Fer06], the potential of sound in giving feedback is explored.

MEAWS [web08] is an open-source program that creates simple games for

music students to practice rhythms and violin intonation. The system mainly deals with the research problems of automatic exercise creation, audio analysis, and visualization of errors. Being a violin teacher himself, the author also lays out some general principles for musical instrument learning and CAMIT systems in his Master thesis [Per08], which is really insightful and practical.

## 2.2 DVT: The Predecessor of iDVT

In particular, I would refer to DVT (Digital Violin Tutor)[BWL06] [LWB06] [YDHW04] [YWH05], the predecessor of the iDVT system to be presented in this thesis. Aiming at providing useful tool for violin practice, DVT actually tries to tackle two problems, music transcription and feedback.

Being most essential in performance evaluation, music transcription is one main concern of DVT. A fast music transcription algorithm is proposed which is specially adapted for violin and home application. In addition, video, piano roll notation, 2-D animation of the fingerboard and 3-D animations are provided as meaningful feedback.

DVT lays the foundation for iDVT in three ways. Firstly, it puts a narrow yet valuable user scenario, unsupervised violin practice, at the core of the research. It clarifies the scope of the research in this direction and pioneers in providing useful solutions. Secondly, it points out the two main components in the systems related to unsupervised musical instrument practice, transcription and feedback. It incubates the general framework proposed in this thesis, which in turn guides the design and implementation of iDVT. Thirdly, it offers a fast and accurate audio processing algorithm for transcription, which is further improved by the method used in iDVT.

## 2.3 Summary

From the review of the current CAMIT systems and work listed above, we can find that unsupervised musical instrument practice is not emphasized in these systems significantly enough. Unsupervised musical instrument practice is cursorily touched, vaguely presented in concept or simply omitted. A general framework is really in need to clarify the important factors of improving unsupervised practice, to study the needs behind it and to guide the design and development of practical systems.

# Chapter 3

## GENERAL FRAMEWORK

In this chapter, the beginning learners' needs during unsupervised practice are analyzed. A general framework for CAMIT system in unsupervised musical instrument practice is proposed considering such needs. Some basic criteria in design and development are also discussed.

### 3.1 What Is Needed in Unsupervised Practice

Before proposing a general framework for unsupervised practice, it is absolutely necessary to get insight into the needs of learners in real unsupervised practice scenario. The needs of learners can be summarized by the following three aspects.

#### 3.1.1 Verification

Verification is the evaluation of the violin player's performance, which is then fed back to the player for adjustment and improvement in the subsequent playing. The verification of musical instrument playing typically includes two aspects: sound and gesture. The basic criterion for verification is the correctness, which investigates whether the performance is within a tolerable *threshold* compared to the standard

reference. The advanced criterion for verification is the expressiveness, which is more or less subjective and differs from person to person.

Verification is needed in unsupervised practice, since it is the foundation for error discovery and correction. A lack of verification will make the practice totally a waste of time, since the player will have no judgment of his performance and cannot improve accordingly.

Verification can be either internal which is made by the instrument player himself, or external which is given out by professionals who are supervising the player. In the following two subsections, these two kinds of verifications will be discussed in detail.

### **3.1.1.1 Self Verification**

Self-verification is the mainstream during the course of unsupervised practice, since there are generally no professionals in companion during the unsupervised practice, as described in Section 1.2. It naturally counts on the player himself to do the verification.

However, amateur learners are usually not able to make accurate self verification. On the one hand, beginners are usually too busy to analyze their performance carefully during the practice. Controlling an unfamiliar musical instrument requires a vast amount of concentration. Since the beginner is already fully occupied by memorizing the rhythm, keeping up with the tempo and coordinating both hands, they simply do not have any cognitive power left to carefully listen to what they have played, let alone critical analysis.

On the other hand, good sense of music is required to make accurate self-verification, which takes long time of training. As most beginners are inexperienced, they merely do not have the ability to evaluate their play. It is highly possible that the learner honestly believe that he played at the right pitch when in fact it is

way out of tune. Even if they do feel something wrong, they would not be able to articulate what the problem is and figure out where the error occurs. Thus, the effectiveness of unsupervised practice is deeply hampered.

#### **3.1.1.2 External Verification**

In view of the incapability of self-verification for beginners, external verification is really in need in unsupervised practice. But due to the unavailability of professionals as external verification sources, it becomes natural to call for an easily accessible substitution that can do the monitoring, evaluation and feedback during unsupervised practice.

### **3.1.2 Instructions**

During unsupervised practice, instructions telling learners what to do and how to do with the instrument are very commonly needed, especially among amateurs. On the one hand, as new knowledge keeps pouring in during the early days of instrument learning, it is natural for learners to miss important points here and there during the lecture. It is also very common for them to forget what was taught as the interval between two consecutive lectures usually spans one week. Thus, the presence of instructions could serve as a good recap that consolidates the concepts and theories taught during the lectures.

On the other hand, beginners are not experienced enough to put what was learned into real practice. Proper instructions can help learners quickly get on the right track. It not only accelerates the learning process, but also ease the anxiety and frustration that usually plagues beginning learners.

However, it should be clarified that instructions, being more theoretical than practical, are just complementary in unsupervised practice. The main focus is always the practical training instead of the theoretical learning. This is also the



most important point that distinguishes practice from lectures.

In current musical instrument tutoring, two kinds of instructions are most common.

### **3.1.2.1 Descriptive Instructions**

The most conventional instructions are in the form of words and sentences describing the actions to take and the things that need attention. They are familiar to learners as they are frequently seen in text books and heard from teachers.

### **3.1.2.2 Demonstrations**

During the early stage of musical instrument learning, the aim of practice is to mimic the standard play as closely as possible. Thus, a clear demonstration is very essential to set up a good example for the reference of learners.

Moreover, being highly demanding in body control and coordination, learning to play a musical instrument is quite different from learning academic subjects which mainly involves mental work. Compared to reading lines of descriptive words on textbooks, learning by example will be much more concise and understandable in most cases.

Demonstration can take various forms relating to different human perceptions. Currently, it can be visual in the form of pictures or video clips showing the playing gestures of professionals. It can also be aural in the form of audio clips showing the reference melody.

As technology and music pedagogy advance, more forms and perceptions may be adopted in demonstration. One possible breakthrough may be the tactile perception. As was briefly introduced in Section 1.1, the bowing of violin is tricky for beginners. If the bow is not pressed hard enough, the violin may produce an unacceptable sound called “surface sound”. However, if the bow is pressed too

hard, the violin may produce a raucous “graunch” noise, which is also undesirable. If the demonstration can simulate the pressure on the hand in correct cases, it is definitely useful to help learners command the correct bowing method.

### 3.1.3 Motivation

Human beings are fickle in their affections. Therefore, they hate dull and repetitive things. Humans beings are social animals, too. Thus, they also fear loneliness. But unfortunately, practice is inherently a combination of both repetitiveness and loneliness. Months and years of such practice may readily wear out one’s passion for the instrument, which renders any further practice meaningless. Thus, motivation is what learners need to make practice not only effective but also enjoyable.

There are many ways to motivate learners in education, which are also good references to be applied in practice. Three of them are most common. The first one is to *attract* learners. The learning content is presented in an interesting and entertaining way to hold the learners’ attention longer. The second one is to *comfort* learners. Words of encouragement and appraisal like those from teachers usually achieve this goal well. The third one is to *offer companions* to learners. Compared to studying alone, group learning or collaborative learning usually have better results.

However, it should also be clarified that motivation has lower priority than verification in CAMIT system design. After all, the ultimate goal of CAMIT system is education rather than entertainment.

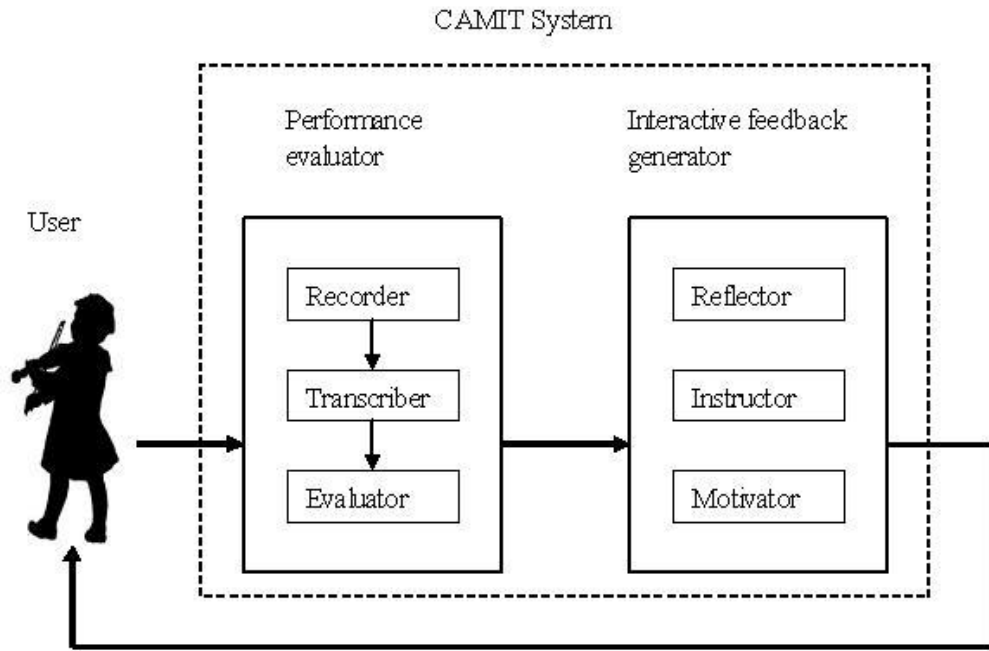


Figure 3.1: General framework for CAMIT system assisting unsupervised practice.

### 3.2 General Framework for CAMIT System in Unsupervised Practice

With the needs of learners in view, the general framework for CAMIT in unsupervised practice can be illustrated in Figure 3.1.

The framework consists of two major components, performance evaluator and interactive feedback generator. Performance evaluator focuses more on the technical part of the system, which tackles the problem of offering external verification with the help of computer technology. Interactive feedback generator focuses more on the human-computer interaction, which tackles the problem of presenting interactive feedback that ensure the system usability and promote the learning effectiveness. They are the most essential building blocks for CAMIT systems in un-

supervised practice. They can be further decomposed into smaller modules, serving more specific needs summarized above. This section will explain them in detail.

### **3.2.1 Performance Evaluator**

As described in Section 3.1.1.2, the most essential need of beginners is the external verification of their performance. Performance evaluator is the core of the framework which aims at addressing this problem using computer technology.

Performance evaluator consists of three modules, the recorder, the transcriber and the evaluator.

#### **3.2.1.1 Recorder**

The recorder records the user's performance in digital formats that can be further processed by computers. With the maturity of sensors and digital media, recording is no longer confined to audio and video, which provides powerful arms and much potential for CAMIT to come up with novel methodology that would push forward music education.

#### **3.2.1.2 Transcriber**

The transcriber extracts useful information from the raw data and transforms it into certain representations convenient for subsequent evaluation. Depending on different aims of verification, different representations may be adopted. For example, the verification aiming at pitch accuracy of sound probably needs representation that contains aural information, while the verification aiming at gesture correctness may adopt representation that holds kinetic information.

### 3.2.1.3 Evaluator

The evaluator compares the transcription results with the reference to provide the evaluation of the performance. The evaluator is an indispensable module in the performance evaluator.

The above three modules constitute a typical performance evaluator and also form the technological core of a CAMIT system for unsupervised musical instrument practice.

## 3.2.2 Interactive Feedback Generator

Interactive feedback generator is the component of the framework that provides users with informative and interactive information during unsupervised practice. It lays more emphasis upon improving the user experience and aims at boosting the usability and the effectiveness of the CAMIT system in serving music educational purposes.

In contrast with the performance evaluator which focuses on solving one particular problem and meeting one specific user need, interactive feedback generator is a hodgepodge that incorporates miscellaneous user needs.

As opposed to the three modules of the performance evaluator, which are highly correlated and appear concurrently in the system, the three modules described in interactive feedback generator are relatively independent of each other. Different CAMIT systems can selectly implement one or more of them according to their emphasis on users' needs.

### 3.2.2.1 Reflector

Reflector is the module in interactive feedback generator that provides the user with a clear picture of his own performance. It is an extension of the *mirrors* used

in conventional musical instrument tutoring.

Mirrors have been a common property in musical instrument tutoring to help learners get a better view of their own gestures. Leveraging the modern computer technology, the reflector can do much more than what mirrors can. The reflector is powerful in the following three aspects.

**Break the time constraints** The reflector can improve the mirror in breaking the time constraints. The mirror reflects the player's gestures when the performance is in progress, which means the player should keep an eye on the mirror while playing to check his gestures. However, this practice is not effective because concentration can hardly be split between playing the instrument and checking gestures through the mirror, especially for beginners.

The reflector keeps tracks of the player's performance and makes it possible for the checking to be carried out after the whole performance. This enables the player to concentrate on the playing while performing, while investigate more carefully about the gestures when self-checking.

**Break the visual constraints** The reflector can improve the mirror in breaking the visual constraints. The mirror only provides visual information to the user, which is just a fraction of the whole picture of the performance: aural and tactile information, for example, also provide invaluable information about the user's performance.

With the maturity of digital cameras and sensors, the reflector can do much better in recording and presenting the player's performance from more meaningful aspects of perceptions.

**Break the feedback constraints** The reflector can improve the mirror in breaking the feedback constraints.

It is true that the ideas behind previous two points have mature counterpart in real practice such as cassette recorders and cameras. However, the most important point that distinguishes the reflector from these counterparts is that, instead of merely recording and revitalizing the performance, the reflector receives analyzed results from the performance evaluator and feed back to users in more intuitive ways.

Remember the end goal for recording and reproducing the performance is for verification. Conventional recorders honestly reproduce the performance and leave the user to make verification through it. But the reflector has the potential to put it a step forward, which not only feed back the performance, but also the verification results to the user. As described in Section 3.1.1, this is really in need to amateur learners.

In this framework, the reflector can be regarded as the front-end of the performance evaluator in 3.2.1 and is usually indispensable.

### 3.2.2.2 Instructor

Corresponding to the user need mentioned in Section 3.1.2, the instructor provides instructions to guide the users during unsupervised practice.

Following the categorization of the instruction in Section 3.1.2, the instructor can take the form of descriptive instructor or demonstrator accordingly.

**Descriptive Instructor** Generally speaking, the descriptive instructor in CAMIT system provides instructions in words, which describe what to do, when to do it and how to do it during the course of practice. It is similar to the conventional text books in serving this end except that it may adopt more interactive features. Instead of waiting for the users to search and browse for the instructions, as in the case of text books, the descriptive instructor may analyze the context of the user considering his performance and progress, and give instructions accordingly. The

most primitive implementation of this idea can be displaying hints and instructions for each etude or practice session the user comes to. This implementation can already be seen in some existing music educational systems. However, it remains under-explored to give instructions more intelligently and interactively with better analysis of the user's context and needs.

**Demonstrator** The powerful multimedia capability of computer technology makes CAMIT a perfect carrier for multi-modal demonstration.

It is true that traditional recording devices like record and cassette tapes have already been used as storage media to preserve audio and video demonstrations for students' repetitive reference. However, with the maturity of digital media, a personal computer can provide all-in-one solution combining all these old technologies. In addition, it is rather cheap and convenient to create, to distribute and to preserve such contents.

### 3.2.2.3 Motivator

Motivator is the component that CAMIT systems can incorporate to enhance unsupervised practice. The popularity of computer games and the thrive of edutainment have laid good foundation for CAMIT system to achieve motivation goals.

However, one thing should be clarified beforehand is that such incentives should not go too far from the true goal of CAMIT systems: musical instrument tutoring. Instrument focus should always be guaranteed. Here the meaning of instrument focus is two-fold.

Firstly, the user should really be playing the instrument. Adapted instruments like the game consoles in the popular music game *Guitar Heroes* [web09] are not plausible to be used in CAMIT systems, since adapted instruments and real instruments are totally different. The experience of practicing on these *fake* instruments has nothing to do with improving real instrument playing.



Secondly, the user should be able to develop musical capability through using the system. Take Guitar Heroes again as an example, instrument play has somewhat been mutated into a shooting game in this case. Instead of training musical acumen, musical sense and proficiency of instrument playing, Guitar Heroes is more of training motor reflex and memorization. The music educational contribution of it is really limited.

I do not mean to blame the design of Guitar Heroes when taking it as the example. After all, Guitar Heroes is just a successful game for entertainment purposes rather than a CAMIT system for music educational purposes. My point is to alert what the consequence will be if the motivator goes blindly too far.

#### **3.2.2.4 Attention Points for Interactive Feedback Generator**

**Timing and Method** When and how to introduce interactive feedback are subtle. As has been discussed previously, the concentration and cognitive power of learners are very limited during the practice. Feedback appearing at improper time and in improper manner may distract and confuse learners instead of helping them. Thus, the timing and method adopted in providing feedback should be carefully considered in the design of interactive feedback generator.

#### **Relationship between Interactive Feedback Generator and User Interface**

Interactive feedback generator is a term I improvise to illustrate and emphasize conceptually the essential feedback component in the design of CAMIT systems for unsupervised practice. In practical system development, interactive feedback generator is melt down into the user interface design and implementation in order to adapt to the integrity and overall style of user interface.

### 3.3 Additional Criteria for A Good Design

There are some additional criteria for a successful CAMIT system focusing on unsupervised practice.

#### 3.3.1 Low Cost

Although computer technology is developing at an ever-increasing speed and have made extraordinary achievements for the human civilization, it is safe to say that human teachers cannot be replaced by computer systems, at least in the foreseeable future. A teacher's role in music education not only includes the teaching of knowledge, but also includes human-to-human communication and interaction, which involves mood and psychology etc. Unless artificial intelligence is powerful enough to simulate human mind and behavior, CAMIT can only be an auxiliary providing limited functions.

Therefore, currently speaking, one important factor that justifies the feasibility of CAMIT system is the comparable low cost. If the cost of a CAMIT system is far beyond that of a teacher, why would learners bother to use a computer program instead of to hire a home tutor?

#### 3.3.2 Simplicity

Simplicity is beauty. A practical CAMIT system should be as simple as possible, because what end users care most is not how complicated the system is, but whether it can get the work done or not. Besides, it should always be made clear that the focus of users in practice is the instrument play, not the CAMIT system. Instead of digging deep into sophisticated algorithms or technologies, a retrospect of how to better serve the users needs is more beneficial.

The meaning of *simple* here is comprehensive. Firstly, the system should be

simple to setup. The setup is most preferable to be fully automatic and everything is done once for all. Secondly, the system should be simple to use. Few users will go through manuals before start. Neither do they bother to try functions only achievable with the presence of manuals. Lastly, the system should be simple to understand. This means all the results should be as self-explanatory as possible.

# Chapter 4

## iDVT: AN IMPLEMENTED EXAMPLE

Following the general framework outlined above, we have developed interactive Digital Violin Tutor (iDVT), a practical CAMIT system aiming at assisting amateur violin players in unsupervised violin practice.

### 4.1 Overview

The pedagogical foundation of iDVT is educationist David Perkin's *Theory One*[Per95], which summarizes four essential aspects of effective learning:

- Clear information
- Thoughtful practice
- Informative feedback
- Strong intrinsic or extrinsic motivation

Inspired by *Theory One*, iDVT aims to be an intelligent practicing companion providing amateur violin learners with these four essence and build a genuinely new learning environment which is both fun and effective.

iDVT has the following three main benefits. Firstly, it provides informative feedback which boosts the learning efficiency of beginners during unsupervised practice. Secondly, it is convenient for students to access in home environment, which gives learners more flexibility over the time and place they learn and practice. Thirdly, the hardware configuration of the system is low and cheap, which is affordable and cost-saving for general public.

As a complete system following the framework proposed previously, iDVT illustrates the capability of the framework in guiding the development of CAMIT system in unsupervised musical instrument practice. It is immediately foreseeable that the framework can be extended to other string instruments like viola and er-hu. It also has the potential to be applied to musical instruments in a wider scale.

The system is jointly developed by Zhang Bingjun and me under the supervision of Assistant Professor Wang Ye at Sound and Music Computing group, National University of Singapore. My contribution in developing the system will be clarified at the end of Chapter 4 and Chapter 5 respectively.

## 4.2 Hardware Setting and System Work Flow

The hardware setting and technical work flow of iDVT system are shown in Figure 4.1.

iDVT system is used when the learner practices a violin piece following a reference notation. The system has two ordinary webcams and one microphone as peripherals, recording the audio of the playing as well as the videos from the front view(focusing on the bowing) and bird's eye view(focusing on the fingering) of the learner.

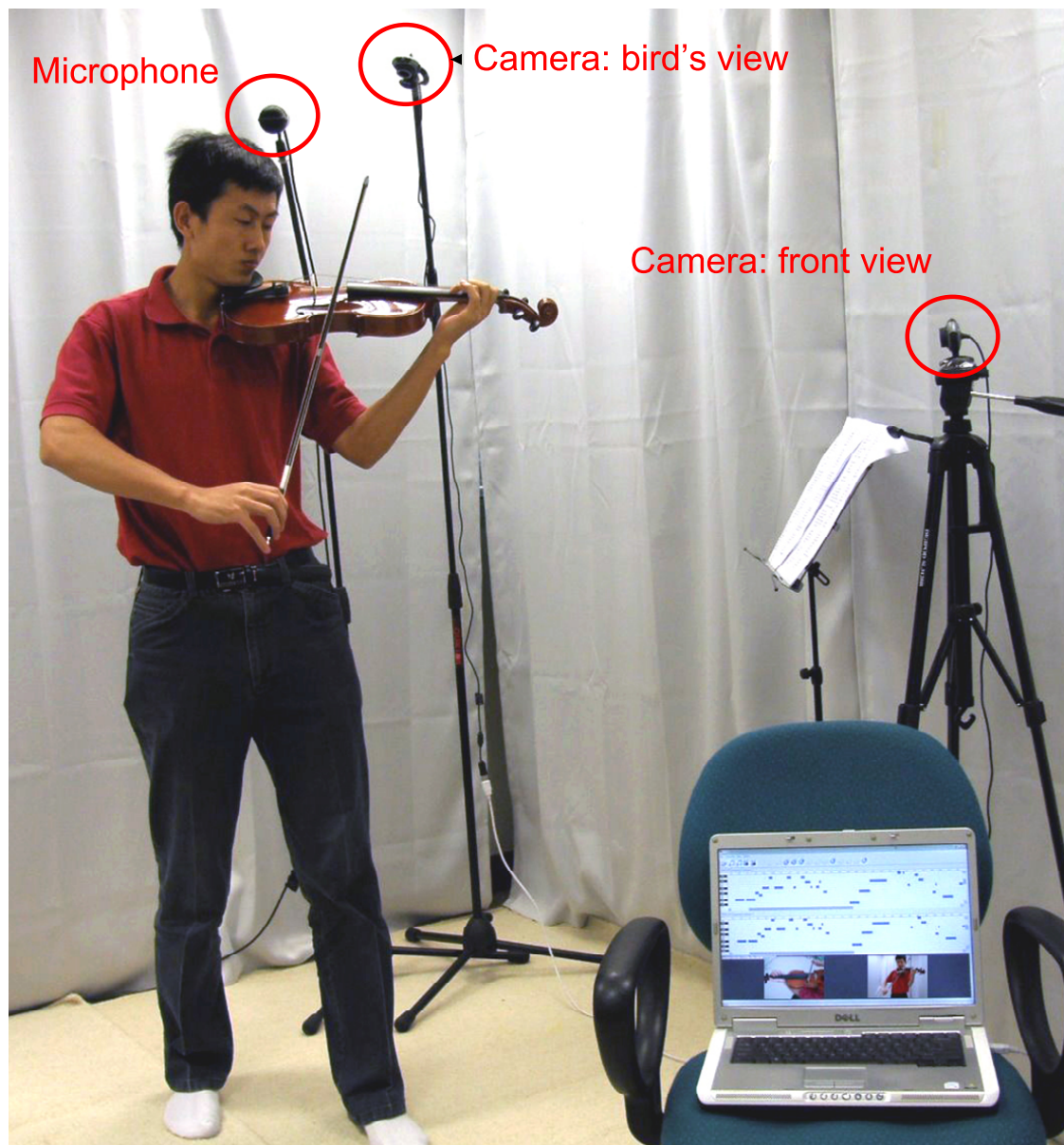


Figure 4.1: Hardware setting of the system.

After the whole recording has completed, the audio and video processing units of the system extract indicative features of onsets (detection functions) from the above three inputs respectively. Subsequently, features derived from audio and videos processing are fused together to obtain a more accurate onset detection result than state-of-the-art audio-only processing. After the onset detection, pitch estimation is conducted at last to produce the MIDI (piano-roll) notation of the played violin music. Through the comparison of the transcribed results and the reference notation (which is prepared beforehand in MIDI), the system manifests every note the violin learner played and indicates which notes are played correctly/wrongly.

## 4.3 Technical Details

iDVT follows the framework described in Section 3.2 and incorporates its two major components, the performance evaluator and the interactive feedback generator, in the design and implementation of the system.

In the remainder of this chapter, the technical details of the system will be described mainly concerning the back-end performance evaluator. In the next chapter, the user interface of the system will be introduced, which mainly embodies the essence of interactive feedback generator. iDVT fully implements the performance evaluator as the technical core, which consists of a recorder, a transcriber and an evaluator (Figure 4.2).

### 4.3.1 Recorder

One audio recorder and one video recorder are implemented for aural and visual recording of the user’s performance respectively.

The audio recorder is implemented using windows SDK, especially the winmm library. By default, the audio is captured in mono, 16bps, 44kHz PCM.

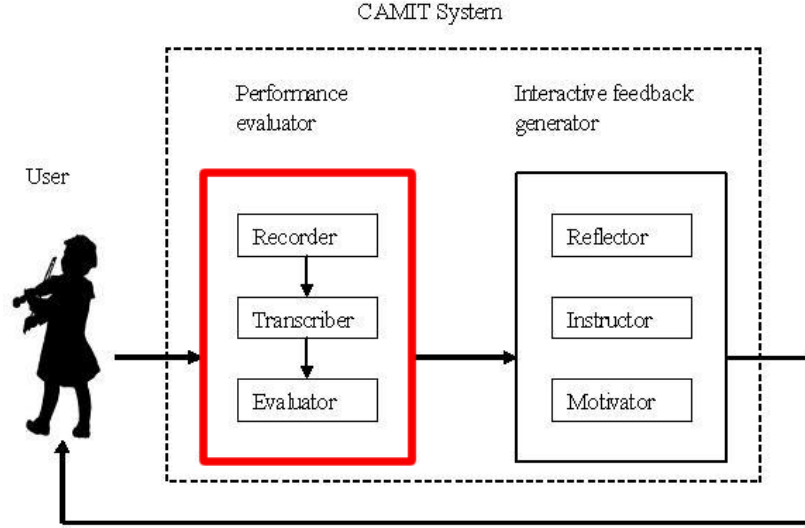


Figure 4.2: iDVT fully implements the performance evaluator as the technical core.

The video recorder is implemented using OpenCV library, especially the cv-cam library. By default, the video is captured with frame rate 30 fps and compressed using DIVX codec.

All the captured data are saved on the hard disk for further analysis.

### 4.3.2 Transcriber

Violin transcription is the main issue iDVT tackles in implementing the performance evaluator. iDVT basically re-implements the state-of-art violin transcription algorithm described in [WZS07]. The work flow of the transcriber is illustrated in Figure 4.3.

In the analysis and understanding of music, the *note* is a basic event. Finding the pitch of notes of pitched non-percussive (PNP) sound such as that from a violin is relatively easy, but identifying the precise beginning and end of specific notes and correlating them with the pitch (note segmentation) automatically is a challenging



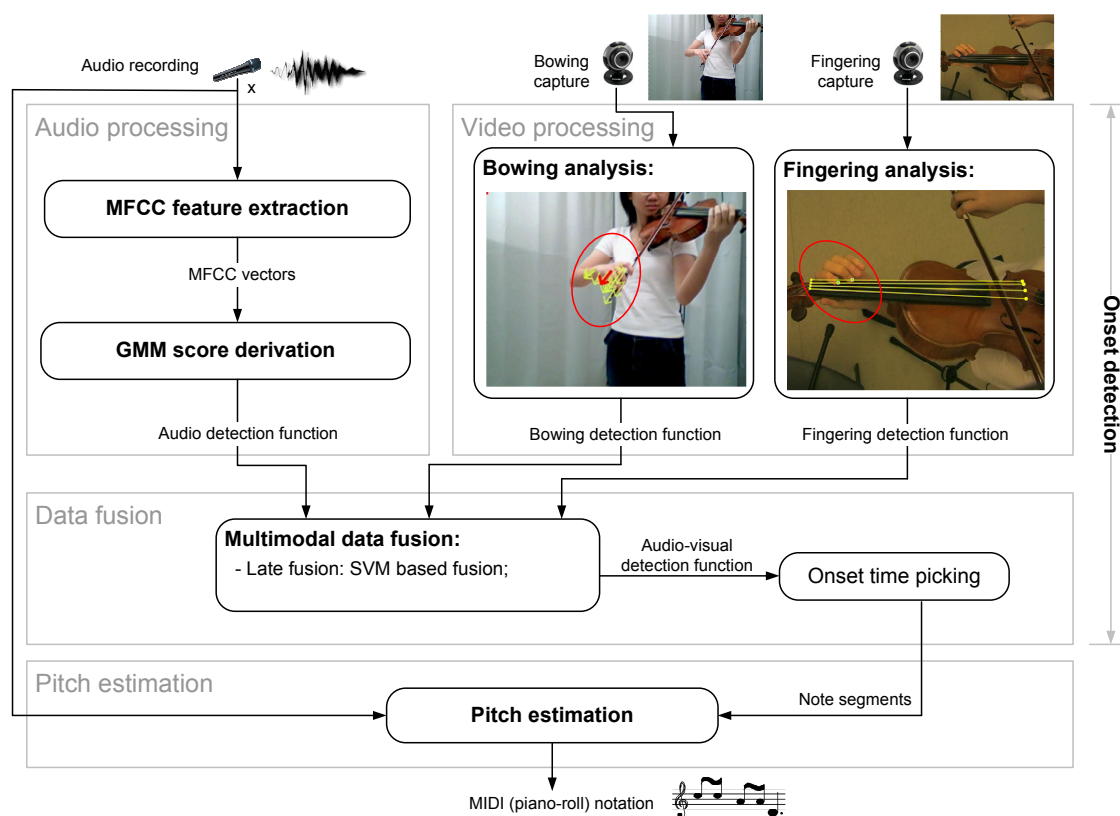


Figure 4.3: Work flow of the transcriber.

and critical task for CAMIT at home [YWH05].

Inspired by [BDA<sup>+</sup>05], which points out a promising combination of cues from different audio detection functions for onset detection, [WZS07] enhance it by fusing detection functions from both audio and video. According to [WZS07]’s experiment, this method is very promising in application oriented violin transcription.

The transcriber consists of three components: audio processing, video processing and audio visual fusion. They will be introduced separately as follows.

#### **4.3.2.1 Audio Processing**

In the audio processing part of the system, a supervised learning approach for onset detection is implemented using Gaussian Mixture Models (GMM) to classify onset and non-onset frames based on Mel-Frequency Cepstral Coefficients (MFCCs) [Log00] of the input audio. One audio-only onset detection function is derived in this phase.

#### **4.3.2.2 Video Processing**

The video processing is motivated by the observations that:

- The bow stroke reversal(right hand) and vertical movements are associated with note onsets;
- The trajectories of fingers(left hand) are associated with note onsets.

These visual cues offer important assistance for note segmentation task.

In the video capturing the front view of the learner, the right hand conducting bowing is tracked in each frame using Kalman filter framework with measurements obtained by optical flow and a skin color Gaussian model. Through the hand tracking, the bowing direction at any given time is obtained. Moments when the

bowing reverses directions are considered as onset times. The bowing detection function can be derived in this phase.

In the video capturing the bird’s eye view of the learner, the fingers of left hand are detected using a two step algorithm. Four violin strings are detected first, after which finger positions are searched along each string using the pre-calculated skin-color Gaussian model. Moments when a sudden change of finger positions occurs are considered as onset times. The fingering onset detection function can be derived in this phase.

#### **4.3.2.3 Audio-Visual Fusion**

In the audio-visual fusion part of the system, the detection functions obtained from audio and video processing are combined to produce an audio-visual detection function more indicative of onsets.

Since the audio and video are recorded simultaneously and time stamped in software level, they are assumed to be synchronized. The three detection functions derived in audio and video processing are interpolated respectively conforming to the same sampling rate and normalized into  $[0,1]$ . Subsequently, onsets are obtained after the detection functions are fed into Support Vector Machines (SVM) [Bur98] for decision level fusion.

After onset detection, the violin audio is segmented into individual note segments and the audio-only pitch estimation is carried out. The pitch estimator evaluated in [WZS07] is employed in our system.

#### **4.3.3 Evaluator**

The evaluator of iDVT is relatively simple. After the transcriber finishes its task, the player’s performance is represented by MIDI in the form of a sequence of pitches associated with onsets. The evaluator compare the transcription with the reference

MIDI obtained beforehand and points out the difference between the transcription and reference literally.

iDVT adopts a coloring scheme that combines the evaluation with the user feedback. The detailed evaluation algorithm will be presented in Section 5.4.3.

## 4.4 My Contribution

The algorithm of the core transcriber was implemented by Zhang Bingjun as presented in [WZS07]. My contribution regarding the transcriber is the migration and integration of his C,C++ and Matlab code into the iDVT system, which includes the incorporation of audio processing and pitch estimation, the incorporation and refinement of video processing and the re-implementation of data fusion. The recorder and the evaluator were also implemented by me.

# Chapter 5

## USER INTERFACE DESIGN

In order to make the system really useful in the everyday practicing of beginning learners, user interface design plays a fairly crucial role. Following the framework described in Section 3.2, the user interface incorporates the interactive feedback evaluator in its design, mainly including the reflector and the instructor (Figure 5.1).

For usability issues, the user interface is organized according to the functionality in the real using scenario rather than literally follows the structure of interactive feedback evaluator. However, major essence of the interactive feedback evaluator has been embodied in the user interface. This chapter will introduce how we design the user interface and why we do that.

### 5.1 Overview

The user interface of the iDVT is shown in Figure 5.2, which mainly consists of three panels. From top to bottom, the three panels are named the reference panel, the performance analysis panel and the video-analysis panel. The first two panels display the reference piece and the transcription result of the user's playing re-

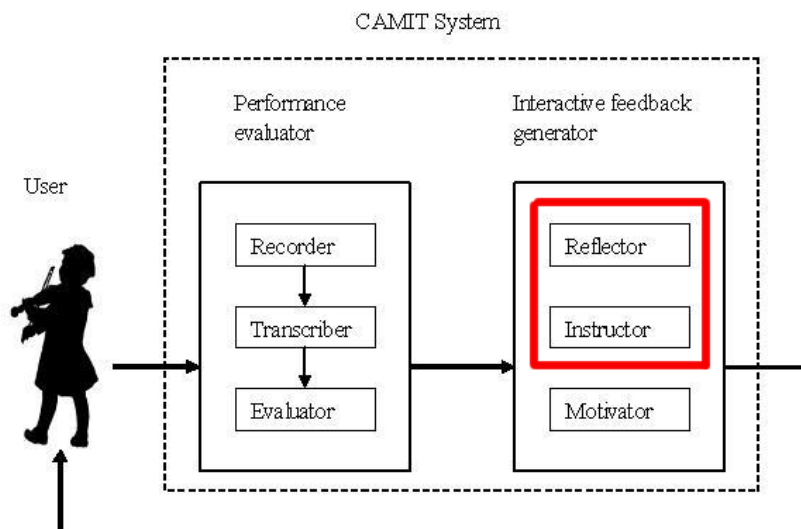


Figure 5.1: iDVT incorporates the interactive feedback generator in the user interface.

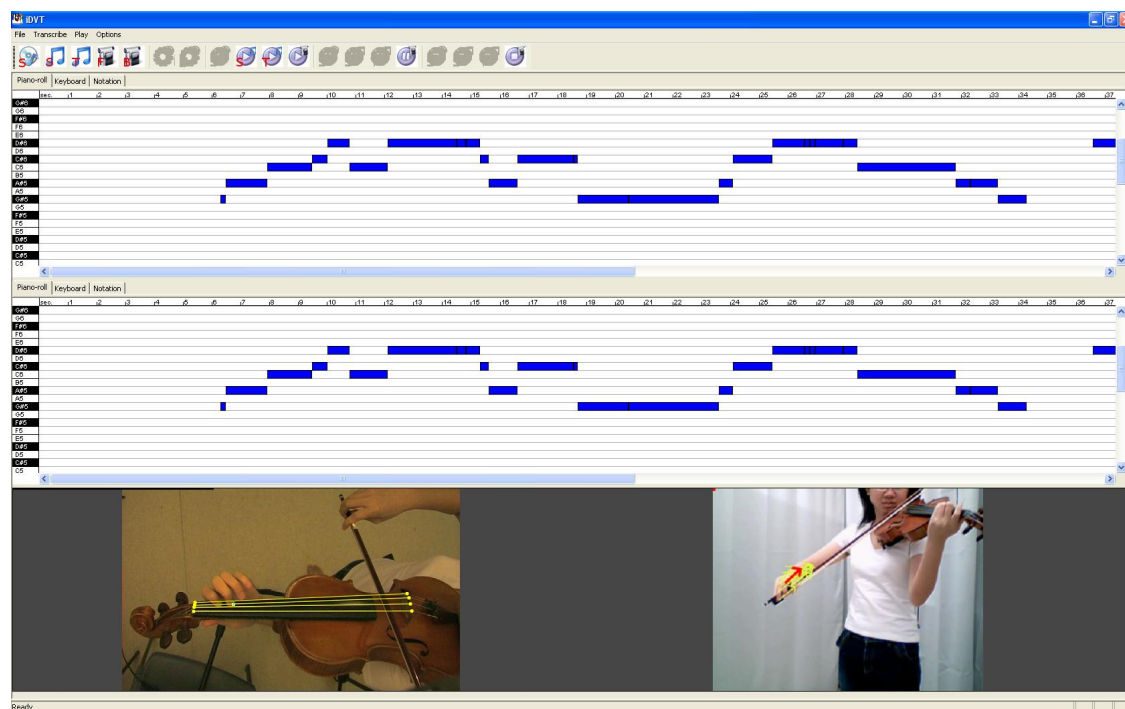


Figure 5.2: User interface of iDVT.

spectively. They are intended for showing how correctly the user played through comparison between the two. The third panel reflects the user's gesture of playing from two angles and at the same time displays the video processing results. Audio-only processing and audio-visual processing are both supported in the system for performance discretion of the two. All the audio/video raw data and processing results can be evaluated through playback supported by the system.

## 5.2 Functionality

### 5.2.1 Reference Panel

The purpose of the reference panel is to display the reference music pieces played by teachers or violin masters. It plays two roles in real application scenario:

Firstly, it serves as an improved substitution for paper-based sheet music. Before the learner begins practicing, he/she can choose the corresponding music file of the piece to play. The five-line staff of the music will be displayed in the panel in the same way as traditional paper-based sheet music. Moreover, we have two additional improvements which paper-based sheet music fails to accomplish. Once started, the system will highlight the correct note to be played according to the tempo of the music piece. In this way, the beginning learner will not only have a clearer view of which note to play next, but also gradually build the correct sense of tempo by following the *flowing notes*. Besides, it will automatically scroll the page if the playing comes to the page's end. Although relatively minor, this improvement avoids the annoyance of flipping the pages and let learners focus more on the playing.

Secondly, it serves as a clear reference when evaluating the performance of the learner. When the learner finishes practicing and wants to check his/her performance, he/she can switch the display from five-line staff to piano roll by

clicking the tab at the top of the panel. The piano roll offers a more natural and intuitive pitch-time layout to evaluate the performance than five-line staff (This will be discussed in detail when describing the layout of the piano roll in Section 5.3.2).

Last but not least, it integrates the functionality of a audio player, which is pretty handy and useful for users to learn through listening.

### 5.2.2 Performance Analysis Panel

The purpose of the performance analysis panel is to display the actual playing of the amateur learner, compare it with the reference and indicate the wrong parts played.

Performance analysis panel is much similar to the reference panel in terms of the audio playback and piano roll display functionality. However, it also has some distinguished features.

The most distinct one is that it incorporates a comparison display mode, through which the difference between the learner’s playing and the reference piece can clearly visualize by combining them in one panel. Once the learner’s audio is transcribed into MIDI using the method described in [WZS07], the system will automatically compare it with the reference and indicate the correct/wrong/missing parts using different colors (This will be elaborated in Section 5.4.3). A convenient option is also provided to switch between showing and hiding the comparison, so that the user will have better control and clearer view of the visualization by changing the comparison mode back and forth.

### 5.2.3 Video Analysis Panel

The video analysis panel is the *mirror* which reflects the motion of violin players for the purpose of demonstration and self-verification.



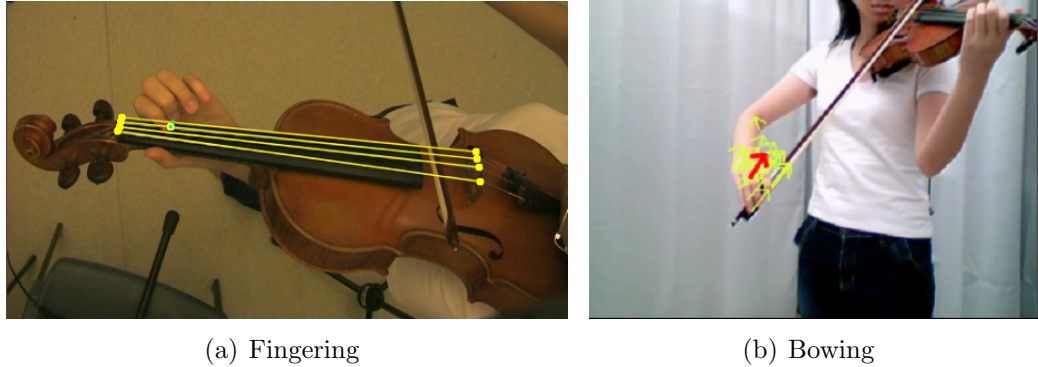


Figure 5.3: Fingering and Bowing.

This follows a common practice in music education that many violin tutors bring along a mirror in the classroom. Once set beside the tutor, the mirror offers students a better view of the tutor's playing gestures from different angles. Once set beside the playing student, the mirror provides the opportunity for the students to investigate their own gestures.

However, mirrors have their inherent shortcomings in fulfilling the demonstration and self-verification tasks, especially for amateur learners. On one hand, when the learner is practicing alone (which is the common case for most people and for most of the time), the demonstration ability of the mirror is invalid due to the absence of tutors. On the other hand, the beginning learner is already in a flurry, having little attention to spare on the mirror: they need to look at the sheet music, memorize the rhythm, pay due attention to both hands and grope for the proper fingering position. In this situation, adding one more thing to take care of is no doubt additional burden for them.

With the incorporation of the video player functionality, the video analysis panel is capable of demonstration if the reference video is available. Moreover, with the recording functionality, the video analysis panel can record the playing gesture of the user. Synchronizing the video and the audio, the panel can reproduce the whole performance. It enables the learner to check the gestures in a relaxed manner

after the playing is done. It also offers the possibility for the tutors to monitor the performance of the learner's practicing and better diagnose the problems of the learner.

In addition, if the user starts the video processing and wait a few minutes for it to finish, the strings of the instrument, the fingering position and the bowing motion of the playing can be highlighted as in Figure 5.3, which gives the user a clearer view of the performance in the whole self-verification process.

#### **5.2.4 Embodiment of Interactive Feedback Generator**

The interactive feedback generator is embodied in these three panels. The reference panel fulfills the role as an instructor in supporting audio playback, which enables the user to play the reference audio for demonstrations. The performance analysis panel embodies the reflector, which manifests the user's performance compared to the reference in the form of piano roll with contrasted colors. The video-analysis panel incorporates both the reflector and the instructor through the support of video playback. If the reference videos are loaded, the video-analysis panel becomes an instructor giving demonstration visually. If the users' videos are loaded and processed, the video-analysis panel becomes a reflector presenting the users' own performance with highlights on their fingers and hands.

### **5.3 Layout**

In this system, two layouts are considered for the music representation regarding different application purposes.

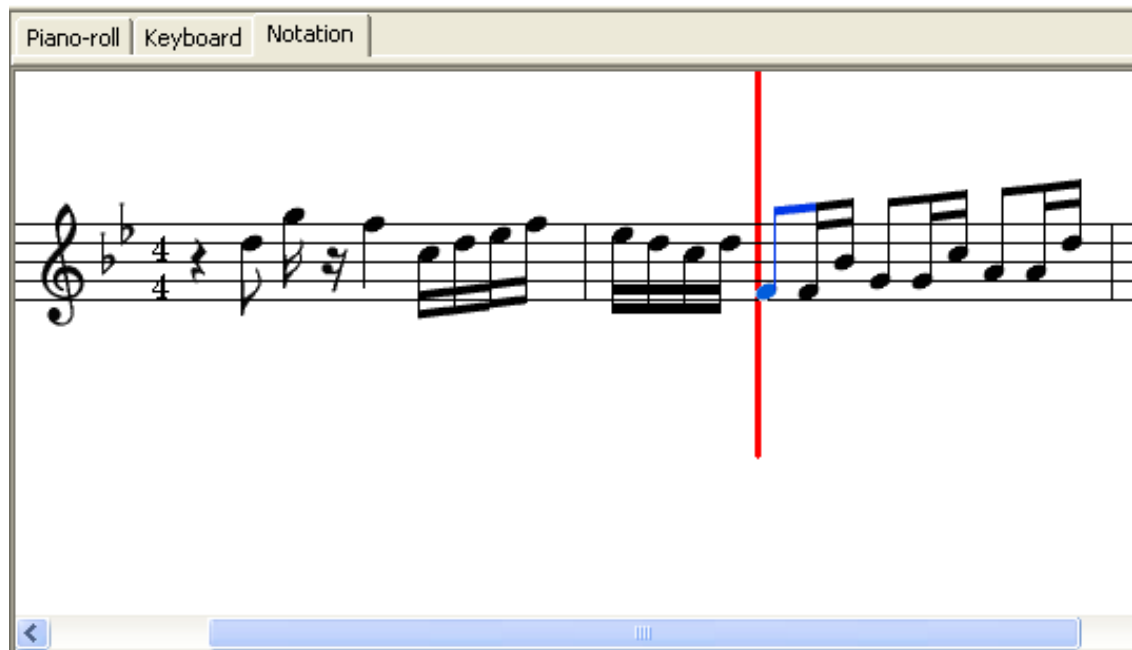


Figure 5.4: Five-line Staff.

### 5.3.1 Five-line Staff

The first and foremost one is the five-line staff layout, which is the most natural and commonly-used music notation in music education. It is a good option for reference displaying since no learning overhead is introduced for the user to receive traditional tutorial and practice with the help of our system at the same time.

As can be seen in Figure 5.4, the five-line staff is rolled out horizontally with a progress bar following the flow of the music. Auto-scrolling is functional when the end of the display area is reached.

Five-line staff layout is applied in the reference panel but abandoned in the performance analysis panel. The reason is that five-line staff is a music representation meant for *perfect* music, well-structured and rigorously conforming to music theories and rules (the property of a reference). However, No one can play the music exactly the same as the notation (consider how hard it is to play a note with duration 0.25 seconds, no more and no less), let alone amateur players whose playing is

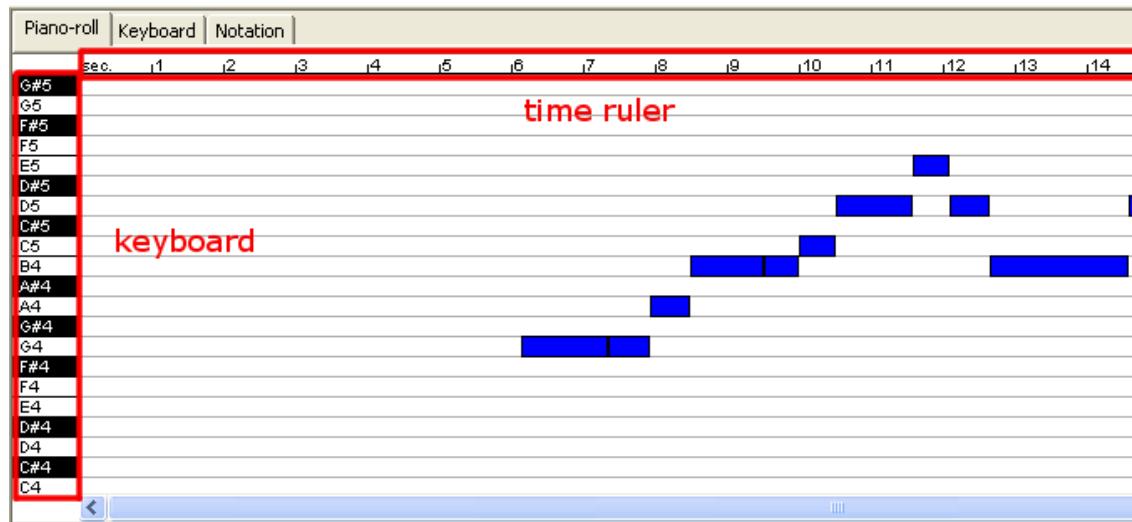


Figure 5.5: Piano Roll.

highly error-prone in nature. The playing may be wildly erratic in both time and pitch, which makes the transcribed five-line staff too messy to be readable, not to say to visualize the comparison and evaluation.

### 5.3.2 Piano Roll

The second one is the piano roll layout (as shown in Figure 5.5), which is an essential element in computer-based music visualization. A time ruler extends across the top of the layout showing the time line of the playing. A piano keyboard goes down the left hand side with corresponding notes displayed on the keys. Horizontal gray lines are drawn to separate neighboring pitches. Each note is represented by a blue rectangle with its vertical position in the canvas indicating the pitch and its width indicating the time duration of the note. One progress bar will show the current timing and pitch during the play. The piano roll layout is implemented in both reference panel and performance analysis panel.

### Reference Panel

Although experienced players might feel uncomfortable with the piano roll music notation, beginners may find it useful in showing a reference piece. Especially for a layman in music who easily loses tempo (very common in undertrained amateurs), the piano roll indicates clearer durations of notes compared to five-line notation. Five-line notation, even with the help of a progress bar, requires the player to interpret the music symbol into temporal context. It depends on both the player's reading ability and sense of tempo to play correctly. But with poor self-verification, the beginner may easily go astray and keep practicing the wrong thing. On the contrary, by following the progress bar in the piano roll which hits the left and right edge respectively for the beginning and ending of the note, the beginner can use visual clues to help verify his/her playing and gradually cultivate the correct sense of tempo.

### **Performance Analysis Panel**

In order to compare the playing of the user with the standard reference to see how he/she performs, the piano roll in the performance analysis panel can highlight the comparison result using different colors.

As can be seen in Figure 5.6, starting from the conventional all-blue visualization, where the system finds a note played in wrong pitch, it will print a gray rectangle to substitute for the original blue-colored one. The system will print out the corresponding reference note using a red rectangle and add dotted gray line to indicate their correspondence as well. In special cases, if the user plays a note where there should be silence, only the gray rectangle will be present with no red correspondence. Likewise, if the user misses one note somewhere, only the red rectangle will be present with no gray correspondence.

Following this simple scheme, it is clear and intuitive to visualize all kinds of possible errors on the piano roll using just three colors.

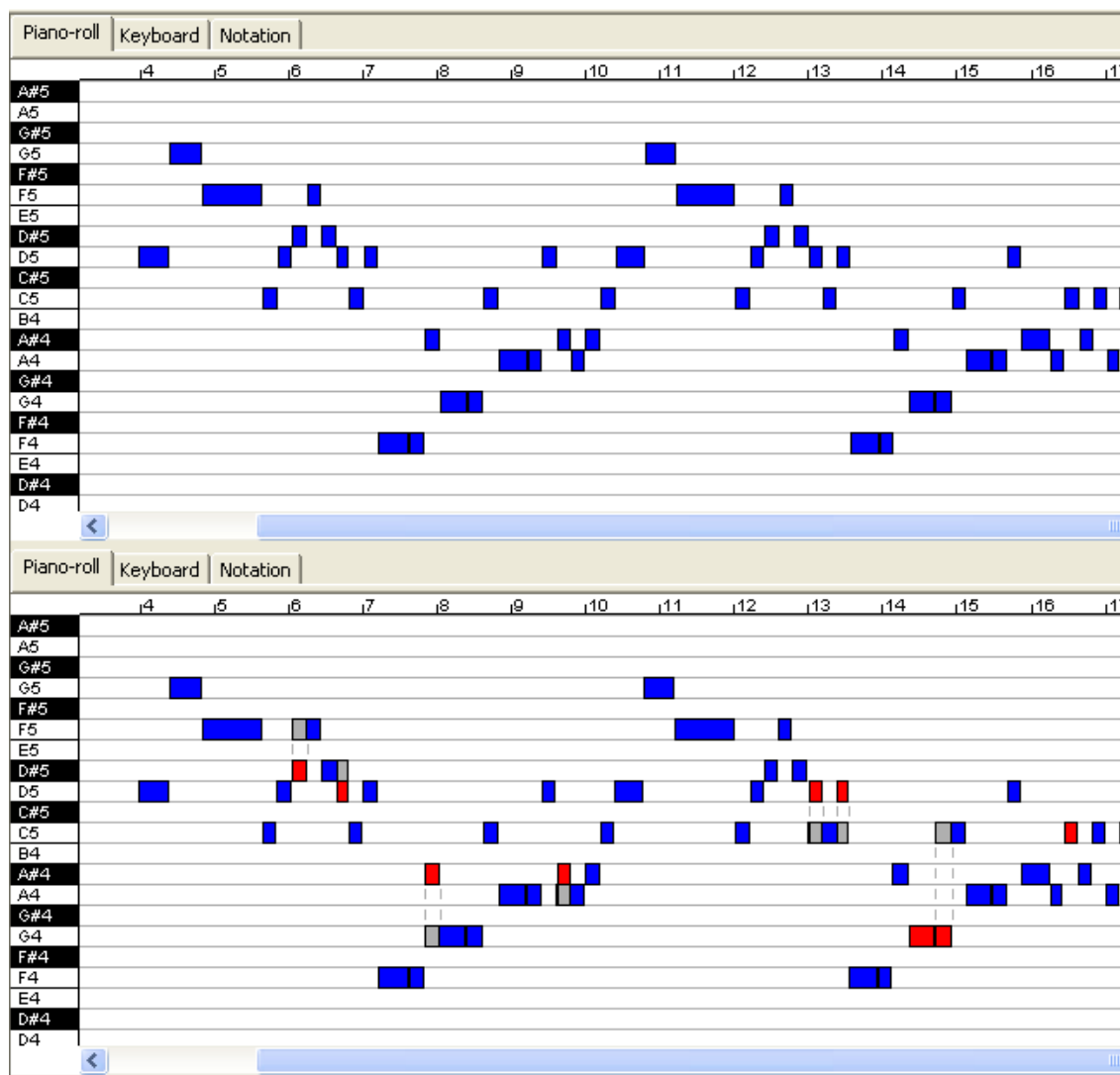


Figure 5.6: Upper panel:Reference Piece; Lower panel:Piano Roll Comparison.(Blue for correctly played notes. Gray for wrongly played notes. Red for corresponding reference for wrongly played notes.)

## 5.4 Implementation

In this section, some details of the user interface implementation will be provided.

### 5.4.1 Overview

The majority of the user interface components in our system are implemented using Microsoft Foundation Classes (MFC), including the framework layout, the menus, the piano roll display and the video playback. The five-line staff is the only exception.

### 5.4.2 Five-line Staff

In order to render a decent five-line staff given the MIDI of one piece of music, we refer to the source code of Rosegarden(version 1.7.2) for the implementation of this part instead of writing everything from the scratch.

Rosegarden is a well-rounded audio and MIDI sequencer, score editor, and general-purpose music composition and editing environment. It is open-source and is implemented under Linux using Qt. Since Rosegarden is a gigantic project with many functions beyond the need of our system, only the module related to the rendering of five-line notation(basically the ones under `src\gui\editors\notation` of Rosegarden's source code folder) is picked out and incorporated into our system.

Since the graphical user interface of Rosegarden using Qt, part of the code related to the rendering of five-line notation was rewritten to fit into the MFC framework while the inner logical structure of the module is maintained.

### 5.4.3 Performance Analysis

The inner representations of both the reference piece and the student's piece are in MIDI format, which record the time stamp (start time, end time) and pitch of

each note. Therefore, the performance analysis is actually a comparison with two MIDI files and the consequent visualization of the difference. A simple algorithm can be adopted to fulfill the task:

---

**Algorithm 1** MIDI comparison and visualization algorithm.

---

```

1: Truncate the silence before the first notes of both pieces;
2: Draw a time line with length equal to the duration of the longer piece;
3: Mark the time line with all the time stamps ( $t_0 \dots t_n$ ) of both two MIDI files;
4: for all the time segments  $\overline{t_i t_{i+1}}$  ( $i \in N, i \in [0, n]$ ) on the time line do
5:   if  $\overline{t_i t_{i+1}} < 0.1$  second then
6:     continue;
7:   end if
8:   Compare the pitches of corresponding time periods in both MIDI files ( $p_r$  for reference,  $p_s$  for student's play );
9:   if  $p_r = p_s$  then
10:    Draw a blue rectangle with start time  $t_i$ , end time  $t_{i+1}$  and pitch  $p_s$  on the piano roll;
11:   else
12:    Draw a red rectangle with start time  $t_i$ , end time  $t_{i+1}$  and pitch  $p_r$ ;
13:    Draw a gray rectangle with start time  $t_i$ , end time  $t_{i+1}$  and pitch  $p_s$ ;
14:    Link the two rectangles with gray dotted line if they are far apart;
15:   end if
16: end for

```

---

For simplicity reasons, this algorithm overlooks score alignment issues [DR06]. It only makes sure that the beginning of the reference and the student's pieces are aligned (through Algorithm 1 Line 1).

Score alignment is meaningful in CAMIT, especially in the evaluation of real performance. Human beings cannot play exactly what the symbolic music (music notation, MIDI etc.) indicates. Missing several notes or the accumulation of small timing errors may lead to misplacement of the whole subsequent notes on the time line. If the comparison rigidly looks into one time segment after another, the ultimate evaluation will be far away from human's judgment, which has tolerance for such *blemishes* to some extent. Consider a simple example, when someone tries to play a sequence notes each lasting 1 second. If he/she misses the first note, but



play all the other notes with correct pitch and tempo, human judges will think the play has a relatively small error(missing one note). But since the whole sequence is misplaced, the naive comparison will think it is completely wrong. Score alignment is thus introduced to make proper alignments in time so that computers can evaluate more reasonably.

However, our simplification is feasible in two senses. Firstly, the beginners' etudes are usually short and simple. Thus, the accumulation of errors can be neglected if each etudes is within reasonable tolerance level (Line 5 of Algorithm 1 is actually neglecting such tolerable errors). Secondly, since the etude is short, missing notes or mistakes in tempo can no longer be regarded as trivial. Imposing stricter constraints during the practicing of such fundamentals is actually good for beginner's further study.

But to make the system robust and useful for advanced usage, score alignment techniques should be considered in the future work.

## 5.5 My Contribution

My contribution regarding this part of the system includes the user interface design, audio playback, video playback and five-line staff display. The piano-roll was originally developed by Zhang Bingjun and was further modified by me to display the evaluation results.

# Chapter 6

## ITERATIVE USABILITY EVALUATION

As soon as an initial prototype of iDVT was completed, we conducted a series of evaluations to test the usability of the system as well as iteratively improving its design. We attempt to address the following goals with these evaluations.

- Receive suggestions on additional features desired for iDVT system
- Test the usability of the interface

### 6.1 Participant

We invited several teachers and students to evaluate the system. The teachers invited were either from music instrument tutoring background or from computer science background. We expected them to give critical and insightful comments for the system improvement. The students were violin learners from music schools with several years of learning experience. We expected them to feedback on the system usability in real application scenario.

## 6.2 Evaluation Strategy

In the evaluation session, each participant was invited individually and gave feedback independently. The whole process of using the system for practicing one etude was demonstrated to the teachers or students. In order to know the usability of the system and look for possible problems in user's real practice, the students were further encouraged to try using the system themselves. The feedback from both the teachers and the students were collected afterwards.

## 6.3 Evaluation Sessions

### 6.3.1 Teachers' Session

After the very initial version of iDVT was completed, several teachers were invited for the evaluation of the system, who offered invaluable suggestions to improve the system.

One enhancement they proposed was displaying the reference with five-line staff instead of the original piano roll. This would make the reference more natural, which is identical to the one commonly used in music education and real violin playing. Another enhancement suggested was highlighting the comparison result explicitly using contrasting schemes instead of simply displaying the reference and transcription. The third suggestion was accelerating the processing speed and boosting interactivity. The observation was that the original version did audio and video processing one after another, which left users idly waiting for several minutes.

In view of these suggestions, we worked out the second version of iDVT with these problems addressed. The five-line staff was implemented as discussed in 5.4.2 and the color scheme for comparing results was adopted as discussed in 5.4.3. The audio processing and video processing code were rewritten to improve

speed and interactivity using multi-threading, which not only reduced the overall processing time, but also enabled the users to use other functions while waiting for the processing result.

### 6.3.2 Students' Session

After the improvement was done after Teachers' Session, two students with three and five years of violin learning experience were invited for the second-round of evaluation to test the system in real practicing scenario.

While watching the demonstration, they thought that the functions provided were "useful" and "considerate" in real application. They especially liked the finger tracking and hand tracking display. As they said, "It is really awesome to see my own playing so closely and highlighted. I can pick out each and every mistake which I would not notice myself. No mistake can escape the camera!"

While trying out by themselves, they had little difficulty in completing the whole process. They thought of the user interface as "straightforward" and the operating process clear to go through.

However, they also revealed some problems in the system which remained to be improved.

Firstly, the comparison result display sometimes looks messy if too many mistakes are present, which tends to scare users off. This is mostly due to the limitation of the current system that it just literally indicates the errors, but cannot give corresponding instructions more intellectually. Knowing what is wrong is critical but not sufficient. Knowing how to correct the error will be a higher demand for users. Moreover, if teachers are present in this kind of situation, they would not stick to each and every mistakes made by the student, but put the most serious one or two mistakes in priority for the students to erect. Improving in a step-by-step manner will serve the learner better, especially those inexperienced.

Secondly, the initial hardware configuration, especially the setting of cameras, is somewhat difficult for the user to get good finger tracking and hand tracking results. Since the quality of the tracking result is related to the background color, the shooting angle of the camera and the distance from the camera to the object etc., it is not always so that a common user will easily get an optimized setting to ensure a good result. Although the tracking algorithms are robust to some extent, it still appears tricky for those totally unfamiliar with the setting without clear guidance.

Thirdly, the work flow of the system could be further simplified. Currently, the recording and the processing modules are not streamlined yet. The user needs to explicitly save the recordings on the hard disk and then load them for further audio and video processing. This design gives good archives of each practicing session, which keeps track of the development of skills and performance accessible to users as well as their tutors. However, in real practice, it includes additional operations (save and load operations), slows down the processing time (compared to real time processing) and hard disk consumption (processed and unprocessed recordings all need to be saved).

## 6.4 Summary of Evaluation

After two rounds of evaluation, the evaluation goals we set earlier were mostly fulfilled. We received very positive feedback of the system from professionals and end users, which acknowledged both the system's feasibility and usability. We also got invaluable suggestions for the improvement of the system, which will be considered in later improvement of the system.

However, it should be pointed out that the sessions conducted above are only the initial steps taken for the evaluation of the system. On the one hand, the participants were basically experienced players and teachers, which were not the

exact targeted user of the system. Beginners and preferably children will be the focus in the future sessions. On the other hand, due to the constraints of resources, the evaluation conducted is limited to relatively small scale and short duration. In the future, we will invite more beginners to participate and allow more time and freedom for them to try out the system. We will include questionnaires for better quantitative analysis of their feedback as well.

## Chapter 7

# CONCLUSION

This thesis proposes a general framework for designing Computer-Assisted Musical Instrument Tutoring systems focusing on unsupervised musical instrument practice. It puts into consideration both the beginners' needs in unsupervised practice and computer system development. The framework consists of the back-end performance evaluator and the front-end user interactive feedback generator, which are further broken into six modules with their functions and significance discussed respectively.

The thesis also presents interactive Digital Violin Tutor (iDVT), a practical Computer-Assisted Musical Instrument Tutoring system following the framework proposed, which aims at assisting amateur violin players in unsupervised practice. iDVT provides accurate music transcription leveraging the fusion of audio and video processing and informative and intuitive feedback with considerate user interface design. The algorithms and designs are discussed in detail.

The iterative usability evaluation was carried out to assess the system and help improving it. The system received very positive feedback of the system from professionals and end users, which acknowledged both the system's feasibility and usability. Suggestions were also raised for future improvement of the system.

## 7.1 Future Work

The thesis has identified a number of important components in a CAMIT system focusing on unsupervised practice through the framework proposed. They are mostly embodied in the structure of the iDVT system. However, some parts of the system remains blank or preliminary which remain to be improved. Combining the feedback from the evaluators, the improvement can be carried out mainly in three directions corresponding to the two major components proposed in the general framework.

### 7.1.1 Performance Evaluator

One possible improvement direction relates to the performance evaluator, which mainly involves the evaluator module.

Currently, the evaluator adopts a naive comparison algorithm and a colored scheme for representation. It is eligible to be applied in simple using cases, but needs to be improved in complex ones.

In the short run, the comparison algorithm can be refined to indicate the errors in the user's performance more accurately and robustly. Score alignment, for example, will be considered for such purposes to tackle complicated situations.

In the long run, evaluations in more sophisticated forms will be incorporated in the system. Besides the current comparison-based evaluation, which merely indicates the discrepancy between the performance and the reference, more objective and subjective evaluation measures can be adopted regarding the correctness and expressiveness of the performance. Furthermore, the evaluation results can be presented in various quantitative and qualitative ways such as scores and comments.



## **7.1.2 Interactive Feedback Generator**

Another possible improvement direction relates to the interactive feedback generator, which mainly involves the instructor and the motivator module.

### **7.1.2.1 Instructor**

Currently, the instructor mainly uses audio-visual demonstrations with highlights to fulfill its function.

In the short term, the instructor can provide more descriptive instructions and hints during the demonstration. This improvement is less technical since it can be easily included by music professionals when the demonstration is recorded. However, it serves better educational purposes.

In the long term, the instructor can be more interactive and active during the practice. Instead of preparing fixed instructions beforehand, the instructor can explore online instruction, which gives instructions according to the user's instantaneous performance and in real time or near real time. This function will make the system more intellectual and make more sense in real application scenario.

### **7.1.2.2 Motivator**

Last but not least, the motivator, which is totally untouched in the iDVT system, can be included in the future to make the system more fun and attractive. Common motivation schemes such as performance scoring and RPG(Role Play Game) storyline can all be adapted to the application scenario to stimulate the attention, passion and motivation of the users in using the system as well as musical instrument practice.

## 7.2 Further Usability Evaluation

Besides the improvement of the system summarized above, further usability evaluation will also be conducted in the future.

We will seek cooperation with music institutions or schools in carrying out the further usability evaluation. Regarding the deficiency of the previous sessions of evaluation mentioned in Section 6.4, three aspects will be emphasized in the future. Firstly, the evaluation will be mainly targeting on beginning violin players. Secondly, more participants will be involved and each of them are allowed to use the system in the real application scenario, i.e., during every day practice and in home environment. Last but not least, we will carefully design the questionnaires for better quantitative analysis of users' feedbacks and preferences.

# Bibliography

- [BDA<sup>+</sup>05] JP Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and MB Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5 Part 2):1035–1047, 2005.
- [Bur98] C.J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [BWL06] W. Boo, Y. Wang, and A. Loscos. A violin music transcriber for personalized learning. In *IEEE Inter. Conf. on Multimedia Expo*, 2006.
- [DR06] R.B. Dannenberg and C. Raphael. Music score alignment and computer accompaniment. 2006.
- [DSJ<sup>+</sup>90] R.B. Dannenberg, M. Sanchez, A. Joseph, P. Capell, R. Joseph, and R. Saul. A computer-based multi-media tutor for beginning piano students. *Journal of New Music Research*, 19(2):155–173, 1990.
- [DSJ<sup>+</sup>93] R.B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, and P. Capell. Results from the piano tutor project. In *Proceedings of the Fourth Biennial Arts and Technology Symposium*, pages 143–150, 1993.
- [Fer06] S. Ferguson. Learning musical instrument skills through interactive sonification. In *Proceedings of the 2006 conference on New interfaces*

*for musical expression*, pages 384–389. IRCAMCentre Pompidou Paris, France, France, 2006.

- [FLO<sup>+</sup>04] D. Fober, S. Letz, Y. Orlarey, A. Askenfeld, K. Hansen, and E. Schoonderwaldt. IMUTUS—an interactive music tuition system. In *Proceedings of the Sound and Music Computing conference (SMC)*, pages 97–103, 2004.
- [FLO07] D. Fober, S. Letz, and Y. Orlarey. VEMUS-Feedback and Groupware Technologies for Music Instrument Learning. In *Proceedings of the 4th Sound and Music Computing Conference SMC’07-Lefkada, Greece*, pages 117–123, 2007.
- [FMC05] S. Ferguson, AV Moere, and D. Cabrera. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 97–102, 2005.
- [HSD06] D. Hoppe, M. Sadakata, and P. Desain. Development of real-time visual feedback assistance in singing training: a review. *Journal of computer assisted learning*, 22(4):308–316, 2006.
- [Log00] B. Logan. Mel frequency cepstral coefficients for music modeling. In *International Symposium on Music Information Retrieval*, volume 28, 2000.
- [LWB06] A. Loscos, Y. Wang, and W.J.J. Boo. Low level descriptors for automatic violin transcription. *Proc. of ISMIR2006*, 2006.
- [Per95] D. Perkins. *Smart schools: Better thinking and learning for every child*. Free Press, 1995.

- [Per08] G.K. Percival. Computer-assisted musical instrument tutoring with targeted exercises. 2008.
- [PWT07] G. Percival, Y. Wang, and G. Tzanetakis. Effective use of multimedia for computer-assisted musical instrument tutoring. In *Proceedings of the international workshop on Educational multimedia and multimedia education*, pages 67–76. ACM New York, NY, USA, 2007.
- [SAH05] E. Schoonderwaldt, A. Askenfeld, and K. Hansen. Design and implementation of automatic evaluation of recorder performance in IMUTUS. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 97–103, 2005.
- [SHA04] E. Schoonderwaldt, K. Hansen, and A. Askenfeld. IMUTUS—an interactive system for learning to play a musical instrument. In *Proceedings of the International Conference of Interactive Computer Aided Learning (ICL)*, 2004.
- [SWK95] S.W. Smoliar, J.A. Waterworth, and P.R. Kellock. pianoFORTE: a system for piano education beyond notation literacy. In *Proceedings of the third ACM international conference on Multimedia*, pages 457–465. ACM New York, NY, USA, 1995.
- [web07a] i-maestro. <http://www.i-maestro.org>, 2007.
- [web07b] Vemus: Virtual european music school. <http://www.vemus.org>, 2007.
- [web08] Meaws. <http://percival-music.ca/software/meaws/index.html>, 2008.
- [web09] Guitar heroes. [hub.guitarhero.com](http://hub.guitarhero.com), 2009.
- [WZS07] Y. Wang, B. Zhang, and O. Schleusing. Educational violin transcription by fusing multimedia streams. *Proceedings of the international*

*workshop on Educational multimedia and multimedia education*, pages 57–66, 2007.

- [YDHW04] J. Yin, A. Dhanik, D. Hsu, and Y. Wang. The creation of a music-driven digital violinist. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 476–479. ACM New York, NY, USA, 2004.
- [YWH05] J. Yin, Y. Wang, and D. Hsu. Digital violin tutor: an integrated system for beginning violin learners. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 976–985. ACM New York, NY, USA, 2005.