Effective Use of Multimedia for Computer-Assisted Musical Instrument Tutoring

Graham Percival
Departments of Computer Science and Music
University of Victoria
British Columbia, Canada
gperciva@uvic.ca

Ye Wang
School of Computing (SoC)
National University of Singapore
Singapore 117543
wangye@comp.nus.edu.sg

George Tzanetakis
Departments of Computer Science and Music
University of Victoria
British Columbia, Canada
gtzan@cs.uvic.ca

ABSTRACT

This paper presents a survey of recent work in computer-assisted musical instrumental tutoring and outlines several questions to consider when developing future projects. In particular, we suggest that the area in greatest need of computer assistance is enhancing daily practice: both motivating students to practice through games and multimedia, and providing an objective analysis of the students' performance. Many existing projects attempt to replace human teachers by providing lessons during daily practice; in most cases, these "daily lessons" are not necessary.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing; K.3.1 [**Computing Mileux**]: Computer and Education

General Terms

Design, Experimentation, Human Factors.

Keywords

Computer-assisted musical instrument tutoring, music edutainment, support system, music transcription, visualization, animation.

1. INTRODUCTION

A number of computer-assisted musical instrumental tutoring (CAMIT) projects have been attempted in the past fifteen years. This paper will discuss some previous projects and present guidelines for future work in this area.

Teaching a musical instrument is quite different from teaching an academic subject such as mathematics or history. In these subjects, there are clear answers to questions, which can be easily graded by a computer – for the grade-school level at least; these subjects become much more difficult at university. We ask a child to complete `4 + 2*8 = __` or "When was the Magna Carta signed?", and the computer can check whether the answer is correct. We cannot grade a musical instrument performance in the same way; there is no clear-cut "correct" and "incorrect" answers. By performing pitch estimation and note segmentation, we can judge that a particular note was unacceptably out of tune (*every* note is "out of tune" to some degree). However, this fact may not be obvious to a music student, even after informing the student that the note was incorrect. We must employ visualization techniques to ensure that the student understands *why* the note was not sufficiently "in tune".

In addition, learning a musical instrument involves a great deal of physical control. Since musicians are rarely covered in sweat while performing, non-musicians seldom consider musicians to be athletes, but in reality musicians must perform physical activities (finger movement, exhaling, adjusting lips) within very tight tolerances. A music student may know perfectly well that he made a mistake; the problem was purely physical. By analogy, consider an athlete training for a long jump: the athlete *knows* how to perform the jump – he knows the optimal distance to begin running, he can pace his footsteps such that his jump always begins within the allowable limit – but this does not guarantee that the athlete can jump 7 meters. In many cases, a music student has all the knowledge he needs; he simply needs to train.

We begin by presenting a survey of related work in Section 2. This is followed by identifying potential goals of CAMIT projects in Section 3 and discussing their relative merits in Section 4. Next we describe a typical session of individual music practice in Section 5, followed by a discussion in Section 6 about how computer programs might help or hinder this process. Then we present our own ongoing projects in Section 7. Finally, we discuss future research in Section 8 and give our conclusions in Section 9.

2. RELATED WORK

Work on CAMIT projects may be split into two categories: projects which have a specific goal, and projects which attempt to provide a complete learning environment by providing a "virtual teacher" for the students' private practice.

2.1 Specific-goal projects

There have been many CAMIT projects which have a narrow focus. These projects aim to provide one tool which music teachers may use to solve a particular problem.

An excellent example of this focused approach is the work of Robine et al [18]. Saxophone players (of any level of ability) were asked to play five notes – three notes with constant loudness, one note which gradually became louder and then quieter, and one note with vibrato. By analyzing the stability of pitch and amplitude of these five notes, they could accurately predict the overall ability of each student as determined by a professional saxophone teacher. This program may be used by saxophonists to practice their control of airflow. This work is a good example of using multimedia analysis to enhance technical exercises.

Many musicians consider chamber music to be the pinnacle of music: playing music with a few other musicians allows the greatest combination of flexibility and structure. Oshima et al [17] make chamber music much easier for beginning piano students with their Family Ensemble software. This project provides an easy way to play piano duets: non-musicians (such as parents or siblings of a student) may perform the secondary part by pressing keys on a piano keyboard. The correct notes are substituted for the secondary (non-musician) player, so the two parts will always be in unison. In addition, score following techniques are used to compensate for mistakes of the student. Although this work does not provide direct feedback to the student, the extra motivation and sheer 'fun' of music it gives students is extremely valuable.

Research is progressing in developing visualization techniques for music. Ferguson et all [8] investigate ways to present multiple streams of data in a single, easily-understandable display. In [7], Freguson investigates using realtime sonification to provide feedback to musicians. This may seem counterintuitive – the main focus of a student should be his own sound – but with sufficient care, these techniques may be applied in certain situations.

The PianoFOTRE system by Smoliar et al [22] attempts bridge the gap between simply playing the notes (which MIDI synthesizers can do beyond any human ability) and performing music (by adding the tiny variations in tempo and dynamics which makes music seem "alive"). The system provides visualizations for the dynamics, tempo, articulation, and synchronization (between hands) of a piano performance. This may be used to increase communication between teacher and student by providing easily-viewable representations of these expressive parameters.

2.2 General projects

There are a few large CAMIT projects which aim to provide general instruction: the student plays a complete piece of music, the computer analyzes the performance, and then provides feedback. These systems are primarily intended for self-learning or distance education.

The first major such project is Piano Tutor [5, 6]. This project used score-following software to analyze a student's performance, then used an expert system to judge which mistakes were in greatest need of help. The system would then use a combination of graphics, voice, and video to inform the student of these mistakes and how to correct them. The system may also choose simpler tasks for the student, so that they may concentrate on improving specific skills.

The IMUTUS project [20, 10, 19] is the spiritual successor to Piano Tutor. It is designed to be a complete, autonomous tutor with no human teachers (although the authors note that it will be more successful when used in conjunction with a teacher), but in this case the target instrument is the recorder. This system operates by providing feedback after each performance: the system prioritizes mistakes based on discussions with over 40 recorder teachers. For example, mistakes in articulation are less important for beginning recorder players than control of air flow. To avoid overwhelming the student, the system only informs the student of a few mistakes. Students may also request hints to view extra annotations made by teachers. This project has now ended, but similar work continues with the VEMUS [2, 9] project, now investigating other wind instruments.

i-Maestro [1, 16] also provides an interactive self-learning environment, but this project is also investigating the user of new gesture-based interfaces. An elaborate framework of server software, client software, music exercise authoring tools, P2P techniques, and 3d motion capture visualization software is planned, to allow students to learn more effectively. This project is still in progress; we look forward to reviewing their results.

3. IDENTIFYING GOALS

Before beginning work on any complex system, we should have a clear concept of the goal(s). In the area of music education, there are three areas: enhancing the teacher's lesson with the student(s), enhancing the student's individual practice, and motivating the student. Most projects will pursue two of these goals (motivation, and either enhancing lessons or enhancing individual practice), but in some cases it may be useful to pick a single goal. For example, when dealing with highly-motivated students (either competitive teenagers or adult beginners), the problem of motivation might be a non-issue. Conversely, for some students (intelligent, talented, but easily bored), motivation may be the *only* problem that needs addressing.

3.1 The teacher's lessons

The study of computer-assisted music education is relatively young, but it is safe to say that we will not be replacing human teachers in the foreseeable future. Teaching humans – especially children – requires a mixture of subject-specific knowledge, communications ability, psychological skills, and creativity.

Using technology to enhance music lessons is nothing new. Some teachers use mirrors so that they can easily monitor the student's movements from multiple angles (or use the same mirrors to demonstrate their own movements to the student). Many teachers use recording devices (cassette tapes, minidiscs, or computers) to record their students and then play the student's performance with the teacher's commentary.

We can easily apply these same examples to educational multimedia. Instead of setting up a single mirror so that a student in the same room can view the teacher's demonstration, we could set up multiple video cameras so that the demonstration can be viewed by many students in various geographic locations and potentially tens of years in the future. We could further improve our archiving of performances by using body sensors (often used by dancers to produce computer animation [21], and currently being investigated by CAMIT researchers). The resulting data can produce a computer animation which may be viewed at any position, angle, or speed.

3.2 Individual practice

The vast majority of a student's time with their instrument is spent on individual practice. Individual practice is less effective than lessons with a teacher, but due to economics and practicality, most students have one lesson per week. The effectiveness of individual practice is therefore absolutely vital to a student's progress. Effective individual practice is particularly difficult for young children; for this reason, several approaches to music education (notably the Suzuki method) stress parental involvement in lessons and supervision of home practice.

There are a number of existing technologies to improve the effectiveness of individual practice. Two early inventions were the tuning fork (a metal device which vibrates at a fixed, known frequency) and the metronome (a device which plays a sound at regular intervals; generally between 30–240 beats per minute). In the late 20^{th} century, electronic tuners replaced tuning forks – students could see an electronic device's estimation of their current pitch, displayed along with pitches of nearby notes.

With multimedia computer programs, we can significantly improve on these old technologies. Instead of comparing pitches (audible pitch vs. pitch of nearby real notes) at individual moments in time, we could compare the pitches in an entire piece. A student could perform a set exercise, and a computer could compare the student's pitch with the expected pitch. The computer could then highlight the three worst notes and inform the student, who would then perform the exercise again to fix those mistakes.

3.3 Motivation

Humans are immensely lazy creatures. Unfortunately, we are also extremely creative. We are extremely skilled at finding excuses to avoid anything that resembles work – practicing musical exercises, fixing mistakes in said exercises, or even taking our instruments out of their cases.

Some people may consider student motivation to be outside the specific area of computer-assisted music education; motivation is a general problem in education and "edutainment" computer programs. There are certainly many problems we can research without regard for motivating students – multimedia analysis, creating multimedia feedback for students, etc. However, the single most useful factor in any practical multimedia system for students is motivation. If students could be motivated to play their assigned musical exercises every day, that would far outweigh the benefit of the fanciest multimedia feedback systems.

There are many ways to motivate students; first we must identify our target audience. For young children, it might be appropriate to display brightly-colored stars. Older children may enjoy the notion of "gaining experience" and "going up levels" – possibly within the framework of a game where the user is attempting to save a princess or defeat an evil wizard. If the target is adult males, then perhaps the ability to compare their scores competitively would motivate them to practice their scales.

4. DISCUSSION OF GOALS

Although there are a few ways that multimedia tools may enrich the lessons of a teacher, we believe that research in this area is likely to be less effective than work in the other two areas. First, the time with a teacher is much less than time spent without a teacher. Many teachers suggest that their students spend an equal amount of time in daily practice as they do in lessons (ie a one-hour weekly lesson would result in one hour of daily practice). Depending on the seriousness of the student, the amount may be double or even quadruple this amount (ie four hours of daily practice; a one-hour lesson per week). Second, music teachers are already investigating this area. As new technology becomes available, music teachers incorporate it in their lessons.

The question of motivation is a current area of research in Education and in the design of computer games. Millions of children (and adults) spend hours each day playing computer games; many game-players even pay a monthly fee to play online games. If we could design a computer-assisted music education program that was half as addictive as the leading online game, this question would be solved.

As experts in multimedia analysis, it seems that our time is best served by investigating the problem of improving individual practice. Music students rarely use technological aids – they may use a tuning fork or electronic tuner at the beginning of a session to tune their instrument, and they may *occasionally* use a metronome, but those are infrequent. Many music students find it difficult to accurately judge their own performance. This is quite problematic because, as the famous phrase goes, "practice makes permanent". If a well-meaning child spends 90% of his time practicing mistakes, his teacher will have a hard time correcting those mistakes.

There are a few reasons for the difficulty of self-analysis. To some extent this is a physical difficulty (the location of the musician's ears compared to the audience's ears). However, the primary difficulty lies in the student's inexperience. Music students lack the "ear training" that professional musicians have; a student may not notice subtle errors in the sound, or even if they are aware of the presence of an error, they may be unable to pinpoint the source. Finally, controlling an unfamiliar musical instrument requires a vast amount of concentration; beginners simply have no cognitive power left to listen to their sound, let alone critically analyze their performance. With more experience, instrumental control becomes subconscious (much like learning how to walk or drive a car), but this process takes years.

5. DAILY PRACTICE

If we are to improve the daily practice of music students, we should describe how students practice. This description is typical for Western instruments such as violin, oboe, or piano. Some portions of this daily routine may be omitted for certain instruments, but the basic framework is the same.

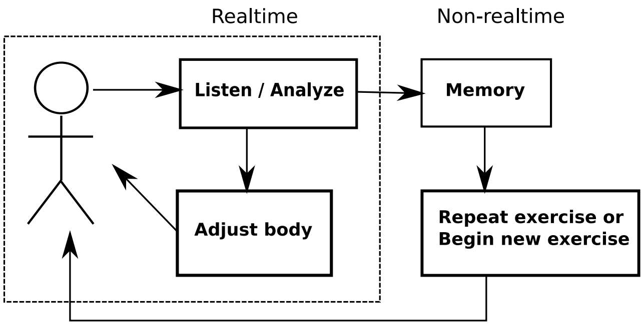


Figure 1: *Schematic of feedback cycles in daily practice*

The practice begins by playing scales. On most instruments, these test intonation (producing the correct pitches), speed, and good sound quality in all ranges of the instrument (high notes and low notes).

This is followed by technical exercises. These are generally quite short; many exercise are between two and ten seconds. These exercises are rarely "musical" (aesthetically pleasing); they are the instrumental equivalent of weight training to improve team sports. Many technical exercises involve playing notes very rapidly (yet still with accurate pitch), some involve playing notes which are quite distant, and other exercises simply ask the student to play a long note with steady pitch, loudness, and tone quality. Depending on the seriousness of the student and the teacher's approach to music education, these technical exercises may be omitted entirely – very few students enjoy performing these exercises, and many teachers consider fostering an enjoyment of music to be more important than improving a student's ability as quickly as possible. The analogy to weight training is also applicable here: if one simply wishes to play sports for fun, weight training is not necessary; if one wishes to play competitively, extra physical training is required to improve quickly.

Next a student will play a 'study' or 'étude'. These lie somewhere between technical exercises and normal music: each study is specially composed to stress certain technical skills (playing high notes, producing a certain kind of sound, etc.), but a study *is* musical. Studies are also much longer than technical exercises; most are between one and ten minutes long.

Finally, a student will begin playing pieces of music. Depending on the student's age and ability, there will generally be two to five pieces of music.

At every stage of this daily practice, the student should be analyzing his own performance as shown in Figure 1. There are two feedback cycles: realtime and non-realtime. In the realtime feedback cycle, the student is performing tiny adjustments to his body in response to the sound – adjusting fingers, lips, air flow, etc. After a student has finished playing, he must decide whether to continue practicing the same material (and if so, what part(s) he should fix next), or whether to move on to new material (another exercise, another piece of music, or ceasing practice).

In his previous lesson, the teacher will have pointed out a few mistakes of the student, and directed the student to correct these mistakes. However, in the six days between weekly lessons, the student may develop other mistakes – or, once the initial mistakes were fixed, the student should concentrate on fixing smaller mistakes which the teacher did not mention during the previous lesson. In addition, these mistakes are not always obvious to the student, even once the teacher has discussed the problem.

For example, suppose that the student was warned that he always played a certain note too high. In some cases – particularly with advanced students – the student can fix the mistake himself. However, less experienced students may lack the ability to hear the difference in pitches. The student may honestly believe that he played the note at the correct pitch, when in fact it was horribly out of tune. When the teacher is present, he may analyze the student's sound and notify the student of mistakes, but without the trained ears of the teacher, the student is helpless to fix the mistake.

6. ENHANCING INDIVIDUAL PRACTICE

In Sections 3 and 4, we identified individual practice as the most important area for computer enrichment; here we examine this area in greater depth and identify directions for useful research.

6.1 What we shouldn't do

No interactive tool should distract the student from listening to themselves. Developing realtime self-analysis skills – maintaining a constant feedback cycle between sound and action – is absolutely vital for playing a musical instrument. Most instruments have variable pitch (even an instrument with relatively-fixed pitch, such as a saxophone, can play bad pitches due to air flow) and immensely variable sound quality. Music students must learn to adjust their bodies (be it fingers, arm positions, air flow, lip position, etc) automatically in response to the sounds they produce.

Consider an analogy to a baby learning to walk. Standing on two legs is an immensely non-trivial feat of balance; a baby must learn to make thousands of tiny adjustments to their bodies in response to sensations in their inner ear. Now suppose that we displayed some colors on a computer screen for the baby – blue if the baby was leaning too far forwards, red if they were leaning backwards, etc. A baby might learn to associate the visual colors with their ability to remain upright, instead of using the sensations in their inner ear. Once we remove the computer display, the baby cannot walk.

For this reason, we suggest that realtime interactive tools should be used with caution. In general, computer analysis and interaction should occur *after* a student has finished playing their instrument. The computer should be used to confirm (or correct) a student's judgement, not as a replacement for realtime self-analysis. This may come as a disappointment to researchers interested in pushing the boundaries of digital signal analysis – realtime processing is much more challenging than offline processing, after all – but realtime multimedia tools may be counter-productive if used indiscriminately.

It should be clear that we suggest caution, not a complete ban. There are some cases where it is appropriate to introduce technological aids for short-term gain at the expense of long-term development. For example, many violin teachers place pieces of tape on the instruments of young students to show them where to place their fingers – in effect, adding frets to a fret-less instrument. Although violinists must learn how to play without tape, many teachers feel that using the tape for a few months is a worthy trade-off. This belief is not universally shared; some prominent violinists argue against the tape. In the same way, a particular realtime computer program may be useful in the short term despite distracting the student from concentrating on his sound.

There is one area which is safe for virtually indiscriminate use of realtime tools, however: technical exercises. As discussed in Section 5, these are short exercises which are aimed at developing specific skills. A realtime visualization tool which is used for specific technical exercises is unlikely to subvert a student's long-term development of his realtime self-analysis skills – especially since different technical exercises will probably use different visualization techniques. It is unlikely that a student will use one particular technical exercise tool often enough that this tool replaces his own ears.

Finally, we should emphasize that our caution refers only to realtime interactive tools – programs which provide feedback to the student while the student is playing their instrument. We have no concerns about multimedia tools which provide feedback after the student has finished playing their instrument (i.e. in the "repeat exercise or begin new exercise" part of Figure 1), even if this data is gathered via realtime digital signal processing algorithms.

6.2 What would be useful

Given our discussion so far, there are two areas that would benefit the most from multimedia computer programs: computer analysis to provide objective self-testing tools, and motivation to practice technical exercises.

In the near future, we suggest focusing on technical exercises. Since each exercise has a specific purpose, we can develop algorithms to analyze sound for those particular features. Analyzing a five-second audio file for one or two features (such as pitch stability or a gradual increase in amplitude) is much easier than analyzing a five-minute piece of music to find all musical mistakes. This also simplifies the task of giving feedback, since the student is only expecting one or two metrics about their performance.

6.3 Motivating exercises: Educational games

In addition to having a clear focus to each computer analysis tool, technical exercises have the greatest need of the extra motivation and 'fun' that multimedia tools can provide. Here we shall examine three successful educational games which provide inspiration for musical edutainment projects.

6.3.1 Project LRNJ: memorizing Japanese kana

LRNJ [12] is a simple downloadable game which teaches users the Japanese alphabets (*hiragana, katakana,* and *kanji*).

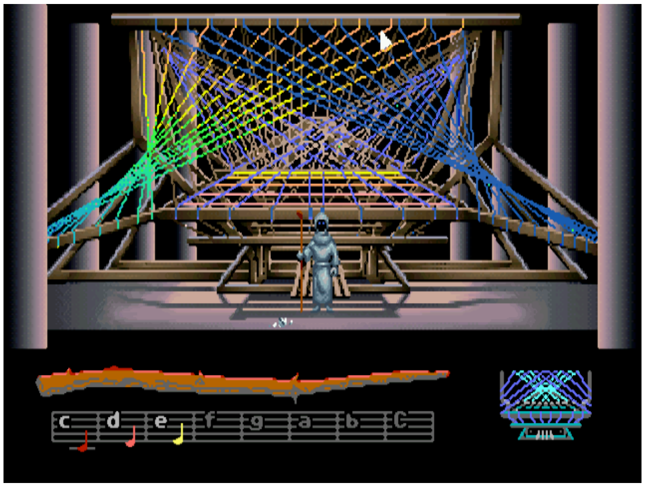


Figure 2: *Screenshot of LOOM; only three notes (CDE) are available. As the game progresses, more notes are unlocked.*

The game is fundamentally a simple flash-card memorization tool: the game displays a character, and the user must type the English equivalent of the character (for *hiragana* and *katakana*) or the meaning (for *kanji*). If the user guesses incorrectly, the game displays the correct answer, and asks another question. However, this game "dresses up" these flashcard questions in the guise of a 1980-style role-playing game. Slime monsters have kidnapped the princess of the kingdom, and the user (playing a poor farmer called Jenk) must rescue the princess. Jenk must visit towns, talk to villagers, and fight slimes. But instead of simply attacking a slime and doing damage (as is customary in RPGs), the user is presented with a Japanese character. If the user correctly identifies the character, then the slime is damaged or killed; if the users makes a mistake, the slime is healed.

By "wrapping" a boring task (memorizing over a hundred characters, not counting the *kanji*) in the guise of a nostalgic computer game, the task became much more fun. The primary author of this paper had attempted to learn *hiragana* and *katakana* in the past, but having very little patience for pure memorization, had given up after only five minutes. When he discovered LRNJ, he played the game for six hours straight. We should emphasize that the true value of LRNJ was not any sophisticated interactive tutoring system (although LRNJ *does* use some intelligence in deciding which character to display next); it was the extra motivation.

6.3.2 LOOM: a musical adventure computer game

LOOM [3] was an innovative adventure computer game created by LucasArts, published in the early 1990s and shown in Figure 2. In this adventure game, the main character is a Weaver: a person who can cast spells by performing short melodies. Some melodies may be read in books, but others must be learned from the environment. For example, to learn the "see in the dark" spell, the user must listen to the song that an owl sings, then try to replicate the notes. Certain spells must be cast in order to progress through the game – for example, the player cannot navigate a maze in the dark without casting the "see in the dark" spell.

6.3.3 *Guitar Hero: graded guitar karaoke*

The Guitar Hero series of games [4], as well as their predecessors Frequency and Amplitude, are musical games for gaming consoles. The basic premise is similar to karaoke: the user must ‘perform’ certain preset songs. In the earlier games, this performance was created with the standard console game controller; with the Guitar Hero games, the user plays on specially-made mock guitars.

6.4 A Game for Classical Instruments

Guitar Hero and LOOM both offer a clear vision of education games for musical technical exercises. Instead of ‘performing’ music with a replica guitar, why not perform it on a real acoustic guitar, oboe, or violin? Instead of typing “`C G F D`” to cast a magic spell, we should have the user perform the melody on their instrument.

We could easily add some practice / repetition to this task: suppose that the hero in our adventure game must climb over a wall. He knows the “create ladder” spell, so he performs the exercise (playing a steady pitch with constantly-rising amplitude) on his violin. The computer analyzes the sound and gives the user a score of 70%. Instead of giving this number to the user as feedback, the game draws a ladder growing from the ground – but stopping before it reaches the top of the wall. The user must then perform the melody again – perhaps ten or fifteen times in a row – until he receives an acceptable score and the ladder reaches the top of the wall. Depending on the test, the user may even be required to score above 70% three times in a row.

Designing the human-computer interface for such a game is a non-trivial task. We do not want users to have their hands on the keyboard (as is the case with LRNJ, LOOM, and many other edutainment games). Users should be playing their instrument as much as possible. There are two options for HCI: we could use a controller which does not require the user to release their instrument, or we could control the game via the sound of the instrument itself.

The advantage of using a controller (such as a typical console game controller) is that the interface is easily recognized by our target audience. The buttons on the controller (up, left, etc.) are easily mapped to the actions in the game. The disadvantage of such a controller is that the user is not playing his instrument. A one-handed controller minimizes this problem – the user could remove a single hand from the instrument while still holding it, provided that the actions within the game were short – but the problem remains.

We could also control the game via the sound of the instrument. By performing pitch detection (and possibly the whole music transcription chain, involving onset detection and note segmentation), we could control actions within the game. The advantage is clear: the user is always playing his instrument, either by playing exercises to advance through the game, or simply providing audio commands to the game. The disadvantage is that the game-controlling sounds are not very intuitive. For example, consider the simple problem of moving a character within the game. A high pitch could move the character up and a low pitch could move the character down, but there is no intuitive mapping between particular sounds and left/right movement.

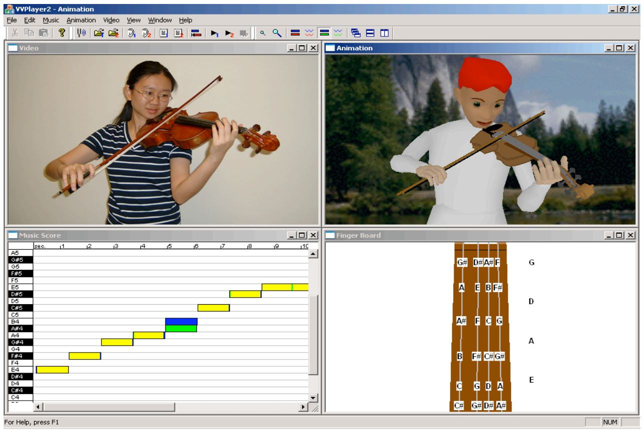


Figure 3: *User interface of the Digital Violin Tutor.*

Finally, as we saw in the discussion about the LRNJ game, edutainment software does not necessarily need to include sophisticated analysis and teaching algorithms. We repeat our earlier statement for emphasis: If students could be motivated to play their assigned musical exercises every day, that would far outweigh the benefit of the fanciest multimedia feedback systems.

7. OUR WORK

We are developing two projects to enhance daily practice.

7.1 Digital Violin Tutor

The Digital Violin Tutor (DVT) [23, 25, 11, 15] provides feedback in the absence of human teachers. DVT offers different visualization modalities – video, “piano roll” graphical displays, 2-D animations of the fingerboard, and even 3-D avatar animations. We present an example of this interface in Figure 3.

DVT consists of several interconnected modules, as illustrated in Figure 4. The student’s audio is transcribed and compared to the transcription of the teacher’s audio. If mistakes are detected, then the proper actions are demonstrated by the 2-D fingerboard animation, video, or the 3-D avatar animation.

The music transcription system in DVT is customized for use with violins in student’s homes. This audio is quite noisy – the microphone will be quite cheap, it will not be placed in an optimal position, and the recording levels will not expertly set. The transcriber must be quite robust against such problems. Fortunately, we can tailor our transcription algorithm to our specific instrument [24].

We aim to develop an intelligent digital tutor that should adapt to student’s knowledge, interest, motivations. Our animated characters (playmates in virtual worlds) will learn and do things jointly with student. Furthermore, network support will be added to DVT. In the future, DVT should balance personalized and peer learning, derive social networks from personalized intelligent aides, and adapt knowledge to either the personal or group level.

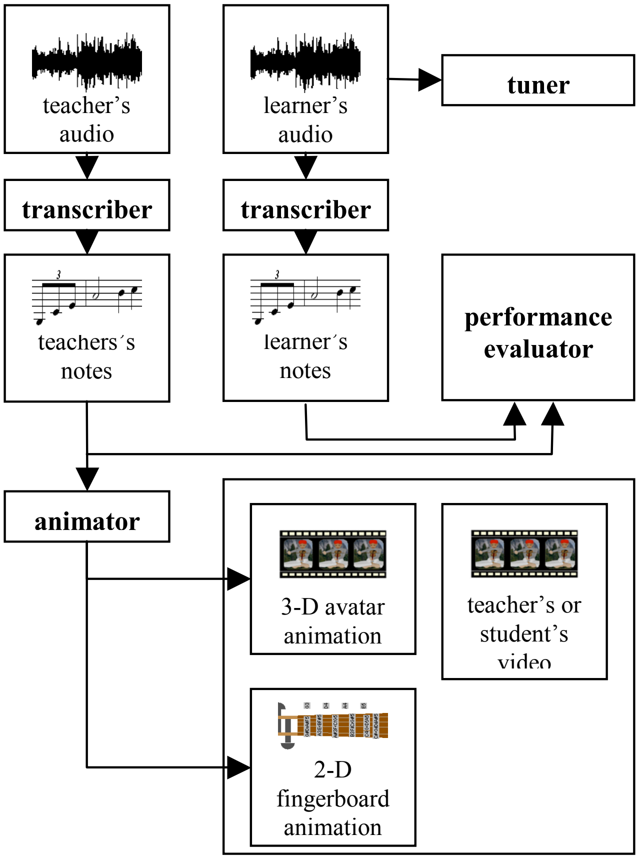


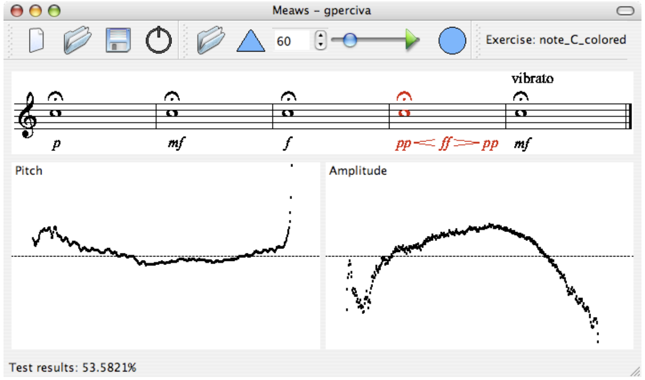
Figure 4: *System design of the Digital Violin Tutor.*

7.2 MEAWS

Another project is MEAWS (Musician Evaluation and Audition for Winds and Strings; it may be pronounced as either "Muse" or "Miaos"). MEAWS focuses on technical exercises: it has a rapidly-increasing number of tests (intonation, rhythm, steady loudness on a long note, etc). Students may select a particular test, perform the exercise, and then receive a grade. Previous exercises may be stored, so students – or their parents or teachers – may easily check previous exercises to view progress (or lack thereof).

MEAWS is intended to be used in conjunction with a larger curriculum and/or under the direction of a music teacher: a teacher may assign a particular exercise to the student, possibly with a minimum grade. For example, a student may be assigned the "long tone with constant pitch but *cresc / decresc* dynamics" exercise illustrated in Figure 5. MEAWS provides no other feedback (no lessons, nor any sparkling stars or happy faces) – this tool is focused on the specific problem of enhancing individual practice, not motivation.

The teacher may ask the student to achieve as high a grade as possible, or may direct the student to achieve a specific grade before the next lesson. While the student is practicing the technical exercise at home, he should have no doubts whether he is performing the exercise correctly or not, since MEAWS provides an objective self-testing tool.



Test results: 53.5821%

Figure 5: *MEAWS: results of a* cresc / decresc *exercise. The amplitude has the right shape, but the pitch should remain constant.*

In addition to increasing the number of technical exercises, we plan to implement an online "grade submission" system. If a student wishes, he may submit his anonymized results to a server. The grade would be compared against students with a similar background (i.e. 12 years old and playing their instrument for 3 years), so that the student could see how he compared to similar students. For students enrolled at a large music school, this may be a small gimmick ("Hey, are you bored? Let's check out how we compare to kids halfway across the world!"), but for music students in more remote locations, this could provide extra motivation or simply lessen a feeling of isolation.

8. FUTURE TECHNOLOGIES

CAMIT is an application which is supported by four different technology areas:

1. *Human-Computer Interface*: before any computer analysis can be performed, the computer must receive information about the student. Most information will be transferred with a microphone and the student's sound, but are there more natural ways of transferring the other information?
2. *Music transcription*: once we have the information (a student's performance), the computer must understand what the student played. In some cases, we need only produce a list of notes and rhythms (the traditional transcription problem), but in other cases we need more information.
3. *User modeling*: after analyzing the student's performance, the computer must interpret this information in conjunction with knowledge about the student. Does a student need to watch another lesson, or should the student simply practice the exercise again? Optionally, how does this student compare to his classmates and/or students of with similar experience?
4. *Visualization*: once the computer has decided what information (if any) the student should receive, this information must be presented in the most intuitive manner possible. This may changed based on content, context, and the output device.

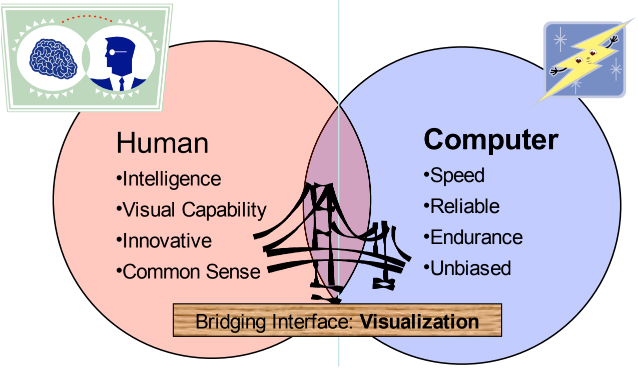


Figure 6: *Bridging the gap between human and computer*

8.1 HCI and Visualization

So far, most CAMIT projects receive most of their information with a microphone and the student's sound – but the normal "keyboard + mouse" tools are still used to set up the software, ask for feedback, or move on to the next exercise. A future digital tutor should enable users to interact with the computer more naturally with microphones (controlling exercises with voice recognition, instead of merely gathering audio data for student performances), video cameras, and haptic sensors.

When designing these systems, we suggest a user-inspired approach, which includes music educators and prospective students/users early in the design loop in order to closely integrate the educational model into our system. Through a continuous spiral process of improvement, it is possible to make a digital tutor truly useful and satisfying users' needs.

8.2 Music transcription

We propose a new approach on music transcription which is based on instrument model and multimedia fusion. Our new method is an instrument-specific approach combining information from audio, video and other sensors.

Despite decades of research effort worldwide, a practically applicable, general-purpose transcription system does not exist at the present time [13, 14]. To move forward, we believe that a potentially very successful approach is to focus on modeling a specific musical instrument incorporating meaningful constraints [24]. For our intended applications in personalized music education, it would be sufficient if the transcriber can extract the pitches and other relevant audio features specific to the targeted instrument (e.g. voice track or violin). With this approach, we expect to achieve significantly better performance in comparison with the state-of-the-art general-purpose transcribers. Furthermore, we propose to incorporate video cues to improve transcription performance.

As digital video cameras are now commonplace, our immediate next step is to extract information from the associated video to assist music transcription. Similar approaches have been applied to speech recognition, but not in the context of music transcription. We will also attempt to try machine learning methods to improve the system performance.

Our user-centered approach assumes that the user usually needs a music transcriber for a specific application (e.g. transcription of solo singing/violin for education). With this approach, we can impose meaningful constraints on the problem to improve transcription performance. Once the transcription performance reaches a certain threshold, it becomes useful for real-life applications. Another challenge is the transcription speed. For our projected applications, transcription speed is a critical requirement. If students must wait for thirty seconds after playing each exercise, they will abandon the tool very quickly.

8.3 User modeling and Network support

The most difficult part of any computer-assisted tutoring project is deciding how to deal with errors. We cannot simply inform the user of all errors – with non-fixed-pitch instruments, *every* note is "out of tune" to some degree; students would become horribly discouraged if they always received a grade of 0. When a human teaches a musical instrument, they perform enormous filtering of the student's errors, for both musical and psychological reasons. If a child is feeling very discouraged, it may be useful to praise the child (deserved or not) and omit any discussion of mistakes entirely!

We already see some effort in the direction of musical error filtering with IMUTUS [20, 10, 19] categorizing mistakes into different priority levels. There is room for much more research in this area. We propose to follow IMUTUS' lead by consulting with numerous experienced music teachers to compile a list of the 'seriousness' of various mistakes. We also propose one more step: to allow a student's real-life teacher to adjust the relative priorities of certain mistakes. For example, a teacher may direct a student to "forget about intonation and rhythm; concentrate on smooth slurs." In this case, it would be counter-productive if CAMIT software continued to nag the student about poor intonation.

As the world becomes more interconnected, exciting new opportunities for collaboration arise. In addition to playing games or writing academic papers with people from across the globe, we can also cooperate in learning how to play musical instruments. Playing and studying music together is far more enjoyable than sitting at home alone, but many students do not have family members who play an instrument (or these students do not play piano and thus cannot take advantage of Oshima et al's [17]. If we could allow such behavior through CAMIT projects (possibly leveraging audio streaming technologies), this would certainly be of great interest to many students all over the world.

One could go one step further and insert the teacher/student relationship into an online learning game. For example, a user could put the program into a mode where he can see other online users. If a teacher and student are both online, they could team up and form a temporary teacher/student session. Teachers could earn points by being graded by students and thus earn some kind of 'fame' as a popular teacher. It could even go as far as enabling teachers to earn money with the system – the more 'fame' a teacher receives, the more money they get. This would enable people in totally remote places of our planet to earn money with their teaching skills and expertise, and likewise for students in remote locations to learn.

9. CONCLUSION

We have discussed recent work in computer-assisted musical instrument tutoring, and examined potential goals of such projects. We identified daily individual practice as having particular need of assistance, and the most productive way to enhance practice is to increase the motivation and efficiency of technical exercises.

This motivation may take several forms, but we feel that many students would benefit from targeted edutainment games. Musical instrument edutainment games present some special challenges: first, we do not want students to be sitting in front of a computer with traditional HCI tools; they should be playing their instrument as much as possible. Second, we must use novel visualization techniques to provide intuitive feedback for the students. Finally, the potential for online collaborate learning is an extremely exciting area that should be investigated.

10. REFERENCES

[1] i-maestro. http://www.i-maestro.org/, 2007.

[2] Vemus: Virtual european music school. http://www.vemus.org/, 2007.

[3] Loom (graphical adventure game). http://en.wikipedia.org/wiki/LOOM, article: 2007, original game: 1990.

[4] Guitar hero (series). http://en.wikipedia.org/wiki/Guitar_Hero_%28series%29, article: 2007, original game: 2005.

[5] R. B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, and P. Capell. A computer-based multi-media tutor for beginning piano students. *Journal of New Music Research*, 19(2-3):155–173, 1990.

[6] 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.

[7] S. Ferguson. Learning musical instrument skills through interactive sonification. In N. Schnell, F. Bevilacqua, M. J. Lyons, and A. Tanaka, editors, *NIME*, pages 384–389. IRCAM - Centre Pompidou in collaboration with Sorbonne University, 2006.

[8] S. Ferguson, A. V. Moere, and D. Cabrera. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In *IEEE Proceedings of the International Conference on Information Visualisation*, pages 97–102, London, 2005.

[9] D. Fober, S. Letz, and Y. Orlarey. VEMUS - une école de musique européenne virtuelle. In *Proceedings of the Journées d'Informatique Musicale*, pages 57–63, Lyon, France, 2007. in french.

[10] 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, Paris, 2004.

[11] W. Jie, J. Boo, Y. Wang, and A. Loscos. A violin music transcriber for personalized learning. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 2081–2084, Toronto, ON, Canada,, July 2006.

[12] D. Johnson. Project lrnj: Learn japanese rpg. http://lrnj.com, 2007.

[13] A. P. Klapuri. Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–282, 2004.

[14] D. M. Klapuri A. *Signal Processing Methods for Music Transcription*. Springer, 2006.

[15] A. Loscos, Y. Wang, and W. J. J. Boo. Low level descriptors for automatic violin transcription. In *ISMIR*, pages 164–167, 2006.

[16] B. Ong, K. Ng, N. Mitolo, and P. Nesi. i-Maestro: Interactive multimedia environments for music education. In *i-Maestro 2nd Workshop on Technology Enhanced Music Education*, ICSRiM - University of Leeds, 2006.

[17] C. Oshima, K. Nishimoto, and M. Suzuki. Family ensemble: a collaborative musical edutainment system for children and parents. In H. Schulzrinne, N. Dimitrova, A. Sasse, S. B. Moon, and R. Lienhart, editors, *ACM Multimedia*, pages 556–563. ACM, 2004.

[18] M. Robine and M. Lagrange. Evaluation of the technical level of saxophone performers by considering the evolution of spectral parameters of the sound. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 79–84, Victoria, Canada, 2006.

[19] 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, Barcelona, Spain, 2005.

[20] 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)*, pages 143–150, Villach, Austria, 2004.

[21] T. Shiratori, A. Nakazawa, and K. Ikeuchi. Synthesizing dance performance using musical and motion features. *2006. ICRA 2006. Proceedings 2006 IEEE International Conference on Robotics and Automation*, pages 3654–3659, May 2006.

[22] S. W. Smoliar, J. A. Waterworth, and P. R. Kellock. pianoforte: a system for piano education beyond notation literacy. In *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, pages 457–465, New York, NY, USA, 1995. ACM Press.

[23] J. Yin, A. Dhanik, D. Hsu, and Y. Wang. The creation of a music-driven digital violinist. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 476–479, New York, NY, USA, 2004. ACM Press.

[24] J. Yin, T. Sim, Y. Wang, and A. Shenoy. Music transcription using an instrument model. *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, 3, Mar. 2005.

[25] J. Yin, Y. Wang, and D. Hsu. Digital violin tutor: an integrated system for beginning violin learners. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 976–985, New York, NY, USA, 2005. ACM Press.