April 27th 2012

Dr. Andrew Rawicz
School of Engineering Science
Simon Fraser University
Burnaby, British Columbia
V5A 1S6

Re: ENSC 305/440 Post Mortem for the Automatic Music Transcriber

Dear Dr. Rawicz,

The attached document, *Post Mortem for the Automatic Music Transcriber,* outlines the process or team went through when designing and implementing our system for ENSC 305/440. Our project was to design an all-in-one device that would not only act as a tuner/metronome for musicians, but also have the ability to produce sheet music from a live or saved recording.

This document details the current state of the device, deviations from our original plans, and our future plans for the device. In addition, we outline some of the budgetary and time constraints we encountered and explain the inter-personal and technical experience gains from working on this project.

ScribeWare Inc. consists of one sixth-year student and three fourth-year students: Mike Tyson, Henry Huang, Patrick Wong, and Shu Hui Wong respectively. If you have any questions or concerns about our proposal, please feel free to contact me by phone at (778) 886-0523 or by e-mail at mjt4@sfu.ca.

Sincerely,

Mike Tyson
President and CEO
ScribeWare Inc.

Process Report for the

Automatic Music Transcriber

Revision: 1.0

Prepared for: Dr. Andrew Rawicz – ENSC 440
Steve Whitmore – ENSC 305
School of Engineering Science
Simon Fraser University

Prepared by: Mike Tyson
Henry Huang
Patrick Wong
Shu Hui Wong

Contact Person: Shu Hui Wong
shw4@sfu.ca

Date: April 27th 2012

Table of Contents

List of Figures

List of Tables

Introduction

For the past thirteen weeks, the idea of an Automatic Music Transcriber has drawn together 4 outstanding individuals – Mike Tyson, Patrick Wong, Henry Huang and Shu Hui Wong – who have worked tirelessly towards the realization of the AutoTab. This report re-examines the process that took this dream from a concept to a reality and documents the experiences of each of the four members of ScribeWare Inc.

Motivation

The role of music transcription has had a tremendous impact in allowing music to be shared between people. It has allowed for the documentation of music pieces to be used for the generations to come. However, the act of transcribing itself is a long and arduous task for the creative mind and many solutions have been provided to help simplify this process but they do not actually carry out the actual transcribing. Our project aims to provide a solution to this problem by simplifying the music transcription process.

Current State of the Device

As described in the project proposal, the AutoTab will not only perform automatic music transcription, it can also act as a playback/recording device and a tuner/metronome. Figure 1 below shows the system overview block diagram of the entire design.

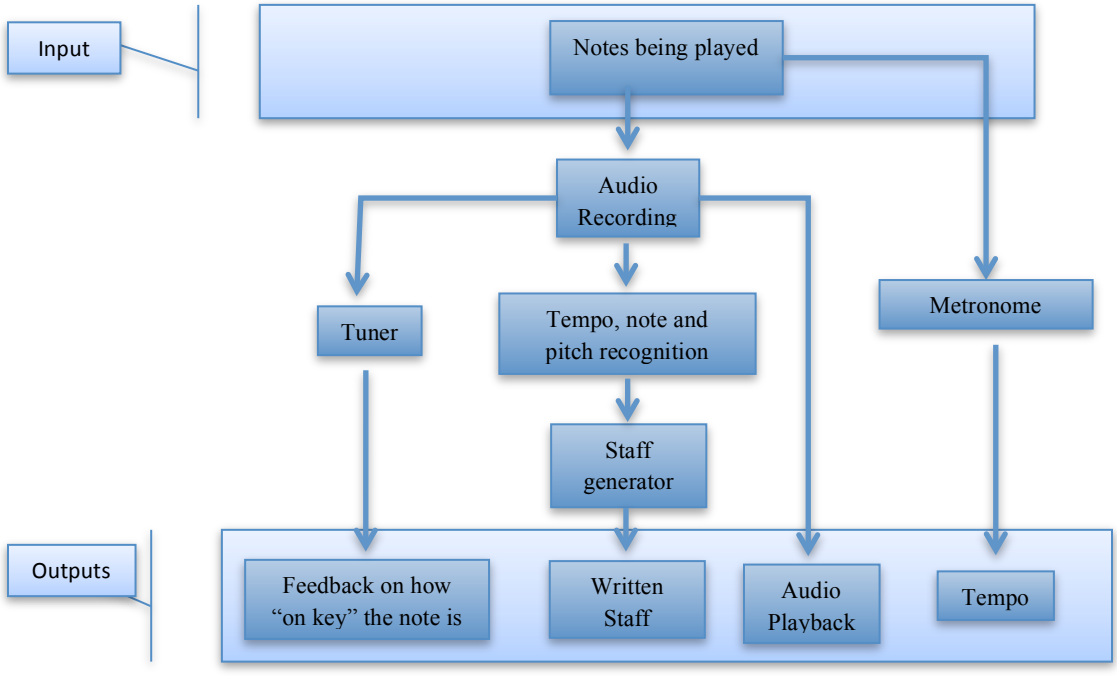

Figure 1: System Overview

As illustrated by Figure 1, the AutoTab contains several subunits. The current state of the device will be explained by reviewing each of the stages in Figure 1.

In the tuner mode, the algorithm involved with tuning the guitar will be implemented by sampling the audio of the string being tuned, using an FFT to process the recording, and identifying the fundamental frequency of the sampled recording. Once the fundamental frequency is identified, the difference between it and the closest matching note will be calculated and feedback will be provided to the user as to whether or not the string is tuned too high, too low, or in tune.

For the metronome, our original plan was that the user would input the tempo they wish to be playing at. Once this information is received, the system will play an audible click or tone to keep the user in time. A timer will be used to keep track of when the sound should be played. However, in our current state, our metronome beats at a pre-determined tempo of 120 beats per minute.

And lastly, for the music transcription, we have implemented a genetic algorithm as described in the document, *Design Specifications for the Automatic Music Transcriber,* and using a third-party software, successfully transcribed our music transcriptions onto staff.

Signal acquisition is performed through a direct connection from our audio source into a line-in jack, while the FPGA board handles both signal conditioning and processing.

Deviation of the Device

Overall System

In terms of functionality, we achieved what we had set out to do. Due to time constraints, we were unable to finish packaging the device, to implement the UI, to power the device using batteries and to make the device portable as described in the functional specifications.

Currently, our prototype completely runs off the FPGA board. Further deviations will be discussed in the upcoming sections.

Tuner

The tuner stage has not deviated much from our original plans. Through testing, we have had to down-sample our audio clips from 48 kHz to 4 kHz to greater improve the accuracy of our results. Also, due to time constraints, we were only able to get the tuner working on Matlab and not on our prototype.

Metronome

Due to time constraints, Instead of using audible sound, we will be using the LEDs to flash in accordance with the metronome's timing. Also, instead of allowing the user to choose what tempo they would like to play at, we have hard set that to 120 beats per minute.

Music Transcription

Due to the nature of the music transcription algorithm, there was no clear methodology in which to achieve our goals. Instead, our genetic algorithm that handles this process was constantly changing as needed and thus, there was very limited deviation from our original plans.

Future Plans

The AutoTab has great potential for further research and improvement. As we reexamine the development of the device, we have the following suggestions for future development.

Overall System

- *Package the device*

  As a prototype to prove the concept, the device in its current state is not portable. One solution is to implement our algorithm onto an ASIC, to power the entire design by batteries and to enclose the device in a plastic case.

- *Determine and Optimize power consumption*

  We hope to determine and optimize the power consumption of the entire system when we develop our own signal acquisition, conditioning and processing methodology.

- *Minimal disturbance to the user*

  One solution is to reduce the size of the device by making it more compact

- *Allow for recording through a microphone*

  We hope to allow for audio recording through the microphone in addition to the line-in jack option that the prototype currently has.

Tuner

- *Provide better visual feedback*

  We hope to implement a display that is similar to current marketable guitar tuners that will provide feedback on how in tune the note is.

- *Improve tuner functionality*

  Currently, our prototype is only able to tune a note that has been played. One solution to this is to give a user the ability to play back a fundamental note to allow for alternate tunings.

Metronome

- *Provide better feedback*

  We hope to implement a timing mechanism that will provide feedback through audible sounds.

- *Improve metronome functionality*

  One solution is to allow for the users to choose the tempo at which they would like to play at.

Music Transcription

- *Allow for multi-instrument transcriptions*

  One solution is to implement an adaptive dictionary so users will be able to calibrate the AutoTab to whatever instrument that is being played.

- *Refine algorithm for better results*

  The genetic algorithm can be refined for greater accuracy and efficiency. We can also implement chord recognition.

In the future, we will need to invite our target users, musically inclined people, to test the device and obtain their feedback.

Budgetary and Time Constraints

Budget

Table 1 contains the estimated cost and the cost of the project up to April 10th, 2012.

Table 1: Estimated Cost vs. Actual Cost

Equipment List	**Estimated Cost**	**Actual Cost**
Altera Cyclone II – DE2 University Dev Board (Terasic)	$269	Sponsored
LCD Display	$50	-
Buttons and Caps	$12	$10
Switch	$2	$2
Batteries	$10	-
microSD Card Socket	$4	-
16 GB microSD	$30	Borrowed 2GB
Microphone	$5	-
Speakers	$5	-
LED	$5	-
Audio Jack	$2	-
Casing	$30	-
Total Cost (15% contingency)	$424 ($488)	$12

As one can see, our estimated cost was more than 30 times the actual cost of the project as of April 10th 2012. One of the major reasons was that Terasic sponsored us and gave us an Altera Cyclone IV development board. We were also able to borrow certain items, such as the SD card from our own team members and thus, the cost of the project is significantly reduced.

Time

A Gantt chart is shown in Figure 2 below in which the expected time to completion is indicated in blue and the slippage is indicated in gray.

ID	*Task Name*	*Start*	*Finish*	*Jan 2012*	*Feb 2012*	*Mar 2012*	*Apr 2012*
1	Project Proposal	1/6/2012	1/16/2012				
2	ESSEF Proposal	1/9/2012	1/11/2012				
3	Research	1/6/2012	2/16/2012				
4	Research (Actual)	1/6/2012	2/23/2012				
5	Functional Specification	1/9/2012	2/6/2012				
6	Oral Progress Report	1/9/2012	2/13/2012				
7	Design Specification	2/6/2012	3/5/2012				
8	Coding/Testing	2/6/2012	4/2/2012				
9	Coding/Testing (Actual)	2/6/2012	4/27/2012				
10	Written Progress Report	2/20/2012	3/19/2012				
11	Integration/Debugging	3/6/2012	4/2/2012				
12	Integration/Debugging (Actual)	3/15/2012	4/27/2012				

Figure 2: Gantt chart

It is interesting to note that algorithm research took longer than expected and in turn, it pushed the completion dates for the coding/testing and integration/debugging stages to a later date. One reason was that we had underestimated the availability of in-depth research for music transcription and the complexity of the proposed algorithms.

Despite being unable to meet all of our deadlines as initially proposed, integration of the entire system was completed successfully prior to our final demonstration on Friday, April 27th 2012.

We all felt that it was very important that we tried our best to follow our project schedule and complete the project as planned. However, even though we tried to adhere to our schedule as best as we could, we misjudged how long it would take to convert all our Matlab code to C for our FPGA but were fortunate enough to get our final demonstration pushed to April 27th from April 23rd.

Inter-Personal and Technical Experiences

Mike Tyson – Chief Executive Officer

My contribution to our project was predominantly in forming the overall direction and concepts relating to our project, and in algorithm development. As with the other team members, I grew into my role organically: in my case, I had a very clear image of our end product, and was able to scale this down into proof-of-concept deliverables that could realistically be achieved on our timeline. In addition, as it is one of my predominant strengths, I was able to brainstorm many of the algorithms achieving the end functionalities we defined, which led to most of my work being performed on paper or in MATLAB.

To quantify, I was responsible for defining the different stages of development and their deliverables, as well as many of the features and requirements for our product, defined in our functional spec. I defined the priorities for the device as a product (intuitive user experience, accurate results, rapid control of features), and designed both the physical and menu interfaces to address these concerns. In terms of the actual software modules, I designed and implemented the tuner, the gridding and quantization algorithms, the MIDI file generator, and the fitness function for the genetic algorithm. My experience with audio production saw me generating much of the reference audio the code was tested with, as well as providing tools with which to test our transcriptions.

In addition to planning and coding project elements, I represented our group in several exchanges with external organizations: I was able to secure our sponsorship from Terasic, from whom we received our hardware free of charge, I met with professors and other experts for advice, and I am currently soliciting companies and investors to see if there is any interest in sponsoring our research towards developing a commercial product.

There were many lessons learned over the course of this semester. In the realm of facts and numbers I have developed a superior understanding of the theory surrounding our project: musical/audio theory, the digital representation of audio and its processing, and the MIDI and WAV specs, for example. More generally, I have established an appreciation for the research and development process: how to go about planning the phases of development from an idea to a product, the point of a proof-of-concept model and how it differs from a prototype, and the different stages in implementing an algorithm into a finalized hardware solution. Most of all, I now have an idea of what's involved when something is invented - how the industry works, and how the work that goes into it can be made profitable.

One of the main challenges in undertaking this project has been adjusting our assumptions (and algorithms) in the face of unexpected results or conclusions. An example of this was in how we evaluated how similar one audio sample was to another: initially, we would subtract one waveform from the other and use the remaining energy to rank its potential as a match. Much later, we came to realize that the signals could effectively be characterized by the frequencies of their most significant peaks. This simplified our calculations significantly, but completely

changed the way in which the algorithm was configured, as well as its strengths and weaknesses.

Another example was in evaluating the overall results: these were much worse than anticipated, but with consistent errors: we kept seeing the correct notes getting paired with additional harmonics, or single frame deviations on longer notes. The problem, as we came to eventually understand it, was that the genetic algorithm by itself assumes that every noise coming from the guitar is deliberate. This is not a realistic model, as there are many impurities created by guitar play that we do not normally consider: the tone created when releasing a string from its fret, or when sliding ones fingers along the string, or just brushing up against one or more extra strings when strumming. Thus, the results needed to be adjusted, taking considerations as to the context of the results.

Henry Huang – Chief Technical Officer

As a member of ScribeWare, my main task was to setup the hardware system on the FPGA. I took on this task because I felt it was where my strengths were, and where I could best help the team. Over the course of setting up the hardware, I've learned many things, both about myself and about the project.

The first lesson that I've learned was that the hardware demands respect. At no time can you forget minute details and think it won't be important later on. While setting up the hardware, I had to worry about memory allocation, clock speeds, timing and routing issues, where the linker script was placing my program, and other problems that I usually wouldn't need to deal with in software. I went into this project knowing about these potential problems beforehand, but after spending time debugging them, I now have a better idea of how to fix them.

Secondly, a personal lesson I've learned was that I should ask for help more often. There were times when I was worked up about something and became lost in my own world, trying to solve a problem. There's a certain level of satisfaction that I get from solving a problem on my own, but when I start spending too much time on an issue, I should really be asking for external help. This problem was most apparent when I was setting up the fast Fourier transform (FFT) in hardware. I spent over a month setting it up and I felt that if I had only asked a professor for help, I would have solved my problems quicker.

The last lesson that I've learned is that a "good enough" approach is far better than the "best" approach. Relating to the setting up the FFT, I had originally tried to use a memory controller. This would have facilitated incredible performance and throughput for reading and writing data to and from the FFT. Unfortunately, after spending weeks trying to get it to work, I gave up and ended up using a FIFO which would have meant lowered throughput, but it involved a more straight forward implementation in comparison.

Overall, I appreciate the hard work that the team has put in to this project. Once the project really got started, you start to realize how much work is actually involved and there is no way that one person can do everything. I'm glad that I had the opportunity to work with such a high caliber group for my capstone project.

Patrick Wong – Chief Financial Officer

Over the course of the past four months, I have learnt many things, both technical and non-technical. As a part of ScribeWare, my role was primarily in dealing with the software algorithm used in our note recognition project, and eventually system integration.

While working on the algorithm, I quickly realized over the course of research and development that many of the things which were seemingly straightforward are in fact not. Given the wide range of existing research, it was obvious that there wasn't yet a foolproof approach to the problem of music transcription. The result of this was that we would wind up developing a fair portion of the concept using a mix of existing research as well as our own ideas.

The group dynamic as it was felt very conducive to the problem solving and creative process that was required. Often during our meetings we would share a few very good ideas as to how we should tackle a particular issue.

The technical skills that I found to be the most beneficial in aiding me with my tasks would have to any of the software programming courses such as but not limited to ENSC 250, 350, 351. Over the course of this project, I have developed a deeper understanding and reached a higher level of comfort when it comes to working with software.

I would describe the organization of our group as mostly flat. We were for the most part self-motivated and driven by our own respective tasks, there was no manager by the usual sense of the word and most importantly, everyone was willing to help one another and got along well. This was particularly wonderful because it meant that we were able to learn from each other.

The toughest part of this project for me personally would have to be the system integration where we married the software and hardware together. It grew become a source of great frustration as I learned to truly appreciate the value of well-designed development tools. It is suffice to say that the Altera Nios II Software Build Tools are not something I would be too eager to work with again soon.

Would I do this again now knowing what the experience is like? The answer would be both yes and no. The four months that we were given wasn't enough for what we would have liked to have accomplished with our project which, as a result, makes for a fairly stressful final month. On the other hand, the people that I've had the opportunity to work with and the opportunity to finally exercise the skills that I've acquired over the years in school made it all worthwhile.

Shu Hui Wong – Chief Operating Officer

From an overall perspective, I have learnt that the design cycle is very much an iterative process. Modifications are constantly being done to not just improve the initial design, but also to compensate for something that might not have worked at all to begin with. I have also learnt that one of the best ways to help yourself understand something is to try and explain it to someone who does not know anything about what you are trying to achieve.

My main contribution to our project was in the form of documentation and ensuring that the communication channels between team members were open. I also had the great fortune of being the person to bounce ideas off of. One of the most important attributes I've learnt is that you truly need to have a firm grasp of what everyone in the group is doing in order to bring it full-circle. Due to the nature of our project, and how open-ended the current research available is, I've learnt that it wasn't as simple as programming an algorithm and porting it over to hardware. Weekly meetings were essential, as was working together, even if we were doing completely different tasks. I've also learnt that it's important to ask when you're uncertain. There is truly nothing to lose and all to gain by doing so.

In retrospect, we as a group initially spent a lot of time on the project brainstorming before growing into our respective roles. However, at no time did anyone needed to be prodded along to complete their tasks or forced to take on a certain role. We played to our strengths and helped out whenever something needed to be done.

Our individual personalities complemented one another in such a way that though we were all pretty easy-going, we knew when to be firm and to not take things personally. Looking back, I don't think I could have asked for a better experience. Where else could you bring 4 students who started out the semester not really knowing one another, ask them to go through a 13-week engineering cycle and have them come out successful and firm friends?

Conclusion

The AutoTab was a complex and challenging project. Over the past 4 months, members of ScribeWare Inc. have developed a viable prototype that fills a specific niche in the music industry. Healthy team dynamics and exceptional teamwork were the keys to the success of the project. All members are proud of their accomplishment and will continue working on the prototype for future improvement, as they are confident that the AutoTab is a marketable product.