Functional Specification 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: February 6th 2012

Executive Summary

> *"Music is enough for a lifetime, but a lifetime is not enough for music" – Sergei Rachmaninov.*

Music, like tradition, can be passed down through many generations. However, the process of manually archiving it to do so has proven to be an arduous task and whilst many efforts have been made to automate it, they are more often than not cumbersome and inefficient. The Automatic Music Transcriber (AutoTab) seeks to give artists a portable, hassle-free, "all-in-one" solution with its ability to record and produce staff, while also doubling as a tuner/metronome for the user.

Development of the AutoTab will occur in three phases. Upon completion of the first phase, AutoTab's frequency/signal processing algorithm will be fully integrated on the software side. It will be able to process a recording, determine what notes have been played and also recognize chords.

Our second phase of development would be the physical realization of our device on a NIOS soft processor. The bottlenecks that are identified in our algorithm will also be deferred to hardware for greater efficiency. Once this stage is completed, MIDI files generated by our device should be able to be used in a digital audio workstation (DAW) or be converted onto written staff.

Finally, in our last phase, we aim to defer all hardware (with exception of our processor) to an external casing for an encapsulated prototype for portability.
The AutoTab will conform to all applicable standards and guidelines, including those of the CSA and ISO/IEC.

Our proposed 13-week engineering cycle for this project will encompass research, design and integration. Our scheduled completion date for an operational prototype is April 2nd 2012.

Table of Contents

List of Figures

Glossary

BIN Value	Derived from the number of samples in an FFT divided by two metric used to describe resolution
CSA	Canadian Standards Association
DAW	Digital Audio Workstation – system designed for playback, recording, and editing of digital audio
DE2	Altera Development and Educational FPGA equipped board
MIDI	Musical Instrument Digital Interface – industry specification for the encoding, storage, transmission, and synchronization of instrument recordings and control data for MIDI compatible devices
Monophonic	Describes music consisting of single notes or a single melodic line
NIOS	Embedded-processor architecture designed for the Altera family of FPGAs
Polyphonic	Describes music with multiple melodic lines or multiple notes played simultaneously such as chords
Staff	Musical notation consisting of five horizontal lines and four spaces each representing a different musical pitch
Tempo	Musical notation describing the pace or speed of the piece
Transcription	The process of writing down music that otherwise was not – such as an improvised piece
Tuning	The process of adjusting the pitch of a musical instrument such that each key/string is separated by a known pitch interval
Quantization	The process of converting a continuous signal to a discrete sample set

1. Introduction

The Automatic Music Transcriber (AutoTab) is a portable lightweight device that will allow for the conversion of audio recordings into written staff. Users will also be able to use it as a metronome, a tuner and also for audio manipulation. The requirements for the AutoTab, as proposed by ScribeWare Inc. are described in this functional specification in the order that they will be met throughout the development process.

1.1 Scope

This document describes the functional requirements that must be met by a functional AutoTab. It fully describes the proof-of-concept device and partially describes the production device. This set of requirements will drive the design of the AutoTab and will be traceable in future design documents.

1.2 Intended Audience

The functional specification is intended for use by all members of ScribeWare Inc. The project manager shall refer to this document as a concrete measure of progress throughout the development phase while design engineers shall refer to it for overall design goals from product design to implementation. Test engineers shall use this document to assist in the similarity of function in the actual system with the documents' described functionality and also to aid in the design of test trials.

1.3 Classification

Throughout this document, the following convention shall be used to denote the functional requirements:

[Rn]

where n represents the functional requirement number.

2. System Overview

The AutoTab is a system that can be modeled as shown below in Figure 1.

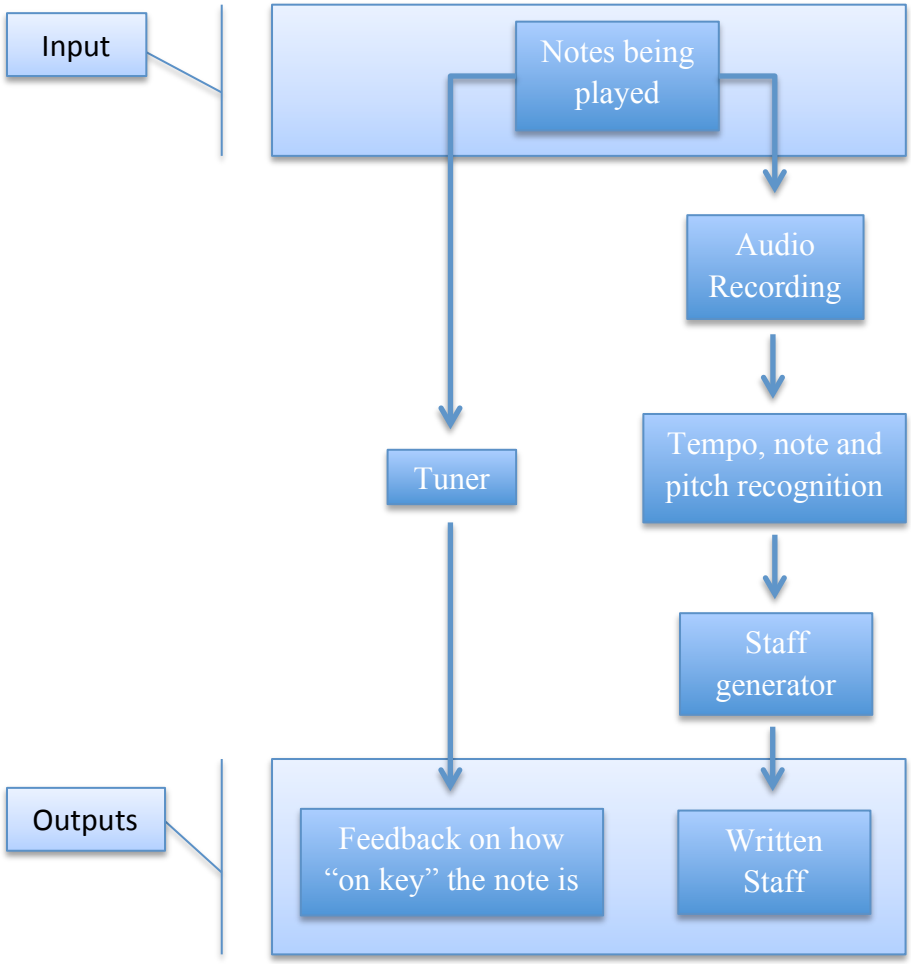

Figure 1: System Overview

We will initially be working on monophonic guitar sounds and given sufficient time and success, we will expand our algorithm to handle more complex situations like polyphonic guitar sounds.

3. Algorithm Deliverables

The algorithm deliverables represents the development of transcription and tuning algorithms in MATLAB, their respective benchmarks, as well as the creation of software to convert the AutoTab's MIDI output to written staff. In using MATLAB for implementation, we can analyze the algorithms for bottlenecks, which will later be delegated to hardware.

3.1 Transcription Requirements

[R1] The device shall convert stored audio files into the MIDI file format.

[R2] The device shall quantize notes up to a resolution of 1/64 of a beat.

[R3] The device shall register a new note at ¥[1] dB above the ambient noise level.

[R4] The device shall register the release of a note once it drops below its initial amplitude by ¥ dB.

[R5] The device shall partition notes according to Western scale with a resolution or bin size of one semitone (i.e. the device shall round a given pitch to the nearest note).

[R6] The device shall identify single notes correctly in both time and pitch for our basic set of instruments[2].

[R7] The device shall correctly identify notes for complex instruments within some tolerance.

[R8] The device shall correctly identify chords for simple and complex instruments within some tolerance.

[R9] The device shall track notes and note lengths for a maximum of ¥ simultaneous events.

[R10] The device shall trim out empty bars at the beginning and end of the recording.

[R11] The device shall analyze one file at a time.

[R12] The device shall provide consistent results.

[R13] The device shall operate smoothly. It should not crash or put undue burden on the computing system.

3.2 Tuner Requirements

[R14] The device shall provide visual feedback relating how close a single note is to the true 'on-key' pitch, to be achieved by showing the disparity in frequency between the pitch registered and the pitch of the nearest note (to some resolution).

[R15] The device shall identify single-note events ¥ dB over the ambient noise level.

[R16] The device shall run continuously, holding the results for the last event for 2 seconds before resetting.

[R17] The device shall operate using live (sampled) audio.

[R18] The device shall be consistently accurate for simple instruments.

[R19] The device shall be accurate to some tolerance for complex instruments.

[R20] The device shall respond and identify a given pitch within ¥ second(s) or less.

---

[1] Denotes a value that will be specified in the future

[2] Both a basic and complex set of instruments will need to be defined. Pertains to the discussion of monophonic vs. polyphonic sounds in the previous section

3.3 Staff Generator Requirements

[R21] The device shall take a MIDI file and convert it to written staff.
[R22] The device shall operate in software on a personal computer.
[R23] The device shall allow note editing to fix errors in transcription.
[R24] The device shall consistently and accurately recognize simple note events[3], based on time signature and tempo.
[R25] The device shall correctly identify complex note events to some tolerance.
[R26] The device shall interpret subjective or complex events consistently.
[R27] The device shall allow the file to be saved as a PDF and printed.
[R28] The device shall allow certain edits to be made such as adding a title or text.
[R29] The device shall take ¥ or less seconds to convert each minute of recorded audio.
[R30] The device shall flag areas of transcription containing potential errors.

4. Proof of Concept Deliverables

In this phase, the proof of concept deliverables will define the different states and how they relate to the hardware on a DE2 board. Regardless of implementation, the device will use an LCD display, LEDs, a switch, and four buttons for its interface. Additionally, a microphone and line-in jack will be used for audio collection, a line-out jack will be used for playback, and a speaker will be used to produce clicks, providing metronome functionality while recording. Finally, the device will contain a micro SD socket to house its memory. Depending on the capabilities of both the display on the DE2 and the one used for the encapsulated prototype, some or all of the LED duties can be absorbed by the LCD display. Ideally, we will use a display capable of custom graphics, allowing us to create icons for each state, a graphical tuner, and other forms of visual feedback.

[3] Simple note events are single/grouped notes. Complex events require a stylistic re-interpretation of the content based on context, and include flans, ties, triplets, etc.

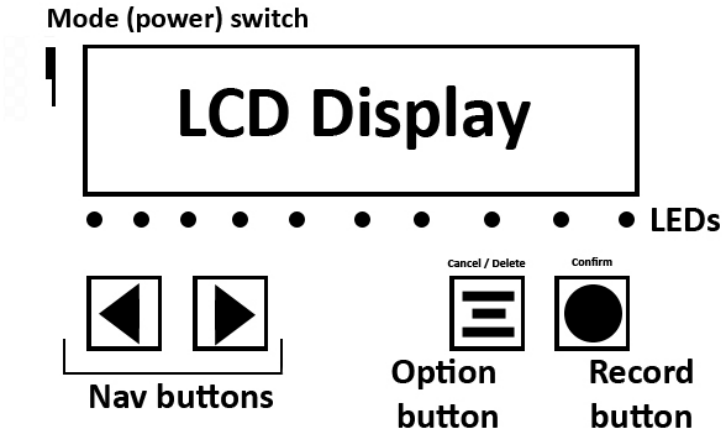

Figure 2: Product Interface

4.1 Mode of Operation

Off
In this mode, the device is powered down. The position of the switch will determine if the device is in tuner mode or recorder mode.

Tuner
In this mode, the device performs real-time, continuous engagement of the tuning algorithm using the line-in or microphone (microphone being disabled when line-in engaged). The audio being analyzed will be passed through to the line-out jack. The red LEDs can be used to signal pitch position, while the green LEDs can signal a tuned note. The LCD display will indicate both that the device is in tuner mode and the note currently being compared. The navigation (Nav) and record buttons will not have any function, however the option button will cause the tuner to enter the Options state.

Recorder
In recording mode, the device collects audio then transcribes it to MIDI. When the recording state is idle, the display indicates that the device is ready. Pressing the Option button will engage the Options state, and pressing either Nav button will cause the device to enter the tempo-editing dialog directly.

Recording is engaged by pressing the record button at which point the device will count-in a bar then start collecting audio from the microphone or line-in. During recording, metronome functionality is provided by speaker clicks and LED flashes, with some form of accent to indicate the musician's position in a given bar. Both the metronome clicks and the incoming audio are passed through the device to the line-out. In addition, the display will change to indicate that recording is in progress, and will specify both the length of the current recording and the remaining time available as a function of free memory on the SD card. Both the Nav and option

buttons do not function during recording; pressing the record button again terminates the session.

Once recording is complete, the AutoTab will transcribe the WAV file to MIDI - the display will indicate that the device is processing, and ideally will be able to provide some form of progress bar or estimation as to the remaining time for this process.

Options
The following options will be made available in a menu:

a. Tempo: the user can adjust the recording tempo in BPM
b. Metronome: the user can toggle the metronome between on, silent (LED only), and off
c. Browse: the user can play back or delete sessions stored on the SD drive

For all menu functions, the Nav buttons will be used for navigation, the option button will exit the menu, and the record button will confirm a selection. In the browser, pressing the option button will return the user to the options menu, holding the option button will delete the selected session (pending confirmation), and the record button will initiate playback of the selected session. During playback, the record button will pause play, the Nav buttons will fast forward/rewind, and the option button will exit.

4.2 Proof of Concept Requirements

[R31] The device shall emulate the end-device interface as much as possible: use the same LEDs, display, and buttons for all functionalities.
[R32] The device shall operate off the DE2 power supply.
[R33] The device shall not use an excess number of controls: four buttons for operation, a switch to determine the mode, and a display/LED bank to provide feedback.
[R34] The device shall have all functionalities available within three command/button presses (excluding mode selection).
[R35] The device shall by default, be primed to start recording or tuning immediately once engaged.
[R36] The device shall retain its settings between uses.
[R37] The device shall have essential options (tempo) available within a single button press.
[R38] The device shall indicate the remaining time available for recording, taking into account overhead for MIDI generation etc.
[R39] The device shall save both a WAV and a MIDI file in a unique identifiable folder after recording (e.g. 'Session 1').
[R40] The device shall indicate an error if no SD card is inserted when in recording mode.
[R41] The device shall successfully terminate recording and generate a MIDI file if the SD card runs out of space while recording.
[R42] The device shall finalize or delete corrupted files due to power failure, SD card ejection, etc. while recording.

[R43] The device shall process sessions after recording is complete. It should indicate both processing state and remaining time if possible (using a status bar). The processing time must not exceed ¥ seconds per minute of audio.
[R44] The device shall be compatible with ¥ (manufacturer) SD cards.
[R45] The device shall not allow clicks generated from the metronome to interfere with the recording.

5. Encapsulated Prototype Deliverables

In this last phase, we describe the deliverables pertaining to the physical product held by consumers.

5.1 Standards

[R46] The device shall comply with CSA Standard CAN/CSA – C22.2 No. 60065-03. [1]
[R47] The device shall comply with ISO Standard ISO 16:1975. [2]
[R48] The device shall be in compliance with FCC Part 15 Rules. [3]
[R49] The device shall be RoHS compliant. [4]

5.2 General Requirements

[R50] The retail price of the device shall be under CDN$40.
[R51] The weight of the final product shall be less than 1 kilogram.
[R52] The device shall not exceed 10cm x 5cm x 5cm.
[R53] The device shall have all design components separable and modular.

5.3 Physical Requirements

[R54] The device shall operate reliably within a temperature range of 0^{o}C to 50^{o}C.
[R55] The device shall operate reliably within a humidity range of 0% to 90%.
[R56] The device shall operate reliably within an elevation range from sea level up to medium altitudes (5,000 – 10,000 feet).
[R57] The internal components shall be protected and insulated.
[R58] The enclosure shall be sturdy.
[R59] The display unit screen shall be visible under normal operation.
[R60] The device shall be able to withstand at least a 5 ft. drop and a compressive force of at least 200 lbs.

5.4 Connectivity

[R61] The device shall have USB connectivity for file transfer and renaming from the device to the computer.
[R62] The device shall have two ¼ in. mono jacks for input and output and come with adapters for 3.5mm jacks.

5.5 Electrical Requirements

[R63] The device shall be powered by standard alkaline 9V batteries that have an average capacity of 565 mAh.
[R64] When powered by batteries, the device shall be able to operate for at least 6 hours in recording mode.
[R65] When powered by batteries, the device shall be able to operate for at least 30 hours in tuner mode.

5.6 Environmental Considerations

[R66] The device shall not interfere with surrounding electronic devices.
[R67] The device shall not contain any of the six restricted substances listed by the RoHS: lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls, and polybrominated diphenyl ether. [4]
[R68] The components of the device shall be independently serviceable to minimize unnecessary waste and minimize warranty costs
[R69] The device shall be mostly recyclable and biodegradable at the end of its lifecycle.

5.7 Reliability and Durability

[R70] The life of the push buttons shall be at least 100,000 cycles.
[R71] The device shall be able to withstand the forces outlined in section 5.3 Physical Requirements.
[R72] The enclosure shall be scratch resistant.
[R73] The mean time between failure for the device shall be at least 2 years or 17532 hours.

5.8 Safety Requirements

[R74] The device shall not harm the user in the event of hardware failure.
[R75] The electric components of the device shall be enclosed and isolated from the user with clear safety markings where applicable.
[R76] The enclosure shall not be a physical hazard to the user.
[R77] The enclosure shall not become a hazard in the event of it being broken.

[R78] The battery compartment shall have clear markings indicating polarity and orientation.

5.9 Usability Requirements

[R79] The operation buttons shall be placed apart from each other to minimize accidental button presses.
[R80] The orientation of the buttons shall be intuitive for the user to use.
[R81] The power button shall be placed away from operation buttons.
[R82] The power button shall be recessed to prevent accidental triggering.
[R83] The device shall be easy to switch between modes (Off/Tuner/Recorder).
[R84] The device shall be limited to one button press at a time.
[R85] The device shall ignore button presses that have been held for longer than two seconds.
[R86] The batteries shall be easily accessible for replacement.
[R87] A volume control shall be used for controlling the output volume unless operating as a pass-through for the audio signal.
[R88] The device controls shall provide good tactile feedback to the user to indicate button presses.

5.10 User Documentation

[R89] The user manual shall be written for an audience with minimal knowledge of electrical devices and music.
[R90] The user documentation shall be provided in French, Spanish, German, Traditional Chinese, Simplified Chinese and Japanese to satisfy product language requirements for international markets.
[R91] A detailed service guide for technicians and vendors shall be created.

6. System Test Plan

Due to the modular nature of the design of the AutoTab, a bottom-up approach will be used for testing the product. The testing process will be divided into three separate stages: algorithm tests, proof of concept tests, and production tests.

6.1 Algorithm Tests

In this phase of the project, the algorithms used for transcribing music and note detection will be settled upon. Different algorithms will be simulated in MATLAB and the one which satisfies the performance requirements will be chosen.

The evaluation process through which this will be done involves importing WAV recordings of musical guitar scales to see which algorithm correctly identifies the notes of each scale. Through this process, each algorithm can be analyzed for their strengths and weaknesses and the chosen algorithm can then have its bottlenecks implemented as hardware for the proof of concept phase.

Any modules pertaining to the chosen algorithm will be tested to ensure it is functional prior to integration and moving forward with implementation on the NIOS processor and the Cyclone II FPGA.

6.2 Proof of Concept Tests

At this stage of development, the device will be assembled with the DE2 development board, buttons, LCD display, and SD card reader. Here much of the user interface for the device will be tested. The device will also be evaluated for general system stability and how it handles the situations outlined in the proof of concept requirements.

6.2.1 Focus Group Tests

Prior to moving the development of the project forward for production, the device will be assembled with the DE2 development board and a ribbon cable running to an enclosure the approximate size of the finished product. This will then be tested with musicians in order to further refine the usability design and features of the device using feedback received from potential future users.

Typical Usage Scenario

1. The user places the device on a surface, such as a table, and plug in the audio source (guitar) into the line-in.
2. The device is activated and is switched to the tuner mode.

    a. The user strums a string and the device gives feedback whether the string is in tune.
    b. The user adjusts the tuning of the string, strums it again, receives feedback from the device, and readjusts as necessary.
3. The device is switched to the transcription mode.
    a. The user presses the record button to start a recording session.
    b. The user plays music and presses the record button to stop the recording.
    c. The user can then use the option and Nav buttons to save, delete, or name the finished recording.
    d. The user can use the Nav and option buttons to start the transcription. The device gives feedback as to when the transcription has finished.
4. The user takes the device and plugs in the USB cable to a computer to transfer the recording and transcription files.
5. When finished, the user turns of the device.

6.3 Production Tests

After implementing the suggestions received from focus group testing the device will be constructed identically to the final end product that will then be run through a battery of tests for safety, reliability, durability, and usability.

7. Conclusion

The functional specification clearly defines the capabilities and requirements of the AutoTab. Development of the final production device will take place in three distinct phases. The proof of concept model is well underway and expected to meet the functional requirements outlined above by the target completion date of April 9th 2012.

References

[1] CAN/CSA-C22.2 No. 60065A-03, Audio, Video and Similar Electronic Apparatus - Safety Requirements.

[2] ISO 16:1975, Acoustics -- Standard tuning frequency (Standard musical pitch).

[3] FCC, Part 15 Title 47 Telecommunications.

[4] "Department for Business Innovation & Skills," 2010. [Online]. Available: http://www.bis.gov.uk/assets/bispartners/nmo/docs/rohs/support-literature/producer-support-booklet.pdf. [Accessed 05 02 2012].