Automated study assistance for musical instruments

Marc Went

MSc Information Sciences student
UVANetID: 10905847

VUNetID: mwt680

Table of Contents

1. Abstract

Research on automatic support for exercising on music instrument. This research combines not only the guidance and evaluation of student's progress, but also included the optimization of such a system for use in the music schools. Using the new technologies like voice and face recognition, the system could become a very efficient tool not only for individuals but also for large music schools or online classes.

This paper will describe several techniques in more details. These techniques are already developed, but have not been combined into one system yet.

The results of the research show that it is possible to design a system capable of teaching students how to play a musical instrument.

By addressing the struggles the student experiences, this paper is able to address several problems subsequently improving the study environment.

2. Introduction

Musical instruments exist for a very long time, people who decide to play an instrument first have to master the technique before they can perform. It is well known that playing instrument has long-time positive influence on the brain.

The positive impact of playing an instrument should be recognized and more supported. In reality many people wish to play an instrument but do not have to time or the patience to exercise, let alone be taught how to play this instrument. The most played instruments are currently the piano, drums and guitar. (The Most Popular Musical Instruments, n.d.)

This paper describes a system that will aid a student who is trying to learn how to play an instrument. The application uses visual feedback to report how the student is progressing with the training and allows him / her to study whenever it is convenient. The student is no longer bound to the schedule of a teacher, who is till today the only possibility for student to get feedback and improve the quality of the training.

To help both the student and the teacher we envisioned a system that can help at any time to improve the student's playing skills. The system will give feedback to students on their performance. This is based not only on the correction of the mistakes, but the system adapts to the learning speed of said student and offers a comfort of right exercise support.

The application receives a performance as an input, which can be either digital or analog. Then the audio input is analyzed by the use of several parameters, like tone duration, tone height, tone frequency. The analysis serves as base to check whether or not the student is playing the thought lesson properly or not.

When the system is finished analyzing the performance an evaluation will be made based on how many mistakes the student made a certain score is calculated and shown to the student. With this score, student can monitor his improvements and progress over time. Above that the system will structure the next steps in the exercising of the student, create a feedback loop on how they should improve and order the exercises.

This system will start off with every student as if they have no previous experiences with the instrument. When a student shows a rapid procedure through the exercises the system will skip to a harder part, this will keep the student challenged.

The feedback is given to the student as visual feedback, this to ensure the student can learn better, auditory feedback is lost after being played, whereas the visual feedback can stay on the screen as long as necessary.

Instrumental lessons are commonly given by a teacher usually for an hour and it has to be scheduled to suite the teacher and the student. They both need to be in possession of the instrument being thought. And travel time for the student to the teacher is also crucial to the motivation of student. This system will try to minimize all these inconveniences as much as possible and will improve the training in the private time of the student.

This system tries to help the student with not only melody but also the recognition of notes as well as the keys to play them on. Therefore it will not become a replacement of the current student-teacher model but an addition to the home study of the student. The feedback the system creates should try to be as good as the feedback given by the teacher. Because of the ability to replay a piece of music at any speed as often as necessary. The student will learn by repetitive training at home, the teacher then can proceed to the next level of training with the student.

3. Related Work

3.1 Audio sensors

The main aim of the system is to recognize one or more musical notes at the same time, and compare the result with the expectation. This will allow the system to observe the student and learn where the biggest struggle lies.

(Lefèvre & Vincent, 2011) shows that when an audio track is segmented into multiple pieces, a learning algorithm can evaluate with very high precision. The audio segment was classified with the K-Mean algorithm and multidimensional Hidden Markov Models.

This Hidden Markov Model allows the system to observe the audio segments, it is proven that these Hidden Markov Models are really powerful in analysis. Multidimensional Hidden Markov Model is an extension to the single dimensional, it is dedicated to multi component data.

This gives a very proper indication that the goal of analyzing an audio track that contains instrumental music, is feasible. To be able to implement at least part of the above mentioned system into this system, will improve the quality of the audio analysis a lot.

To be able to expand the system, and aiding not only pianists but also many other musicians including guitarists etc., the system has to be able to recognize the instrument currently played. Even though (Kubera, Wieczorkowska, Rás, & Skrzypiec, 2010) shows that recognition of individual instruments in a multi instrument track is possible. The precision is not nearly high enough to let it work in a live band. This will suffice for the system because in combination with (Lefèvre & Vincent, 2011) a model can be build where not only analyzes an audio track but also associates the track to a specific instrument.

(Broenink, 2011) describes in his paper how he tested note picking. He tested three algorithms, Gaussian Peak Picking algorithm, McLeod algorithm and Tartini.

These algorithms each have their specialty, Tartini is an algorithm as well as a program. It calculates the fundamental frequency that is being played by recording sound through a microphone.

The Gaussian Peak Picking algorithm samples the recorded audio into buffers, on these buffers a Gaussian function is applied after which a transformation done. Finally the fundamental frequency is identified.

The McLeod Pitch Detection algorithm doesn’t work with transformations like the Gaussian Peak Picking algorithm but by modifying the square difference method. After this a peak is chosen that is the fundamental frequency.

All of these algorithms had trouble to register lower octave notes, but in the upper bound the note picking was done well enough by Gaussian Peak Picking Algorithm. This system could use the data found in this paper to combine de Gaussian and the McLeod algorithm to have the highest probability of picking the correct note.

These four papers show that it is possible to create a system that detects and classifies tones generated by musical instruments. The auditory sensors are important to the system to be able to analyze and classify the students play. Yet latency is a big problem, since audio processing happens in segments.

There are currently extensive methods of voice recognition, these methods can distinguish many languages as well as dialects. A proper example of voice recognition is Google Voice Search[1]. This system is deployed since 2011 and allows the user to search Google through speaking their query (Schalkwyk, et al., 2010). This can be very useful because the Voice Recognition algorithm does not only checks for sentences but needs to convert audio waves into manageable data. Which is useful for this system to convert audio signals to data in real time.

3.2 Vision

Every student learns at a different speed and difficulty. To be able to aid each student individually a profile is necessary. This would also allow multiple students to use a single system, which will aid the teacher as well. The teacher can switch between students faster, and allow a more personalized lesson.

The main task will be to recognize the face of the student. Facial recognition is already widely used in many systems, like border security and even Facebook (Pinto, Stone, Zickler, & Cox, 2011). By using facial recognition the system can learn how to interact with the student, there would be no more hassle with student accounts, logins and password. This will allow the student to start practicing faster than before.

The application of facial expression is an important aspect of our project to research. This expression will indicate the current emotional state the student is in. When a student is playing a musical piece properly, he will be content or even happy, but when the student makes mistake after mistake, he will become frustrated. The system will have to adapt to the emotion of the student.

Facial expression detection is more resource intensive so this recognition is not possible as frequently as the audio recognition. Overall this method can achieve proper results according to (Valstar, Jiang, Mehu, Pantic, & Scherer, 2011). Although it does not result in perfect score, the system can improve when proper calibration is applied. Facial expression detection is possible not the best idea to do in real time on the user's device, but rather in the cloud. In the cloud a database can be kept, which improves facial expression detection because of more calibration options.

3.3 Ambient Intelligence

The system will also be sensitive and responsive to the presence of people. It will learn where and when a student is playing the instrument and what level, this allows for a more detailed student profile. Which results in better practice sessions.

A student playing at night may play mellow music rather than fast rhythmic music. When the system learns from the patterns the student has, the adaptation will become more accurate. This will hopefully help the student to play with more passion than before.

(McKenna, Chauncey, Arnone, Kaarst-Brown, & McKnight, 2014) looked into the possibility of "*learning through playful interactions*" The paper examines next generation of technologies in combination with serious games. By not viewing studying as studying but as playing a game, the participants were more playful and showed improvements in their skill.

3.4 Intelligent Visualization

Recognition and analysis sets a base for the student, yet the student cannot study if he does not know what to look at. Visualization is for this reason very important.

When a student is improving their skills a more positive vibe can be given, this to sustain the current mood and motivation of the student. A frustrated student should lose this frustration as soon as possible, so the system should adapt and show supportive tips and tricks. A student will hopefully get back on track and continue with playing.

[1] http://www.google.com/insidesearch/features/voicesearch/

The system should also be able to detect when a student is struggling, then the view will change to assist the student to find the proper chord or note to strike. A comprehensive example for the piano is the project: P.I.A.N.O. (Weing, et al., 2013), it visualizes the notes that the student has to play on the keyboard. This gives immediate feedback to the student.

4. Interaction Design

4.1 Design Description

4.1.1 Target group

The system envisioned has a target audience of children between the ages of 6 to 12. Yet any aged student can use the system due to the ease of use.

Since the target group is not able to read fluently, the system must be easy to understand. Therefore key features in the system are often simplified to buttons with images, or simple words.

4.1.2 Interaction

The key interaction with the system is using a tablet, either Android or iOS. This is because tables have a screen large enough to display all the necessary information as well as a better processor that their mobile phone equivalent.

The student will use this tablet as any kind of note sheet, but without the necessity to turn over the paper. It has to be put on the piano or keyboard.

4.1.3 Scenario

"Johnny (12) has been playing wordfeud with his friends but was not really good at it. Since his 9th birthday he started playing piano, yet he always had difficulty with reading the music notes. Several weeks ago he had downloaded the application "Musiq-Buddy". After the welcome page, Johnny is able to select a song he wants to practice. Today, unlike yesterday, Johnny does not want to study hard and decides to select a fairly simple song to play.

While practicing he sees a notification popup, it says his crush from class has challenged him to play a game. Here he can put his practice to the test, by playing against his crush. She has selected a song he is not very familiar with, but he is willing to give it a shot. As expected he lost, but it does not matter since he gained some experience, but more importantly, his crush has won and he will be able to have a conversation starter.

After this game he wants to play another game but no one is online anymore to play a game, so he starts a game against a stranger, this time a song he knows well, this time winning the game.

After an hour of practicing and challenging other people, it is time for Johnny to start doing his homework.

4.2 Lo-Fi Prototype

4.2.1 Human-Computer Interaction

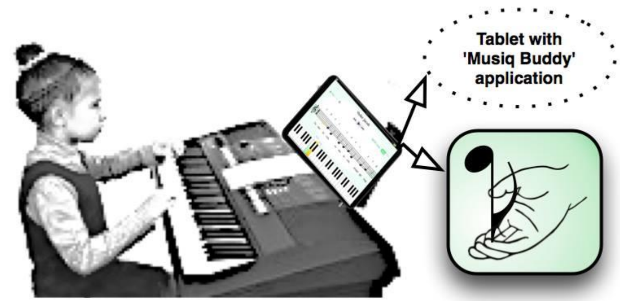


Figure 1: Human-System interaction

While the user is playing the piano, a tablet is placed, where normally the notebook goes. Here the student has an overview of the notes whilst still being able to play on the piano.

(Appendix A) Shows the UI Flowchart. It shows that the main interaction of the user in the Song Library is, here the user can select the desired song, as well as settings. In the desired song, the user can select multiple options to play the song.

4.2.2 Welcome

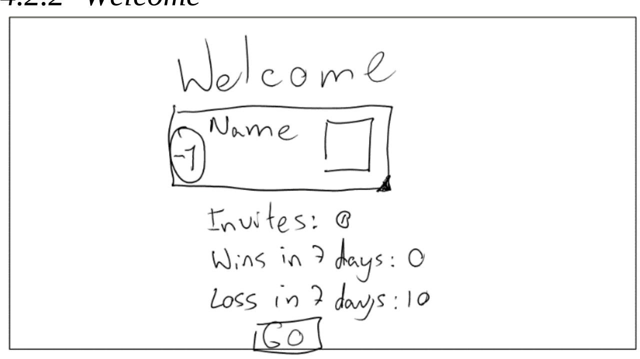


Figure 2: The welcome screen

Figure 2 is the opening page, here the user can select their desired profile. The last user will be selected by default. With this profile the user also get to see some primary statistics of his game(s), whether or not he has any game invites, etc.

4.2.3 Home

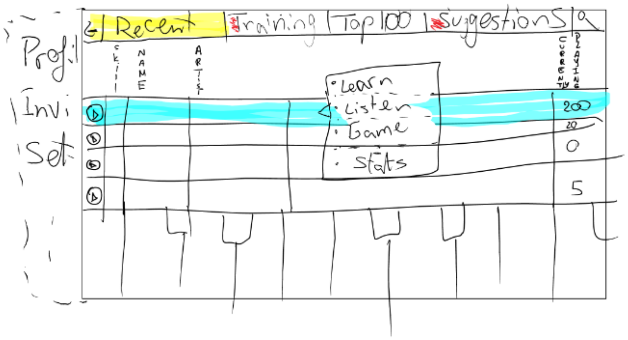


Figure 3: User based song list, after login

When the user profile is selected, a list of songs is shown. Here the user can select a song they want to play. There are 4 tabs that allow the user to select either a recently played song, a song they are currently training on, a top 100 most played songs and a suggestion of songs the user might like based on previous selections.

To the left of these tabs, there is a pop-up list where the user can select the profile window, invitations to play a game, settings etc. To the right of the tabs is a search button so the user can easily search a song they want to try out.

The list of songs consists of a preview of the song, the assumed skill needed to play the song, the song name, artist, genre etc. and finally it shows how many players are currently training or playing this song. This would allow the user to choose a popular song and to invite a random person to play a game.

When the user has chosen the song he wants to play, he long-presses the song. This shows a menu of options.

4.2.4 Listen

If the users chooses listen, the window showing shown in Figure 4 is shown.

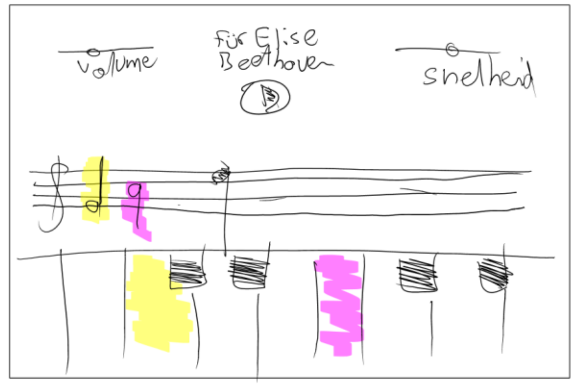


Figure 4: Listen screen where a song is shown

Here the song name is shown, with the controls to play/pause the song, increase or decrease the volume as well as in and decreasing the playback speed. Below the controls the music notes are shown, where the current note or chord is marked in color A and the next part is marked with color B. This allows the user to easily map the notes to the keyboard. When a song features not just the right hand but also the left hand, a second note line will appear to show the notes or even chords for the left hand.

The keyboard at the bottom of the screen is marked with the same colors to indicate how the chord or note is positioned on the keyboard.

4.2.5 Learn

The learning screen layout and functionality is based upon the listening screen with the playback speed control replaced with a button. This allows the user to set whether or not the system has to wait until the user plays the note correctly. If this is disabled the player will continue with the next note regardless the note is played correctly or not.

Feedback will be given that the note is played wrong, but the system continues with the next note.

When this button is enabled the system will wait until the note is played correctly. This way the user know when he played a note incorrectly and which note to strike properly.

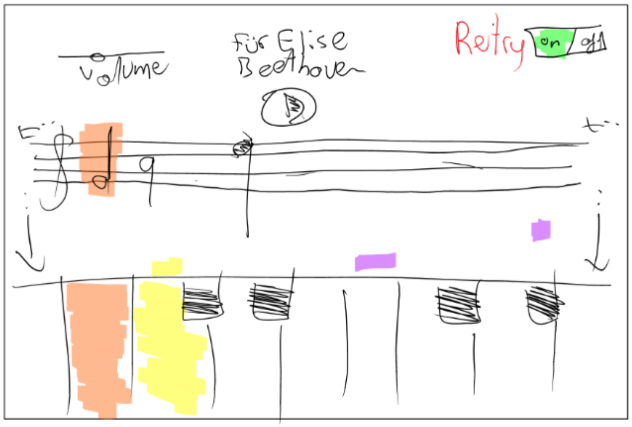


Figure 5: Play screen, where the user gets feedback

When the user finally plays the correct note, positive feedback will be given. In form of the note being highlighted in green.

4.2.6 Game

Since a song is already selected, the user will see a screen where he can choose between single and multi-player. When single-player is selected a similar screen to the Listen screen will show. Here the user can test his skills against himself, by getting as much points as possible.

The multi-player screen has the possibility to select either a friend or a random person. They will each play a game to get the highest score. This is not a real-time game, since another user might be sleeping at the other end of the world.

When the user invites someone, this person will receive an invitation. This invitation will pop-up as a notification to the user, as well as reminders when this notification is dismissed.

5. System Design

5.1 Hardware

The system is designed based on several hardware specifications. The main hardware requirement is, as mentioned earlier in the paper, a piano or a keyboard to work with the system. These are the first instruments supported by our system. Future releases can extend support to other used instruments as well.

Next necessity for the system to work is a tablet. It allows the student to replace the usual note sheet, since the tablet has a screen that is large enough to show all the fundamental information. A mobile phone is discouraged because the screen of these devices are a lot smaller and would not be able to show all the necessary information in a comfortable way. A notebook on the other hand would be too big because of the keyboard and touchpad. Therefore the placement of this device would be awkward if possible at all.

The system can link the piano and tabled by recording the played music and converting it. A keyboard has two possibilities, either record audio like the piano, or by using the MIDI connection most keyboards have on the back. This MIDI connection is not built in any tablet device and therefore would require a MIDI to Micro-USB or Lightning connector, depending on the tablet used.

5.2 Software

5.2.1 Operating system

As of now the biggest two operating systems for tablets are Android and iOS. If this system were to be implemented at this moment, two possibilities would be considered. The first option would be Android, it is based and JAVA, which is one of the most used programming languages at the time. The second option would be iOS which is a very popular operating system as well. The biggest drawback is that it requires Apple computer to be able to write code for it. Yet android operating system has hardware level driver support for MIDI (MIDI Manufacturers Association, 2014). Therefore a custom built operating system is another option, yet not as viable as already existing operating systems. This system would be based on the most popular system at the time, and extend it to support MIDI interfaces. Currently many custom Android distributions are available, making a stock Android with hardware based MIDI support is an option.

The advantage of a custom operating system is that there is low level access to the hardware which subsequently will result in a more efficient driver to run MIDI on, resulting in low latency and support for connecting MIDI interfaces through USB for example.

Yet building a custom operating system has disadvantages as well, mainly the building time of this system can be too long to be feasible. An even bigger disadvantage would be that the user would have to have extensive knowledge on how to install a custom operating system on their tablet.

5.2.2 Algorithms

There are several algorithms to choose from to calculate notes played on an analog device. Several algorithms have been discussed earlier in the paper (Broenink, 2011).

Gaussian Peak picking algorithm takes in a buffer of a certain size, it applies Gaussian smoothing to smooth the outer sides of this buffer – making them less relevant. Then Fast Fourier Transformation is applied which transforms the smoothed data into frequencies, from which the key frequency can be extrapolated.

McLeod Pitch Detection method does not use the Fourier Transformation but combines autocorrelation and the square difference method to calculate the frequency.

After normalization several peaks are found, to proceed, the peak with the highest amplitude are chosen. This peak is multiplied with a value, between 0.8 and 1.0. This results in the fundamental frequency, since the piano has intense overtones. To filter these out, the variable of 0.8.

Another interesting algorithm described by (de Cheveigne & Kawahara, 2002). YIN is an algorithm that calculated the fundamental frequency with very low latency. Another advantage of this algorithm is that, where most other algorithms need an upper and lower bound for their search boundaries, YIN only requires a lower bound parameter.

To analyze MIDI input, the instrument data must be segmented into parts after which it can be analyzed using Hidden Markov Models. This is useful because, as described in (Raphael, 1999), the instrument data can be segmented into parts where the frequency changes. This model is useful because it allows to process chords, where multiple notes are played at the same time. This Hidden Markov Model also allows the checking of the note length and accuracy when a chord is played.

To allow song suggestion, machine learning algorithms are the most suitable to use. The change according the user feedback as well influence from other users around the world. The easiest machine learning algorithm to apply here is Bayes Naïve classification. According to a larger dataset, other Musiq-Buddy users, a selection of songs can be made to generate a personalized list.

This system can also detect mood swings and adapt according to machine learning. This allows the system to change the task for the user, allowing the user to feel better after this task. When a user is frustrated, an easier part will be trained, this allows the user to gain confidence and improve faster. Whereas when the user is already feeling confident more rapid training is possible.

To determine the level of the student, the system will look at the score the student gets while playing a song. This algorithm will look at how much overlap the played song has with the original. This overlap contains the note length, accuracy as well as tempo. When a certain threshold – properly played notes – is reached several times in a row, the student will proceed to the next level. The exact algorithm to compare the played song and the original is not yet selected for this system, the expectation is that in the upcoming years multiple algorithms will be created for this purpose.

Finally, a recommendation system will be implemented. This will give the user suggestions on the songs suitable to play. This is dependent on the user his skill level, as well as previously played songs. When the system recognized the user enjoys playing a specific genre, it will suggest more songs of the same, or similar, genre.

There are several systems that use this recommendation system, one can think of Amazon, Google as well the music streaming service Spotify. Each system has a specific recommendation system, Amazon for example, recommends buyers, as well as future buyers what others have purchased as well. They will also inform the user by mail if the system has more suggestions. Google personalizes the search results, therefore the results of the same query will differ (slightly) based on the user.

Spotify suggest a song based on a playlist, this allows the user to listen to more of the similar songs. If the user dislikes a certain song, the system adapts to this preference. A combination of these three recommendation systems would be an ideal recommendation system for this system. It will recommend songs based on previous search queries, based on previously played songs as well as what other users liked to play.

5.3 Input

There are several ways to interact with the system and give it input. The first kind of input is sound, which can be categorized into two subcategories. The first sound input is by using the MIDI interface. This requires an electric keyboard with an MIDI interface as well as a MIDI to tablet converter, since no tabled supports the MIDI connector a transformer is necessary. This input sends the pressed keys to the system after which they are processed.

The more difficult approach, system-wise, is to use the built in microphone to record the sound waves and covert this into a frequency and pitch, this audio analysis happens in real-time, which means all the calculations must be done in less than 50ms.

Finally another input method is the user interface, whenever the user wants to interact with the system, it will receive the actions as input. This input is processed and proper action is taken. This interaction includes selection of music. In a later version even playing piano on the system itself can be built in, which will require the system to process the touch of the user.

5.4 Output

As described earlier, user interaction is an input to the system, yet is works as output as well, after processing the system will generate proper feedback to loop back to the user. This is either playing a song, or showing feedback to other user actions.

One of those feedback actions is the response to the audio generated by the user while playing the piano. The system will show the user if any mistake was made and if so, which mistake it was. This feedback will help the user study more effectively.

Finally the algorithms described earlier output the digitized frequencies that were recorded earlier while the user was playing.

(Appendix B) shows a diagram explaining the system flow. All the components come together inside the system, where processing of the data happens. Finally the feedback gotten from the input and processed in the system will be shown in the User Interface.

5.5 The Music

The system cannot function without a proper music library, yet license fees with big labels are expensive, a solution to this is to use free to use music. Most of the really old songs, like Mozart and Beethoven are license free. This is because of the copyright laws which state that the term of the copyright is as long as the author lives and can be extend by 50 or 70 years. Other songs like We Wish You a Merry Christmas are also freely available. When the system proves successful, the license fees for popular songs will be able to be covered, like Spotify did with their music library.

With most sheet music printed on paper, digitizing would require OCR technology described by (Bellini, Bruno, & Nesi, 2001). Music notes are segmented into parts after which they are transformed to a digital format. By normalizing this formal, digitally delivered sheet music can also be transformed into this digital format. The biggest issue described in this paper is that with complex sheet music, the system does not work perfectly. Because of this, this sheet music must be digitalized manually.

5.6 User Profile

The system is built in a way that multiple users can use the system on a single device. By storing their personal data with Google or Facebook, the user progress is saved in the cloud. This data would contain the play history as well as received achievements. This not only allows multiple users on a single system, but also single user on multiple systems. This improves the user experience since a user can show off their skill at a remote location without the need to carry everything with them.

There also exist single sign-on for multiple services, but this is not widely used since not everyone want to trust a single service with multiple logins. Also the extensive API's that Google and Facebook offer make it more pleasant for the developer as well as the user.

6. Discussion

During the research several problems arose, some of them were major design discussions, yet others were easier to solve.

One problem which arose was how to design the system properly, this system is designed with a piano or keyboard in mind, yet there are other instruments as well.

Another hurdle the research faced were background noises that are recorded during playing on a piano. To make the audio analysis more manageable, the decision was made to ignore the possibility of background noise. Mainly because the assumption was made that a tablet is used and placed on the piano, making the possible background noise diminishable.

Multiplayer was the next difficulty where some assumptions were made. The first presumption was to not overcomplicate the system, yet without the multiplayer the system would be too simplistic. Next problem faced was how to design the multiplayer to integrate as seamlessly as possible with the single player learning mode. Of the several ideas that came around, this system implements it in a way that allows the user to switch from single player to multiplayer in an instant and play with either friends or random people.

Another problem faced was the fact that popular music requires licensing fees due to copyright and author rights. Therefore the first version of the system only license free songs are used. This prevents the system from being sued and shut down.

7. Conclusion

This paper shows that this system has the ability to help users with their training, the user can play whenever he feels like it. Not only does this system has the ability to help the user but also the teacher, since the assignment given can be trained at home at any time of the day.

The related work discussed in chapter 3 proved to be very helpful, first of all the audio sensors are the most important, there were 4 algorithms that proved to work well, when combined they prove to be even more powerful, but they cannot be executed every time at the same time due to real-time constrictions.

But not every interface needs an algorithms to calculate the notes, MIDI for example already sends audio data to the device. This is of great help because it allows the system to run real time without any hiccups.

Secondly the paper on emotion recognition, it is not possible to combine audio recognition algorithms and emotion recognition in real-time, but emotion recognition could be run in the cloud. Emotions of humans do not change every second. By running it in the cloud the speed of the recognition can be increased as well because of dynamically allocated resources.

Intelligent Visualization showed that that there is a similar system like this paper describes but requires a keyboard and beamer. This paper used the techniques covered in the paper how to show musical notes interactively, but squeezed into one relatively small device.

By designing this system with simplicity in mind, the interface became coherent but not cluttered. Allowing users of any ages to operate this system. The system is based on 3 primary interfaces, first of all the song library, where the user can select the song to play. Second the listen and learn interface, where the song is played and the user can listen to the song. And finally the multiplayer section where the user can challenge a friend or a stranger to play a certain song. This interface improves the skills of the user as well enjoying the training by creating a competition between friends.

This system is able to assist the user, by following the learning speed as well as the emotions of the user. This allows the user to study faster by using the system to adapt, leading to less frustrated user.

8. Future Work

This system can be extended to support not only the piano but other instruments as well. The algorithms described in this paper allow this system to extend relatively easily.

Furthermore, the system can also be extended using another interface. The 3.5mm jack inside all tablets can be used to record audio directly from the piano or keyboard rather than through air waves. This would prevent recordings with background noise, which happens when the systems uses the built in microphone to record the audio.

The next version can also improve on the emotion recognition. The primary version will contain a simplified version to be able to recognize simple emotions. More difficult emotions require more hardware and will definitely have to be calculated in the cloud, since it allows for a more complicated calculations because of dynamically allocated resources.

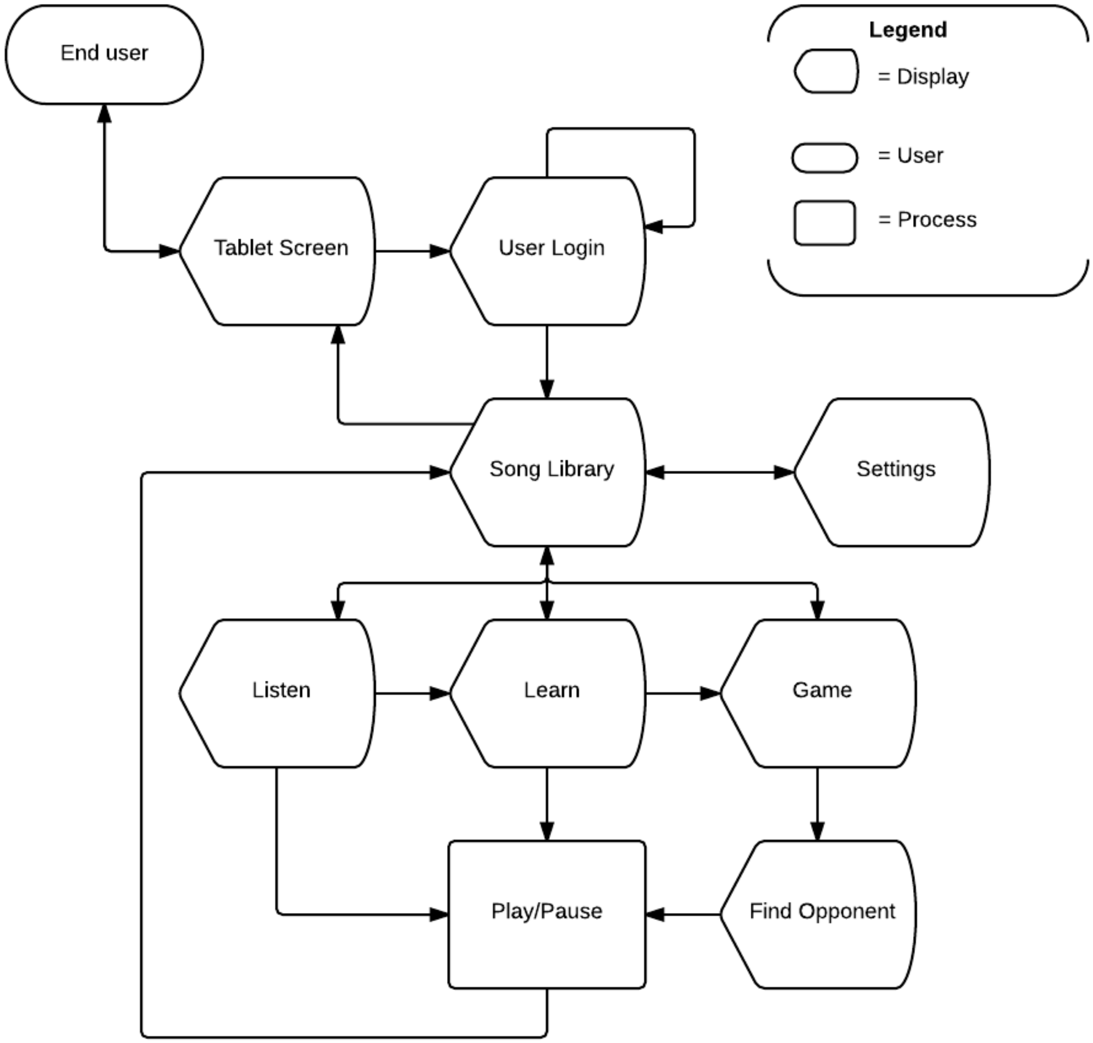
If the system catches on, more songs will necessary to keep the user base happy. More songs mean deals with record labels. A popular model at the time is Spotify, users pay 10 euro a month and wit that money Spotify can pay the licensing fees to record labels. By applying a similar business model deals with record labels should be possible to achieve. Another option is to make music packs available as extra additions, these would charge the user a small amount of money to cover the licensing fees of these songs.

The multiplayer can be extended as well, mainly the scoring system. The first version will compare scoring just using a percentage. The scoring can be spread over more categories than just a percentage. The feedback can also be improved upon, visual feedback could for example show stimulating messages like “Good Job” or “Almost there”. This is not the only improvement feedback can get,

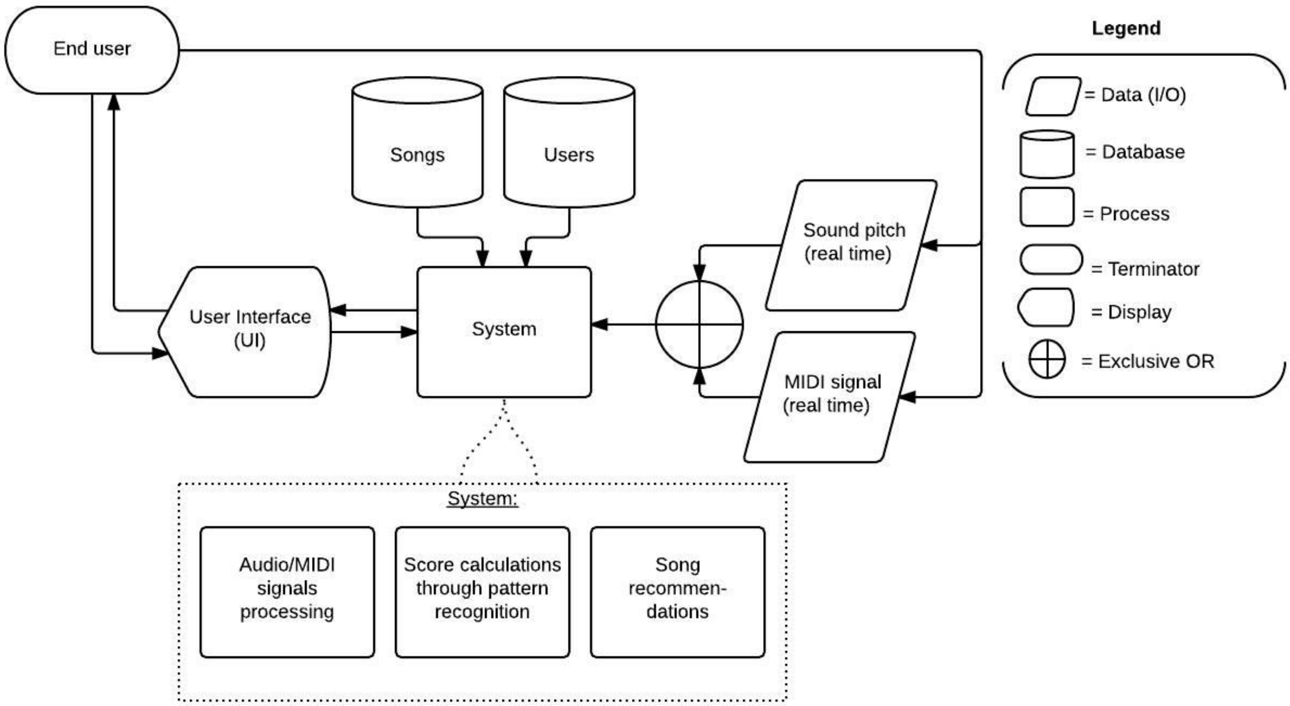
Linking the application to social media will also improve the social aspect of the game, when a user wins a game or attains a good score in his training, this should be able to be shared. By sharing the score not only does the user boast about his skill but also promotes to system to the public. Friends can be invited as well, which will lead to a larger user base.

9. Appendix

9.1 Appendix A



9.2 Appendix B



10. References

Bellini, P., Bruno, I., & Nesi, P. (2001). Optical Music Sheet Segmentation. *Proceedings of the First International Conference on WEB Delivering of Music*.

Broenink, D. (2011, 12 13). *Real time pitch detection*. Retrieved from Dirk Broenink: http://www.dirkbroenink.com/downloads/Dirk%20Broenink%20-%20Dissertation.pdf

de Cheveigne, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Acoustical Society*, 1917-1930. doi:10.1121/1.1458024

Kubera, E., Wieczorkowska, A., Rás, Z., & Skrzypiec, M. (2010). Recognition of Instrument Timbres in Real. *Machine Learning and Knowledge Discovery in Databases*, 97-110.

Lefèvre, S., & Vincent, N. (2011). A Two Level Strategy for Audio Segmentation. *Digital Signal Processing*, 270-277.

McKenna, H., Chauncey, S., Arnone, M., Kaarst-Brown, M., & McKnight, L. (2014). Emergent learning through playful interactions and serious games when combining ambient intelligence with wireless grids. *INTED2014 Proceedings*, 5080-5090.

MIDI Manufacturers Association. (2014). *Android OS MIDI and Audio*. Retrieved from MIDI Manufacturers Association: http://www.midi.org/aboutmidi/android.php

Pinto, N., Stone, Z., Zickler, T., & Cox, D. (2011). Scaling Up Biologically-Inspired Computer Vision: A Case Study in Unconstrained Face Recognition on Facebook. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 35-42.

Raphael, C. (1999). Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models. *IEEE Transactions on pattern analysis and machine intelligence*, 360-370.

Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., . . . Strope, B. (2010). Google Search by Voice: A case study. *Advances in Speech Recognition*, 61-90.

The Most Popular Musical Instruments. (n.d.). Retrieved from Indigoboom: https://indigoboom.com/most-popular-musical-instruments/

Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., & Scherer, K. (2011). The First Facial Expression Recognition and Analysis Challenge. *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 921-926.

Weing, M., Röhlig, A., Rogers, K., Gugenheimer, J., Schaub, F., Könings, B., . . . Weber, M. (2013). P.I.A.N.O.: enhancing instrument learning via interactive projected augmentation. *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 75-78.