Music Education using Augmented Reality with a Head Mounted Display

Jonathan Chow[1] **Haoyang Feng** [1] **Robert Amor**[2] **Burkhard C. Wünsche**[3]

[1] Department of Software Engineering
University of Auckland, New Zealand
Email: jcho205@aucklanduni.ac.nz, hfen020@aucklanduni.ac.nz

[2] Software Engineering Research Group, Department of Computer Science
University of Auckland, New Zealand
Email: trebor@cs.auckland.ac.nz

[3] Graphics Group, Department of Computer Science
University of Auckland, New Zealand
Email: burkhard@cs.auckland.ac.nz

Abstract

Traditional music education places a large emphasis on individual practice. Studies have shown that individual practice is frequently not very productive due to limited feedback and students lacking interest and motivation. In this paper we explore the use of augmented reality to create an immersive experience to improve the efficiency of learning of beginner piano students. The objective is to stimulate development in notation literacy and to create motivation through presenting as a game the task that was perceived as a chore. This is done by identifying successful concepts from existing systems and merging them into a new system designed to be used with a head mounted display. The student is able to visually monitor their practice and have fun while doing so. An informal user study indicates that the system initially puts some pressure on users, but that participants find it helpful and believe that it improves learning.

Keywords: music education, augmented reality, cognitive overlap, human-computer interaction

1 Introduction

Music is an important part of virtually every culture and society. Musical traditions have been taught and passed down through generations. Traditionally, Western culture has placed a large emphasis on music education. For example, the New Zealand Curriculum (New Zealand Ministry of Education 2007) defines music as a "fundamental form of expression" and states that music along with all other forms of art help stimulate creativity.

Traditional music education focuses on individual practice assisted by an instructor. Due to time and financial constraints most students only have one lesson per week (Percival et al. 2007). For beginner students, this lesson usually lasts half an hour, and the majority of time spent with the instrument is without any supervision from an instructor. Sanchez et al. (1990) note that during these unsupervised practice times students may play wrong notes, wrong rhythms, or simply forget the instructor's comments from previous lessons. These issues all hinder the learning process and "provide a source of frustration to both teachers and students". Sanchez also notes that "much of the joy that should accompany the discovery of music dissipates during practice time".

Duckworth (1965) reports that a lack of motivation and interest has been a common problem through history. Problems such as neglecting to teach critical skills such as sight-reading (i.e., playing directly from a written score without prior practice) were already a major concern as early as 1901.

The use of multimedia to enhance practice has been explored previously. Percival et al. (2007) present guidelines for making individual practice more beneficial and note that "by 'wrapping' a boring task ... in the guise of a nostalgic computer game, the task becomes much more fun". The authors give several examples of games that achieve this goal, including some of their own work. They note that due to the subjective nature of the quality of music, computer-aided tools are more suitable for technical exercises where quality and performance can be objectively measured.

Most computer supported music education tools use a traditional display to convey information to the user. Augmented Reality (AR) can be used to create a more direct interaction between the student and the system. Azuma (1997) describes augmented reality as creating an environment in which the user "sees the real world, with virtual objects superimposed upon [it]". The author goes further to explain that "virtual objects display information that the user cannot directly detect with his own sense" and that "the information conveyed by the virtual objects helps a user perform real-world tasks ... a tool to make a task easier for a human to perform".

Azuma (1997) presents an overview of a broad range of disciplines that have used augmented reality, such as medical training, military aircraft navigation and entertainment. The review suggests that AR has been successfully used in a wide variety of educational applications. The main advantage of AR is that a *perceptual* and *cognitive* overlap can be created between a physical object (e.g., instrument) and instructions on how to use it.

A head mounted display can be used to combine real and virtual objects and in order to achieve an immersive experience. Two types of devices exist: optical see-through and video see-through. An optical see-through device allows the user to physically see the real world while projecting semi-transparent virtual objects on the display, while a video see-through device uses cameras to capture an image of the real

Copyright ©2013, Australian Computer Society, Inc. This paper appeared at the 14th Australasian User Interface Conference (AUIC 2013), Adelaide, Australia, January-February 2013. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 139, Ross Smith and Burkhard C. Wünsche, Ed. Reproduction for academic, not-for profit purposes permitted provided this text is included.

world which is processed with virtual objects and the entire image is displayed on an opaque display. An optical see-through device is preferable in real-time applications due to its lower latency and facilitation of a more direct interaction with real world objects.

This work is an attempt to overcome some of the deficiencies in the traditional music education model by using augmented reality to create a perceptual and cognitive overlap between instrument and instructions, and hence improve the end users' learning experience and motivation.

Section 2 reviews previous work using visualisations and VR/AR representations for music education. Section 3 presents a requirement analysis, which is used to motivate the design and implementation of our solution presented in sections 4 and 5, respectively. We summarise the results of an informal user study in section 6 and conclude our research in section 7. Section 8 gives an outlook on future work.

2 Related Work

A review of the literature revealed a number of interesting systems for computer-based music education. Systems for piano teaching include *Piano Tutor* (Dannenberg et al. 1990), *pianoFORTE* (Smoliar et al. 1995), the *AR Piano Tutor* (Barakonyi & Schmalstieg 2005), and *Piano AR* (Huang 2011). Several applications for teaching other instruments have been developed (Cakmakci et al. 2003, Motokawa & Saito 2006). We will review the *Digital Violin Tutor* (Yin et al. 2005) in more detail due to its interesting use of VR and visualisation concepts for creating a cognitive overlap between hand/finger motions and the resulting notes.

The *Piano Tutor* was developed by Dannenberg et al. (1990) in collaboration with two music teachers. The application uses a standard MIDI interface to connect a piano (electronic or otherwise) to the computer in order to obtain the performance data. MIDI was chosen because it transfers a wealth of performance related information including the velocity at which a note is played (which can be used to gauge dynamics) and even information about how pedals are used. An expert system was developed to provide feedback on the user's performance. Instructions and scores are displayed on a computer screen placed in front of the user. User performance is primarily graded according to accuracy in pitch, timing and dynamics. Instead of presenting any errors directly to the user, the expert system determines the most significant errors and guides the user through mistakes one by one.

Smoliar et al. (1995) developed *pianoFORTE*, which focuses on teaching the interpretation of music rather than the basic skills. The authors note that music "is neither the notes on a printed page nor the motor skills required for the proper technical execution". Rather, because music is an art form, there is an emotional aspect that computers cannot teach or analyse. The system introduces more advanced analysis functionalities, such as the accuracy of articulation and synchronisation of chords. Articulation describes how individual notes are to be played. For example, staccato indicates a note that is separate from neighbouring notes while legato indicates notes that are smoothly transitioned between with no silence between them. Synchronisation refers to whether notes in a chord are played simultaneously and whether notes of equal length are played evenly. These characteristics form the basis of advanced musical performance abilities. In terms of utilised technologies, *pianoFORTE* uses a similar hardware set-up as *Piano Tutor*.

The *AR Piano Tutor* by Barakonyi & Schmalstieg (2005) is based on a "fishtank" AR setup (PC+monitor+webcam), where the physical MIDI keyboard is tracked with the help of a single optical marker. This puts limitations on the permissible size of the keyboard, since for large pianos the user's view might not contain the marker. The application uses a MIDI interface to capture the order and the timing of the piano key presses. The AR interface gives instant visual feedback over the real keyboard, e.g., the note corresponding to a pressed key or wrongly pressed or missed keys. Vice versa, the keys corresponding to a chord can be highlighted before playing the chord, and as such creating a mental connection between sounds and keys.

A more recent system presented by Huang (2011) focuses on improving the hardware set-up of an AR piano teaching system, by employing fast and accurate markerless tracking. The main innovation with regard to the visual interface is use of virtual fingers, represented by simple cylinders, to indicate the hand position and keys to be played.

Because MIDI was created for use with equipment with a rather flexible form of input (such as pianos, synthesisers and computers), a purely analogue instrument such as the violin cannot use MIDI to interface with a computer. The *Digital Violin Tutor* (Yin et al. 2005) contains a "transcriber" module capable of converting the analogue music signal to individual notes. Feedback is generated by comparing the student's transcribed performance to either a score or the teacher's transcribed performance. The software provides an extensive array of visualisations: An animation of the fingerboard shows a student how to position their fingers to produce the desired notes, and a 3D animated character is provided to stimulate interest and motivation in students.

3 Requirements Analysis

An interview with an experienced music teacher (Shacklock 2011) revealed that one of the major difficulties beginner students have is translating a note from the written score to the physical key on the keyboard. Dirkse (2009) notes that this fundamental skill can take months to develop. None of the previously reviewed systems addresses this problem. Furthermore, with the exception of the *Digital Violin Tutor*, none of the reviewed systems addresses the problem of lacking student interest and motivation. This issue is especially relevant for children who are introduced to music education through school curricula or parental desires, rather than by their own desire. Our research focuses hence on these two aspects.

Augmented Reality has been identified as a suitable technology for the above goals, due to its ability to create a perceptual and cognitive overlap between instrument (keys), instructions (notes), and music (sound). The association of visuals with physical keys enables users to rapidly play certain tunes, and hence has the potential to improve the learning experience and increase motivation. In order to design suitable visual representations and learning tasks, more specific design requirements must be obtained.

3.1 Target Audience

Similar to the *Piano Tutor*, we target beginner students, with the goal of teaching notation literacy and basic skills. This is arguably the largest user group, and is likely to benefit most from an affordable and fun-to-use system.

3.2 Instrument choice

From the various available music interfaces, the MIDI interface is most suitable for our research. It provides rich, accurate digital information which can be used directly by the computer, without the signal processing required for analogue input. MIDI is also a well-established industry standard. In order to avoid any analog sound processing we choose a keyboard as instrument. In contrast to the work by Barakonyi & Schmalstieg (2005), we do not put any size restrictions on the keyboard and camera view, i.e., our system should work even if the users sees only part of the keyboard.

3.3 Feedback

The system should provide feedback about basic skills to the user, i.e., notation literacy, pitch, timing and dynamics. The feedback should be displayed in an easily understandable way, such that improvements are immediately visible. This can be achieved by visually indicating the key each note corresponds to. One way to achieve this is by lighting up keys using a superimposed image, as done in the *Augmented Piano Tutor* by Barakonyi & Schmalstieg (2005).

3.4 Motivation and Interest

It is important that the system fosters motivation and interest, as this will increase practice time and hence, most likely, learning outcomes. One popular way to achieve this is by using game concepts. Percival et al. (2007) cite several successful educational games in areas other than music. They note that the game itself does not have to be extremely sophisticated; merely by presenting a seemingly laborious task as a game gives the user extra motivation to persevere. Additional concepts from the gaming field could be adapted, such as virtual “badges” and “trophies” to reward achievements (Kapp 2012).

4 Design

4.1 Physical Setup

Based on the requirements, the physical setup comprises one electronic keyboard, one head mounted display with camera, and one computer for processing. The user wears the head mounted display and sits in front of the keyboard. The keyboard connects to the computer using a MIDI interface. The head mounted display connects to the computer using a USB interface. The head mounted display we use for this project is a Trivisio ARvision-3D HMD1. These are video see-through displays in that the displays are not optically transparent. The video captured by the cameras in front of the device must be projected onto the display to create the augmented reality effect. The keyboard we use for this project is a generic electronic keyboard with MIDI out. Figure 1 illustrates the interactions between these hardware components.

4.2 User Interface

As explained in the requirements analysis, the representation of notes in the system must visually indicate which key each written note corresponds to. We drew inspiration from music and rhythm games and Karaoke videos, where text and music are synchronised using visual cues. In our system each note is represented as a line above the corresponding key, where the length of the line represents the duration of the note. The notes approach the keys in the AR view in a steady tempo. When the note reaches the

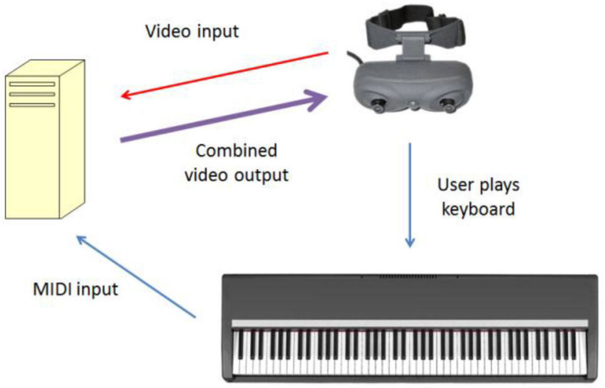

Figure 1: Interactions between physical components of the system.

keyboard, the corresponding key should be pressed. Similarly, when the end of the note reaches the keyboard, the key should be released. This is drawn on a virtual overlay that goes above the keyboard in the augmented reality view as illustrated in Figure 2.

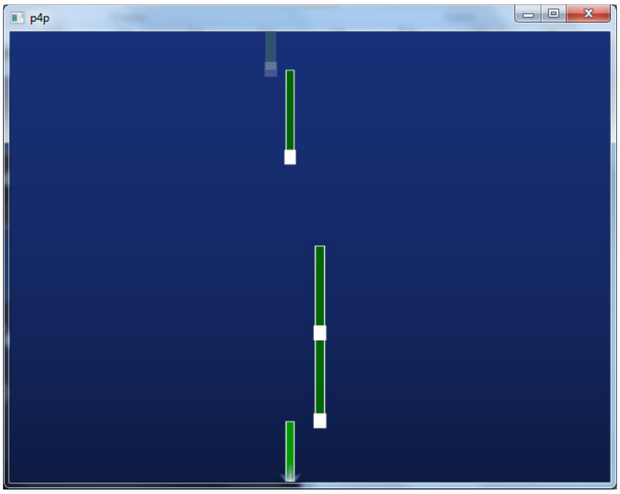

Figure 2: Lines representing virtual notes approaching from the top.

At the same time, the music score is displayed above the approaching notes. In order to help improving notation literacy, a rudimentary score following algorithm follows the notes on the written score as each note is played. The music score and virtual notes are loaded from a stored MIDI file. This MIDI file becomes the reference model for determining the quality of the user’s performance. This means that users must have an electronic version of the piece they want to practice, either by downloading one of the many MIDI music templates available on the Internet, or by recording an instructor playing the piece.

A MIDI file contains timings for each note, which are strictly enforced by our system. This is in direct contrast to *Piano Tutor*, which adjusts the music’s tempo to suit the user. We decided to force timings, since maintaining a steady tempo despite making mistakes is a skill that musicians need (Dirkse 2009). However, the user has the option to manually adjust the tempo of a piece to suit their ability. This makes an unfamiliar piece of music easier to follow, since there is more time to read the notes. Slow practice is a common technique for improving the fluency of a piece of music (Nielsen 2001). This feature encourages the user to follow these time-tested processes towards mastery of a piece of music.

We added a *Note Learning Mode*, which pauses each note as it arrives and waits for the user to play the key before continuing to the next note. This takes away any pressure the user has of reading ahead and preparing for future notes. By allowing the user to step through the notes one by one, the user gets used to the hand and finger motions, slowly building the dexterity required to play the notes at proper speed.

4.3 Augmented Reality Interface

Creating the augmented reality interface requires four steps:

1. Capture an image of what the user can see.
2. Analyse the camera image for objects of interest (Feature detection).
3. Superimpose virtual objects on the image (Registration).
4. Display the composite image to the user.

Steps 2 and 3 are the most complex steps and are explained in more detail.

4.3.1 Feature Detection

The feature detection step can be performed by directly analysing the camera image using computer vision techniques. An alternative solution is to use fiduciary markers and to define features within a coordinate system defined by the markers. Feature detection using markers is easier to implement and usually more stable, but often less precise and requires some user effort for setting up the system (placing markers, calibration). In our application a markerless solution is particularly problematic, since the camera view only shows a section of the keyboard, which makes it impossible to distinguish between keys in different octaves. A unique identification of keys would either require global information (e.g., from the background) or initialisation using a unique position (e.g., the boundary of the keyboard) followed by continuous tracking. We hence chose a marker-based solution based on the ARToolkit software. The software uses markers with a big black border, which can be easily identified in the camera view and hence can be scaled to a sufficiently small size. NyARToolkit is capable of detecting the position and orientation (otherwise known as the pose) of each marker and returns a homogeneous 3D transformation matrix required to translate and rotate an object in 3D space so that it is directly on top of the detected marker. Because this matrix is a standard mathematical notation, it can be used directly in OpenTK.

4.3.2 Registration

A critical component of the AR interface is to place the visualisations of notes accurately over the correct physical keys of the keyboard in the camera view. While detecting the pose of a single marker is simple using the ARToolkit, placing the virtual overlay is more difficult. The first problem comes from the fact that the user is positioned very close to the keyboard when playing, and hence usually only sees a section of the keyboard. Hence multiple markers must be laid out along the length of the keyboard such that no matter where the user looks there will still be markers visible to the camera.

We decided to use identical markers for this purpose, since our experiments showed that detecting differing markers at the same time significantly reduced performance. This was slightly surprising and might have to do with problems with the utilised libraries, the hardware, or the set-up (e.g., insufficient size of the markers).

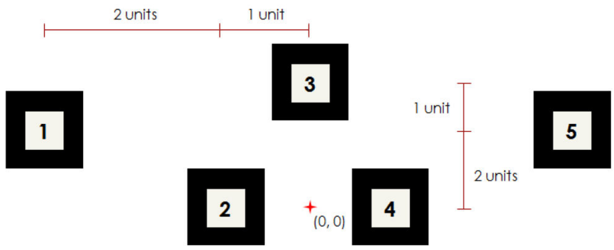

Figure 3: Marker configuration with unique relative distances between every pair of markers.

We overcame this problem by using identical markers and devising a pattern for the markers such that the relative distance between any two markers is unique. Figure 3 illustrates this set-up. If two markers are visible in the camera view, then the distance between them (in units) can be computed from the size of the markers, and be used to identify the pair.

Figure 4 shows an example. Marker 3 is at position $(0, 3)$ and marker 4 is at position $(1, 0)$. If the camera can only see the area within the orange rectangle, the algorithm will calculate the positions of the markers and determine that they are 1 unit apart horizontally and 3 units apart vertically. The only markers in the figure that satisfy this constraint are markers 3 and 4. Since we know that the user is positioned in front of the keyboard, there are no ambiguities due to orientation, and the origin of the marker coordinate system can be computed and used to position the virtual overlay onto the keyboard in the camera view.

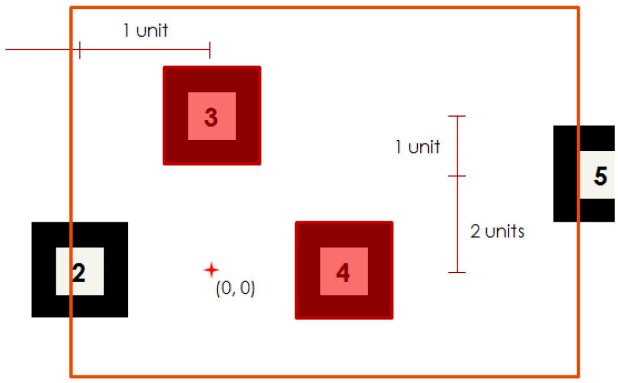

Figure 4: Example of deducing the origin based on a limited camera view.

A further problem encountered for the registration step was jittering and shaking of the overlay, due to noise and numerical errors in the markers' pose calculation. This not only makes it difficult to associate note visualisations with physical keys, but it is also very irritating to the user. We found that this problem was sufficiently reduced by taking a moving average of the transformation matrix for positioning the overlay into the camera view. The optimal kernel size of this moving averages filter depends on the camera quality. A larger kernel size results in a more stable overlay, but reduces response time when the user changes the view direction. Higher quality cameras with higher frame rates achieved stable registration even for small filter kernel sizes.

Figure 5 shows the augmented reality view of the keyboard.

Figure 5: Augmented reality view of virtual notes aligned with physical keys.

4.4 Performance Analysis and Feedback

The MIDI interface is used to obtain the user's performance for analysis. MIDI is an event-based format; each time a key is pressed or released, a digital signal containing information about the way the note was played is sent to the computer. This information includes the note that was played and the velocity at which the note was played. A high velocity indicates a loud sound, while a low velocity indicates a soft sound. The time at which the note was played can be inferred from when the event was received. MIDI also supports information about other keyboard functionalities, such as pedals or a synthesiser's knobs, but this information was outside this project's scope.

The user's performance must be compared against some reference model in order to assess it. Since MIDI is capable of storing such detailed information, we decided to use recorded MIDI files of the music pieces as reference models. This allows evaluating the user's note accuracy and rhythm accuracy. Other information, such as dynamics or articulation, can be added, but as explained previously, were considered too advanced for beginners.

Feedback is important as it allows the user to learn from mistakes and to set goals for future practice. Real-time feedback on note playing accuracy is provided by colour coding the note visualisations in the AR view as illustrated in figure 6. Colour is the most appropriate visual attribute for representing this information, since colours are perceived preattentively (Healey & Enns 2012), colours such as red and green have an intuitive meaning, colours do not use extra screen space (as opposed to size and shape), and colour changes are less distracting than changes of other visual attributes (such as shape).

Figure 6: Colour changes in the note visualisation for real-time performance feedback.

At the end of a performance additional feedback is given summarising how many notes were hit and missed (see figure 7). The summary allows users to monitor improvements, to compare themselves to other students or expected standards, and to set goals for subsequent practices.

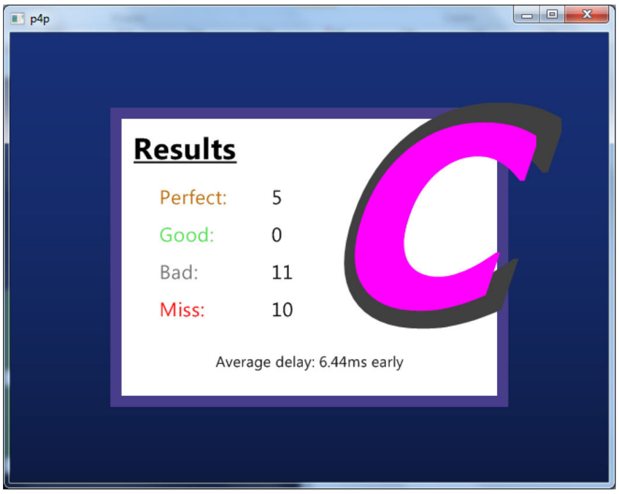

Figure 7: Summary feedback at the end of a user's performance.

5 Implementation

The application was written in C#. Although many graphics-related libraries are written in C++, the advantages of having a rich standard library, garbage collection, and simplified interface design were considered important for rapid prototyping. For capturing images from the camera, we used a .NET wrapper for OpenCV called Emgu CV. For detection and tracking of virtual markers, we used NyARToolkit, a port of ARToolkit. For displaying and drawing graphics, we used OpenTK, a .NET wrapper for OpenGL. For interfacing with the MIDI device, we used midi-dot-net, a .NET wrapper for the Windows API exposing MIDI functionality.

6 Results

6.1 User Study

A preliminary evaluation of the system was performed using an informal user study with seven participants. All users were students with a wide range of piano playing skill levels, ranging from no experience at all to many years of experience. Users were asked to learn a piece using the system. Open-ended questions were asked of each subject about likes and dislikes, how beneficial they believe the system is and an overall rating of the system.

Four participants (57%) liked the representation of the notes in the AR view, while two (29%) criticised that it was difficult to look at the written notation and concentrate on the virtual notes at the same time. Three users (43%) admitted that they did not look at the written notation at all. Six users (86%) said that keeping up with the approaching notes was very intimidating and the pressure from trying to find the following notes caused them to miss even more notes.

The feedback system, especially the summary at the end, was found to be very helpful. The display of quantitative results allowed users to set goals for improvement. In addition, the game-like nature of the system resulted in participants competing with each other on achieving higher summary feedback scores. All participants believed that the system would be helpful for starting to learn playing piano, and all participants enjoyed using the system.

6.2 Discussion

The goal of this research was to design a system for improving piano students' notation literacy and their motivation and interest. The results of the preliminary user study are very encouraging, and both experienced and inexperienced piano players enjoyed the use of the system and did not want to stop using it. Having a competitive element has proved advantageous, and for home use the integration of multiplayer capabilities and online hosting of results should be considered.

The display of music notations proved distracting. Many users just wanted to play the piano based on the indicated keys in the AR view, rather than learning to read notes (notation literacy). A possible explanation is that most participants of the study were not piano students, i.e., had no motivation to learn reading of notes, but were keen to learn playing the instrument. More testing is required to investigate these observations in more detail. We also want to explore alternative AR visualisations and user interfaces, especially for combining written notation with the virtual notes.

The responses about the forced timings creating pressure led to the development of the tempo adjustment feature and the note learning mode described earlier. Both of these modes slow down the rate at which notes have to be played, giving the user much more time to decide what to do next.

Evaluating our application with regards to game psychology uncovers the following shortcomings (Caillois 2001, Hejdenberg 2005):

- **Game width:** a game should address multiple basic human needs such as self-esteem, cognitive needs, self-actualisation, and transcendence (the need to help others). Game width could be improved by giving more feedback during piano practice, such as praise, encouragement, and corrections; having social interactions (practicing in pairs or in a group); and by increasing the level of difficulty (adding time limits, obstacles, random events).
- **Imitation:** a game should enable the player to constantly learn. This could be achieved by ranking music pieces by difficulty and by increasing requirements or using different types of visual hints.
- **Emotional impact:** common ways to achieve an improved emotional impact are visual and sound effects and rewards (high score lists, virtual badges).

7 Conclusion

Our preliminary results indicate that the proposed application is useful to budding musicians. As a game, it breeds interest in music and the learning of an instrument. As an educational system it motivates users to practice and improve. With the exception of improving notation literacy, the requirements have been met. We have demonstrated that real-time augmented reality using head mounted displays is a viable way to convey instrument playing skills to a user. Head mounted displays are becoming increasingly available and affordable to consumers, and proposed devises such as Google's "Project Glass" (Manjoo 2012) demonstrate that such equipment might soon be as common as mobile phones.

8 Future Work

Necessary future developments include performance analysis, such as incorporating dynamics and articulation. This could eventually be integrated into an expert system. A more comprehensive feedback summary would benefit users by narrowing down specific areas for improvement. The score following system can also be improved. Research into techniques for improving the efficiency of learning notation literacy would be beneficial to the music community since this problem has existed for a long time.

A formal user study needs to be performed to determine the usability and effectiveness of the system for piano education. Of particular interest is the effect of wearing AR goggles, and how the effectiveness of the application compares with human piano tutors and computerised teaching tools using traditional displays.

References

Azuma, R. T. (1997), 'A survey of augmented reality', *Presence: Teleoperators and Virtual Environments* **6**(4), 355–385.

Barakonyi, I. & Schmalstieg, D. (2005), Augmented reality agents in the development pipeline of computer entertainment, *in* 'Proceedings of the 4th international conference on Entertainment Computing', ICEC'05, Springer-Verlag, Berlin, Heidelberg, pp. 345–356.

Caillois, R. (2001), *Man, Play and Games*, University of Illinois Press.

Cakmakci, O., Berard, F. & Coutaz, J. (2003), An augmented reality based learning assistant for electric bass guitar, *in* 'In Proc. of the 10th International Conference on Human-Computer Interaction'.

Dannenberg, R. B., Sanchez, M., Joseph, A., Capell, P., Joseph, R. & Saul, R. (1990), 'A computer-based multi-media tutor for beginning piano students', *Journal of New Music Research* **19**, 155–173.

Dirkse, S. (2009), A survey of the development of sight-reading skills in instructional piano methods for average-age beginners and a sample primer-level sight-reading curriculum, Master's thesis, University of South Carolina. 85 pages.

Duckworth, G. (1965), 'Piano education', *Music Educators Journal* **51**(3), 40–43.

Healey, C. & Enns, J. (2012), 'Attention and visual memory in visualization and computer graphics', *IEEE Transactions on Visualization and Computer Graphics* **18**(7), 1170–1188.

Hejdenberg, A. (2005), 'The psychology behind games'. `http://www.gamasutra.com/view/feature/2289/the_psychology_behind_games.php`, Last retrieved 15th September 2012.

Huang, F. (2011), Piano AR: A markerless augmented reality based piano teaching system, *in* 'Proc. of the International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC 2011)', pp. 47 – 52.

Kapp, K. M. (2012), *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*, Pfeiffer.

Manjoo, F. (2012), 'Technology review - you will want google goggles'. http://www.technologyreview.com/review/428212/you-will-want-google-goggles/, Last retrieved 25th August 2012.

Motokawa, Y. & Saito, H. (2006), Support system for guitar playing using augmented reality display, *in* 'Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality', ISMAR '06, IEEE Computer Society, Washington, DC, USA, pp. 243–244.

New Zealand Ministry of Education (2007), 'New zealand curriculum online - the arts'. http://nzcurriculum.tki.org.nz/Curriculum-documents/The-New-Zealand-Curriculum/Learning-areas/The-arts, Last retrieved 25th August 2012.

Nielsen, S. (2001), 'Self-regulating learning strategies in instrumental music practice', *Music Education Research* **3**(2), 155–167.

Percival, G., Wang, Y. & Tzanetakis, G. (2007), Effective use of multimedia for computer-assisted musical instrument tutoring, *in* 'Proceedings of the International Workshop on Educational Multimedia and Multimedia Education', Emme '07, ACM, New York, NY, USA, pp. 67–76.

Sanchez, M., Joseph, A. & Dannenberg, R. (1990), 'The Frank-Ratchye STUDIO for Creative Inquiry - The Piano Tutor'. http://studioforcreativeinquiry.org/projects/sanchez-joseph-dannenberg-the-piano-tutor-1990, Last retrieved 25th August 2012.

Shacklock, K. E. (2011), 'Personal communication'. March 2011.

Smoliar, S. W., Waterworth, J. A. & Kellock, P. R. (1995), pianoFORTE: a system for piano education beyond notation literacy, *in* 'Proceedings of the third ACM international conference on Multimedia', MULTIMEDIA '95, ACM, New York, NY, USA, pp. 457–465.

Yin, J., Wang, Y. & Hsu, D. (2005), Digital violin tutor: an integrated system for beginning violin learners, *in* 'Proceedings of the 13th annual ACM international conference on Multimedia', MULTIMEDIA '05, ACM, New York, NY, USA, pp. 976–985.