

Motion and Audio Capture of Lectures for Use in Digital Game Based Learning Environments

Neil Gannon, Paul Powell, Tony McCabe
Institute of Technology Sligo, Ireland

Abstract

This paper outlines the idea and methodology of using low cost motion and audio capture to present content by an educator during a theory based lecture with the intention of embedding and augmenting the recorded lecture session in a virtual space. Released in 2010, the Microsoft Kinect™ removed the development barrier for software developers and educational professionals interested in developing natural user interface (NUI) and voice user interfaces (VUI) learning applications. Traditional online learning methods replicate current passive learning methods such as theory lectures which are recorded and offered online via streamed video through packages such as Moodle and Adobe Connect™. These more established distance teaching approaches do not cater for new methods for learning or a new game-orientated student cohort, despite the availability of technology which excels at interactivity and requires a user's active attention for successful interactions. Initial demonstrations of the augmented lectures to students illustrated an acceptance of the technology for use in teaching and learning and the augmented lectures were received positively. These augmented lectures could foster a feeling of inclusiveness and social inclusion for the online learner which could be absent from video based alternatives.

1. Introduction

E-learning has arguably become the biggest revolution in education in generations such that in 10 years from 2002 to 2011, the number of US students enrolled in online programs as a percentage of fulltime conventional students has risen from 9 to 32% [1]. The current estimates for annual expenditure growth in e-learning between now and 2017 is 23% [2]. This growth in online learning has unmasked a number of issues and concerns from students undertaking such educational programs.

Providers of education are under immense competitive pressure to provide online distance

education due to the global capture potential and also the relative low cost provision compared to fulltime, conventional teacher-led classroom education which by its nature comprises additional costs by way of significant overheads.

On the other hand, current providers of online education are under pressures of their own, reflective of the global competition in the online education market. Current providers must be cognisant of the needs and requirements of students in an open global market, failure to deliver a high quality, cost effective and innovative online programs and you risk your customer base going elsewhere [3], [4].

One aspect of this is innovative methods for teacher-student interaction and dissemination of the material. Technologies such as Moodle and Adobe connect are common place and they provide platforms for delivery of both content and online live and recorded lectures. These vehicles, and similar packages, are almost the accepted minimum norms nowadays for online delivery and they have expanded to include options such as student discussion groups, they have effectively become the online versions of the lecture rooms.

Other methods used in the online world include wikis, blogs, podcasts and vlogs (blogs with videos). With little more than internet and computer access required, these technologies can be used whereby the web address if provided to the students and they access the sites for information on modules from the tutor/lecturer.

One of the key determinants of a student's success in an online education environment is the feeling of social inclusion. Social inclusion is defined as the student's feeling of being there or not, feelings of not being there can leave a student feeling disconnected, isolated and removed from their education [5], [6].

In today's entertainment world, the use of computer or console gaming is as accepted an entertainment medium as any other. Gamers (defined as persons who plays computer or role playing games) when questioned, have been found to enjoy a particular social inclusion when taking part in (particularly) role playing games, this despite the

outward looking opinion is that it is a solitary activity. Research has shown that gaming is in fact an increasingly social and inclusive activity and an activity resulting in the formation of new offline and online interactions and relationships [7], [8], [9] and the larger massive multiple player online role playing games actually thrive on the formation of these virtual relationships being formed, Gamers prefer to interact in the online/virtual world and one would imagine that they similarly thrive in an online/virtual learning environment.

Far from feeling socially excluded (an attribute associated with online learners), Kaye and Bryce and others have shown that social gaming experiences were found to be generally enjoyable and fun. These were facilitated by feelings of social belonging, networking and interactions with friends in social gaming contexts. There are social networks created in online games, similar to 'traditional' social networks on platforms such as Facebook. These are the ideal and typical attributes and commentary we would wish to see from students engaged in online learning and this suggests that gaming environments should be considered ideal platforms for online and distance learning to foster social inclusion among learners [10], [11], [12].

A series of experiments were carried at the Institute of Technology Sligo into real-time low cost motion and audio capture of lectures during the between 2012 and 2014 using the Microsoft Kinect™ sensor [13]. The purpose of these of experiments was to explore the potential usage applications for the Kinect sensor for new educational digital game based learning experiences. Released in November 2010 the Kinect™ sensor was part of a motion gaming push in the home video game console market. The sensor differed from existing motion based solutions offered by other console manufacturers which utilized local motion sensors such as accelerometers, gyroscopes and the use of either 2D RGB cameras or IR sensors for approximation via triangulation of a user's position in a 3D space [14], [15].

The Kinect™ sensor utilizes some of the existing approaches while introducing a new paradigm in user interactions through body part identification and machine learning [16]. The sensor combines numerous technologies to offer multiple sources of usable data including skeletal tracking, audio, depth, colour and infrared data streams. These are combined to offer developers the ability to perform real-time skeletal tracking and speech recognition.

Of particular interest to the research was the capability to perform real-time motion capture through the skeletal tracking functionality of the Kinect™ sensor. The device is capable of tracking up to twenty joints on up to two people at once. Each

tracked joint has a position in metres relative to the sensor and is capable of assigning a tracking status to each of the twenty joints indicating whether a joint is track, inferred or not tracked Combined with the sensors microphone array for audio capture which features a multichannel echo cancellation system and beam forming for the reduction of external noise in the recording space the sensor offers an all in one solution to explore new methods of capturing and repurposing educational material.

We tested the premise that the skeleton and audio data could be saved and utilized in existing educational and training game based applications also under development. The low cost of the sensor and the availability of the Kinect™ for Windows SDK were a perfect fit for our existing .NET Framework based game engine.

2. Methodology

The following section outlines the methods developed to capture and process the motion and audio data from the Kinect sensor. At the original time of development all software was custom developed to achieve the results we desired. The experimentation was performed and delivered to students. Since the initial research a range of professional commercial game development engines became available that allowed us greatly improve the process and quality of recordings performed.

2.1. Setup and Recording

Experimentation began once rooms had been assigned for fulltime theory classes. This restricted the available space in which we had to record but also presented real world conditions where space and layout would not be ideal for motion capture purposes. Figure 1 shows the layout that was used throughout the testing period. The device needed to capture the full body of the lecturer and was placed to the side of the classroom as to be able to see behind a fixed desk obstruction. All skeletal recordings were saved as XML in a simple custom data type.

Multiple modes of audio capture were available to the lecturer. In the live tests carried out, the Kinect™ sensors microphone array was used but other tests utilized widely available wireless gaming headsets to guarantee high quality audio recordings of the lecturers voice. In the case where an external audio recording device was used the Kinect™ was still used to record audio as a backup recording and to capture any in-class questions fielded by the students. All audio recordings were saved using the WAV file format, this was the native type supported by the recording devices.

The Kinect™ for Windows SDK was used to develop a custom Windows application which handled all aspects of the recording. The lecture room was equipped with a Windows PC but did not meet the requirements for running the recording software and an additional laptop which met the minimum hardware requirements for recording was used during sessions. Setup for each recording session was manual and required the Kinect sensor to be placed in the same location each session. The recording software could be remotely controlled using a wireless Xbox 360 controller, allowing the lecturer to start, pause, resume and stop recording at any time.

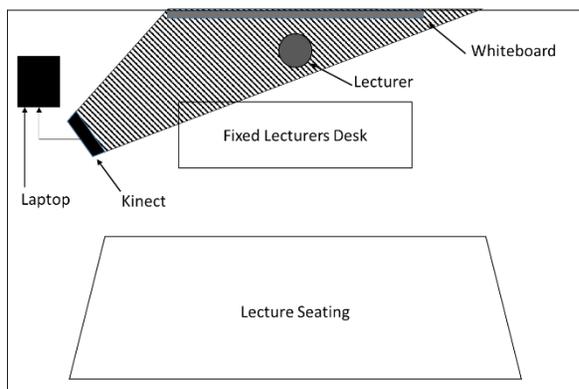


Figure 1. Original Recording Setup: Lecturer standing at top of class and Kinect™ camera to the side

2.2. Class Progression

Each lecture consisted of a presentation of relevant PowerPoint slides. It was necessary for these slides to be integrated into the final recording. An issue that arose was in determining when the slides changed in the class so that they could be matched with the slides progressing in the digital lecture. As the recordings were performed on a separate computer than the one used for the presentation it was not possible at the time to synchronise the two and capture the required input. As an alternative, three methods capable of catching these changes were implemented:

2.2.1. Wireless Xbox 360 Controller A wireless controller that was connected to the laptop running the recording software was used by the lecturer. When a slide change occurred the lecturer pressed a button on the controller which saved the time relative to the recording at which the slide needed to change.

2.2.2. Gesture Detection. The recording application could detect a limited set of gestures and a horizontal hand swipe could be used to signal the change in slides. Like the previous method a timestamp was created. It was possible to add gestures post-recording by swapping recorded frames with a new set containing the required gesture.

2.2.3. Voice Recognition The recording application could also detect a limited set of voice commands which could be used to control the application. One of those commands supported the creation of a timestamp for slide changes.

2.3. Post Processing

All recording sessions resulted in two files, an XML file containing the raw captured skeletal motion data and a WAV file containing the audio captured from the session. To guarantee the highest quality, each of the two files needed to be manually examined. In the cases where issues were identified these files then needed to be edited to produce a final working recording. Two separate methods were developed for processing each of the data types.

2.3.1. Skeletal Data Processing A custom tool was developed for analysing and editing recorded skeletal data. The tool was tasked with identifying bad skeleton frames in a recording session using the tracking state of each joint. This might occur if a joint had a state of Not-Tracked or inferred then this frame was marked as possibly containing flaws and would require manual editing. Two options were developed for editing frames which were deemed unusable after manual inspection:

2.3.2. Repositioning of Joints in Bad Frames Each joints position is represented as single vector. If a joints position was not as expected in a frame then this vector could be manually moved to the correct position. This method was only used in situations where continuous bad frames were small in numbers. This method would be time consuming for a large number of frames and lead to the development of the next solution capable of editing larger segments of a recording.

2.3.3. Substitution of Skeleton. In the cases where a large number of continuous frames were found to be unusable then either portions or the entirety of the skeleton was replaced using frames from elsewhere in the recording or from previous recordings. This method lessened the requirement of having a perfect view of the lecturer during recordings and enabled the

addition of pre-recorded gestures into a recording where the gestures did not originally exist.

2.3.4. Audio Data Processing. The audio data processing was a significantly easier process. The devices used to record the audio were tested to ensure that the resulting audio recording was of a high quality before been used in the recording sessions. The only significant changes to the original recordings were the trimming to the start and end of the recording and slight adjustments to the audio levels. Once both data types had been processed, two files were produced and were ready for use in the intended 3D application.

2.4. Commercial Game Engines

Recent changes in the access to commercial game engines used for the development of high quality computer and video games has provided access to professional tools developed for motion capture. Using Epic Games Unreal Engine 4 and the Kinect 4 Unreal plugin we have been successful in recreating our virtual lectures using modern commercial tools.

The skeletal post processing detailed in 2.3 is greatly simplified with recording data now saved in a widely used industry animation format. Animations can now be loaded into commercial modelling and animation software where alterations can easily be made.

Using the proposed recording setup shown in Figure 12 a virtual lecture was created but not tested outside of experimentation.

3. Results

Once processed the recordings were delivered through a virtual 3D environment. Utilizing a previously developed 3D game engine we were able to easily integrate the recordings into existing 3D environments as a proof of concept. The presentation slides used in the lecture embedded in the environment and were updated using the methods described earlier.

Three sample applications were developed that used the data produced after processing. A Windows based 3D application, a Silverlight 5 3D application and a Windows Phone 3D application (see Figure 2 and Figure 3). The Windows and Windows Phone applications had their data embedded while the Silverlight application streamed the data from a web server to reduce the initial loading time of the application (see Figure 4).

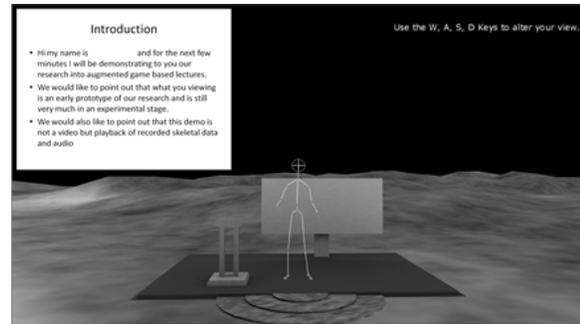


Figure 2. Early Motion Capture Prototype 2012.

Here a lecturer has his skeletal movements captured presenting a lecture and the data is inserted into a digital background scene.



Figure 3. A Virtual Classroom is created with content slides embedded into the digital space

Live Streamed Skeletons

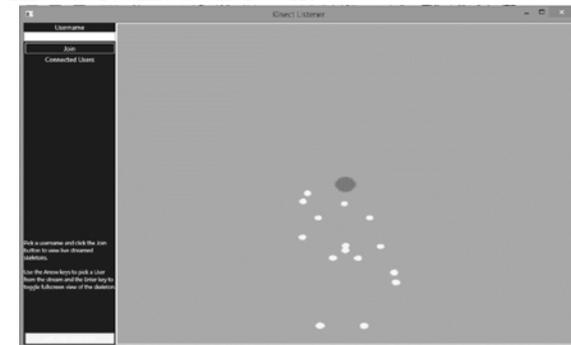


Figure 4. Live Skeleton Broadcast of a presentation by a lecturer was demonstrated and illustrates the ability to deliver live digitally augmented content

3.1. Student Feedback

The test lectures were opened to a fulltime class and they were asked to comment on the technology. 10 students responded to the questionnaire (out of 50). Of the students tested, there was a 50:50 male female split as shown in Figure 5. 90% of those who

viewed the content were in the 18-35 age bracket, (see Figure 6). Asked if they had technical issues with software usage, 90% of the respondents successfully managed to get the virtual classroom software running on their own computers. The delivered prototype only supported devices running the Windows operating system. One user's Windows based machine did not have hardware capable of running the prototype application (see Figure 7).

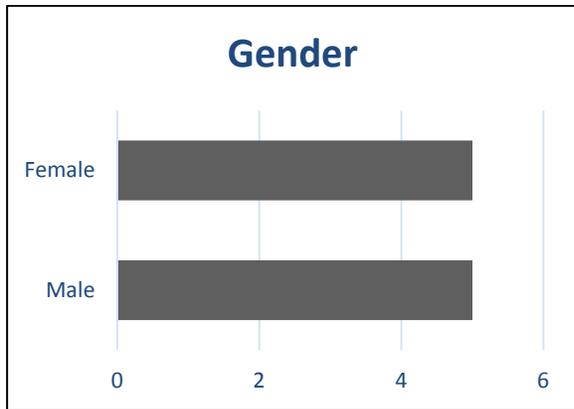


Figure 5. Gender

Figure 5 shows a 50/50 split between the genders of the respondents to the survey on the usage of the virtual classroom prototype software in 2014.

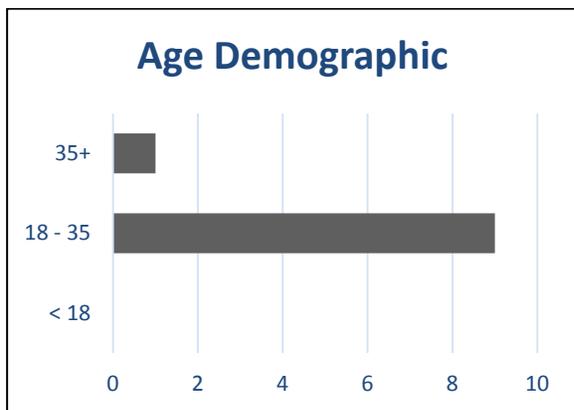


Figure 6. Age demographic

Figure 6 shows that 90% of the respondents were in the age demographic of 18 – 36 (inclusive) with only 10% in the 36+ category and none of the users were below the age of 18.

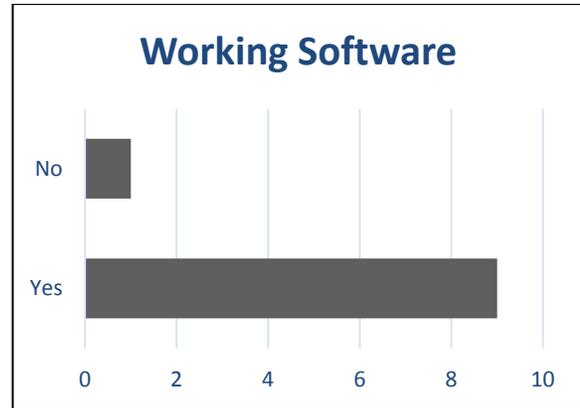


Figure 7. Respondents who successfully downloaded and ran the software on their own hardware

When asked if the augmented lectures were more engaging than conventional video or video-streaming of content, 60% agreed that it was more engaging against 30% who did not (see Figure 8).

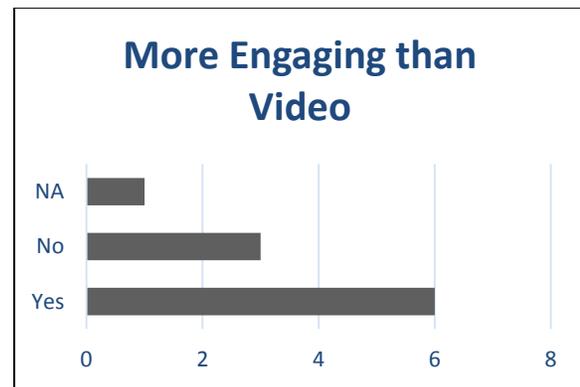


Figure 8. Respondents Views on Engagement

The 10% Not Applicable applies to users who were unable to run the virtual classroom prototype.

When asked if they considered the platform a valid teaching methodology, 60% agreed whereas 40% answered maybe, no student responded to these questions (see Figure 9). 40% of students also suggested that they would like to see more content delivered in this manner, 40% were neutral on this delivery approach and only one student would not like to see this technology used more (see Figure 10) and this was a student who had technical difficulties with running the software.

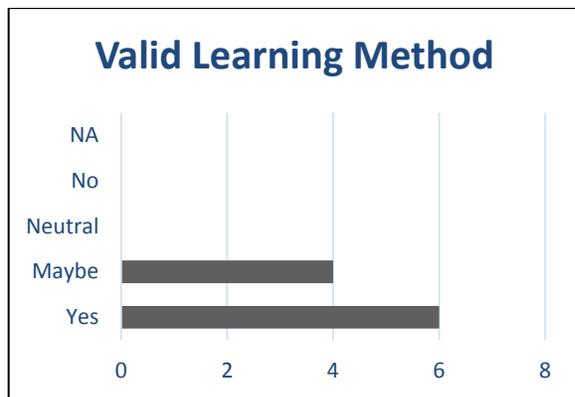


Figure 9. Respondents view of virtual lectures as a valid learning method

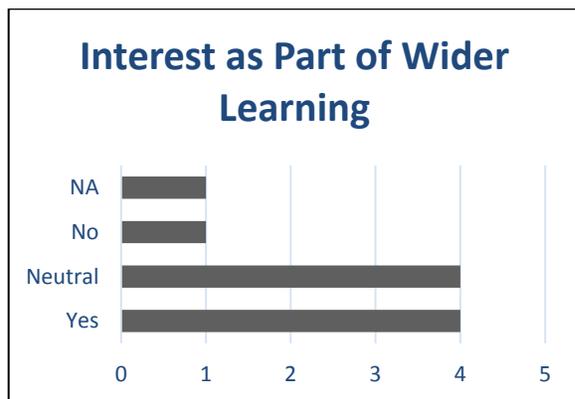


Figure 10. Respondents Interest in more game based learning of the type demonstrated

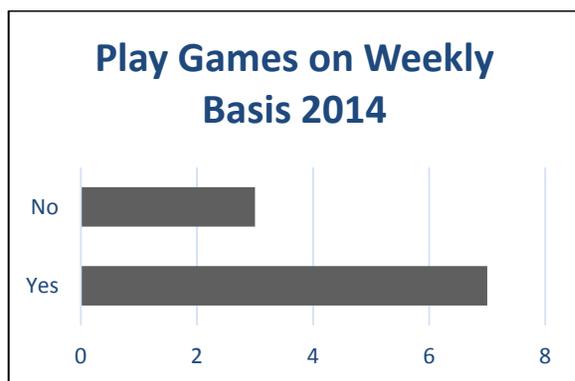


Figure 11. Responses when asked if games are played on a weekly basis in 2014

It is important to note that none of the respondents chose neutral or no, implying that students are open to the use of new learning technologies.

In order to put the data into some kind of gaming perspective, student participants were surveyed on their game playing habits. Figure 11 shows that 70%

of the students questioned play games on a weekly basis, this does not exclude the possibility that some of the remaining students play games on a less than weekly basis, this was not determined during the questionnaire.

4. Discussion

This article presents the possibility of integrating aspects of online gaming to classroom teaching with a particular target of online teaching. The future student demographic is one of gamers, and it is likely that a significant number of these students may even classify gaming as one of the pastimes or interests given the fact that 91% of 4-17 year olds play computer games [17]. In order to attract this cohort of students, or at the very least engage with these students, we aim to integrate virtual lecturers in a 3-dimensional lectures and laboratories so as to distribute and disseminate academic content. In this study we investigated the usability of the Microsoft Kinect™ platform to create virtual learning environments hosted by avatars (as opposed to the real life lecturer) similar to those environments experienced in computer gaming, particularly those immersive, online and interactive games in which online gaming communities are created and share the virtual world of the game.

The traditional lecture room was replaced by a virtual representation of a lecture room. The virtual setting could be whatever is desired by the lecturer but here we stuck to the traditional classroom, which comes with computer, desk, whiteboard and slides. However, in one test, a 5 minute presentation on cell toxicity testing was recorded in a virtual lecture room followed by examples of cell staining being shown to the students in a virtual laboratory setting, this virtual laboratory had been previously developed by our research group to deliver and assess laboratory based science content. The slides are provided by the lecturer and contain the academic content. The lecturer is replaced by an avatar whose movements are tracked using the Microsoft Kinect™ system.

Once developed, this content was shown to a class to whom the content would be relevant. The class were asked to answer a quick survey on their thought on the delivery approach, 20% of the students watched and completed the survey, which was a voluntary exercise.

One of the most striking statistic from the data (or at least striking to those outside of the gaming industry) is that 70% of the age group 17-35 play computer games on a weekly basis and this encompasses a 50:50 male/female split (see Figure 11 and Figure 5). This is reflective of the global usage in

computer game playing and is expected to rise over the coming years as the younger populations move into this age bracket, currently 91% of 2-17 year olds play computer games on a weekly basis (NPD, 2011). Currently (ESA, 2013) 40% of all computer game users play on either 'Persistent Multi-Player Universe' or 'Action, Sports, Strategy, Role-Playing' type games, both these types of gamers could be seen as potential users of our proposed kinect-enabled virtual online delivery mode and this is reflected in our finding that 40% of those students questioned would welcome this type of delivery of content (see Figure 9).

None of the students surveyed suggested that this approach was not a valid approach to teaching with 60% of them agreeing that it was a valid method of teaching with the remaining 40% answering 'maybe'. This indicates an acceptance on the part of the student body of virtual environments as a way of imparting information relevant to their academic learning and achievement. It is this acceptance that suggests that in order to capture the imagination and also capture the attention of students, teaching institutes should seriously consider the development of virtual environments to disseminate academic content. This approach may help engagement and inclusion of students involved in online learning, particularly those who may categorize themselves as gamers.

The application of this augmented lectures and accompanying 3-dimensional laboratory environment could be well suited to online learners, particularly those who have a familiarity with gaming and online gaming environments. It may encourage interaction with their academic content and interaction with their online peers who may also be logged in to the virtual classroom at the same time; this could also foster social inclusion among online and distance learners. However, as demonstrated here, the technology could equally be used for delivery of content to conventional fulltime students.

While the study demonstrates the ability to deliver academic content which is received positively by students, it does possess drawbacks in its current guise.

4.1. Technical Issues

In its current form the post processing means that the audio tracks from existing recorded video lectures used in online learning could be extracted and repurposed to be used in conjunction with skeletal recordings. This would significantly reduce the effort in creating new augmented virtual classes and allow educators to reuse existing teaching materials for experimentation in game like teaching methods.

As versions of the application had been developed for web browser integration with existing web based online learning platforms such as Moodle and blackboard. No additional custom extensions of these existing products would be required other than linking to an external web address of where the material would be hosted. This is in line with current practice on these platforms.

Due to the flexible nature of the recordings, both of the data types could be re-recorded at a later date if they were not deemed to be of a usable quality. The range of motion recorded in the tests was minimal allowing the possibility of using pre-recorded skeletons in situations where motion capture may not be possible or post processing of the content was not possible or feasible due to technical deficiencies.

The original recordings were large and resulted in roughly 40MB of audio and 400MB of motion capture data generated per hour. However, optimization of the XML data could be employed in an attempt to reduce the large motion capture files. When using the commercial engines motion recordings are saved in a standard animation format which is significantly smaller, more flexible and reusable. The recorded audio can be encoded to MP3 or another compressed audio format during the processing stage. This would significantly reduce the audio file sizes for consumption.

In our previous research we developed a fully interactive and explorable 3D environments (a biotechnology laboratory) that directly relate to the lecture subject be developed. This environment (and others in development) should permit the lecture to be paused at selected intervals and allow learners to partake in carefully designed and related active learning tasks within the 3-dimensional environment.

Attaching the Kinect™ sensor to an overhead projector used in the lecture hall would produce a higher quality motion capture recording due it facing the lecturer, as proposed in (Figure 12). The possibility of losing the lower half of the skeleton is more likely with this but could easily be resolved in post processing outlined in section 2.3.3.

The live streaming of skeletal data was tested and is a feasible option for the delivery of live lectures. In testing a Microsoft Windows Azure® web service was deployed to broadcast skeletal data from a single source to multiple clients that were using both traditional PCs and smartphone devices. While this option performed well it would however lose the benefits of the post processing and having a customized 3D interactive environment developed around the lecture content.

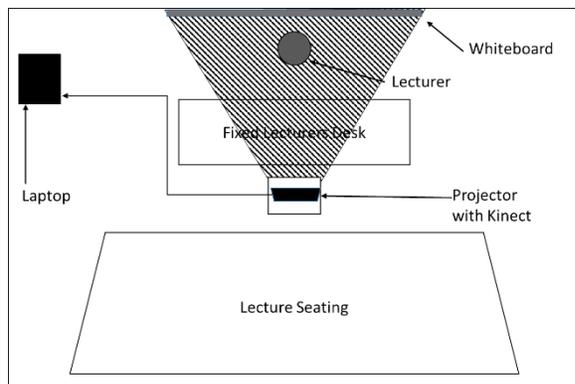


Figure 12. Proposed Recording Setup with Kinect™ sensor placed overhead.

The resulting offering would deliver little benefits when compared to traditional live streamed video lectures. Ideally any recordings made in a term would be used in the following college term. This time period would allow for a longer development cycle permitting lecturers to collaborate with developers in helping to deliver the most interactive and relevant learning experience possible.

When using some of the advanced features in the commercial development tools there is the need to introduce recommended hardware requirements. While the games can be configured by the user to achieve the best possible performance there is always the concern of some users being unable to play the games due to hardware limitations on their own personal machines.

5. Conclusion

Interactive technologies are evolving at a rapid pace and offer new opportunities in developing unique teaching and learning methods at a low cost with minimal interference to traditional methods. Finding new ways to repurpose existing educational content such as video lectures and offer interactive digital learning environments to a gaming demographic. Using commercially available games development tools it's now possible to create experiences of a quality on par with the games were students users are playing for entertainment. The features offered by these tools greatly improve the design process and speed at which virtual environments can be developed and deployed.

The Kinect™ sensor offers a relatively cheap motion and audio capture option and this recording system was easy to setup and had little or no impact on the lecture class. Once recorded the captured data was flexible for editing and use in interactive environments. These environments could be

specifically designed around the subject material as well as embedding content presented in class.

The learning environment created would allow for data to be collected on how learners interact with the virtual environment and what elements are the most engaging to them but also the elements with which they have learning issues. This data could be used to inform educators, designers and developers during the design process of games as to what can be done to maximize the teaching/learning potential of augmented lecture.

Once developed the environments are also portable across platforms including desktop, mobile devices and through modern web browsers with no changes to recordings.

6. References

- [1] Elaine Allen, I and Seaman, J, 2013. Changing Course: Ten Years of Tracking Online Education in the United States. Babson Survey Research Group and Quahog Research Group, LLC. Retrieved April 22nd 2014 from <http://files.eric.ed.gov/fulltext/ED541571.pdf>
- [2] IBIS capital, 2012. Global e-Learning Investment Review. Retrieved April 22nd 2014 from <http://www.ibiscapital.co.uk/resources.aspx>
- [3] Capra, T. (2011). Online Education: Promise and Problems. *Journal of Online Learning and Teaching*. 7(2), 288-293
- [4] Hanover Research Council, 2009. Best Practices in Online Teaching Strategies. Retrieved April 22nd 2014 from <http://www.uwec.edu/AcadAff/resources/edtech/upload/Best-Practices-in-Online-Teaching-Strategies-Membership.pdf>
- [5] Joyce, K and Brown, A, 2009. Enhancing Social Presence in Online Learning: Mediation Strategies Applied to Social Networking Tools. *Online Journal of Distance Learning Administration*, 12(4)
- [6] Online learning Insights 2012, The LMS Divide – Social Presence in Online Learning. Retrieved April 22nd 2014 from <http://onlinelearninginsights.wordpress.com/2012/02/27/the-lms-divide-social-presence-in-online-learning/>
- [7] Bryce, J., & Rutter, J. (2003). The gendering of computer gaming: Experience and space. In S. Fleming & I. Jones (Eds.), *Leisure cultures: Investigations in sport, media and technology* (pp. 3–22). Eastbourne: Leisure Studies Association.
- [8] Yee, N. (2006). The demographics, motivations and derived experiences of users of Massively Multi-User Online Graphical Environments. *Presence: Teleoperators and Virtual Environments*, 15(3), 309–329.
- [9] Yee, N. (2007). Motivations of play in online games. *CyberPsychology & Behavior*, 9, 772–775.

[10] Kaye, L. K and Bryce, J. (2012). Putting The “Fun Factor” Into Gaming: The Influence of Social Contexts on Experiences of Playing Videogames. *International Journal of Internet Science*, 7 (1), 23–36.

[11] Gajadhar, B.J, de Kort, Y. A. W. and Jsselsteijn, W. A. (2009a). Rules of engagement: Influence of co-player presence on player involvement in digital games. *International Journal of Gaming and Computer-Mediated Simulations*, 1(3), 14–27

[12] Gajadhar, B.J., de Kort, Y.A.W., & Jsselsteijn, W.A. (2009b). See no rival, hear no rival: the Role of social cues in digital game settings. *Proceedings of the 13th CHI Nederland Conference*, (pp25 – 32). CHI-NL publishing

[13] Microsoft Corporation, 2014. *Kinect for Windows*. Retrieved May 23rd 2014 from <http://www.microsoft.com/en-us/kinectforwindows/>

[11] NPD, 2011. Kids and Gaming 2011. Retrieved May 23rd 2014 from https://www.npd.com/wps/portal/npd/u/news/press-releases/pr_111011/

[15] Nintendo Co. Ltd. , 2014. *Wii Official Site*. Retrieved May 23rd 2014 from <https://www.nintendo.com/wii/what-is-wii/#/controls>

[14] Sony Corporation, 2014. *Playstation move*. Retrieved May 23rd 2014 from <http://us.playstation.com/ps3/playstation-move/>

[16] Shotton, J., Fitzgibbon, A., Cook, M, Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. 2013. *Real-Time Human Pose Recognition in Parts from Single Depth Images*. Retrieved May 23rd 2014 from Available at <http://research.microsoft.com/pubs/145347/bodypartrecognition.pdf>

[17] ESA, 2013. 2013 Sales, Demographic and Usage Data. Retrieved May 23rd 2014 from http://www.theesa.com/facts/pdfs/esa_ef_2013.pdf