Drexel University
Electrical and Computer Engineering Department

Multilevel Tracking for Dynamic Augmentation of Live Performance

ECE Senior Design Group 06
Final Report
March 23, 2013

Team Members:
Matthew Zimmerman, ECE, mzimmerman@drexel.edu
Manu Colacot, ECE, m.colacot@drexel.edu
Julian Kemmerer, ECE, jvk@drexel.edu

Advisor:
Dr. Youngmoo Kim, ECE, ykim@drexel.edu

Abstract

Performances aim to captivate an audience. In recent years, this is often been accomplished using various forms of technology. A simple and nearly ubiquitous example of this is visualizations shown on projector screens behind performers [1]. Such visualizations are often pre-planned and scripted, leaving little room for creative expression during the performance itself. In order to address this issue, we have developed a system that allows for dynamic manipulation of various performance effects: both audio parameters and visual parameters.

To give context, many performers are already very emotive while on stage. They naturally make expressive gestures during the course of their performance. We have developed a system that takes these motions, and uses them to augment the performance in an innovative way. This type of motion based effects system is similar to the V Motion Project, a Kinect based project aimed at turning motion into music [3].

In most current performances, an individual's movement has a minimal, if any, range of impact on sound and visual effects. There are also certain audio and visual effects that are underutilized or completely ignored in the context of live performance - the simplest example being left/right audio panning. Audio and visual effects are typically relinquished to someone like an audio engineer or stage lighting director.

We have given some of this ability back to the performer through an audio and vision based tracking environment. This tracking consists of two levels, stage position and gesture/motion. The stage level tracking is accomplished through audio multilateration via hidden signals within live audio. Gesture/motion tracking is completed through vision based Microsoft Kinect devices.

As of now, stage level tracking is able to determine the location of a performer within a lecture hall size room. The accuracy of this tracking is approximately 1-3 feet. Motion/gesture tracking is able to control relatively simple audio effects and allows for gesture based control of effect selection.

At this time we perceive ourselves to be on schedule, though behind our original project goals. Due to only partial progress, our budget was significantly lower than expected. Major breakthroughs include the transmission of gesture/motion information to a central computing location and fine tuning of the stage level tracking algorithm.

Contents

List of Figures

List of Tables

1 Problem Description

As technology progresses, it makes its way into various performance applications. Much of which is for musical purposes and includes everything from electronic devices to computer software such as effects pedals and Pro Tools, respectively. In general, many performers seek different ways to creatively augment live performances.

Most of these systems rely on a single sensor or set of sensors to provide measurement of one physical parameter. However, we have developed a system that combines multiple large and small scale tracking methods to provide a unified real-time multi-parameter data source which is available for audio and visual effects mappings.

Currently methods for tracking performer stage position and gestures are underutilized. Therefore, the two goals for this project were to:

1) Design an accurate position and gesture tracking system for both the performer and stage scale that was as non-destructive as possible.

2) Combine various sensor information from the different tracking methods to provide real-time multi-parameter mapping data for audio and visual effects manipulation.

The purpose of the stage position tracking was to determine a performer's position on an indoor stage. This is a difficult problem, as current technology like GPS only works outdoors. Even the best indoor position location systems can take up to a minute and a half to successfully track a person's position - completely impractical for our purposes [5]. Therefore we were required to develop a method for indoor localization.

In addition to stage level tracking, our system also utilized the Microsoft Kinect in order to perform gesture/motion tracking. Because most performances usually contain more than one individual on stage, our system had to be able to accommodate multiple performers. Thus, we developed a way to get tracking data from multiple Kinects and then map that to various effects.

Both methods of tracking had to be integrated into one easy to use unit that made it possible to control both audio and visuals. To better understand the desire for an augmented live performance consider a simple example of the common concert venue. There is a static panning of instruments done prior to the start of the performance typically by a sound engineer employed by the venue. Once the performance begins it is difficult, if not impossible, for the performers to dynamically alter even simple parameters such as stereo panning. Another example is that of a disc jockey (DJ), whose live remixes of audio require creative control of several parameters simultaneously. A system that yields additional control options like gestures and continuous motion tracking would vastly expand the capability of the DJ and would arguably allow for greater control than the discrete 'button pushing' of live remixing.

Mappings of simple and complex parameters like the ones aforementioned exist, however, these are poorly utilized or complicated enough to deter natural creativity. The emphasis of this project is not the development of specific mappings of positions/gestures to effects but rather the design of the system that make such vast real-time data sets available to the performer(s) for enhanced expression. While some mappings will be designed for initial analysis and demonstration, the majority of work will be put towards combining multiple data capture systems into a single interface, rather than the implementation of novel effect.

2 Progress Toward a Solution

2.1 Stage Level Tracking

When the term began, we were in the midst of resolving various problems with timing and signal accuracy for our audio localization algorithm. Because of this, the system was not fully functional and still very error prone. Since then, a fair amount of progress has been made by using a simple Fast Fourier transform (FFT) to confirm the presence of an audio signal, followed by appropriate windowing, and cross-correlation (Equation 1) to accurately locate the origination of the signal in the time domain.

$$(f \star g)[n] = \sum_{m=-\infty}^{\infty} f^*[m] g[n+m] \tag{1}$$

As with the other portions of this project, a system that integrates Open Sound Control (OSC) [4] packets has been implemented such that a central server can receive the information from any device that is capturing audio localization information. OSC is a convenient networking format thats acts as a simple and cross platform container for encoding and decoding information. These packets are sent over a wireless network to the server, which then processes them in accordance with the functionality desired by the performer.

2.2 Gesture/Motion Level Tracking

At the start of this term a stripped down version of an "off-the-shelf" gesture recognition system, implementing dynamic time warping [6], for the Microsoft Kinect was in use. This reduced-complexity system was combined with the continuous motion portion of this project. The resulting system allows performers to change effect parameters by using gestures alone.

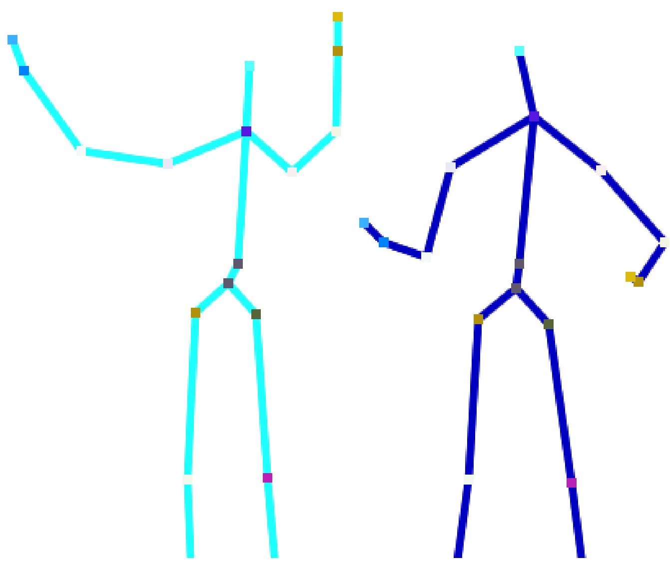


Figure 1: Microsoft Kinect skeleton data visualization.

The information captured through the Microsoft Kinect vision system was extracted using Microsoft's Kinect Software Development Kit (SDK) [2]. This SDK makes skeleton information readily available for programmatic use. Figure 1 is a visualization of the skeleton data as provided by the SDK. This information is then wrapped within OSC packets and transmitted to a central processing PC. The OSC protocol is capable of utilizing User

Datagram Protocol (UDP) packet transport - optimal for streaming data in as close to real time as possible. Currently the gesture/motion tracking system can stream live low latency skeleton information to a central computing PC for effects control.

2.3 Integration

At the end of last term, this portion of the project was largely completed. Communication between various devices and platforms was finalized and implemented. The major task that remained was processing the transmitted information. As mentioned earlier, communication to and from the stage level tracking is completed using OSC. It was important to select a coding framework that allowed for easy use of OSC libraries as well as had significant audio and visual capabilities. OpenFrameworks was selected for its simple, cross platform, and capable libraries. Currently a relatively powerful desktop computer serves as the central server where information streamed from all levels of tracking is processed as well as where all audio and visual effects are generated.

The graphical user interface (GUI) capable of manipulating these incoming data streams has undergone significant changes. A user is now able to control both audio/visual effects as well as the 'chirps' used in the audio localization. As an update from last term, we have purchased an audio receiver that will be used for demoing our final solution.

3 Constraints

3.1 Economic

The economic constraints of this project originate mostly from the relatively high cost of computers, audio amplifier systems with multiple channels of audio, and the multiple Kinects needed. A detailed description of exact costs is given in the budget portion of this report.

3.2 Manufacturability

This project does not require any additional items to be manufactured or any additional manufacturing steps. The custom portions of this project are all software based.

3.3 Sustainability, Environmental

This project requires slightly more electricity than a typical consumer home theatre setup. However, it is possible that if this system was in constant use that the processing resources needed by the central processing server could cause increase the required power of the system (as compared to a typical desktop).

3.4 Ethical, Health and Safety, Social, Political

It is difficult, but not impossible, to construe social or ethical impacts from this project specifically as the impacts have yet to be truly evaluated outside of the prototyping environment. However, it is not outside of reason to speculate on the societal impacts of a live performance enhancing system in general. Of course these societal impacts are limited to the musical performance but, in a larger sense, this system could forever change the expectations of performance goers and begin a new era of experimentation with live performance. It is possible that performance venues will begin to install permanent replications of our system in an attempt to provide the option of these augmented effects to every performer. It is even possible that the audience could be tracked and their actions collaboratively included into the artists performance.

4 Budgets

The industrial budget for the Spring 2013, Table 1 differs only slightly from the budget proposed in Winter 2013, Table 3. Most major components have been identified and purchased. The first major expenditure was the Denon AVR890 amplifier system. This device is used to easily support multi-channel audio routing. This is required for both proper audio effect reproduction and audio multilateration.

Table 1: Proposed Budget for Spring Term 2013

Spring 2013 - Industrial Budget				
Item	**Units**	**Price (per unit)**	**Price**	**Purchased**
Denon AVR-890	1	$299.99	$299.99	Yes
Microsoft Kinect	2	$109.99	$219.98	Borrow
Macbook Pro - 15inch 2.3 GHz	1	$1699.99	$1699.99	Yes
Central Desktop PC	1	$1300.00	$1300.00	Borrow
Cables/Connectors	1	$60.00	$60.00	Yes
Apple Airport Extreme	1	$179.99	$179.99	Borrow
Apple Developer License	1	$99/year	$99/year	Borrow
Microsoft Kinect SDK	1	$ 0	$ 0	N/A
Microsoft Visual Studio	1	$ 0	$ 0	N/A
Xcode	1	$ 0	$ 0	N/A
		Total	$3858.95	$2059.98

Instead of getting a desktop computer, a new laptop was purchased. This laptop computer is used to control one of the two remote Kinect devices. The central processing PC was borrowed from a team member for the duration of this project. This central PC is able to take in data from both the stage level tracking and gesture level tracking and map that to certain audio and visual effects. The current budget includes 2 Microsoft Kinect devices as presently this is all that is necessary. It also includes various smaller components such as cables, a router, and software, all of which were relatively low cost and were purchased or borrowed on an as need basis during testing.

Table 2: Personnel Expenditures

2013 - Personnel Budget				
Position	**# Employees**	**Hourly Wage**	**Total Hours**	**Salary**
Electrical Engineer	3	$40	600	$24,000
Project Manager	1	$47	150	$7,050
Senior Manager	1	$70	60	$4,200
			Total	$35,250
Fringe Benefits (≈35% of salary)				$12,337.50
Total Expenditure				$47,587.50

We have also included a break down of the expenditures that would be necessary if the project were to have been done with hiring employees, managers, etc. This can be seen in Table 2.

Table 3: Proposed Budget for Winter Term 2013

Winter 2013 - Industrial Budget			
Item	**Units**	**Price (per unit)**	**Price**
Yamaha YHT-597 BL Home Theatre System	1	$495.95	$495.95
Microsoft Kinect	4	$109.99	$439.96
Desktop for Audio/Visual Processing and Control			
Intel Core i7-2600K	1	$319.99	$319.99
Gigabyte GA-Z68XP-UD3P	1	$149.99	$149.99
CoolerMaster Elite 350 Case	1	$56.99	$56.99
Western Digital 500 GB Hard Drive	1	$59.99	$59.99
Corsair XMS3 16 GB RAM	1	$99.99	$99.99
EVGA GeForce GTX 660	1	$226.49	$226.49
Asus Xonar D1 7.1 Channel Sound Card	1	$226.49	$226.49
		Total	$1939.34

5 Schedule

5.1 Stage-Level Tracking *(Matthew Zimmerman)*

The challenge when accomplishing stage level tracking currently is obtaining a reliable and clean signal from a microphone, while also keeping the microphone mobile and easy to use. By collecting recordings from several locations within Mitchell Auditorium in the Bossone Research Center, it was determined that simple iPhone microphones performed as well as stand-alone USB microphones connected to a laptop. Previous versions of the multilateration algorithms were unable to deal with the iPhones automatic leveling/equalization.

However, the most recent implementation of the algorithms are not affected. As mentioned, the multilateration algorithms have been refined and the most recent versions rely on both the time and frequency domains in conjunction to determine the source speaker and location (in time) of the audio signal. Future work will focus on the remaining refinements needed as well as streamlining the computational complexity of these calculations for easier use on less capable devices like iPhones. As of now, stage level tracking is able to determine the location of a performer within a stage sized room. The accuracy of this tracking is approximately 1-3 feet.

5.2 Watermarking *(Manu Colacot)*

In order to implement audio triangulation without interfering with the performance, audio watermarking must be utilized. This is the process of embedding audio information into a signal so that the embedded audio is indiscernible through normal auditory processes [7]. In this way the speakers can output both the audio produced by the performer and also the 'silent' chirp, that is needed for the triangulation. A brief literature review on this topic was performed. Though live demo worthy code has not been fully implemented, a proof of concept MATLAB script showing that audio can be hidden in an actual song was developed. Several simulated tests were completed using this system in cooperation with multilateration algorithms.

5.3 Gesture/Motion-Level Tracking *(Julian Kemmerer)*

Several challenges facing gesture and motion tracking were overcome. The first being the computational complexity of gesture tracking. Rather than continuously tracking many possible gestures to control numerous audio effects, gestures are utilized to switch between audio effects. This limits the number of gestures that need to be compared within the gesture recognition algorithms that run at a minimum of 30 times per second (as to keep up with the Kinects' data stream). This decreases the amount of processing required at any given point.

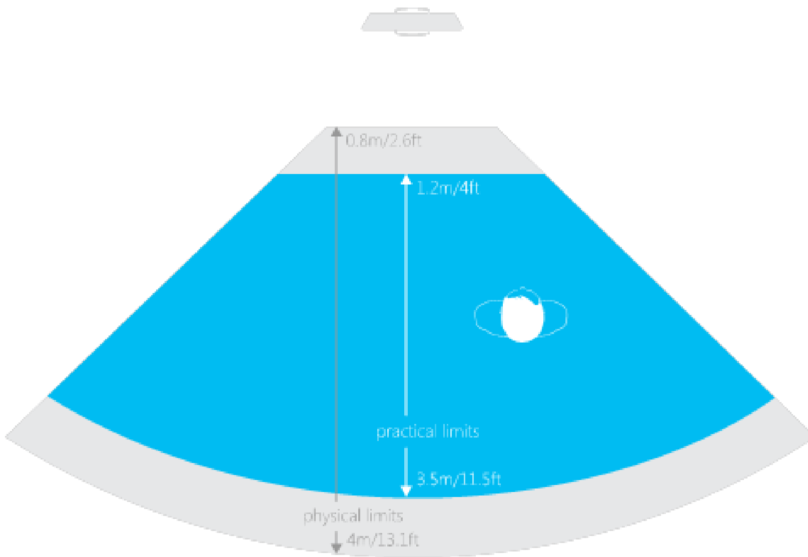


Figure 2: Microsoft Kinect field of view

The second major hurdle was dealing with several Kinects simultaneously. Figure 2 shows the relatively limited field of view for a single Kinect device. In addition, there is a Microsoft implemented restriction limiting the tracking to two skeletons simultaneously (however, to maintain the greatest level of detail, we limit our Kinect devices to only one performer). Thus, our system is able to support multiple Kinects for multiple performers, simultaneously.

Two options for handling the massive amounts of streaming data from multiple devices were evaluated. The first option was having an extremely powerful central server with direct USB buses from each Kinect device. Even if the system was able to handle those data rates and process efficiently, there was still the issue of USB cable length. These Kinect devices will need to be spread across stage-sized areas. The maximum length for USB cables is a bit over 15 ft - which is not suitable for our task. The second option evaluated was passing Kinect data over the network via OSC packets. The possible issue with this technique is the delay incurred via network transfers. However, after testing the delay in transit was observed to be negligible. Therefore, the second option was chosen.

5.4 GUI *(Manu Colacot)*

A GUI was designed and implemented in order to organize both stage and gesture level tracking. The GUI allows the user to choose joints he or she wants to track on the Kinect, and what that tracking will map to. Several GUI frameworks were evaluated with the most important factors being, 1) compatibility/communication with C++ (used for openFrameworks audio/visual effects and OSC network communication), 2) appearance, 3) ease of use. Originally Objective C was considered, though communication with openFrameworks became an issue. C# along with the Windows Presentation Foundation (WPF) framework was used as the GUI structure is familiar to most users, the codeset is well documented, and communication with C++ code is easily implemented.

5.5 Gesture Mapping *(All members)*

Though not the focus of development within this project, mappings are certainly necessary to make it successful. Having a set of ‘demo features’ that the team is able to pick from, but are not entirely limited to, will allow for better testing and also help others understand the overall design.

As previously mentioned, integration of all levels of tracking is quite important and is for the most part completed. The team is currently working to make final changes to the design to allow for data to be seamlessly transferred between Kinect devices, iPhones, and the central processing server. The challenge in this is that these two systems are, and provide, very different sets of data, so communication between them, in an orderly fashion, is integral to the overall success of the project. Along with this, live audio input is being integrated and manipulated with the overall system.

6 Summary and Conclusions

We have developed a system that gives performers the ability to control various audio and visual parameters live while on stage. This was accomplished by developing technology that performed indoor localization while also tracking human motion data, and then mapping both parameters to specific effects.

By creating a system that provides performers with multi-parameter real-time data, we believe that we have developed a platform to not only encourage, but enhance creativity in live performances. Performers now have extra degrees of freedom that include both their physical location on stage as well as different joints on their body to control audio as well as visual parameters. This increases the creative possibilities while on stage.

References

[1] Matthew Neil Bain. Real time music visualization: Real time music visualization: A study in the visual extension of music. Master's thesis, The Ohio State University, 2008.

[2] Microsoft Corporation. Kinect development sdk. `http://www.microsoft.com/en-us/kinectforwindows/develop`, 2012.

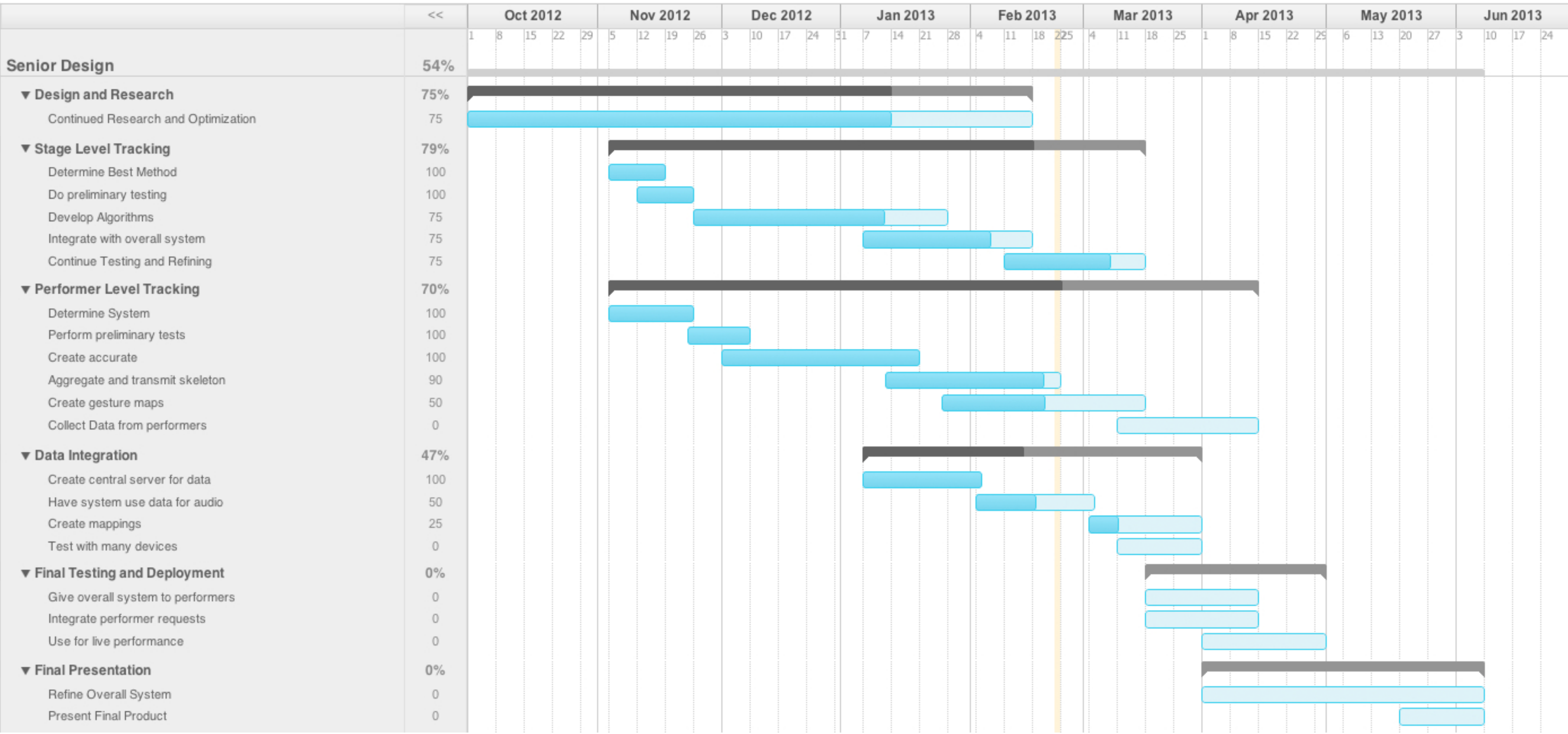
[3] V Energy. The v motion project. http://www.v.co.nz/, 2012.

[4] Adrian Freed and Andy Schmeder. Features and future of open sound control version 1.1 for nime. In *NIME*, 04/06/2009 2009.

[5] Apple Insider. Apple's 20m purchase of wifislam .... http://appleinsider.com, March 2013.

[6] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions of Acoustics, Speech, and Signal Processing*, ASSP-26(1):43–49, February 1978.

[7] Sadasivam Subbarayan and S.Karthick Ramanathan. Effective watermarking of digital audio and image using matlab technique. *2009 Second International Conference on Machine Vision*, 2009.



<<
Oct 2012
Nov 2012
Dec 2012
Jan 2013
Feb 2013
Mar 2013
Apr 2013
May 2013
Jun 2013
Senior Design 54%
Design and Research 75%
Continued Research and Optimization 75
Stage Level Tracking 79%
Determine Best Method 100
Do preliminary testing 100
Develop Algorithms 75
Integrate with overall system 75
Continue Testing and Refining 75
Performer Level Tracking 70%
Determine System 100
Perform preliminary tests 100
Create accurate 100
Aggregate and transmit skeleton 90
Create gesture maps 50
Collect Data from performers 0
Data Integration 47%
Create central server for data 100
Have system use data for audio 50
Create mappings 25
Test with many devices 0
Final Testing and Deployment 0%
Give overall system to performers 0
Integrate performer requests 0
Use for live performance 0
Final Presentation 0%
Refine Overall System 0
Present Final Product 0

APPENDIX A

Design Constraints Summary

Team Number: ECE - 06

Project Title: Multi-level Tracking for Dynamic Augmentation of Live Performance

1) Summary of the Design Aspects of the project:

In this project a multi-pronged tracking environment was designed to augment live performances. This system was developed to track both a performer's location on stage as well as his gestures, and then combine and map this data to various audio and visual effects. The main design goals for this project were to:

1. Provide an accurate stage position and performer gesture tracking system that minimally hiders the flow of a natural performance.
2. Combine various sensor information from the different tracking methods to provide real-time, multi-parameter mapping

2) Summarize how Realistic Constraints that were evaluated for the project, including:

Economic:
The economic constraints revolve around the need for multiple Microsoft Kinects as well as an audio receiver that supports multiple channels with at least 5.1 surround sound.

Manufacturability:
Nothing had to be manufactured for this project because our development was software based.

Sustainability:
As long as a power source is available to support multiple Microsoft Kinects, a desktop computer, and a theater sound system (multiple speakers and a Receiver) this system should be sustainable.

Environmental:
Aside from those required to generate electricity (which our system runs on), there are no environmental constraints.

Ethical, health and safety:
There are no ethical, health, or safety constraints.

Social:
While this has not been tested yet since our system has not been fully implemented in any performance venues, it is conceivable that this technology can be used to blur the line between performers and audience, and thus result in an entirely new social experience during live performance.

Political:
There are no political constraints.

3) Cite list of Standards/Regulations that were used or evaluated for the project (use IEEE Reference-style):

Ethernet: *Xerox Parc, IEEE 802.3, 1974*
Universal Serial Bus 2.0 (USB 2.0): *Hewlett Packard, Intel, Microsoft ,NEC, et al, 1996*
User Datagram Protocol (UDP): *Reed, David, RFC 768, August 28,1980*
Wi-fi: *Wi-fi Alliance, IEEE 802.11 a,g , 1999-2003*

Drexel Writing Center: Appointment made successfully!

Drexel Writing Center <writingcenter@drexel.edu> Thu, May 9, 2013 at 8:22 PM
To: mjc344@drexel.edu

Manu Colacot,

You have successfully made an appointment on 05/13/2013 between 10:00am and 11:00am. The appointment is with Prof. Henley ESL Specialist (One Drxl Plaza GL106) at Drexel Writing Center.

You can make, cancel, or modify appointments by logging into the scheduling system at https://drexel.mywconline.com.

Please note that sessions are 45 minutes long. The last fifteen minutes of the session are reserved for the reader to complete paperwork.

Drexel University
Electrical and Computer Engineering Department

Multi-level Tracking for Dynamic Augmentation of Live Performance

Senior Design Group 6
Final Report
May 17, 2013

Team Members:
Manu Colacot, ECE, m.colacot@drexel.edu
Julian Kemmerer, ECE, jvk@drexel.edu
Matthew Zimmerman, ECE, mzimmerman@drexel.edu

Submitted to:
Dr. Youngmoo E. Kim, ECE, ykim@drexel.edu

Abstract

are

Performances aim to captivate an audience. In recent years, this is often accomplished using various forms of technology. A simple and nearly ubiquitous example of this is visualizations shown on projector screens behind performers. Such visualizations are often pre-planned and scripted, leaving little room for creative expression during the performance itself. In order to address this issue, we have developed a system that allows for dynamic manipulation of various performance effects: both audio parameters and visual parameters.

To give context, many performers are already very emotive while on stage. They naturally make expressive gestures during the course of their performance. We have developed a system that takes these ~~movements and~~ motions, and uses them to augment the performance in an innovative way. This type of motion based effects system is similar to the V Motion Project [3].

In most current performances, an individual's movement has a ~~narrow~~ minimal, if any, range of impact on sound and visual effects. ~~If any.~~ There are also certain audio and visual effects that are underutilized or completely ignored in the context of live performance - the most simple example being left/right audio panning. Audio and visual effects are typically relinquished to someone like an audio engineer or stage lighting director.

We have given some of this ability back to the performer through an audio and vision based tracking environment. This tracking consists of two levels, stage position and gesture/motion. The stage level tracking was accomplished through audio multilateration via masked signals within live audio. Gesture/motion tracking is completed through vision based Microsoft Kinect devices.

As of now, stage level tracking is able to determine the location of a performer within an lecture hall size~~d~~ room. The accuracy of this tracking is approximately **<NUMBERS!>**. Motion/gesture tracking is able to control relatively simple audio effects and allows for gesture based control of effects selection.

At this time we perceive ourselves as on schedule, though behind our original project goals. Due to only partial progress, our budget was significantly lower than expected. Major breakthroughs includes the transmission of gesture/motion information to a central computing location and fine tuning of the stage level tracking algorithms.

Table of Contents

*Done in LaTEX

List of Figures

*Done in LaTEX

List of Tables

*Done in LaTEX

Problem Description

As technology progresses, it makes its way into various performance applications. Many of which are for musical purposes and include everything from electronic devices such as effects pedals for guitars to computer software such as Pro Tools for music synthesis and manipulation. In general, ~~many~~ performers seek different ways to creatively augment live performances.

Most of these systems rely on a single sensor or set of sensors to provide measurement of a single physical parameter. However, we have developed a system that combines multiple large and small scale tracking methods to provide a unified, real time, multi-parameter data source available for audio and visual effects mappings.

which is

The existing methods for tracking performer position and gestures on stage are often inaccurate and underutilized. Therefore, the two goals for this project were to:

1) Design an accurate position and gesture tracking system for both the performer and stage scales that was as non destructive as possible.

2) Combine various sensor information from the different tracking methods to provide real-time, multi-parameter mapping data for audio and visual effects manipulation.

The purpose of the stage position tracking was to determine a performer's position on an indoor stage. This is a difficult problem, as current technology like GPS only works outdoors. Even the best indoor position location systems can take up to a minute and a half to successfully track a person's position - completely impractical for our purposes **[APPLE ARTICLE]**. Therefore we were required to develop a method for indoor localization using audio triangulation.

In addition to stage level tracking, our system also utilized the Microsoft Kinect in order to perform gesture/motion tracking. Because most performances usually contain more than one individual on stage, our system had to be able to accommodate multiple performers. Therefore we had to develop a way to get tracking data from multiple Kinects and then map that to various effects.

Both methods of tracking had to be integrated into one easy to use unit that made it possible to control both audio/visual effects and parameters.

To better understand the desire for an augmented live performance consider a simple example of the common concert venue. There is a static panning of instruments done prior to the start of the performance typically by a sound engineer employed by the venue. Once the performance begins it is difficult, if not impossible, for the performers to dynamically alter even simple parameters such as static ~~stereo~~ panning. Another example is that of a disc jockey (DJ), whose live remixes of audio require creative control of several parameters simultaneously. A system that yields additional control options like gestures and continuous motion tracking would vastly expand the capability of the DJ and would arguably have greater control than the discrete 'button pushing' nature of live remixing.

Mappings of simple and complex parameters like the ones aforementioned exist, however, these are poorly utilized or complicated enough to deter natural creativity. The emphasis of this project is not the development of specific mappings of positions/gestures to effects but rather the design of the system that makes such vast real-time data sets available to the performer(s) for enhanced expression. While some mappings will be designed for initial analysis and demonstration, the majority of work will be put towards combining multiple data capture systems into a single interface, rather than the implementation of novel effects.

Progress Toward a Solution

Stage Level Tracking

When the term began, we ~~audio localization~~ were ~~was~~ in the midst of resolving various problems with timing and signal accuracy for our audio localization algorithm. Because of this, the system was not fully functional and ~~was~~ still very ~~error prone~~. Since then, a fair amount of progress has been made by using ~~use~~ the Fast Fourier transform (FFT) to confirm the presence of an audio signal ~~that an audio signal is present~~, followed by appropriate windowing, and cross-correlation to accurately locate the origination of the signal in the time domain.

As with the other portions of this project, a system that integrates Open Sound Control (OSC) **[4]** packets has been implemented such that a central server can receive the information from any device that is capturing audio localization information. OSC is a convenient ~~format~~ networking ~~packet~~ format thats acts as a simple and cross platform container for encoding and decoding ~~of~~ information. These packets are sent over a wireless network to the server, which then processes them in accordance with the functionality desired by the performer.

Gesture/Motion Level Tracking

At the start of this term a stripped down version of an 'off-the-shelf" gesture recognition system, implementing dynamic time warping [5], for the Microsoft Kinect was in use. This reduced complexity system was combined with the continuous motion portion of this project. The resulting gesture system ~~allows for~~ allows performers to change effect parameters ~~through~~ by using gestures alone.

Figure 1: Microsoft Kinect skeleton data visualization.

The information captured through the Microsoft Kinect vision system is presented in the form of a software development kit (SDK) **[2]**. This SDK makes skeleton information readily available for programmatic use. **Figure 1** is a visualization of the skeleton data as provided by the SDK. This information is wrapped within OSC packets and transmitted to a central processing PC. The OSC protocol is capable of utilizing UDP packet transport - optimal for streaming data ~~in~~ as close to real time as possible. Currently, the gesture/motion tracking system can stream live, low latency, skeleton information to a central computing PC for effects control and process gestures that control effects mappings ~~themselves~~.

Integration

At the start of this term the integration portion of this project was largely completed. Communication between various devices and platforms was finalized and implemented. The major task that remained was processing the ~~information being transmitted~~ transmitted information. As mentioned earlier, communication to and from the the stage level tracking is completed using OSC. It was important to select a coding framework that allowed for easy use of OSC libraries as well as had significant audio and visual capabilities. OpenFrameworks was selected for its simple, cross platform, and capable libraries. Currently a relatively powerful desktop computer serves as the central server where information streamed from all levels of tracking is processed as well as where all audio and visual effects are manifested.

The graphical user interface (GUI) capable of manipulating these incoming data streams has undergone significant changes. A user is now able to control both audio/visual effects as well as the masked 'chirps' used in the audio localization. As an update from last term, we have purchased an audio amplifier that will be used for demoing our final solution.

Constraints

Economic

The economic constraints of this project originate mostly from the relatively high cost of audio amplifier systems with multiple channels of audio and the multiple Kinects needed. A detailed description of exact costs is given in the budget portion of this report.

Manufacturability

This project does not require any additional items to be manufactured or any additional manufacturing steps. The 'custom' portions of this project ~~all~~ are all software based.

Sustainability, Environmental

This project requires ~~little~~ slightly more electricity than a typical consumer home theatre setup. However, it is possible that if this system was in constant use that, the processing resources needed by the central processing server could cause ~~an~~ increased ~~the~~ required power of the system (as compared to a 'typical' desktop).

Ethical, Health and Safety, Social, Political

It is difficult, albeit not impossible, to construe social or ethical impacts from this project, specifically as the impacts have yet to be truly evaluated outside of the prototyping environment. However, it is not outside of reason to speculate on the societal impacts of a *live performance enhancing system* in general. Of course these societal impacts are limited to the musical performance but, in a grandiose sense, this system could forever change the expectations of performance goers and begin a new era of experimentation with live performance ~~for artists~~s. It is possible that performance venues will begin to install permanent replications of our system in an attempt to provide the option of these augmented effects to every performer. It is even possible that the audience could be tracked and their actions collaboratively included into the artist's performance.

Budgets

*The side by side table comparison should be done in LaTEX, in additon to regular budget tables
Add items for zero cost Microsoft Kinect SDK and openFrameworks
Add Macbook (extra laptop) to budget if not there
Limit the number of kinects and laptops now to just 2?

The industrial budget for the Spring 2013, **Table 1** differs only slightly from the budget proposed in Winter 2013, **Table 2**. Most major components have been identified and purchased. The first major expenditure was the **Denon .....** amplifier system. This device is needed to support multi-channel audio. This is required for both proper audio effect reproduction and audio multilateration.

A new desktop computer was not purchased. Instead, a new laptop computer was purchased. This laptop computer is used to control one of the two remote Kinect devices. The central processing PC was borrowed from a team member for the duration of this project. This central PC is able to take in data from both the stage level tracking and gesture level tracking, map that data to certain audio and visual effects. The current budget includes **<NUMBER!>** Microsoft Kinect devices as it was determined that at this time only **<NUMBER>** are needed.

Work Schedule

Stage-Level Tracking (Matthew Zimmerman)

Task: Tracking performer location within a stage sized space using audio.
Completion Status: Partial.

The challenge when accomplishing stage level tracking currently is obtaining a reliable and clean signal from a microphone, while also keeping the microphone mobile and easy to use. By collecting recordings from several locations within Mitchell Auditorium in the Bossone Research Center, it was determined that simple iPhone microphones performed equally as well as stand-alone USB microphones connected to a laptop. Previous versions of the multilateration algorithms were unable to deal with the iPhone's automatic leveling/equalization. However, the most recent implementation of the algorithms are not affected. As mentioned the multilateration algorithms have been refined and the most recent versions, rely on both the time and frequency domains, in conjunction, to determine the source speaker and location (in time) of the audio signal. Future work will focus on the remaining refinements needed as well as streamlining the computational complexity of these calculations for easier use on less capable devices like iPhones. As of now, stage level tracking is able to determine the location of a performer within a stage sized room. The accuracy of this tracking is approximately **<NUMBERS!>**.

Watermarking (Manu Colacot)

Task: Embed localization signals within live audio.

Completion Status: Partial.

In order to implement audio triangulation without interfering with the performance, audio watermarking must be utilized. This is the process of embedding audio information into a signal so that the embedded audio is indiscernible through normal auditory processes **[6]**. In this way the speakers can output both the audio produced by the performer and also the 'silent' chirp, that is needed for the triangulation. A brief literature review on this topic was performed. Though live demo worthy code has not been fully implemented, a proof of concept MATLAB script showing that audio can be hidden in an actual song was developed. Several simulated tests were completed using this system in cooperation with multilateration algorithms.

Gesture/Motion-Level Tracking (Julian Kemmerer)

Task: Track gestures and continuous motions of performers.

Completion Status: Partial.

Figure 2: Microsoft Kinect field of view

Several challenges facing gesture and motion tracking were overcame. The first being the computation complexity of gesture tracking. Rather than continuously tracking many possible gestures to control numerous audio effects, gestures are utilized to switch between audio effects. This limits the number of gestures that need to be compared within the gesture recognition algorithms that run at a minimum of 30 times per second (as to keep up with the Kinects' data stream). This lessens the amount of processing required at any given point.

The second major hurdle was dealing with several Kinects simultaneously. **Figure 2** shows the relatively limited eld of view for a single Kinect device. In addition, there is a Microsoft implemented restriction limiting the tracking to two skeletons simultaneously (however, to maintain the greatest level of detail, we limit our Kinect devices to only one performer). Thus, our system requires multiple Kinect devices to function. Two options for handling the massive amounts of streaming data from multiple devices were evaluated. The first option was having a extremely power central server with direct USB connections from each Kinect device. Even if the system was able to handle those data rates and process efficiently, there was still the issue of USB cable length. These Kinect devices will need to be spread across stage-sized areas. The maximum length for USB cables is a bit over 15 ft - no suitable for our task. The second option evaluated was passing Kinect data over the network via OSC packets. The possible issue with technique is the delay incurred via network transfers. However, after testing this implementation the delay in transit was negligible and and unnoticed even with processing delays of the central server. This option was chosen.

GUI (Manu Colacot and Julian Kemmerer)

Task: Allow users to easily manipulate effect mappings.

Completion Status: Partial.

A simple GUI was designed and implemented in order to organize both stage and gesture level tracking. The GUI allows the user to choose joints he or she wants to track on the Kinect, and what that tracking will map to. In addition, the GUI can be used to display performer location on stage. Several GUI frameworks were evaluated with the most important factors being, 1) compatibility/communication with C++ (used for openFrameworks audio/visual effects and OSC network communication), 2) appearance, 3) ease of use. Originally Objective C was considered, though communication with openFrameworks became an issue. C# along with the Windows Presentation Foundation (WPF) framework was used as the GUI structure is familiar to most users, the codeset is well documented, and communication with C++ code is easily implemented.

Gesture Mapping / System Integration (All members)

Task: Create and map novel gestures/motions to audio/visual effects.

Completion Status: Partial.

With remaining time on the project the team as a whole is investigating novel mappings for various gestures and motions. Though not the focus of development within this project, these mappings are certainly necessary to make it successful. Having a set of 'demo features' that the team is able to pick from, but are not entirely limited to, will allow for better testing and also help others understand the overall design.

As mentioned, integration of all levels of tracking is quite important for the most part completed. The team is currently working to make final changes to the design to allow for data to be seamlessly transferred between Kinect devices, iPhones, and the central processing server. The challenge in this is that these two systems are, and provide, very different sets of data, so communication between them, in an orderly fashion, is integral to the overall success of the project.

Timeline or Gantt Chart

Teamwork

The teamwork section must clearly and concisely state the responsibilities of each team member and his/her contribution to the senior design project. You can convey this information in a method of your choice, i.e. text, **table**, etc.

***Insert Table Here**

Summary and Conclusions

We have developed a system that gives performers the ability to control various audio and visual parameters live while on stage. This was accomplished by developing technology that performed indoor localization while also tracking human motion data, and then mapping both parameters to specific effects.

By creating a system that provides performers with multi-parameter real-time data, we believe that we have developed a platform to not only encourage, but enhance creativity in live performances. Performers now have extra degrees of freedom that include both their physical location on stage as well as different joints on their body to control audio as well as visual parameters. Ultimately, this increases the creative possibilities of live performance.

References