An Analysis of Audio Mixing Console Designs, Recording Software, and a Look at Future Interfaces Using Intuitive Input Technologies

Benjamin Guerrero

Submitted in partial fulfillment of the requirements for the
Master of Music in Music Technology
in the Department of Music and Performing Arts Professions
Steinhardt School
New York University

Advisor: Agnieszka Roginska

May 3, 2012

ABSTRACT

This thesis examines the intuitive input technologies available today and explores the future of these systems in existing musical environments. The core of this thesis is a proof-of-concept interface design that departs from analog hardware imagery to fully take advantage of these new technologies. Sources will be used to analyze the historical progression of analog equipment and apply that progression towards possible endeavors and evolutions of digital and virtual audio mixing environments. The analysis will discuss the parameters expected by audio engineers in a modern studio, the available intuitive input technologies available today, and the practical application of an intuitive input controller in the studio, addressing both the possible as well as anticipating opposition towards the new technology.

ACKNOWLEDGEMENTS

I would like to dedicate this thesis to my parents, Roland and Karen. Without their influence and guidance, I would not have been able to accomplish as much as I have. I was fortunate enough to grow up in a household where creativity and independence was encouraged. For that, I am incredibly grateful.

I would also like to thank my father and Donald Bosley for editing my thesis. Their input allowed me to better focus my thoughts and words.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

One of the main goals for this thesis is to design and demonstrate a new software interface for audio mixing that could be perceived as a more efficient and intuitive design than what current audio hardware and software offer today. In order to do this, previous and current designs of mixing boards must be researched to understand what works well, what changed over time, and discuss design choices that may no longer be necessary. Next, intuitive input technologies will be examined to decide what should be used for the final prototype. The technology used will fundamentally dictate the overall design and user experience for the prototype. Finally, the new interface design will be discussed critically to figure out its place among current software and hardware setups and future designs will be discussed theoretically.

Ever since Thomas Edison's first recordings on his mechanical phonograph cylinder in 1877[19], people have been fascinated with the idea of capturing and reproducing sound. The mass production and distribution of audio equipment in the 20th century allowed recorded sound to be used for music, art, communication, marketing, propaganda, archival purposes, and much more. Through technology, combining audio signals together has become easier, more accessible, versatile, and complex.

The reason for the added complexity is the intricate signal flow of the audio. Recordings are no longer a simple transducer straight to a medium like vinyl or tape. Today's professional studios use computers, outboard gear, patch bays, MIDI, plug-ins, sample libraries, and analog-to-digital (A/D) converters that must all communicate via cables, software, and protocols. Electrical signals and binary information must work harmoniously together in order to mix audio. For musicians, the process of mixing music can be seen as arduous and unintuitive compared to creating music with other musicians. For producers, it can be seen as a time-consuming, yet creative process of aural exploration. For audio engineers, mixing can be fun and artistic, but overall, it remains a process with a steep learning curve and generally requires years of experience.

One of the many goals of this paper is to demystify the audio mixing process and explore optional avenues for non-engineers to mix audio as well as alternative designs that might be better for everyone at all levels of experience. All of the research done looks at history from a designer's point of view. Ultimately, the design of the equipment influences the creative process of manipulating sound. Similar to the way instruments are designed to produce sound in a specific way, the idiomatic workflow of mixing is designed a certain way.

1.1 Why is design important?

"*Good design should be innovative. Good design should make a product useful. Good design is aesthetic design. Good design will make a product understandable. Good design is honest. Good design is unobtrusive. Good design is long lived. Good design is consistent in every detail. Good design is environmentally friendly. Last but not least, good design is as little design as possible.*"[22]

-Dieter Rams

The words of Dieter Rams, the legendary industrial designer for Braun, are known by many designers as the ten principles of "good design." Design is important for audio equipment because manipulating audio is an abstract process.

Whether someone is a musician with an instrument, a foley artist with a field microphone, or a producer finding that "new sound," both knowledge and creativity are necessary to recreate the sounds that they hear in their mind. Yes, both skill and knowledge are necessary to use audio tools, but the tools should not be a hindrance to the task at hand. This goes for all stages of audio production, from capturing the original signal source, to mixing, collaborating, and to final distribution. Hardware and software can ideally work together to assist in the creative process. This is not always the case, but just like any technology, it remains a work in progress.

The graphical user interface (GUI) of computer programs plays an important role in human-computer interaction. The GUI affects the aesthetic and creative appeal to the user when interacting with the software. The user's workflow is sometimes dependent on the flexibility of the GUI. The inflexibility of a GUI design is an opportunity for competing software companies to introduce new features that may please users. Current recording software is heavily based upon analog audio hardware in terms of GUI design because it has allowed for an easier transition from the physical equipment to the computer.

Over time, the design of analog mixers became more efficient by adding more channel strips in a smaller amount of space. However, this design is not as efficient when implemented in the digital, virtual realm, especially when a keyboard and mouse have to be used. Current recording software implements an audio mixer that graphically simulates the look and feel of an analog mixer. This design choice is known as skeuomorphism. A skeuomorph is a derivative object that retains ornamental design cues to a structure that was necessary in the original. For example, artificial film grain or a hopper on a digital slot machine is considered a skeuomorph.

For the most part, skeuomorphs serve little purpose other than fashion design rather than practical design. Some digital audio effects feature screws and handles in the graphical design. While these efforts demonstrate a great attention to design detail, they serve no functional purpose. Skeuomorphs can make new digital emulations look comfortable and familiar, but they can also hinder or limit a design. They often result in poor use of space, have the potential to look "dated" sooner, and exhibit inflexibility in regards to innovation.

It's like adding a rotary dial on an iPhone. At this point, novelty has taken over function. For the sake of mere familiarity, anachronistic analog imaging has been automatically and unnecessarily adapted into digital interfaces. The satirical rotary dial, shown in *Figure 1*, is a spoof of what would seem to be an obviously absurd adaptation, yet the same kind of adaptations persists in digital audio interfaces without any question.

Figure 1. Rotary dial on an iPhone.[34]

Just as the Internet has changed the way information is exchanged, multi-touch, along with other technologies, are changing the way humans interface with computers. Using fingers and gestures can be a far more intuitive method of input than a mouse, cursor, and keyboard, especially for creative applications. With a little bit of ingenuity, new GUI designs can be created that are conceptually independent of analog equipment and introduce a unique method of mixing that is both efficient and intuitive for the user.

There are a lot of different hardware and software solutions available for audio engineers today and various reasons why one chooses a particular combination over another. Familiarity is a big factor, but new software is always experimenting with different kinds of workflows. We all think differently and, therefore, there is no one design that works best for all circumstances and for all people.

The following analysis is neither meant to be purely empirical nor concrete, but rather theoretical and observational. The main concept that will be discussed is designing an interface that takes full advantage of the technological medium in which it is implemented. There are a number of necessary, evolutionary hurdles that must be crossed before a singular design becomes the norm that people can agree is more ergonomic, efficient, and intuitive.

2. HISTORY OF THE ANALOG MIXER

Throughout history, the design and process of any man-made tool was created within the limitations of the technology and concepts of the time. As scientists better understood the world around them and inventors developed new methods of implementation, designs would progress in a "better" direction, at least, better for some. Looking at the equipment used today, it is difficult to fully appreciate the arduous developmental transitions of the design process to get to where we are today. One can look at a picture of old audio gear and glamourize the past because of the sounds that were made, but at the time, creative engineers using this same audio equipment were constantly faced with technical limitations to the point of having to develop techniques to work around them with "studio tricks." Things that are taken for granted today were eventually added to newer designs so that the creative process would become easier and less laborious. This same creative fluidity is the essence of our contemporary experience of progress.

Mixing boards of the 1960's almost look foreign compared to the design of boards today. A single line amplifier was as large as a tissue box because of the size of the valve amp required for the amount of gain and impedance it was creating.[35] Eight of these would be installed underneath the console on the side. They had to be easily accessible because the valves overheated and/or blew out frequently. The knobs, dials, switches, and sliders used in the early 1960's were quite large and inefficient. The boards only had room for eight microphone input amplifiers, pan pots, equalization (EQ), and a 4-track output. Outboard equipment like compressor/limiters and EQ were more than twice the size of rack mounts today. The machinery used for echo chambers required their own room. The hardware looked similar to equipment used for the Manhattan Project during World War II, as shown in *Figure 3*, but artists still produced excellent recordings with them.

Figure 2. EMI REDD.47 line amplifier.[35]

Les Paul was an important early inventor for analog equipment and recording techniques.[12] He invented the solid-body electric guitar and was known for experimenting with recording techniques such as overdubs, tape delay, phasing effects.

Figure 3. The REDD.51 mixer in Studio Two of Abbey Road Studios, shortly after it was installed in 1964.[35]

He had Ampex build the first 8-track tape recorder in 1954, which laid much of the foundation for analog technology for the following decades.

The Beatles were lucky enough to work closely with EMI engineers while recording at Abbey Road Studios. As their creative minds pushed the limitations of the studio equipment, new workarounds were made and eventually new consoles were created for them. Between the Beatles, their producer George Martin, and EMI engineers Norman Smith, Ken Townsend, and Geoff Emerick, innovative studio techniques were created such as tape loops, double tracking, vari-speed recording, unconventional microphone placements, and various sound effects and unique instrument combinations.[35] Creative minds will always create beyond whatever perceived limitations that are around them.

Meanwhile, in America, recording engineer, Tom Dowd, was pushing for Atlantic Records to install the eight-track recorder, making them the first record company to offer multi-track recordings.[27] Dowd is also credited for commercializing and popularizing stereo sound. Multiple advancements were made during the 1970's including the first 32-channel mixing consoles. These boards featured EQs and FX sends on all channels. Inline mixing consoles gave the ability to route signals from one channel to another. During this time, famous mixing console manufacturers debuted, such as AMS Neve, Solid State Logic, API, Harrison, and Raindirk. In 1977, Neve developed the first computer-controlled moving fader automation called the Neve Computer Assisted Mixdown (Necam).[1] In 1979, their 8108 board featured audio and digital technology and was the first console to have assignable controls and memory faders. These early designs demonstrated the power and flexibility that digital technology could contribute to audio mixing.

With any complex recording process, a clear and efficient workflow is necessary at the professional level. Over the period of a few decades, analog equipment evolved and transformed, as newer and smaller vacuum tubes were used and smaller, more versatile circuits were built. As designs became more compact, more tracks could be installed in a mixing board and more outboard gear could be mounted on racks. All of these changes occurred through the process of critical analysis and adapting to new technology as it became available, all while keeping a balance between form and function. This is all evident in the largest and smallest mixers today. The channel strip on a mixing board became very ergonomic, efficient, and comprehensive compared to early boards. *Figure 4* clearly shows the signal flow from input to output. This standard design is proven and is incorporated into most modern mixing boards.

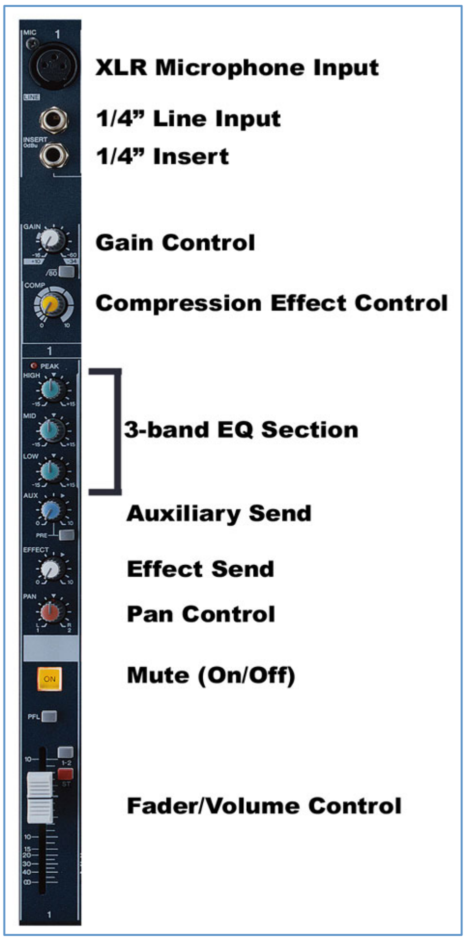

Figure 4. Modern channel strip on a mixer.[6]

Analog equipment has inherent noise in its systems. Transistors and vacuum tubes that are sensitive to things like electrical currents and temperature add artifacts to signals. These anomalous sounds are known as non-linear distortions or "analog warmth." As the compact disc and personal computer became popular in the 1980's, so did the digital audio format for its subjectively "cleaner" sound. As A/D converters were added into the recording workflow, mixing boards had to be able to communicate with computers, recording software, and other digital hardware. To this day, analog and digital effects and components are used together to record music.

3. DIGITAL AUDIO WORKSTATIONS

The main purpose for digital audio workstation (DAW) software is to mix, arrange, and edit digital audio tracks. These are the same functions of a multi-track tape machine, but with digital samples rather than analog tape. The controls are a keyboard and mouse instead of knobs, buttons and knives for cutting tape. Computers started to be implemented in studios in the mid 1980's. They gave recording studios greater flexibility and potentially cleaner signals than analog systems were able to offer. The main DAW software used by professionals today are Apple's Logic Pro, Avid's Pro Tools, Mark of the Unicorn's Digital Performer, and Steinberg's Cubase.[31]

There are many advantages of recording audio using computers rather than analog mediums including visualizing waveforms, non-linear editing, copy & paste, edit-undo, and hard disk drives for storage. The edit/arrange window allows one to quickly combine multiple tracks and "takes," as well as get rid of unwanted audio. Digital effects are more expansive and offer more control than analog effects. The debate over the differences in aesthetic qualities of these systems is a complex issue that will not be covered here.

In the end, all DAW software packages essentially do the same thing: mix audio. They all have very similar layouts as well. In order to make the transition easier for audio engineers and to make the most sense at the time, the software had to imitate the hardware. By maintaining a similar look and feel, the adoption rate for these digital systems rose rapidly. However, even though software programmers went to great lengths to imitate the appearance of analog mixers, the lack of unique designs specifically created for all-digital environments has resulted in user experience issues, which have continued after 25 years of efforts to adapt analog modalities to the digital realm.

To maintain a competitive edge, new features and functionality in updated DAWs are marketed as unique and easy-to-use. It is fair to say that the audio applications available today are their own best versions. The most convenient feature that has been added in the latest versions of the previously mentioned software is creating a comp track. This allows a user to record multiple takes in a folder and linearly select portions of each take. The software automatically crossfades between the selections and flattens it to a single, newly edited track. Even a few years ago, this process was done in the same way that analog tape had to be edited. In the same fashion that razor blades were used to cut and splice sections of tape together, this way of thinking was reflected in the digital process of cutting and pasting individual sections at a time. Automating this process and simplifying the selection to be similar to highlighting text created a quick and intuitive design alteration that users would unanimously agree is "better" than the older way.

Unfortunately, not all new features are easy-to-use or are even necessary. For example, most DAWs attempt to create music notation from MIDI performance data, but they require lots of tweaking with rhythmic quantization and enharmonic or accidental spellings. Sometimes the MIDI performance doesn't directly relate to the way music is notated. Even then, music notation software, like Finale and Sibelius, have more options for customizing the notation. Some new features are good in theory, but not always in practice. These growing pains and the feedback from the community help weed out the changes that are deemed "good" and "bad." No one knows the best way or most intuitive way of doing something because there is always another way, a different way of doing the same thing. However, these subjective measurements do plot trends and movements in the industry, but at what point do these movements run their course and necessitate the exploration of a new direction? Who is to say whether or not the current direction is the "right" direction?

One current problem that most audio engineers could likely agree upon is the limitation of computer screens, especially when mixing without a mixing board. With a large, high-resolution monitor, about 24 tracks can fit in the edit/arrange window and 24 tracks can fit in the mixing window, but not comfortably. The tracks are squished together so text is abbreviated and there is a lack of clarity, particularly in the waveform. The view scale and the size of the tracks constantly have to be resized to move between a micro and macro view of the mix. Extra tracks are hidden elsewhere and the window has to be scrolled left and right or up and down to see everything. Either the edit/arrange window or the mixer can take up the whole screen so an additional monitor is often necessary to see both at once. Even then, it is not enough screen real estate. What is the solution for this issue?

4. RECORDING ENVIRONMENTS

There are 3 main types of recording environments today: the home studio, the professional recording studio, and live production. The setups in these environments vary from simple to complex. The simplest is the home studio environment. This could consist of a computer and an analog-to-digital (A/D) converter. Recording software today is capable of successfully emulating musicians, analog mixers, and outboard gear at a professional level. This has been most helpful for composers that can use high-quality sample libraries to create entire orchestral scores that may be used in movies, video games, and popular music. When home studios are expanded, it is common for equipment purchases to focus on versatility, rather than the sound quality, so that the space required for equipment is at a minimum. Eventually, with enough investment, home studios can look like a scaled down professional studio.

The professional studio market is very competitive and very expensive. Artists go to pro studios not just for the highest quality equipment, but also for the acoustical properties of the studio space, as well as the knowledge and experience of the engineers. In the US, the best studios can most likely be found in New York, Los Angeles, or Nashville. These locations have a high concentration of professional artists seeking the highest production value for their music. These studios spend millions of dollars to get the largest facilities with proper acoustics and update their equipment frequently to have the latest gear for the artist's to find "new sounds." Acoustically treated ensemble rooms and multiple isolation booths can completely change the sound of a mix that no hardware or software can reproduce. The control room usually has a large selection of outboard gear to give the producer the freedom to experiment. The studio usually owns expensive microphones and A/D converters to capture the cleanest sound source possible. Once all of these ingredients are together, a house engineer uses them to lay down the best tracks possible for the artist.

These studios are not available to everyone because of the expense involved. Generally, a record label is required for an artist to afford the necessary studio time. In the movie industry, professional studios are almost always used, especially for multi-channel mixes. The equipment necessary for surround sound mixes is often outside the budget for home studios. This expense gap is one possible source for the slow adoption of 5.1 speaker systems. If it were more affordable to create multi-channel mixes, there would possibly be more content created that would encourage the necessary speakers and receivers for proper listening.

Live audio production is very similar to professional studio production, but with a different focus. Because of the amount of sound that is generated on a stage, live audio engineers must maintain a strict balance between signal-to-noise ratio and feedback. Cheaper, more compact, and more durable equipment is used on the road compared to studio equipment. However, large production tours often strive to recreate the "studio sound" in a live performance. Digital plug-ins are generally the middle ground for this goal. Newer digital boards are able to handle the exact same effects used in the studio. If the effect is not possible live, samples are often incorporated into the production.

When using this equipment in any environment, the audio engineer often feels as if they need to be in two places at once. To solve most audio problems, the engineer cannot be both behind the mixing board and solving the issue at the same time. Yes, assistants and stagehands do help, but it is still a frustrating process. It would seem that a more flexible and/or mobile alternative interface might benefit these recording environments.

5. INTUITIVE INPUT TECHNOLOGIES

Inventions, such as the digital computer and telecommunications, have been able to completely transform the world we living in. The word "invention" is used here to mean the inception of an idea or concept, not just a singular product. The car was invented and changed the way people traveled. A modern vehicle has tens of thousands of patents connected to it. There have been amazing advances in engine designs and new materials that were all created to make a better car, but new industries are not usually created around a new engine or a new material. They are supplemental technologies that are created to solve a specific problem. Entirely new inventions that completely change the way people think can have an uncontrollable amount of momentum with unfathomable possibilities.

One common argument is whether or not a new product is revolutionary or simply evolutionary. Planned obsolescence has existed for a long time and has ultimately stunted the rate of growth for new technologies.[5] It's all a part of calculated business strategies where the bottom line is the top priority. For the most part, these technologies have been on a set path for a very long time, which is "MORE!" The expectation has been more speed, more resolution, more quality, more graphics, more memory, more efficiency, more buttons, more features, more effects, and basically more of everything. Most of this "more" mentality has been created by consumers. They use a product to its limits and come to the conclusion that more of the same thing would solve whatever limitation they had found. This is not always the case. When a certain level has been reached, something entirely different is necessary to overcome an obstacle, not more of the same.

One adage that is attributed to Henry Ford, even though he never said it, is "If I'd asked customers what they wanted, they would have told me, 'A faster horse!'"[24] Instead, Ford gave people the Model T. Steve Jobs often used this quote when asked how he came up with new products. The whole idea behind this statement is vision; having vision beyond the now. People's preferences are based on familiarity and that's fine. By not taking big steps, companies are able to maintain a customer base and create sustainable growth. However, there comes a point when pushing the limits is no longer satisfactory. Sometimes all that is needed is an entirely new design and approach to expand the growth and market. Complacency has historically inhibited progress. New isn't always better, but inadequate designs are worse than not attempting innovation at all. How does one think outside the box and find a possible solution for a problem when some people don't even think the current design is a problem to begin with because it is familiar? Then, how would one convince others that this new, different way of doing something just might be better than what has been done for years or even generations?

Creativity, imagination, and a sense of adventure are necessary attributes for any inventor. The journey for creating something new is long and arduous, often times with many disappointments and frustrations. An invention can go through several prototypes before realizing the entire method is wrong because an unforeseen complication has emerged. Even when an inventor creates something that satisfies the original vision, there's no guarantee that users will appreciate and understand it. Dedicating hours and years of ones life into creating something new ends up being a huge gamble. How can anyone possibly know if the risk is worth taking?

Focus, dedication, and passion; these are some very key elements to successful designs. At the closing keynote of Apple's World Wide Developer's Conference in 1997, Steve Jobs famously said, "Focus is about saying no." Having a sense of direction helps weed out things that do not fit in a designer's vision. Not having focus can lead a designer or even a whole company in many different directions to the point where the total is less than the sum of its parts. Some truly great designs and ideas emit the passion that was put into making them in the first place, which is important. There have been countless new technologies created in the past couple of decades and some of them will never see the light of day outside of a laboratory. If it doesn't have that something special, it's going to be very difficult to convince others to believe in this new thing.

5.1 Naming the Future

In May 2007, Bill Gates and Steve Jobs got together on the stage of the annual All Things Digital D8 Conference, an event hosted by Wall Street Journal's tech columnist, Walt Mossberg. In this rare joint appearance, these two technology pioneers gave their insight on the future of the industry. At this point in time, the iPhone hadn't been released yet, but Jobs still had one in his pocket. When asked about future designs, they both gave vague answers, but what they did say revealed a lot about their scope of vision. Jobs kept using the term "post-PC devices" to describe things like the iPod, the iPhone, and the iPad. By 2007, he had already had the iPad in his lab for 2 years.[24] The iPad had 5 years of research and development before it was released.

"Post-PC devices" was a very controversial term because many assumed that these new devices were meant to replace desktop and notebook computers, but that's not what Jobs meant at all. He simply meant that there was a time where the personal computer was the main device for content creation and consumption. "Post-PC" simply means the period of time after the computer's lone technological dominance in the household. In terms of Apple's future, he was talking about portable, multi-touch devices that worked independently from a computer, as well as in conjunction with.

Gates, however, used terms like "new form factors" and "natural interface" to describe future designs. In terms of Microsoft's future product line, he was talking about products and services like Sync in cars, Microsoft Surface, and the Kinect for Xbox. Mossberg makes a point that all of these new technologies are essentially traditional computers in a different form factor. Then, Jobs makes a point that the purpose of it all is the consumer experience. Changing the computer experience is both evolutionary and revolutionary. After a well-tempered design is established, some experiences will be augmented and others will be replaced.

5.2 What is Intuitive Input?

Intuitive input refers to systems that are designed to gather information from the natural actions and movements of people. It is also referred to as gestural controls. The main purpose is to interact with a computer with the same finesse and nuance people have with any physical object. There is a strong connection between music and movement so it makes sense to use this interactive system for audio production and manipulation. Dancing is an obvious example of this music/movement connection, but when it comes to music performance, the subtlest of movements play an important role. The qualities of the sounds a musician produces while playing an instrument can be directly correlated with their technique. A lot of it has to do with physics and the specific instrument involved, but if a person's technique is poor, there's a good chance that their sound quality will be poor as well.

There are a number of difficulties that these systems must overcome before they can be used practically in any environment, much less a musical one. "First of all, human gesture has to be noise filtered, calibrated and quantized in order to have a normalized representation. Such data can then be analyzed in order to extract pertinent information. Human beings do not reproduce their gesture with precision and there is no formalization of movement that may help us to recognize its meaning. Gesture allows humans to interact with their environment, to modify it, to get information and to communicate. Gesture applied to a musical context may follow the same objectives: modifying musical or sound material, getting information on the performance and communicating information to other processes."[37]

Figure 5 shows a basic signal flow of these input systems. Obviously, a lot of steps are involved within each box and even in the arrows to get from one box to another. There are a number of different sensors that can be used to capture specific aspects of human movements. Some of these sensors can even be used in different combinations to increase the amount of data input from the user. Of the many emerging technologies currently available, MicroElectroMechanical Systems, Nintendo's Wii remote, Microsoft's Kinect, and multi-touch surfaces will be discussed here.

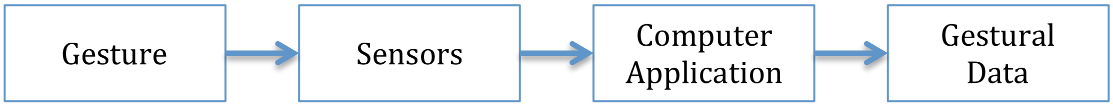

Figure 5. Typical workflow of a gestural based system.[36]

5.3 MEMS

MicroElecroMechanical Systems (MEMS) is a large industry that has grown at a substantial rate in the past few years. MEMS are small devices built at the microscopic level that are typically designed as sensors or actuators. These tiny sensors are built into phones, cameras, printers, projectors, cars, and many other consumer electronics. Their small scale and low power usage makes them ideal for portable devices. The main ones that will be discussed are accelerometers, gyroscopes, and magnetometers.

An accelerometer is, at minimum, a one-axis sensor that measures proper acceleration of an object. In the case of the iPhone, it measures the X- and Y-axis of the device to orient the screen based on gravity so that the user always sees the content upright. Because it measures acceleration, it also detects if the device is shaken forcibly. In modern notebook computers, an accelerometer can detect whether or not the computer is falling to the ground and quickly stop the hard disk drive from spinning to minimize the damage upon impact. Accelerometers are also used to detect collisions in cars to deploy air bags. In rocket ships, they are used to detect the rate of change of velocity.

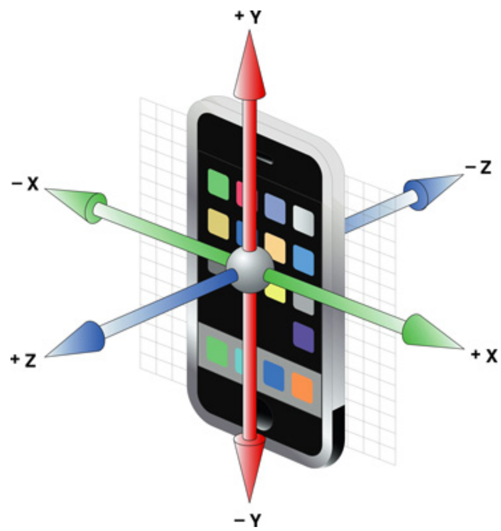

Figure 6. The iPhone divided by a 3-Axis chart.[33]

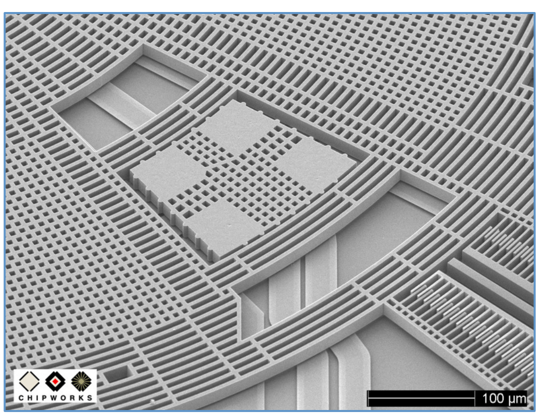

Figure 7. Zoomed image of a MEMS gyroscope.[23]

The iPhone uses a three-axis gyroscope to know its exact orientation in space. Pivoting around the X-axis provides pitch, the Y-axis gives yaw, and the Z-axis gives roll. This is used most often to enhance the video game experience on mobile devices. For example, moving the device could be used to realistically pilot an aircraft or steer a car and control the gas pedal at the same time. "Mechanical engineers measure the sensitivity of MEMS gyroscopes in mV/dps (millivolt divided by degrees per second) so the output of the oscillator (in mV) divided by the sensitivity (mV/dps) provides the angular rate applied to the package, in degrees per second."[23] As a reminder of how small these sensors are, *Figure 7* shows a zoomed image of a gyroscope designed by STMicroelectroincs, one of the largest manufacturers of MEMS technology. The image was taken with an electron microscope.

A magnetometer measures the strength or direction of magnetic fields. They have many archaeological, military, and other exploratory uses, but in mobile devices, they are mainly used as compasses. They can be used to display maps based on the geographical orientation of the user, but they can also be useful by displaying other relevant information on the screen.

5.4 Wii Remote

Nintendo released the Wii in November 2006. It's a video game system that was designed to broaden the demographic of typical gamers. Its selling point was the unique motion controller that allowed movement to be part of the gaming experience. Prior to this system, video games on consoles and PCs generally only used buttons as a form of input from the user. This enhanced interactivity then led to new gameplay designs. The Wii Remote combines buttons, an accelerometer, and an infrared optical sensor to receive multiple types of input from the user. The three-axis accelerometer detects acceleration along its axis to transmit the motion of the user. Triangulation is used to calculate the distance between the infrared optical sensor that is located on the front of the remote and a Sensor Bar, which features 10 infrared LEDs, that is centered above or below a television. The system knows where the user is pointing at on the screen based on how far the user is from the television and the angle of the remote relative to the ground, which can be calibrated. The Wii MotionPlus was released in June, 2009. It's an attachment for the Wii Remote that features a single-axis gyroscope and a dual-axis tuning fork gyroscope to capture complex motions of the user.

Figure 8. Wii Remote.[43]

Wii Music is a simple game released by Nintendo in October 2008. It allows the users to play virtual instruments by imitating the motions of a musician playing the actual instrument. There have been numerous experiments done with the Wii Remote when it comes to music performance and production. One notable research development is using the Wii Remote as a head-tracker for 3D audio.[41]

5.5 Kinect

The Microsoft Kinect is a 3D capture device for the X-Box 360 video game system. It was released in November, 2010. With the combination of software and hardware, the system can interpret a user's movements and translate them into commands. For hardware, the Kinect uses a color VGA video camera, a depth sensor, and a multi-array microphone. The camera is used for video monitoring and facial recognition. The depth sensor is implemented using an infrared projector and a monochrome CMOS (complimentary metal-oxide semiconductor) sensor. By projecting infrared light onto a scene, "the phase of the outgoing light wave may be compared to the phase of the incoming light wave to determine phase shift,"[26] which can be used to calculate distance even without a light source in the room. The microphone array utilizes four microphones to isolate the voice commands from the players from the room noise. This allows the players to use the voice controls speaking comfortably from several feet away.

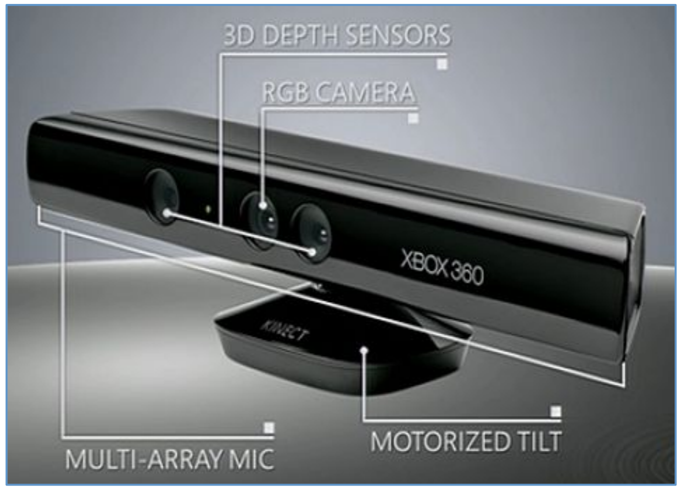

Figure 9. Microsoft Kinect[9]

The "brain" behind the sensor is the software that generates meaningful information from the hardware sensors. Before writing the software for the Kinect, machine learning was used to understand and categorize the human structure and typical human movements. By analyzing millions of people and their movements, using images, special camera rigs, or motion-capture, the computer is able to sift through huge data sets to establish probabilities and statistics about the human form. It's able to recognize a person's essential body parts and even extrapolates the rest of the body if an object or another player hides part of it.

When a user stands in front of the Kinect, it first determines physical distance and applies a rudimentary skeleton inside of the user's frame. Next, it determines the user's body type, current pose, and identifies about 30 body parts including the head, torso, knees, hips, elbows, and thighs. It does this based on its experience and knowledge of other humans it has examined. After establishing the body type, it applies a most probable pre-made skeletal structure based on pre-programmed formal kinematics models and overall experience. Finally, it creates a 3D avatar based on all of its analysis, which can then be dressed up with skin, clothes, hair, and other features. The Kinect does this 30 times a second![15] Since this process takes mere milliseconds, there's still enough time for game developers to use this information for gameplay.

In terms of musical applications, a simple YouTube search for "kinect music" will display thousands of videos of the Kinect being used in very creative ways. A majority of the demonstrations send MIDI data or Open Sound Control (OSC) information to control instruments or effects parameters in music software based on one's hand movements. OSC will be addressed to a greater extent later. Most of the focus in Kinect music development appears to be on interactive performance. Some of the more creative projects, such as "Mau5bot Music Sequencer" or "Mufin Vision," actually create entire 3D musical environments to explore and interact with.

5.6 Voice Control

Ever since Captain James T. Kirk, of the USS Enterprise (NCC-1701), interacted with the ship's computer using his voice on the 1966 TV show, *Star Trek*, people have been dreaming of talking with computers. There are countless examples of this in the science fiction and popular culture, so much so that we expect the characters of futuristic stories to talk with their technology. Why don't we have this technology yet? Why can't we tell our air conditioner to lower the temperature from anywhere in the house? Why can't we tell our computers to create a new Logic project with 8 mono tracks, 4 stereo tracks, and 8 MIDI tracks and see instant results?

Well, the technology is almost there. Microsoft has Sync in cars. Google has Voice Actions for Android. Apple has Siri for the iPhone 4S. Speech to text software has been available since the early 1990s. There has been a lot of progress made in the past 20 years, but there's still a long way to go before scientists and engineers make science fiction a reality.

There are two steps to voice control: speech recognition and translation. Both of these steps require extremely complex and comprehensive algorithms. The biggest hurdle for speech recognition is accuracy. Can the system identify each phoneme and each word? What happens if the user slurs, mumbles, or stutters? Does the system ignore a user's mistake or recognize a completely different word? For translation, once the system has recorded the user's words, does it actually understand what they said? Does it understand the vocabulary, sentence structure, and the context of what the user is saying? Can the system then turn those words into actions or a response that is relevant to what the user said?

Most modern speech recognition systems use Hidden Markov Models (HMM) to give each word or phoneme a value (ceptstral coefficient). The system can be trained automatically and is computationally feasible. Concisely, it connects each word one says to the most likely word in its database. Once the string of words is recognized, it has to interpret what was said. It would be a daunting learning curve if people had to learn a new way of speaking just for a computer to understand us. Ideally, the computer should understand the way we naturally speak. The problem is that natural language is implicit, highly contextual, ambiguous, and often imprecise. Numbers are always the same so arithmetic is easy for a computer, but words can have multiple meanings and can change with context or sentence structure.

IBM's supercomputer, *Watson*, is designed for natural language processing. *Watson* aggregates words into as many contexts as it can. Based on the words it is given, it generates probabilities for candidate answers it surmises. By running key words through complicated algorithms like temporal reasoning, statistical paraphrasing, and geospatial reasoning, context can be generated to narrow down the massive amount of information its server holds.[14] All the while, *Watson* learns from its mistakes to constantly refine its algorithms. This kind of machine learning and natural language processing can break down the communication barrier between humans and machines.

Now, not everyone has room for a server-farm supercomputer in their house, so companies that implement voice control systems into mobile devices send small data packets over the Internet to their own servers to do all the processing. This system works very efficiently to give a near immediate response to the user. As these systems get better at understanding natural language and the delivery process for these services become more widespread and standardized, voice control in future hardware and software will be more prominent.

5.7 Multi-touch

There are two main types of touchscreens: resistive and capacitive. The majority of early smartphones and pocket PCs used resistive touchscreens. To sense touch, they required a fingernail or stylus to apply pressure on the many layers that made up the screen. This often led to a frustrating user experience whenever commands weren't read because not enough pressure was applied or the response was just slow. The first tablet PCs came out in 2001, but never caught on for a number of reasons. In the end, it came down to user experience.

Capacitive touchscreens use a special circuit arranged in a coordinate system. Each point along this grid can sense changes when it comes into contact with a conductive material. This is why a fingertip is required to use multi-touch screens and why a stylus and gloves won't work. In the nanoseconds after the screen registers a touch, raw data around the touch is captured, background noise is removed, pressure points are measured, the approximate touch area is established, and finally, the exact coordinates are calculated. This touch data is then sent as electrical impulses to the computer processor for interpretation.

The advantage of this system is that it can detect multiple touches on the screen as well as movement. This new innovation opens up many new possibilities for user interaction and interface design. Instead of requiring a button for a simple task, it can be conveyed as a simple gesture. The screen can detect multi-finger taps, multi-finger pinches, rotation, multi-finger swipes, panning, and multi-finger long presses. This simple vocabulary can increase screen real estate by removing unnecessary buttons while simultaneously putting the user closer to the content with intuitive gestures. In terms of form factor, most multi-touch products come in the form of MP3 players, smartphones, tablets, or less commercially available as large televisions or tabletop screens.

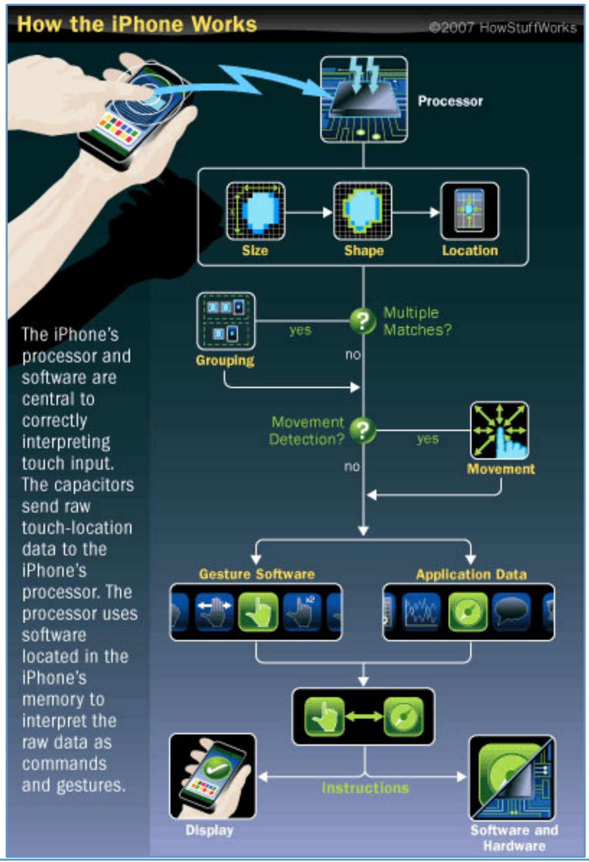

Figure 10. A flowchart of the iPhone's multi-touch screen.[44]

Between the Apple App Store and Google's Android Market, there are thousands of music apps* available for sequencing and performance. Some of these apps are designed to imitate analog hardware and others feature brand-new designs and concepts for creating and manipulating sound. Many popular apps can act as a controller that sends OSC and MIDI data wirelessly to computer software. The multi-touch screen allows a single finger to generate multiple messages at once based on its location on the screen and how it is moving. For example, the length of a piano key can send MIDI CC#11† (expression) and imitating vibrato similar to a violinist can send MIDI CC#33 (modulation wheel – fine). These types of designs give users more control over the sounds generated versus traditional hardware controllers.

* "Apps" is the abbreviation of applications. "Apps" generally refers to mobile applications and "applications" refers to software on a desktop or notebook computer.

† In addition to note value and duration, MIDI also sends Control Changes to give the user more control over the sound that is being generated.

5.7 Augmented Reality

Augmented reality (AR) is the technique of combining hardware and software to overlay virtual elements onto a physical, real-world environment. It is possible to enhance one's current perception of the world by displaying 3D characters or information from the Internet that are relevant to their current location. The idea of computers and technology modifying a person's environment virtually has been around for over 50 years in both science fiction and in early mechanical inventions, such as the failed Smell-O-Vision. Although technology is not at the point to fully escape into another world, like a holodeck[‡], technology is at the point of creating a window into a virtual world.

Multiple pieces of hardware are required to work together in order to create AR. Information from an accelerometer, gyroscope, magnetometer, and a GPS are combined to tell a computer exactly where a user is in the world and what cardinal direction they're facing. Then, a camera can be used to display the natural environment on a screen. Any new information that needs to be added will show up on the screen and correspond with what the user is looking at. For this to be practical, the device must be portable and be able to connect to the Internet to gather new information. At the moment, smartphones are the most cost effective devices with which to experience AR.

This setup can take mundane information and make it immersive and interactive. For example, an Internet search for "Italian restaurants" will give a person a list of places, their contact information, reviews, and even a map of where they are located. With AR, an individual can be in the middle of a city, on the sidewalk, look at a phone and see the restaurants laid out 360^{o} around them. A person can see exactly where they are in relation to the restaurants, touch the one that is closest and see all of its available information. Not only is the same information being used presented in a timelier and geographically relevant manner, it creates an entirely new experience.

The immediate uses for AR revolve around geographical data. Some video games have taken advantage of using the camera and reality to be the level design. For example, the game company, Parrot, has created an AR Drone. The game allows a user to fly a real, helicopter-like toy with a camera attached to it. The video feed is sent wirelessly to the software and allows one to shoot at virtual objects generated on the screen or even to interact with other AR Drones. In terms of music, there have not been many projects that take advantage of AR. However, it might be possible to create a virtual mixing environment.

[‡] A holodeck is a simulated reality facility from the fictional *Star Trek* universe. In this room, matter is simulated to recreate objects, people, sounds, and smells.

5.8 How to choose what technology to use?

With the variety of platforms and technologies available, how does one choose which to develop for? There are already thousands of music software and hardware products available so where would a new design fit in? Just like any design, it comes down to form and function and how it affects the end user experience. The whole setup required of the user has to be taken into account when making design decisions. All aspects of business should also be taken into account, such as target audience, market share, competition, distribution, projections, etc. A thorough business strategy should be in place before investing too much time or money into a new project. After moving through the phases of theory, research, and development, solid decisions must be made about the hardware and final design. A simple place to start is to compare and contrast the possible technologies that can be utilized.

It is very difficult to show all of the features and benefits of these different technologies on a single table so only the important highlights will be pointed out. Simply looking at the sensors side-by-side in *Figure 11*, there are noticeable advantages and disadvantages to using each one.

Technology	Pros	Cons
Accelerometer	• Widely available • Works well with movement	• Does not provide 360º information
Gyroscope	• Accurate 360º 3-axis coordinates	• Less Standard
Magnetometer	• Establish exact heading	• Susceptible to interference
Infrared	• Determines distance • Allows user to move freely	• Less mobile • Requires space • Imprecise • 130-200ms latency[4]
Multi-Touch	• Portable • Precise coordinate grid • As little as 1ms latency[3]	• Most screen sizes < 10” • Smudges on screens
Voice Control	• Convenient • Intuitive	• Often inaccurate • Not widely available

Figure 11. Table comparing sensors.

The devices shown in *Figure 12* utilize different combinations of these sensors.

Device	Sensors	Pros	Cons
Wii Remote	• Accelerometer • Gyroscope (adapter) • Infrared	• Large screen • Buttons w/ motion • Visible pointer • Haptic feedback	• Not portable • Latency • Difficult to distribute apps
Kinect	• Infrared • Camera • Voice Control	• Large screen • Entirely motion-controlled	• Not portable • Requires space • Latency
iPod Touch/ iPhone/iPad (iOS)	• Accelerometer • Gyroscope • Magnetometer • Multi-touch • Camera	• Portable • Large market • AR • Low latency • Multi-purpose	• Small screen • Smudges • Have to hold the device

Figure 12. Table comparing devices.

Obviously, interacting with music software on these devices is going to be a unique experience for the user, but at some point, the line between novelty and practicality has to be drawn. This goes for all new devices and technology. What demographic will utilize it and how broad or narrow is that demographic? Why are they going to use it? How are they going to use it? How will it benefit the user? Is this the "better" way of doing this task? "Better" is a subjective term/goal, but an important one. If the person that designed the software doesn't believe their design is the "better" way of doing something, then they are going to have a hard time convincing others of that prefernce.

Many of the musical projects using the Wii Remote incorporate simple movement recognitions to allow the user to play air drums or air guitar, albeit with noticeable latency. The more sophisticated systems which utilize head-tracking use the Wii remote backwards where the remote is the under the TV and the user has to wear a headset with multiple LED IR emitters. For some, this may be inconvenient. At E3 2011, Nintendo announced the Wii U, which should be released Q4 2012. It is specifically designed for AR by adding a camera and a touchscreen display in the user's hands, but there is still no formal distribution process for independent developers using Nintendo's technology. It is very limiting when only major game developers can afford to create for Nintendo.

Microsoft has the Kinect Fun Labs for developers to distribute, but this requires the Xbox. There is a lengthy process to link the Kinect to specific OSC parameters in certain applications on a computer. Even then, there is a lack of fine control over these parameters. When interacting with interfaces, broad movements work fine for the Kinect, but it is not as accurate as the Wii Remote's pointer or a finger on a touchscreen. Another downside to the Kinect is the amount of space required to use the system. Most home living rooms are not even large enough to use the Kinect properly. The couch usually has to be moved and it becomes an inconvenience. It would definitely not be ideal in a studio environment. At best, the Kinect is used for multimedia performances or installations. The technology represented in the Kinect is not likely to expand to widespread use by audio engineers or professional musicians.

The two largest issues with the Kinect and the Wii Remote are latency and distribution. The human ear can detect as little as 10ms of latency so when an interactive system has as much 200ms of latency, there's a distraction in the user experience. In terms of distribution, Apple's App Store has had over 25 billion downloads on its more than 315 million iOS devices.[1] Microsoft and Nintendo's sales are in the tens of millions, but neither directly supports an independent app developer ecosystem. It is for these reasons that developing an interactive music system for iOS is the most financially and technologically feasible answer.

In terms of iOS versus Google's Android operating systems, it comes down to personal preference. Both have similar technologies and distribution systems. However, Apple has been designing audio software infrastructures since the 1980s. CoreAudio is a much more robust and efficient audio system than Android's OpenSL ES. In terms of processing latency, "at a 44.1kHz sampling rate, a 256-sample buffer is about 5.8 milliseconds [on iOS]... On our latest Android devices, the minimum buffer is 16384 samples long which is about 371.5 milliseconds long."[20] The best Android devices run at 108.8ms. Between the two, iOS is probably the better operating system for specifically developing musical applications.

6. HUMAN-COMPUTER INTERACTION

Human-Computer Interaction (HCI) is the field of study that analyzes the relationship between people and computers. It combines multiple fields of study including computer programming, cognitive psychology, and graphic and industrial design disciplines. Poorly designed interfaces can lead to many mistakes and frustrate the user. Ergonomics are also important to consider whenever a device is used for an extended period of time. The multidisciplinary nature of HCI can benefit from a team of researchers with various backgrounds working together.

There have been multiple studies that have compared or at least addressed the need to compare the speed and accuracy of users with acquisition and manipulation tasks using Tangible User Interfaces (TUI), multi-touch, and a mouse.[16][17] [40] Multi-touch is a better interface than a mouse for bimanual tasks and similar in simpler tasks.[17] However, TUIs are still the best interfaces overall because they offer eyes-free manipulation and haptic feedback.[16] Additionally, multi-touch has a known issue with "exit errors."[40] Due to the angle and reach of a user's arm, the touch area of the finger changes and can introduce unintended extra input as the finger is lifting off the screen. Exit errors do not occur often, but they can still create a small noticeable change that affects fine-grained manipulations. Multi-touch is not the most advanced technology available and is not going to replace TUIs, but when combined with other sensors and technologies, it can offer a high level of versatility.

Apple's iPad has been chosen as the main device to develop the prototype interface design on because of the technologies it combines and its widespread use. Since the iPad's introduction in April 2010, it has reached virtually all levels of practical application in nearly every industry, including music. There have been many creative synthesizers designed for the iPad that professional artists perform with and completely new musical interfaces that have made certain tasks easier than previous technology. Some may see the 9.7" screen of the iPad as a problem that limits the amount of information that can be seen at any given time, but it can also spawn new designs by stressing the need for efficiency. One source of the perceived screen issue comes from superimposing analog hardware designs onto software and expecting it to translate in a digital world. Sometimes it works, but generally there is not enough screen real estate to see all the tracks and buttons at the same time, even on a high-resolution computer monitor. By creating an efficient, virtual mixing environment on the iPad, audio engineers would be able to visualize their mix and implement this new system into their workflow.

6.1 Mixing Intuitively

The Art of Mixing by David Gibson is an excellent first book for many audio engineers. It introduces the idea of visualizing an audio mix in 3D space. Each instrument track represents a virtual bubble that a mixer can control the shape and size of. The height and depth, for the most part, corresponds with the frequency spectrum and audible layers. This visualization method can also account for audio effects. *Figure 13* shows the snare reverb as a large box compared to the focused snare hit in the center of the mix. It is fair to say that most people, not just audio engineers, visualize a musical mix in a similar manner. People even describe sounds using visual adjectives: brighter, darker, wide, fat, tight, etc.

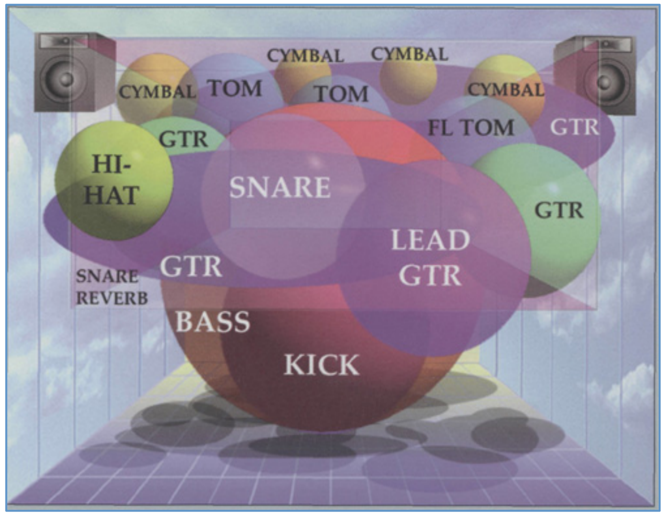

Figure 13. Example of a heavy metal mix from David Gibson's *The Art of Mixing*.[18]

Now comes the rhetorical question, "Why do we not mix audio the same way we hear it?" Looking at a mixing board, there is little resemblance to this visual process. The tracks are stationary, laid out numerically and horizontally. If tracks 1, 2, 9, 13, and 17 are the only audio signals panned to the center, it is difficult to tell by just looking at the mixing board. The panning knobs are small with just a little white tick to indicate the pan degree. With experience and an organized workflow, an audio engineer understands the best way to organize tracks to get a quick snapshot of groups. However, this process gets complicated for anyone when 48+ tracks have to be mixed. The clarity of the mix can be sacrificed because it becomes difficult to simply hear tracks panned on top of one another. A visual mixing guide could assist in seeing the overall stereo separation of the mix in an at-a-glance view.

Simply through creativity, Steven Spielberg was able to capture the world's imagination in his 2002 film, *Minority Report*. One goal of the technology used in the film was to create a realistic not-too-distant future. The iconic scene in the movie, where Tom Cruise is controlling a large curved computer display with hand gestures, influenced the development of real technology, as well as the technological direction

Figure 14. Tom Cruise using a multi-gesture interface in *Minority Report*.[39]

of other movies, as evidenced in the 2008 film, *Iron Man*. Combining the *Art of Mixing* visualization concept along with this touch-holographic computer system would create an excellent mixing interface. While this technology doesn't exist yet and current 3D capture systems aren't as accurate as they could be, multi-touch may be the next best thing for making this type of interface possible.

There are a number of objections with the concept of shapes, colors, and objects representing sound. In a room, sound waves propagate through air particles. It is all around us and cannot be seen by the human eye. By the time sound reaches our ears, multiple sounds sources combine to create a single composite complex waveform. Our ears and brain can then identify the separate sound sources by analyzing their location, timbre, envelope, and other sound cues. Sound is immersive where images are not at the moment. An image can visualize a sound source, but how can it represent its volume or location 3D space? How can images visually display sounds that are coming from all around you? How could these images be manipulated with gestures and provide relevant changes and feedback for the listener? The new interface created for this thesis is called MixPad and may provide solutions to some of these concerns.

7. MIXPAD

MixPad is a proof-of-concept iPad app influenced by the concept of touching and manipulating a visual representation of the audio tracks. It addresses multiple issues with hardware mixing consoles and DAWs and attempts to create a user experience that might be more intuitive and efficient than what has been available before. The main two issues that will be solved with MixPad is the visual representation of the mix and overall use-of-space. These issues are solved using spatial and temporal organization. Currently, mixing boards are organized by tracks numerically. This is important when keeping track of what microphone or instrument cable is plugged into which input in the A/D converter. When displayed on a mixing board or an edit/arrange window, they can be color coded to see groupings of tracks. For example, all the drum tracks are red and all the vocal tracks are blue. When dealing with a large number of tracks, there are often tracks that are hidden and have to be scrolled to because they do not all fit on the screen. Organizing tracks by where they are panned displays tracks in relation to each other and only displaying them when they are producing sound utilizes the space in the screen more efficiently and is conceptually similar to how people hear sound.

DAWs cover a wide range of features so only the main functions of the mixing board will be addressed. The main goals for designing this multi-touch multi-channel mixer was to establish a solid pan, gain, and metering system that was both unique and intuitive. *Figure 14* is an image of the prototype app that has been developed thus far. The images used were taken from Apple's Logic Pro application.

Instead of seeing "Track 1," "Track 2," or even "Kick," "Snare," the tracks are images of the actual instrument being played. These track icons graphically represent what was formerly a track on the mixing board. They can be grouped anyway the user likes, but naturally, the tracks are going to be arranged the way the user wants the mix to be heard or the way the instrument is naturally laid out and mic'd. Experimenting with the interface for only a few moments, anyone with any level of audio experience can learn the control parameters very quickly. Even someone with no audio experience could be taught very basic functions and use those skills immediately. This would be ideal for videographers that have to create their own audio mix. Move the track icon left and right for stereo panning. Pinch the track icon bigger or smaller to change the gain. Press the "Play" button and the user would immediately see the metering in action. These controls parameters are explained in detail below.

Figure 15. Screenshot of an 18-track mix using the prototype app, MixPad.

The iPad 2's screen resolution is 1024-by-768. For a stereo mix, the screen's x-axis is divided in half to represent the main left and right output. In the current design, the y-axis is not utilized, but at a later point it could represent a wet/dry reverb send. For now, the vertical space is just used for organizational purposes. Apple's pan parameter values range from -1 to 1 for a left/right mix. *Equation 1* is used to translate the gestural panning:

$$y = \left(\frac{x}{w/2}\right) - 1 \qquad (1)$$

where x is the x-coordinate of the center of the track icon on the iPad screen, w is the width of the screen, and y is the pan value to send to the mixer. On the iPad 2, 512 pixels are available for either output. The resulting pan value is a detailed floating-point number that gives the user more finesse than setting a degree value, the analog to digital adaptation utilized on most digital mixers. If one moves their finger very slowly, they could actually pan a track one pixel at a time. The new iPad's resolution is 2048-by-1536 so it would provide finer values for panning.

When Apple's multi-touch screen recognizes a pinch gesture, it outputs a scale factor relative to the point of the two touches in screen coordinates. It calculates this many times a second so the scale factor value is normally hundredths away from 1, over and over again. For example, a scale factor of 1.01 would represent a 1% increase and 0.99 would represent a 1% decrease. This actually worked out well for calculating gain because the gain values for each track range from 0 to 1, or -∞ to unity gain when compared to an analog mixer. Because the system is supplying numbers in a small range very rapidly, these values have to be added as an increase or decrease to the existing gain. *Equation 2* does just that:

$$g = f - 1 + c \tag{2}$$

where f is the scale factor of the user's pinch gesture, c is the current gain for the corresponding track, and g is the new gain value. The image gets bigger or smaller, louder or quieter as the user pinches the track icon, but the image snaps back to its original size once the user lifts their fingers off the screen. This ensures that each track icon gets equal real estate on the screen. This may seem counter intuitive compared to *The Art of Mixing* concept, but it was necessary for practical reasons.

Each track has a different signal-to-noise ratio and headroom, so the gain value will not always correspond to the amount of volume that is being outputted. For instance, a rack tom track may need to be at +1dB and the main vocal track could be at -6dB simply because of imbalanced input levels at the time of the recording. This would then make the image of the rack tom much larger than the image of the main vocalist, which would contradict what the user would be hearing. By having all the track icons be a similar size, it forces the user to rely on their ears to balance the mix so that everything could be heard.

For displaying metering, instead of using a traditional VU meter, the opacity of the image changes with the output signal of the track. This essentially animates the track icon so that it appears and disappears along with the waveform of the signal. This animation only happens during playback. This optimizes the screen space so only the tracks that are playing at the moment are visible and tracks that are silent fade away. The actual range is from 5% to 95% opaque so that a track icon behind another one is still visible and the silent tracks appear as a watermark. This simple technique could be used to create a dynamic edit/arrange window for recording software so that a 48-track mix can condense so that only the actively playing and upcoming tracks can be seen, similar to how publishers optimize an orchestral score. *Figure 16* demonstrates a track icon animating with a half-second kick drum waveform.

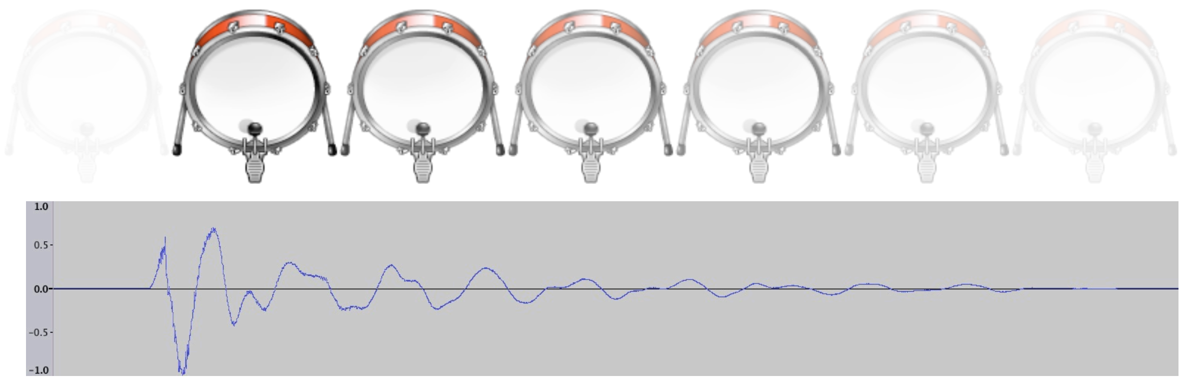

Figure 16. An example of using opacity to display metering of a waveform.

Obviously, there are many more aspects to a mixing board than pan, gain, and metering. There is solo, mute, automation, groups, input enable, record enable, signal chain, buses, playback information and controls, etc. In addition to the mixing board, a comprehensive DAW needs to have an edit window, audio bin, plug-ins, MIDI software instruments, etc. Creating an entire mixing system from scratch is a huge endeavor. How is it all going to fit on an iPad screen? There are a number of different solutions that involve hiding functions that do not need to be visible all the time, getting rid of redundant information, using different gestures to bring up different pop-up menus, sliding drawers, new windows altogether, augmented reality, and other efficient organization techniques. For example, muting tracks could be as simple as touching the track icon and then touching a single mute button. The current prototype does not implement all of these ideas, but *Figure 17* below shows a mock-up of what a more finished design would look like.

Just like other interactive music systems that use intuitive input technologies, MixPad would be designed as a controller for existing DAW software. It would be a supplemental interface that would wirelessly interact with the project file on the computer using OSC messages. OSC is an important protocol for communication between computers and various control surfaces. It is used to send bundles of time-stamped information packets that carry identifier and value information. This protocol works with most popular recording software available and is utilized in many hardware and software controllers. OSC gives musicians and audio engineers the freedom to choose what equipment they want to use when setting up a project and automating data.

Overall, MixPad has a lot of potential because it invigorates the idea of designing interfaces that interact the same way we conceptualize audio. Because the new technologies available today promote intuitive input, this could result in a new wave of interface designs that break the mold from tactile interface designs and focus on creating interfaces that emulate the same way our minds think.

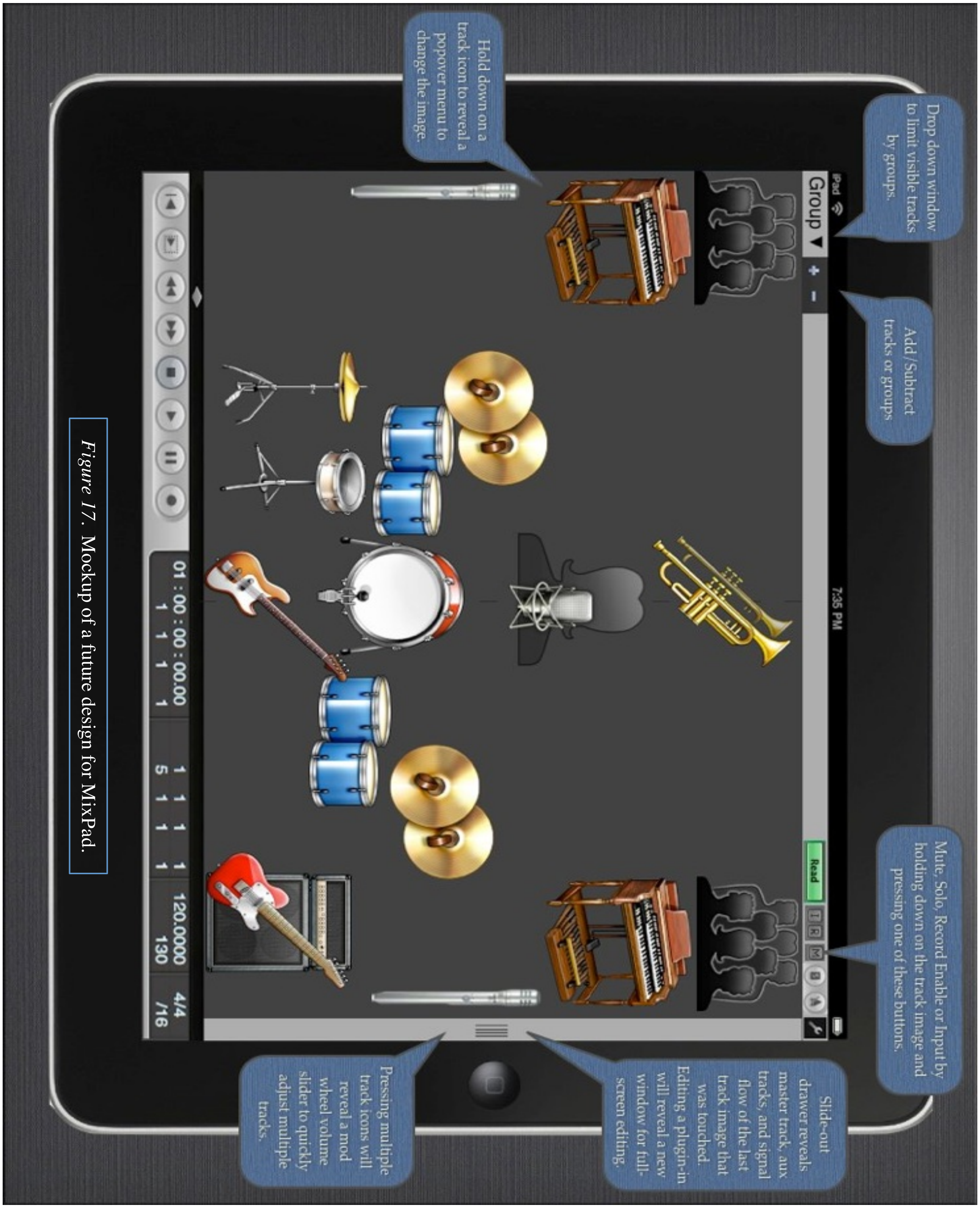

Figure 17. Mockup of a future design for MixPad.

7.1 3D Mixer with Head-Tracking

Between the ever growing video game industry and the rise of Blu-ray, which can support up to 7.1 mixes, there has been steady growth in the need for surround sound and 3D audio mixes. Fortunately, video games use advanced audio engines to handle the audio mix. As long as the original signals are mono, the game's engine handles the stereo or multi-channel output to spatialize the sound source to accompany the player's actions and movements. For movies, professional recording studios have had to upgrade their systems for these new multi-channel formats. 5.1 is widely adopted at audio post-production studios. Various formats can be used for 5.1 output including Dolby AC-3 (lossy), Dolby TrueHD (lossless), DTS (lossy), DTS-HD (lossless), and LPCM (uncompressed). Many studios have yet to adopt 7.1 systems for various reasons including the fact that few movie theaters are set up for it, much less home theater systems. It's still a fairly new format saying that the first movie to be released theatrically with 7.1 was *Toy Story 3* in 2010. Home studios are capable of generating a multi-channel mix because of recording software available today, however, the setup required is still expensive.

Physical mixing boards that support multi-channel formats can use a joystick to pan audio tracks in a surround sound field. This joystick typically sends surround angle and surround radius data to whatever recording software is being used. This information is incredibly difficult to automate with a keyboard and mouse. A simulated joystick on a screen requires fine motor skills that a mouse cannot interpret. One can have better control when using the joystick on a mixer, but in the end, it is an arduous process to automate a 64-track audio production for a feature-length movie in this manner. Even then, it becomes difficult to fully grasp all of the elements in a surround sound mix.

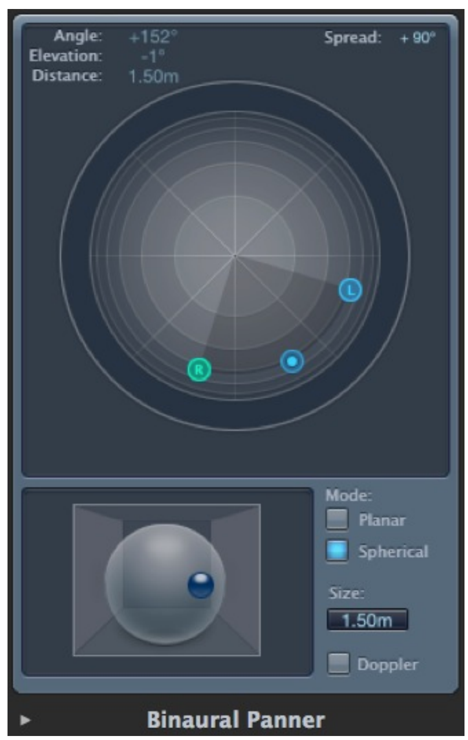

Figure 18. Binaural Panner in Apple's Logic Pro.

A method of 3D mixing for stereo headphones uses a head-related transfer function (HRTF) to create a binaural, virtual surround mix. HRTFs are a series of stereo impulse responses that are designed to capture the spatial characteristics of sound traveling to both ears on a head that came from a specific angle (azimuth) and elevation. These impulse responses could then be convolved with monophonic sound sources to place them in a virtual space while imitating natural difference cues that occur between the ears. The binaural panner in Apple's Logic Pro software uses a graphic interface to assist in HRTF panning. *Figure 18* shows a sound source placed on a 1.50-meter virtual sphere. This system actually works well in terms of visualizing the sound source.

One possible problem with this design has to do with using a mouse and cursor to set all these parameters. Yes, all the numbers can be typed in, but that's time consuming and unintuitive. It requires extra brainpower to process coordinates within a virtual space. What may be more intuitive is simply pointing in the air where one wants the sound to originate from. It would also be nice to set the coordinates as fast as "Track 1 is here" can be spoken. Such a design could be implemented using an iPhone and a gyroscope. The design shown in *Figure* 19 allows a user to set binaural panning information by simply holding the iPhone in their hands towards the intended sound source and pressing the "Track 1" button. Multiple tracks can be positioned in a matter of seconds using this system. Each track can also have its own distance parameter so a sound source could easily be placed in front of another.

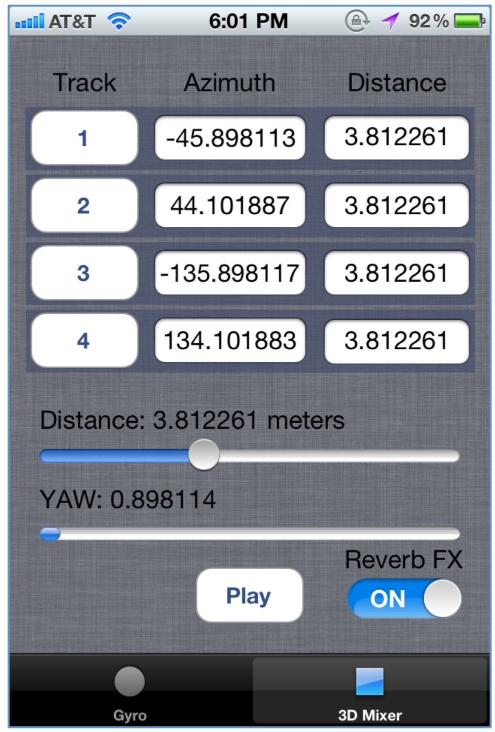

Figure 19. 3D Mixer with head-tracking for iPhone.

This implementation uses Apple's 3D mixer built into CoreAudio. At the moment, the mobile version of the 3D mixer does not include an elevation parameter, but the gyroscope could be used to send the appropriate information to computer software that does utilize elevation information.

One problem with using headphones for 3D audio is the "inside-the-head" effect created by two things: the lack of visual cues and the sounds moving with our head.[11] If we cannot see the sound source and the sound source is moving with us, our brains assume that the source of the audio is inside our own heads. This can create strain on our minds and ears when listening to headphones for a long period of time. It is also difficult to place a sound source directly in front of a listener because of the lack of visual cues, thus the 3D effect is not very convincing. One way to solve this issue is through head-tracking. By being able to analyze a person's head movements, it is possible to calculate the necessary pan values so that virtual sound sources sound as if they are stationary. This would then create an "out-of-head" effect while wearing headphones. In this implementation, the calculation required only simple math to set the gyroscope's yaw to each track's pan value.

However, even the addition and subtraction can be confusing because of a necessary conversion. The iPhone's gyroscope provides pitch, roll, and yaw information in radians from π (left) to –π (right). The mixer's azimuth panning parameter goes from -180º (left) to 180º (right). Not only do radians have to be converted to degrees, but the values increase and decrease in the opposite direction. For example, if "Track 1's" azimuth was set to 45º and the head moves from 0º to -45º (yaw decreases counter-clockwise), "Track 1" would have to be panned to 0º for the listener to be facing the sound source. Once converted, this system correlates to how humans hear and visualize sound, much more so than a joystick. The track buttons and distance slider allows the user to quickly position sound in a virtual room. *Equation 3* demonstrates the head-tracking algorithm where *a* is the current azimuth for the track's, *r* is the gyroscope's yaw in radians, and *p* is the new adjusted pan for the track.

$$p = a + r \times \frac{-180}{\pi} \tag{3}$$

Typically, expensive custom headgear, glasses, or headphones have to be used with custom software to create a head-tracking system. That means they generally aren't accessible to the general population or typical consumers. Our culture has quickly adapted to using earphones and MP3 players or smartphones to be our primary personal listening devices. The gyroscope that is in many of these devices could be used to at least attempt a standard head-tracking system. Since people don't wear iPhones on their heads, the system can be calibrated for a more natural, yet still somewhat accurate location, like a chest pocket.

7.2 Critical Analysis

MixPad is far from being a finished product. It was created to prove the concept that mixing sound could be done with a new interface that is not influenced by analog designs. The panning and metering functionality is responsive and straightforward. It would nice if the panning value could be displayed as the image is touched and if there were a visible centerline and snapping mechanism for images that need to be set to 0º as well as a locking system so other tracks aren't accidently moved. Whenever a track icon is on top of another one, it is difficult to know which one will be moved the next time those images are touched. Currently, the layers are set numerically by track number, but it might be possible to dynamically change the layering priority based on when a track is playing and which track was touched last.

The gain pinch gesture has had the most issues when MixPad has been demonstrated for different people. Finger sizes vary greatly so the size of the bass drum track might work well for one set of fingers, but not for someone else's. Giving the user the ability to scale the size of all the images until they can comfortably pinch track icons could solve this problem. The images used are transparencies so that means its touch area is only where there is an image showing and not the overall dimensions of the icon. Therefore, a skinny hi-hat stand image is difficult to touch and pinch because there is only a narrow area for touching. Eventually, custom images will have to be created to correct this issue by creating a minimum area of touch for the images. The images will also have to animate to show when they are muted, soloed, record enabled, peaked, and then additional customizations for the user will be added such as color and style.

At the moment, the pinch gesture only changes the gain as a percentage from the gain at the initial point of touch. It would actually be smoother if the pinch gesture rather controlled the rate of change over time. For example, a pinch that increases the image size by 25% will linearly add 2.5dB per second to the track. The tracks' gain could be displayed so the user better understand the relationship between the pinch gesture and the gain increase or decrease. The sensitivity (rate of change over time) could then be changed to the user's liking. All of these subtle interface adjustments will then have to be tested with multiple users to better understand the relationship between sounds and touch input. In the end, the best touch interface will be the one that can be greatly customized to fit the needs of each individual user.

The gyroscope in the iPhone is very sensitive and detects the slightest movement in the user's hand, but fast and irregular movements can throw off the gyroscopes information so that 0^{o} is moved in a slightly different direction than what it originally was. By adding a calibration button, a new 0^{o} could be set to adjust to wherever the user wants center to be. This sort of refinement is important because the gyroscope will eventually play a role in the future of MixPad.

8. LOOKING FORWARD

"A big definition of who you are as a designer is the way you look at the world… You are constantly looking at something and asking, 'Why is it like that? Why is it like that and not like this?' And so in that sense, you are constantly designing."[22]

-Sir Jonathan Ive

Sometimes new technology can feel like it is going in too many directions at once or simply forced. 3D movies have been attempted for decades, but it has yet to be fully accepted in the mainstream. The definition of a "better" experience changes but in some circles, definitions of "better" remain static due to preference and the comfort afforded by familiarity. There are still people who tout mono as a better format than stereo, never mind surround sound. Traditionalists exist in all areas where progress could be made to the point where they become an obstacle. C'est la vie. Such is life. Over time newer generations overcome these hurdles and eventually their "new" ideas become the traditionalist view. Paradigm shifts are inevitable.

For the future of music technology, analog hardware is essential. It is the foundation for many digital devices. There will always be audiophiles and enthusiasts that prefer the "analog warmth" that hardware provides over the clean sound of digital systems. Computers merely enabled the home studio market and allowed consumers to engage in audio processing without a traditional mixing board. Professional studios and live audio systems will probably continue to use digital mixing boards for many years. There will be additional progress in trying to recreate the studio sound for a live performance. iPads are already being used in studio and live settings for mobility and convenience. There are already dozens of apps available that utilize OSC messages to talk to mixing consoles. The engineer doesn't have to be in the control room or behind the mixing board to make minor changes. It offers more flexibility and practicality than what has been available before. At some point, these controllers may even do certain tasks better than current hardware and software options.

The next iterations of MixPad will most likely include wireless OSC messages to communicate with various DAW software and augmented reality for enhanced multi-channel and 3D mixing. MixPad would be a versatile visual assistant that would be woven into existing workflows. Eventually, plug-ins would be manipulated in the same way that the track icons would be. Imagine drawing on a frequency band or spectrogram and have MixPad output the necessary notch filters and automation information to the computer. Imagine pinching and panning gestures for a multi-band compressor. This setup could allow engineers to touch sound in a way that has never been done before.

If augmented reality were added to MixPad, it would serve as a portal to a virtual mixing room where a user could control the sound all around them. Whether they are doing a stereo or a 10.2 multi-channel with height mix, this virtual interface would give them an unprecedented level of control. Adding more tracks would not be a problem if the entire wall in front of a user were a virtual projector screen for their stereo pan. The gyroscope and even head-tracking with the front-facing camera would know exactly what the user was looking at and adjust the screen accordingly. Entire windows and interfaces could exist virtually using augmented reality.

Figure 20. Computer generated image of a virtual interface.[28]

In a 3D or multi-channel mix, track icons could be placed 360º in the air and the appropriate pan information would be automatically generated. Similar to *Minority Report*, a user's hand gestures in the air would be controlling whatever they were focusing on. This new virtual real estate would be able to handle many tracks comfortably and give an accurate visual representation of the mix. This system could expand to virtual glasses and gloves so that both hands would be free to make gestures. The possibilities are quite endless when one is working with virtual and augmented reality.

Realistically, MixPad will be a companion tool that will assist the engineer and/or producer. It will be a wireless controller for when they not behind the computer. It can be a visual aid for bleed-through or other undesirable, yet inaudible artifacts. The computer will still be the main workstation because MixPad will be dependent on the setup and organization within the DAW. Very specific functionalities will actually be easier with MixPad than on a computer and could be the main interface for things like pan automation for multiple tracks in 5.1. Its ease-of-use and simple learning curve could immediately benefit anyone that is new to audio mixing. In the end, its adoptability will be dependent on the open-mindedness of those using it.

Ideally, this technology will develop into tabletop multi-touch surfaces. Yes, that technology already exists, but there are other fundamental components that must come into play as well. Full-fledged computer operating systems and comprehensive professional software must make the transition to multi-touch technology. At this point, professional recording studios could install a display as large as a 48-channel mixing board, but it would be doing so much more than just mixing. Imaging an audio engineer being able to touch the waveforms of audio signals on a very large digital cutting board in order to edit and arrange them in any manner they please. This could change everything.

Even then, there's more that could be done. A company named Senseg[38] has developed a haptic feedback system for multi-touch screens that allows the user to feel virtual materials from silk to sandpaper. By creating an attractive electrostatic field, the friction between the finger and glass could be manipulated to simulate the sensory perception of contours and textures. This added dimension to touch screens could allow musicians and audio engineers to not only touch, but also feel sound in a way that has never been done before. This technology is right around the corner and could easily be adapted to large multi-touch screens.

The future is full of promising designs and technology. The biggest hurdle is the transition. Some people will go along for the ride and others will be left behind. Some designs will never go away, others will be adapted, and a few may disappear altogether. Both the way we conceptualize sound and the equipment we use to manipulate audio have a direct impact on the final creative output we call music. These new designs could inspire an entire generation of musicians to create sounds that are not even imaginable today.

8. RESOURCES

[1] "AMS Neve Celebrates 45 years of Innovation, Growth and Technology Leadership" April 28, 2012; <http://www.ams-neve.com/NewsAndEvents/news5.aspx>

[2] "Apple's App Store Downloads Top 25 Billion" April 20, 2012; http://www.apple.com/pr/library/2012/03/05Apples-App-Store-Downloads-Top-25-Billion.html

[3] "Applied Sciences Group: High Performance Touch" *MicrosoftReasearch*, April 29, 2012. <http://www.youtube.com/watch?v=vOvQCPLkPt4>

[4] Bierton, D. "Tech Report: Kinect – The Latency Question" April 18, 2012; <http://imagequalitymatters.blogspot.com/2010/08/tech-report-kinect-latency-question.html>

[5] Bulow, J. "An Economic Theory of Planned Obsolescence" *Quarterly Journal of Economics*; November 1986, Vol. 101, Issue 4, pg. 729-749.

[6] "The Channel Strip" April 24, 2012; <http://articles.georgesmusic.com/319/whats-with-all-those-knobs-a-primer-on-mixing-boards/>

[7] Chaudhary, A., Freed, A., Wright, M. "An Open Architecture for Real-Time Audio Processing Software" *107th AES Convention*, New York. September 24-27, 1999.

[8] Coulter, L., Jones, R., "How to Record Your Own Music and Get it on the Internet" *Chartwell Books*, 2009.

[9] Crawford, S. "How Microsoft Kinect Works" April 16, 2012; <http://electronics.howstuffworks.com/microsoft-kinect2.htm>

[10] Daley, D. "The Engineers Who Changed Recording" *Sound on Sound*; October 2004.

[11] Durlach, N., Rigopulos, A., Pang, X., Woods, W., Kulkarni, A., Colburn, H., Wenzel, E. "On the Externalization of Auditory Images" *Presence*, MIT, Vol. 1, No. 2. Spring 1992.

[12] Edwards, C. "The Original Circuit Bender" *Engineering & Technology*, Vol. 4, Issue 15, 2009. 37-38.

[13] Elliot, Richard. "Workstations into Studios; Studios into Workstations" *IEE Colloquium on Workstations Moving into the Studio*, London Nov. 24, 1994.

[14] Ferrucci, Dr. D. "How Watson Works" April 17, 2012; <http://www-03.ibm.com/innovation/us/watson/building-watson/index.html>

[15] Fitzgibbon, A., Robertson, D., Criminisi, A., Ramalingam, S., Blake, A. “Learning priors for calibrating families of stereo camera” *11th IEEE ICCV*, Brazil. October 2007.

[16] Fitzmaurice, G. W. & Buxton, W. “An Empirical Evaluation of Graspable User Interfaces: towards specialized, space- multiplexed input.” In *Proc. CHI ‘97,* ACM (1997), 43-50.

[17] Forlines, C., Wigdor, D., Shen. C. & Balakrishnan, R. Direct. “Touch vs. Mouse Input for Tabletop Displays” *Proc. CHI ‘07,* ACM (2007), 647-656.

[18] Gibson, D. “The Art of Mixing: A Visual Guide to Recording, Engineering, and Production” *Artistpro*, 2nd edition. 2005.

[19] Gordon M, et al. “Recording.” *The New Grove Dictionary of Jazz, 2nd ed*. Ed. Barry Kernfeld. Grove Music Online. Oxford Music Online. March 28, 2012; <http://www.oxfordmusiconline.com/subscriber/article/grove/music/J371600>

[20] Guillot, J. “Android is far behind iOS” April 18, 2012; <http://www.musiquetactile.fr/android-is-far-behind-ios/>

[21] Hotelling, S., Strickon, J., Huppi, Brian. “Multipoint touchscreen” Patent 7663607. Feb. 16, 2010.

[22] Hutswit, G., dir. *Objectified,* Pixel Productions, 2009. Film.

[23] “iPhone 4 Gyroscope Teardown” April 15, 2012; <http://www.ifixit.com/Teardown/iPhone-4-Gyroscope-Teardown/3156/1>

[24] Isaacson, W. “Steve Jobs” *Simon & Schuster*, 2011.

[25] Laing, D. "Paul, Les." *Grove Music Online*. *Oxford Music Online*. April 23, 2012; <http://www.oxfordmusiconline.com/subscriber/article/grove/music/49128>

[26] Latta, S., Tsunoda, K., Geisner, K., Markovic, R., Bennet, D., Perez, K. “Gesture Keyboarding” Patent Application 20100199228. Filed Feb. 23, 2009.

[27] Moormann, M. “Tom Dowd & the Language of Music” *Language of Music Films*, 2004.

[28] “Multivirtuals: Winners of the Future” April 15, 2012; <http://fairsights.blogspot.com/2010/12/multivirtuals-winners-of-future.html>

[29] Owen, P., Pank, B. “Non-Standard Platforms - Ready for the Digital Era” *IEEE Colloquium on Workstations Moving into the Studio*, London, Nov. 24, 1994.

[30] Payne, J. “The Music Software Revolution, The Revolutionaries, How It Changed Everything” *EQ* 17, 4; April 2006; pg. 18.

[31] Pejrolo, A. "Creative Sequencing Techniques for Music Production: A Practical Guide to Logic, Digital Performer, Cubase and Pro Tools" *Focal Press*, 2005.

[32] Peterson, G. "New Directions in Console Design: It's All About Control" *Mix*; May 2005; 29, 6; pg. 104-106.

[33] Roche, K. "Pro iOS5 Augmented Reality" *Apress*, 2011.

[34] "Rotary Ring" by Panorama Concepts. April 9, 2012; <http://itunes.apple.com/tr/app/rotary-ring/id427785239?mt=8>

[35] Ryan, K., Kehew, B. "Recording the Beatles" *Curvebender Publishing*, 2006.

[36] Sapir, S. "Gestural Control of Digital Audio Environments" *Journal of New Music Research*, 2002, Vol. 31, No. 2, 119-129.

[37] Sapir, S. "Interactive Digital Audio Environments: Gesture as a Musical Parameter" *COST G-6 Conference on Digital Audio Effects (DAFX-00)* Verona, Italy; Dec. 7-9, 2000.

[38] "Senseg Technology" April 23, 2012; <http://senseg.com/technology/senseg-technology>

[39] Speilberg, S. dir. *Minority Report*, Twentieth Century Fox, 2002. Film.

[40] Tuddenham, P., Kirk, D., Izadi, S. "Graspables Revisited: Multi-Touch vs. Tangible Input for Tabletop Displays in Acquisition and Manipulation Tasks" *CHI 2010: Multitouch*; April 10-15, 2010; Atlanta, GA. 2223-2232.

[41] Ubilla, M., Mery, D., Cadiz, R. "Head Tracking For 3D Audio Using the Nintendo Wii Remote" *ICMC 2010: Research, Education, Discovery*. New York, June 1-5, 2010.

[42] Verna, P. "Mackie Designs Owes Console Success to its Founder" *Billboard – The International Newsweekly of Music, Video and Home Entertainment*; Oct. 30, 1999; 111, 44; pg. 35, 38.

[43] "Wiimote" April 17, 2012; < http://en.wikipedia.org/wiki/File:Wiimote.png>

[44] Wilson, T., Fenlon, W. "How the iPhone Works" April 17, 2012; <http://electronics.howstuffworks.com/iphone3.htm>

[45] Whittleton, D., Corkerton, T. "A Computer Environment for Surround Sound Programming" *IEEE Colloquium on Workstations Moving into the Studio*, London, Nov. 24, 1994.