WaaZam! Supporting Creative Play at a Distance in Customized Video Environments

Seth E. Hunter
BA. Cognitive Science and Studio Art, University of Virginia, 2000
MFA. Art and Technology, School of the Art Institute of Chicago, 2007
MS Media Arts and Sciences, Massachusetts Institute of Technology, 2009

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
Media Arts and Sciences at the
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy, Massachusetts Institute of Technology
February, 2014

Author:

Seth Hunter
Graduate Research Assistant
MIT Program in Media Arts and Sciences

Certified by:

Pattie Maes
Alexander W. Dreyfoos Jr. Professor of Media Arts and Sciences
MIT Program in Media Arts and Sciences

Accepted by:

Pattie Maes
Chair, Departmental Committee on Graduate Students
MIT Program in Media Arts and Sciences

WaaZam! Supporting Creative Play at a Distance in Customized Video Environments

Seth E. Hunter

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on October 10, 2013,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

This thesis presents the design, implementation and evaluation of *WaaZam*, a telepresence system for families that supports social engagement through creative play in a unified video space. The goal of the project is to design an environment that facilitates more engaging shared experiences for geographically separated adults and children by placing them together in fantastic worlds and providing tools for them to customize the environment.

Standard video mediated technologies provide a live window between remote spaces but do not support users who want to design and interact together in those virtual spaces. We outline the design an implementation of a merged telepresence environment with assets layered in 3D space. The software includes a scene maker and a scene renderer in which users can quickly design digital sets and then play in them together. It supports both physical and digital customization through gestural tools that allow users to seamlessly layer digital and physical content with their bodies. Within the environment users can also transform their appearance and record video.

We present the result of pilot studies and a formal evaluation of families using *WaaZam* in four conditions: separate windows, magic mirror, provided fantastic sets, and self-made sets. Across these conditions we measure dramatic and narrative play patterns, mutuality, social engagement, and physical engagement. Adults and children are interviewed to determine if personalization increases the richness and depth of the user experience.

A longitudinal design plan outlines a framework for how this platform could scale beyond the laboratory. We provide an outline for implementing scalable web technologies, a video protocol that supports depth and metadata, and design guidelines for installation in the homes of geographically separated families to support shared experiences.

Thesis Supervisor: Pattie Maes
Title: Professor of Media Arts and Sciences

WaaZam! Supporting Creative Play at a Distance in Customized Video Environments

Seth E. Hunter
BA Cognitive Science and Studio Art, University of Virginia, 2000
MFA. Art and Technology, School of the Art Institute of Chicago, 2007
MS Media Arts and Sciences, Massachusetts Institute of Technology, 2009

Thesis Advisor:

Pattie Maes
Alexander W. Dreyfoos Jr. Professor of Media Arts and Sciences
MIT Program in Media Arts and Sciences

Thesis Reader:

V. Michael Bove Jr.
Principal Research Scientist, Director of the Object-Based Media Group
MIT Program in Media Arts and Sciences

Thesis Reader:

Mitchel Resnick
LEGO Papert Professor of Learning Research, Lifelong Kindergarden Group
MIT Program in Media Arts and Sciences

ACKNOWLEDGMENTS

I would like to thank the following invaluable advisors of this research:
Pattie Maes for advising the project, guidance, support and patience all these years.
Kori Inkpen for taking time to help critique the user study and advise the evaluation.
Tony Tang for many wonderful conversations and collaborating on a shared vision.
Sue Hessey for study design and evaluation critique.
Mitch Resnick for inspiring me to support creativity and supporting the project.
Michael Bove for your insightful observations, knowledge, and research perspective.
Ian Kegal for your collegial advice and introducing me to your research community.
Edith Ackermann for dreaming with me at the inception of the project and your "beginner's mind".

Without the assistance a team of developers this research could not have happened:
Jason Moran for developing with creativity, a critical mind, and believing in the vision of the project.
Andrew Farley for teaching me how to do project management and late night coding.
Andrew Otto for being such a good friend and helping us setup our network servers.
Arturo Castro for making this research something that could have an impact beyond the lab.
Bruno Selles and Bridget Palma of Vasava for your creative ideas and design collaboration.
Noura Howell for your energy, attention to detail, and support when I needed it.
Anette Von Kapri for collaborating, working together late nights, and trusting me.
Arjun Narayanan for development of the puppetry algorithms and UI components.
Yuta Kato for your assistance with the UI components and collaboration with Arjun.
Jonathan Trevino for your friendship and taking ownership to bring critical features to the platform.
Kevin Wong for going above and beyond the call to assist with the project.
John Bowler for improving the segmentation and doing exploratory object tracking research.
Linlin Haung for designing new sets and helping me design the overall look and feel.

This research was sponsored by a generous grant from Cisco Systems University Relations:
Stephen Frasier for funding and supporting the project.
Stephen Quaterno for believing in the vision, advising the project, and being a friend and liaison.
Carol McDevitt for listening, your interest, and setting up the project at Cisco.

To Family, friends and colleagues:
Kim Hunter for everything and more: I love you.
Carol Kilcourse for always being there when we needed you.
Amanda Nastari Giacalone for caring for my children, listening, and being so consistent.
Eric Rosenbaum for being a friend and colleague and sharing the same path with me.
Simon Olberding and Jochen Huber for your friendship and encouragement.

Table of Contents

CHAPTER 1. INTRODUCTION

"Distance does not break off a relationship absolutely, only the activity of it." –Aristotle

Our society is progressively more globally distributed, a trend which has blurred notions of home, family and community. Being geographically separated however, does not mean that we can not still enjoy being together. Video mediated communication tools are currently used by families to maintain long distance relationships, but they are not designed to support the shared activities of families. *WaaZam* began by asking: *How can we create environments that support shared experiences at a distance?* and subsequently: *"Could families design and build their own virtual worlds together?"*

Video mediated communication (VMC) systems like *Skype* and *Facetime* have become the de facto form of communication for many families. As our network capacity and hardware capabilities have increased, existing forms of video mediated communication do not sufficiently support users who want to do things together. The focus of most research has been on improving resolution, compression, multi-party bridges and frame-rate. Conventional systems provide a window into remote environments but do not merge remote and local participants into one environment.

WaaZam was inspired by the idea of designing a VMC system that allows people to do things together at a distance in more imaginative and creative ways. Many young children lose interest in standard video calls without engagement in a common activity. When families invent and inhabit their own worlds they quickly move from conversation to shared activities.

This document presents the design, implementation, and evaluation of the *WaaZam* application, which allows families to design and inhabit worlds together at a distance. The purpose of the system is to increase creative play and social engagement during video sessions. Our study provides evidence that being together in the same virtual space increases engagement for adults and children. We found that families with existing creative play patterns create more personally meaningful environments and engage in more active, social, and creative activities in our system.

1.1 Problem

Our families, colleagues and friends are increasingly more geographically separated. Video conferencing technologies provide a live window between remote spaces but do not typically support users who want to design and interact in a shared video space. Although possible with current technologies, there are very few communication platforms that facilitate creative interaction at a distance. Creative telepresence systems have been under-explored in the research literature, particularly in the context of bodily activities such as family play, theatre and dance that require participants to coordinate directly with each other.

Figure 1. Examples of children in video calls having difficulty managing visibility and attention.

During video calls most young children have trouble staying in the field of view of the camera, knowing what to talk about and feel impatient when they have to stay in one place [Ballagas, 2009]. The distant adult has a very limited shared activities to select. What is lacking from conventional VMC systems is support for shared activities that extend beyond communication to the realm of being in the same virtual space. Other researchers have introduced the idea of composited video spaces [Agamanolis, 2002] but no one has done research on the types of shared play activities that would benefit families. We present tools and techniques for parents and children to invent and play together in customized environments.

1.2 Proposed Solution

WaaZam is a VMC system that supports the creation of virtual spaces composed of images, movies and animations that are mapped to a three dimensional space. Remote users can interact in the environment together by moving their bodies and can navigate the environment through a gestural interface. The system allows users to login to a session together, select and modify sets, play together in the environment and save video recordings.

Figure 2. An illustration of the WaaZam environment, a father and son who live apart using the system to share LEGOs.

We designed the application with three goals in mind: **to increase engagement between participants, facilitate more imaginative and creative play, and provide tools that allow users to create and**

personalize virtual environments together. In previous studies, introducing shared activities [Raffle, 2011] dramatically increased engagement. By encouraging a shift from turn taking to more continuous forms of unstructured play families can improvise new activities. We support more active, social and creative types of play by rendering bodies and objects in the same virtual space.

Research focused on online creativity shows that when people are creative it promotes feelings of self-confidence and self-worth that lead to emotional well-being and more healthy social relationships [Bruckman, 2009]. When children construct their own world views through play they develop divergent thinking skills that can ultimately lead to them being more innovative later in life [Sawyer, 2012]. Creative forms of play such as pretending and constructing narratives also help children develop their imagination and increase their capacity to empathize with others and reflect on their own actions.

Papert and Piaget assert that children build knowledge through experience and learn through activities they find personally meaningful. We prioritize customization in order to understand the types of environments parents and children want to be in, support a wide range of play scenarios, and transfer agency and ownership to participants in the system.

WaaZam supports gestural transformation of users, dynamic puppetry, body merging, costuming, effects and other features suggested by users in our pilot study. Each of these features are intended to encourage imaginative forms of play that are only possible in virtual spaces. By focusing on the transformative aspects of play we hope to support children who enact aspects of their identities when pretending together. *WaaZam* provides a unique opportunity for adults to be together with children in spaces that embody the child's imagination.

To evaluate the system we observed the activities, behaviors, engagement, attention, mutuality, and the types of play that are common to adult-child pairs. We examined four environmental conditions: Skype (traditional separate windows), magic mirror (you visit my space or I visit yours), constructed imaginary sets, and customized sets. Our initial results indicate that there is a clear benefit to being in the same space together. Using constructed backgrounds was found to add significant value for many users and increases creative play between participants.

1.3 Usage Scenarios

General usage scenarios are presented in chapter 6 for researchers interested in extending this work to other domains. In this thesis we focus on the benefits of supporting creative play for adults and children ages 6-12. We choose this demographic because we felt they would benefit the most from playing together at a distance. In 2011 38% of families in the United States were geographically separated from their extended families and place of birth by more than 200 miles [Raines, 2011]. In families with children under the age of 12, 40% are divorced with one parent seeing the child less than half of the time [NIMH, 2002]. More than 50% of grandparents living in North America are over 200 miles from their grandchildren. [Davies AARP, 2009]

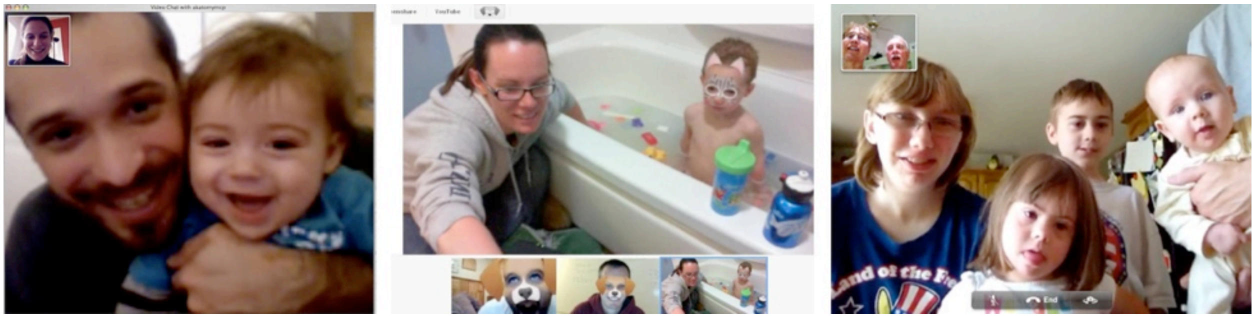

Figure 3. Screenshots of conventional VMC systems: Skype, Google Hangout, and Facetime.

With these trends increasing as people migrate to new cities to find work, children lose touch with family members who are not in their immediate vicinity. Although Google Hangout recently added costuming features, there are very few options for adults and children in separated families to customize environments and play at a distance with each other.

Figure 4. An 8 year old girl and her geographically separated mother using WaaZam to act out a Christmas event they attended, dance together, and record a video in which they pretend they are climbing Mt. Rushmore.

We have designed an application for geographically separated parents, grandparents, and relatives to have creative play sessions at a distance from their respective living rooms. They can design scenes, share and inhabit media from their day, visit each others spaces, and play in imaginary spaces designed by themselves or other users. Typically users play freely in a digital scene by pretending and improvising together. If they like something they have done, they record a video of the episode and share it with family and friends.

1.4 Contributions

This thesis makes unique contributions in several areas related to creative VMC systems for families. These contributions are in the areas of design, implementation, evaluation, a technical toolkit, and an analysis of areas for future research. The specific contributions are:

1) The design of the *WaaZam* system with a set of design guidelines and a overview of the unique features included in the system.
2) The technical implementation of our system with source code for researchers and a downloadable application for families who wish to utilize the software.
3) An analysis of qualitative and quantitative data from a formal study of 24 participants which examines the differences between traditional and merged video, the effect of imaginary backgrounds on engagement and play patterns, and a detailed discussion of the value of personalization with examples from specific groups.
4) A software toolkit with documentation and tutorials for researchers interested in developing networked peer to peer applications that utilize depth cameras and metadata.
5) A roadmap for future research in creative VMC systems and features requested by users which may increase engagement for users and improve the interactive capabilities of the system.

1.5 Thesis Roadmap

Chapter 2 outlines the background and related work. It begin by establishing foundational concepts in the areas of family communication, describes theories on the nature of play and a summary of constructivist theories of creativity. Our design approach builds on areas of remote collaboration systems, interactive art, distributed creativity systems, and previous VMC systems. We describe key projects in these areas with a particular focus on video mediated systems for children and families that support synchronous shared experiences.

The design and functionality of our system is described in chapter 3. It begins with our motivation, specific design principles related to VMC environments, and an explanation of the general functionality of the *WaaZam* environment. This is followed by descriptions of the interactive features in the environment and the functionality supported by the scene maker software.

Our technical implementation is described in chapter 4. We outline three network setups: a local area network, high-definition video communication system, and a broadband solution. Our software development process, the software architecture for the scene maker, the data infrastructure and network protocols we developed are outlined. Researchers interested in designing similar systems may want to refer to this chapter as a guide to our codebase.

Chapter 5 outlines the pilot studies and a formal user study followed by a discussion and analysis of the results. During the initial stages of the project we interviewed experts on creativity, conducted a pilot in our laboratory and installed the software in classroom settings in Hartford Connecticut. Our formal user study and research questions are described in detail here. We discuss both the qualitative and quantitative results from the study in the final section of the chapter. Further details of the study procedure, interview questions, and coding criteria are provided in Appendix A-E.

Chapter 6 addresses application domains outside of family play for VMC systems that support shared activities. We present tools, techniques, and new application domains that we did not explore in depth but were outlined as possibilities early in our process. This chapter may be interesting to anyone interested in extending the range of application domains and capabilities of our system.

The conclusion of the document provides a summary of our research contributions, a detailed outline of possibilities for future work and a vision of how society in the near future will be able create and play together at a distance. The core convictions that have driven this research and the vision that has resulted from it are reflected in the conclusion.

CHAPTER 2. BACKGROUND AND RELATED WORK

"Computers serve us best when they allow everything to change" –Seymour Papert

Designing video mediated systems for geographically separated adults and children requires understanding the nature of play and creativity and how families play together in co-located settings. Since the early eighties collaboration systems have attempted to create shared virtual spaces. More recently researchers have been working on mixed reality and gestural systems specifically designed for parents and children who are geographically separated.

The first part of this chapter will provide an overview of the theoretical and psychological literature on play and creativity. We reflect on current communication habits of families and the growing adoption of video technologies in connected homes. How well do current technologies support shared activities and creative authorship?

The second section of this chapter is a review of research systems which create shared digital spaces in co-located and remote settings. The representation of self and other in virtual spaces for telematic applications has been explored in work settings, interactive art, theatre, and social portals across many different disciplines. Key projects from each area are reviewed to include discussion of the affordances and limitations of each approach and their contributions made with regard to *WaaZam* and similar promising collaborative technologies.

The final section focuses specifically on more recent video mediated systems for families. Commercial products such as the Kinect and Wii have popularized family gaming and support gaming at a distance. The effect has been that people are now more familiar with technologies that are gestural in nature and expect interfaces to relate to their movement. Extending beyond games, researchers have been able to utilize emerging technologies to design new applications for families. Over the last five years, several researchers have been studying how children play at a distance by categorizing mixed reality configurations and approaches bridging physical and digital spaces. Other researchers have studied how to support the social relationships between parents and children by creating a shared context for interaction.

2.1 Family Communication, Play, and Creativity

WaaZam is designed to support playful activities at a distance and provide tools for creative authorship within that context. As a researcher, the values that drive my research are rooted in theory about play and it's role in childhood development. This understanding is based on the work of researchers specialized in the study of children such as: Piaget, Papert, Resnick, Jenkins, and Turkle. These learning theorists advocate a constructivist approach to designing new technologies. If we enable children to be producers of content instead of just consuming new media experiences, we empower them to build the future we will mutually inhabit.

The goal of this section is to define play and understand different play patterns between children and between children and adults. We will discuss how these transfer into our experiences with each other in virtual environments. The tools currently available to families for remote collaboration in commercial settings provide insight into the lengths people are willing to go to be together and the types of activities they would like to do together at a distance. Finally, we provide a theoretical foundation for understanding creativity in the digital age and the communities that currently support digital authorship online.

2.1.1 Research on the nature of play

Play is characterized by psychologists [Garvey, 1977] as spontaneous, self-initiated, self-regulated activities which are relatively risk free and not always goal oriented. An essential characteristic of play is pretending and imagination in which roles and rules govern the symbolic use of representation [Leontiev, 1981] [Nikolopolou, 1993].

Time spent in self-organized, social free play is key to children's social and cognitive development [Johnson, 1977]. Sociocultural theorists discuss the importance of play in childhood development as the most significant activity of the early childhood years. [Vygotsky 1977, Bodrova and Leong 1996]. Play provides a safe context for expressing emotions and gaining a sense of control and facilitates emotional development. [Frued, 1959] [Erikson, 1963]. Play also facilitates more advanced development in children through role play which promotes shared understanding in social situations. [Piaget, 1962].

Vygotsky theorized that play is socially and culturally determined and through play children acquire the tools and meanings of their culture. Several researchers and educators have proposed that children's play differs across cultures, generations, genders, and socioeconomic status [Roopnarine and Johnson, 1994]. Izumi et al found that while play activities are universal across cultures, adult attitudes about playfulness and the role of play in learning profoundly influence the types of intergenerational play that occur. Differences in possessiveness, imaginary play, and social negotiation are culturally correlated but the factors that account for differences can also be attributed to individual and family contexts. Significant research has been conducted on gender differences in pretend play. Boys are more likely to play fight and play competitive games and girls engage more frequently in role playing.

Figure 4a. Seven categories of play (clockwise from top left): Attunement, body play, object play, social play, transformative play, storytelling, and pretend play.

The National Institute for Play identifies 7 patterns that constitute the elements of play: attunement, body play, object play, social play, pretend play, narrative play, and transformative-integrative play (see Figure 1 above). These patterns are seen as part of a holistic framework overlapping and interrelated, but are useful in the scientific literature for understanding how play is linked to development and cognitive benefit. In our research we distinguish between patterns of play to determine the maturity of the child and the character of play in virtual worlds:

1)Attunement play (peek-a-boo, speech imitation) is the ground base for states of play. It is when the mother and child acknowledge each other through eye contact and facial imitation.
2) Body play (swimming, rolling around, dancing) is learning through motion by experimenting with how we move about the world for the pleasure of doing so.
3) Object play (toys, LEGOs, marbles) develops skill in manipulation and relating objects to each other.
4) Social play (tag, games, celebratory play) involves exchanging signals during interactions with others in the beginnings of relating to society.
5) Pretend play (stuffed animals, role-playing) demonstrates the ability of the child to sense their own mind and the mind of others and is considered by many scientists to be the key to innovation and creativity.
6) Narrative play (storytelling) is the beginning of learning abstract concepts and making sense of the world by describing the world around us to others.

7) Transformative play (acting, role-playing, playing music) is about fantasy, becoming and acting out a character through transcendence of reality into a safer realm of the imagination.

As children develop and play with each other they begin to relate in more social ways. One measure of this is mutuality; the extent to which participants are engaged with each other during a play session. Mutuality can be examined in play using the Piaget/Parten Scale [Rubin, 1978], which rates levels of play engagement as solitary, parallel, and associative/cooperative. Parten and Howes [Parten and Howes, 1981] observed that social play between children can be characterized by 5 stages of mutual regard and reciprocity.

Play scales can also apply to adult-child interactions but the literature generally recommends using different scales for the adult and child. Adult-child mutuality scales developed by [Crawley and Spiker, 1983] can be used to understand parent-child play interaction. This instrument is divided into two sections: child ratings (play maturity, social initiative, and object initiative) and guidance ratings (directiveness, elaborativeness, and sensitivity). Each characteristic is assessed on a 5-point scale (1 = low, 5 = high). Directiveness is defined as the degree to which the parent attempts to guide the child's behavior. Elaborativeness refers to the extent to which the adult follows and extends the child's self-initiated behaviors. Sensitivity is defined as the adult's awareness of the child's cues and signals and whether the adult's behavior is in accordance with that of the child [Crawley and Spiker, 1983].

Understanding the nature of play can help researchers describe the patterns of play, measure mutuality, and describe the social dynamics of adult-child play in new systems. In particular on the *WaaZam* platform, children engage in social play that becomes pretend play, narrative play and even transformative play. This is an indication of trust in the other participant, immersion in the platform, and an inclination to be creative and innovative as a result of the system.

2.1.2 Video communication technologies and families

A quantitative investigation of American families and the use of communication technology was conducted by Pew Internet & American Life Project in 2008 and 2010. The project found that 95% of Americans have access to cell phones and a computer and are connected to the internet. Of those surveyed, 66% had a broadband connection. The majority of Americans still use a phone on a regular basis (75%) - with 35% talking on a daily basis and 54% using cell phones on a daily basis. In 2010 approximately 25% of parents and children communicated by email or text message. Of the 74% of Americans with access to the Internet, 24% reported using video chat with their children and 4% reported using it in the last 24 hours. [Raine, 2010]. Recent reports from the Pew Internet Project state that 34% of Americans own a tablet computer and 49% own a smartphone and 19% have tried using video services on their mobile devices. Overall trends indicate a steady increase in adoption on all video enabled platforms.

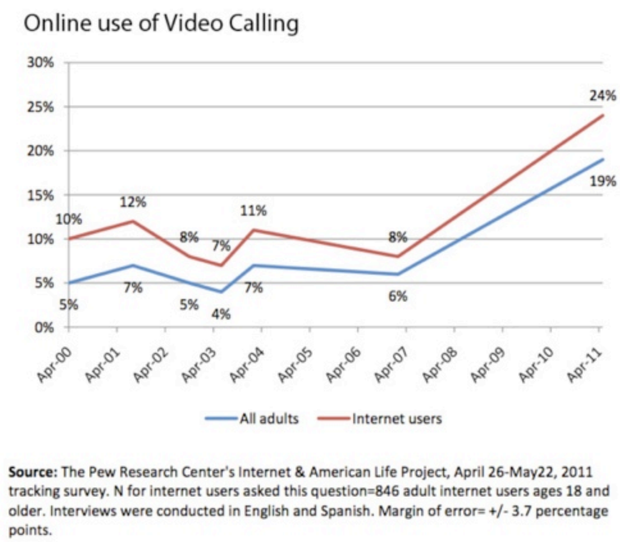

Figure 5. Research shows a steady increase in family use of video calling services.

There are very few investigations of technology use by children 8-12. A recent study [Synovate, 2011] found that 3% of Americans own a cell phone before the age of 10, but as many as 10% own devices capable of connecting to the internet. By the age of 11, 20% own cell phones and 46% of parents say they plan to get phones for their children by the age of 11. The trends indicate that children are getting connected at a younger and younger age although the effects of connectivity at a younger age are still undetermined.

Figure 3 has screenshots of conventional VMC systems that exist today. Google Hangout supports group chat, mutual viewing of web content and sharing screens through the browser using WebRTP. Skype was recently acquired by Microsoft and will be integrated into the core libraries of the Windows platform. Skype supports free peer to peer calls but charges for multi-party calls and calls to mobile devices. FaceTime is integrated into all tablets, phones, and OSX devices and works across devices. Oovoo is a startup that has 75 million registered users, most of whom are teenagers looking for an alternative environment in which to hangout with each other. Although these platforms provide easy options for families, very little has been done to integrate shared activities and more advanced capture devices that would enable foreground/background subtraction. Google Hangout has introduced storytelling applications for bedtime stories at a distance.

Family members living in separate geological locations adopt synchronous communication technologies more quickly than co-located families. Early investigations indicate that the telephone is not effective at supporting family communication because it does not support a conversational context [Evjemo, 2004] with particular difficulty for younger children [Ballagas, 2009]. Several studies have been conducted on the use of video chat as an alternative to audio communication [Judge, 2010] [Kirk ,2010]. These studies are enthusiastic about the potential benefits of the technology to support communication but show that the families who have the greatest need for it typically do not have access or knowledge of how to setup connections between homes. As bandwidth limitations decrease and internet services are bundled with cable services, there has been an increased use of these services in lower income households.

Figure 7. Asynchronous systems: Carenet, Digital Family Portraits, Toy Assembly, and Awareness Portraitss

A majority of the work in the HCI community on family connectedness is focused on asynchronous remote communication techniques. There are many examples of interfaces such as digital portraits [Mynatt, 2001], ambient displays for the elderly [Consolvo, 2004], shared family calendars [Plaisant, 2006], sketching and notes [Fussell, 2004], connected video furniture [Siio, 2002], and even video messaging systems [Zuckerman and Maes 2004]. Most of the asynchronous systems indicate that they provide a low bandwidth and convenient way for families to maintain an awareness of each other to bridge the gap between instant synchronous communication. For this reason it is worth considering how synchronous designs might leverage asynchronous research to create more unified hybrid instantiations such as Kinect based video messaging services that lead to synchronous communication [Vetere, 2009] [Tang, 2013].

2.1.3 Constructivist creativity and authorship

Very few systems support creative authorship at a distance between family members. However in co-located settings through object and social play children and parents often build things together and invent their own worlds. In *WaaZam* we designed and incorporated "Build your own world" tools into a virtual environment guided by constructivist principles.

Developed by Papert [Papert, 1980], constructionism builds on Piaget's [Piaget, 1954] ideas about how people generate knowledge through experience. Individuals construct mental models to understand the world around them and learn best in a context they find personally meaningful. Papert's ideas became known through his book *Mindstorms: Children, Computers, and Powerful Ideas*. *Mindstorms* introduced

LOGO, a multi-paradigm language adapted from Lisp which children understand through the motion of turtle graphics and a set of commands to understand the turtle.

Figure 8. Early LOGO Turtle movements, LEGO Mindstorms, Scratch Building Day at MIT

In *Experience and Education*, John Dewey [Dewey, 1938] argues for social reform that shifts from instruction to experiences and interactive curricula with individual purpose. He stresses the importance of social and interactive processes of learning that were based on a philosophy of experience. Dewey suggested that all learning builds on prior experiences, biases, and knowledge. His work suggests that designers should consider a theory of experience that is open to the individual background of each user and remains adaptable, allowing them to configure their own worlds.

James Gee [Gee, 2003] is a linguist who writes about video games and learning. He argues for an approach to new media design that increases digital literacy in ways that are relevant to the identity of the learner. By empowering children (and adults) to become producers instead of consumers we will move more towards a society of creative innovation. Henry Jenkins [Jenkins, 2006] proposed similar arguments in a "transmedia" context, providing examples of communities like fans of Harry Potter who remixed scenes from movies, books, video games, and internet discussions into customized stories. Both Gee and Jenkins propose a style of media engagement that allows people to remix new media in ways that are personal and relevant to shifts in medium and culture.

Advocating for digital literacy by encouraging ownership, agency, and participation is echoed by Buckingham and Salen [Buckingham and Salen 2007] who provide evidence of the growing divide between the media-rich world of children's lives outside of school and their experiences in the classroom. They debunk fantasies about the negative effects of technological change on learning and provides a constructive alternative forms of digital literacy that are critical and creative. Further, they note that learning in the digital age is a collaborative process between the parent and child. The best way to empower children in the medium is to co-create content with them and co-play games and media that are introduced to them so that the media can be examined and discussed by both the parent and child.

Socially scaffolded learning through experience has its foundations in the writings of the Russian philosopher Vygotsky. *The Social Formation of Mind* [Wertsh, 1985] highlights the importance of collaborative creativity and play in social environments to learning and development. Vygotsky claimed that higher mental functioning in the individual has social origins and believed that tools and signs are socially motivated by psychological desires to relate to others and the world. According to his theory, tools that allow us to negotiate, learn, and create with others are increasingly important in the digital age because of the social capital: thought finds its origin in new media.

Over the last ten years there is clear evidence that these theories have had a deep impact on the HCI research focused specifically on design for children. In 2011, the author and three other collaborators [Yarosh et al., 2011] surveyed the theory values and motivation that informed ten years of work in the Interaction Design Conference. Our research found that 71% of researchers were informed by theory in their papers with 28% explicitly claiming constructivism informed their research and 25% claiming constructionism. Learning through play, games for learning and in-situ learning followed closely behind. 58% of researchers were doing mixed reality work and there was a general trend towards work that increasingly bridges the physical and digital world. The core values that inform the community aligned with embodiment, constructivism and constructionism focused on using system design to support the social, intellectual and creative growth of the child.

2.2 Shared Space Collaboration Systems

The adoption of VMC systems into the workplace has been relatively low despite widespread availability of the technologies in recent years. Although there has been extensive work in research on synchronous Video Mediated Communication (VMC) systems in the workplace the literature has not demonstrated clear benefits to task completion and productivity [Kirk, 2010]. Other studies indicate that video in the workplace can enhance communication with gesture, convey non-verbal information that builds trust, and help manage pauses in conversation more effectively [Isaac, 1994].

Other domains of telepresence have been explored by a select group of researchers, artists, and performers porting the techniques into their work in order to enable synchronous remote experiences. There are very few overviews of generalized telepresence applications in the literature but similarities between platforms are striking between interactive art and theatre and between gaming and social television applications.

In family settings and home use, video has grown much more rapidly because the desire for intimacy and a sense of presence between family members [Romero, 2009]. VMC technologies have been been given a great deal of attention with the availability of competing commercial systems [Harmon, 2008]. The lessons learned from extensive workplace research are beginning to migrate into the home but often there is not a lot of cross-referencing between domains. In the following sections I attempt to find links between emerging VMC technologies for families and remote collaboration systems for the workplace, interactive systems in the arts, dramatic and theatrical telepresence systems, and social television systems; all of which have aspects that have influenced and contributed to the design and development of the *WaaZam* system described in this thesis.

2.2.1 Remote collaboration systems

Research in computer supported collaborative work (CSCW) is classified by place and time. Collaboration can happen at the same or different times in a co-located or a remote setting. For our purposes we are interested in work that is synchronous and remote and therefore has features which are applicable to the *WaaZam* application. These systems have subcategories which include audioconferencing, videoconferencing, telepresence, groupware, and collaborative mixed reality systems [Wolff, 2007]. The works described below are included because they contains a live video image that creates a shared experience in a mixed reality augmented space.

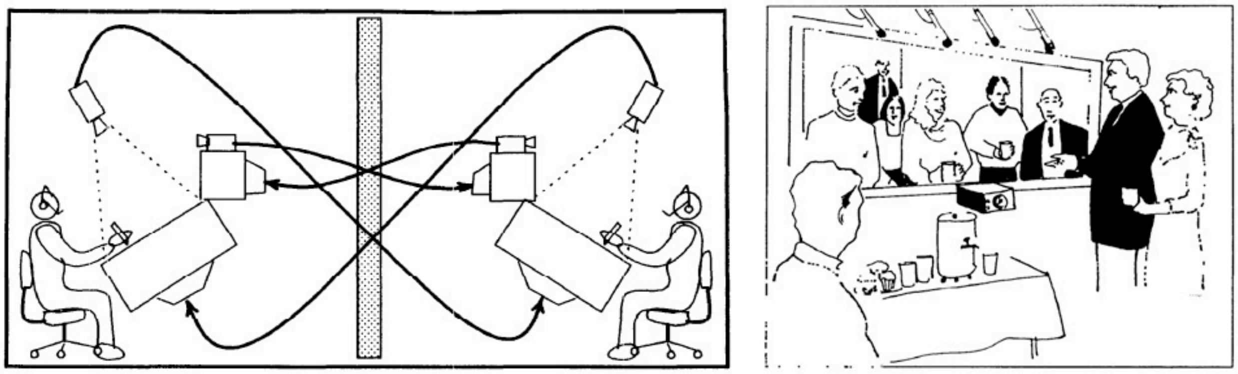

Figure 9. VideoDraw, 1990, MirrorSystem, 1993

The *VideoWindow* system [Fish, 1990] was an early exploration of informal social dynamics in the workspace through a shared window and pointed out many of the human factors concerns that arise when you install such a system in an actual workplace. It was installed in Bellcore NJ. The researchers concluded that for such a system to work it needs to be embedded in a context that would also encourage co-presence in the real physical world in order to mitigate socially awkward situations where participants do not want to interact with others at a distance.

VideoDraw [Tang, 1990] was one of the first systems to support a shared video sketchbook for multiple participants at a distance. The design is based on how people collaborate in co-located shared drawing systems when brainstorming ideas. The project was developed at Xerox PARC and was one of the first systems that supported a shared space for creative engagement.

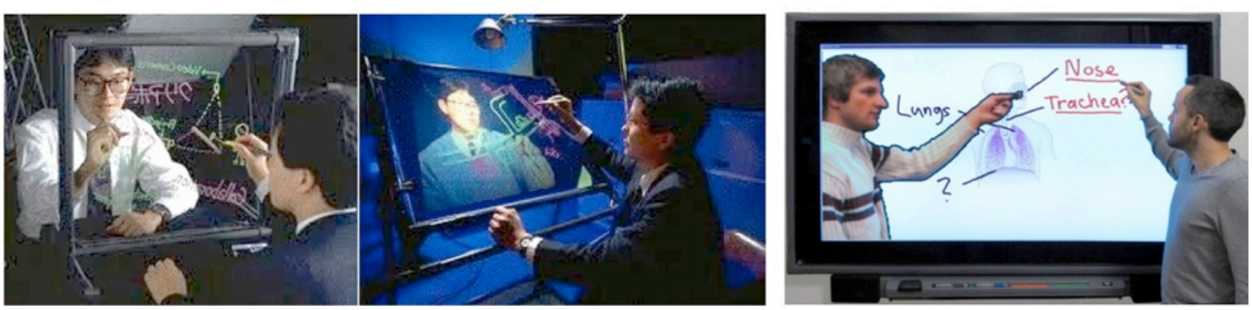

Figure 10. Clearboard, 1992 and Colla-Board, 2011

ClearBoard [Ishii, 1992] allows two remote users to collaboratively draw on a shared virtual surface. The project focuses on the seamless integration of interpersonal spaces and a workspace by merging the viewpoints of both people into a single plane and allowing viewers to augment the surfaces with the TeamPaint application, a multi-user paint editor. This project has become infamous in the HCI community for its elegance and because since its introduction it has been hard to produce projects with such an intimate shared space. Another question that has been raised in conversations among researchers is why such systems still have not made it into the mainstream workplace. It may be that the interpersonal intimacy is not as critical in the workplace as it is in the home where shared drawing interfaces like the *Wayve* [Lindley, 2010] support create drawing over social video between families.

Colla-Board [Nescher, 2011] is a collaborative whiteboard system in which users are projected onto the whiteboard surface. Users can see the other person's expressions and gestures as they collaborate on a shared drawing. Similar to Clearboard, this system uses a virtual representation of the other person in the shared space and manages the compositing between drawings and the user. A similar project, *VideoArms* [Tang, Neustaedter, and Greenberg, 2007] captures and reproduces people's arms as they work over large displays which promotes awareness of remote collaborators without showing their entire bodies in the frame.

2.2.2 Interactive art and the representation of self

The Greek myth of Narcissus is about a hunter who is exceptionally beautiful but is disdainful of those who love him. As divine punishment he is condemned to fall in love with his own reflection in a pool of water. He becomes entranced to the point of fixation with his own beauty, and starves to death.

If Narcissus were alive today and peered into a computer screen he might see himself differently. A technical way of describing Narcissus' dilemma is to say he is trapped in a closed loop of self-reflection and absorption. On the computer our image is formed in a dialogue between the self and the world beyond. David Rokeby writes that interactive media, "operates like a wayward loop of consciousness through which one's image of one's self and one's relationship to the world can be examined, questioned and transformed." [Rokeby, 1996]

Myron Krueger is one of the early pioneers of virtual reality and interactive art. Beginning in 1969, Krueger developed the prototypes for what would eventually be called Virtual Reality. These "responsive environments" responded to the movement and gesture of the viewer through an elaborate system of sensing floors, graphic tables, and video cameras. Audience members could directly interact with the video projections of others in a shared environment. Krueger also pioneered the development of unencumbered, full-body participation in computer-created telecommunication experiences and coined the term "Artificial Reality" in 1973 to describe the ultimate expression of this concept.

Many interactive artists following Krueger [Levin, 2003] [Rozin, 1999] [Utterback, 1999] [Snibbe, 2003] grounded virtual experiences in the real world by incorporating a representation of the viewer into a digital mirror. The use of the silhouetted recordings of the viewer in Myron Krueger's VideoPlace, Daniel Rozin's mechanical mirrors, and Scott Snibbe's *Deep Walls*, are popular examples that illustrate the holding power of seeing yourself echoed in an artwork. This technique has potential for meaningful social experiences at a distance and to encourage creativity in *WaaZam* because of the same experiential characteristics evoked in galleries which show these works.

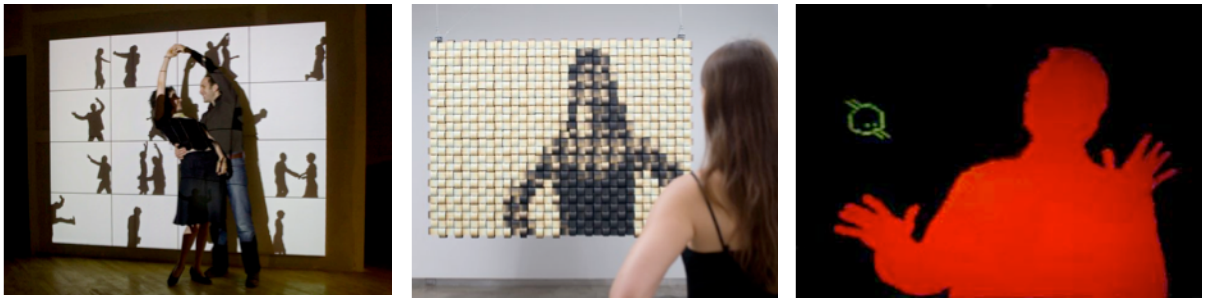

Figure 11. Snibbe's Deep Walls 2004, Rosin's Mechanical Mirror 2005, and Krueger's VideoPlace 1985

Freud's notion of the uncanny [Freud, 1953], when something is familiar and foreign at the same time, (resulting in a tension between states) is one possible explanation for why we find transforming mirrors so engaging. The confusion of distinguishing between the simulated and the real, and playing within those boundaries, make transforming mirrors evocative and mysterious. Freud's theories on the uncanny state that this results from a clash between the id and the superego (our intuitive and cognitive representations of self). This may help explain why the evocative nature of mixed reality interfaces which are both real and virtual.

Sherry Turkle proposes that the screen is a liminal space [Turkle, 1998] where the boundary between the real and virtual is blurred, providing users with a safe space to experiment with possible expressions of their selves. She writes that when we interact with computers we are interacting with our "second selves" an online extension of the self which changes how we think about our personae, how we exchange ideas and feelings, and how we relate to others.

The playful responsive environments first introduced by Krueger are possible today in people's homes with a richer set of creative tools and interactive possibilities. I imagine *WaaZam* to be a safe environment to invite our families and friends to experiment playfully with identity, self and others. Krueger's notion of an "artificial reality" included the notion of distributed interactive theater in a video space composed of images form widely separated sources superimposed to form a single image. He concludes his book: "While the pressure for interactive expression on the part of both the audience and the artist will lead to more experimentation; the seeds for the interactive media of the future have largeely been sown. We will have to wait to see which shall germinate." [Krueger, p. 226 Artificial Reality II, 1984]

2.2.3 Green-screens and distributed creativity with video

The use of green-screening techniques (also known as chroma keying) was developed in the 1930's and is attributed to Larry Butler, who won an academy award for effects achieved with its use in the 1950's while working at Warner Brothers. It is commonly used now for weather forecast broadcasting in which the news presenter appears to be in front of a large map during a live broadcast. It has been used ubiquitously in the film industry for special effects and as processors improved, researchers began to develop systems which were capable of real time chroma keying around 1978.

Figure 12. Installations by Paul Serman: Telematic Vision 1994, Telematic Dreaming 1996, All the World's a Screen 2012

Roy Ascott coined the term telematic art [Ascott, 1968] and defined it as an art form that transforms the viewer into an active participant in creating something which remains in process throughout its duration. Ascott organized a computer-conferencing project between the United States and the United Kingdom called *Terminal Art*. Later he worked on a project called *La Plissure du Texte* which allowed other artists to participate in collectively creating texts that become a story in a networked environment. He coined the term 'distributed authorship' at this time.

In the 1990s Paul Sermon [Sermon, 1991] began research related to Ascott's focused on the creative use of telecommunication technologies. He explored using video conferencing techniques for theatre, design and performance interactive installations. *Telematic Dreaming* was an interface that allowed people to network together augmented projection in the bedroom for intimate exchanges between people sleeping in different locations. *Telematic Vision* connected two sofas a distance in a virtual space as if someone were in the room with you by using blue couches and projecting the results from the other space into the room with the viewers. Sermon has continued to develop over twenty installations over the past two decades. The most recent relevant to this project is '*All the World's a Screen*', a connected live performance space between two gallery spaces in Manchester and Barcelona. The performance occurred realtime and transported actors from both locations into the same shared screen in a reenactment of 'Seven Stages of Man'. It pushes the boundaries of what is possible in telematic art and generative cinema, combining the possibilities of tele-present performance with miniature scale-models and animated scenes.

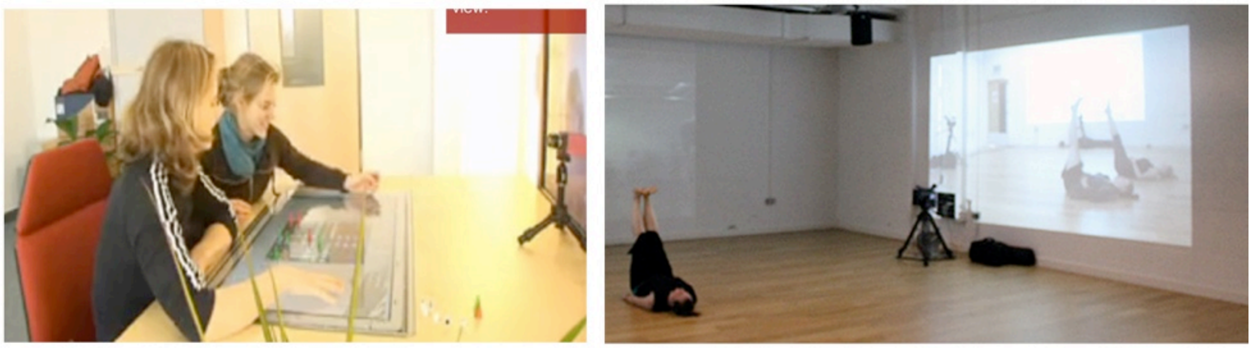

Figure 13. European Commission VConnect Initiative: Families play games together, person doing yoga at a distance.

The European Commission has funded a number of multi-year initiatives aimed at orchestrating video narratives generated by participants at a distance. Marian Ursu from Goldsmith's Narrative Interactive Media Group and Ian Kegal from British Telecommunications have worked on a number of projects that explore how people can be creative at a distance. The *VConnect* project, Video Communication for Networked Communities is funded from 2011-2014 and explores how we can improve video communication between groups. The NM2 collaborative research project united technology and creative experts across Europe to explore community created television media and sportscasting and connecting two orchestras live at a distance.

Figure 14. HyperMirror blue-screen telepresence application.

HyperMirror [Morikawa and Maesako, 1998] projects two distanced collaborators onto the same screen using a full body blue screen setup. The authors describe in their documentation a "what I see is what you see"(WISIWYS) approach. They implemented a dialogue system with a special mixer that chromakeys participants and mixes the results into a single view. The purpose of the system was for cultural exchange between classrooms in geographically separate locations.

2.2.4 Related composite video systems

This section summarizes prior work and work in progress that utilize compositing video techniques to combine views in video mediated communication. I reflect on the different approaches of researchers who merge remote spaces into one view and differentiate *WaaZam* in light of closely related systems.

Reflection of presence (later named Reflexion) [Agamanolis and Bove, 1997] was a pioneering platform developed at the MIT Media lab to explore interpersonal video communication in a "magic mirror" environment. The system used audio cues to fade a person in and out depending on who was speaking and was used for video chat and social TV and for large numbers of participants. It utilized background segmentation algorithms for very still backgrounds, mirror effects so inhabitants feel they are occupying the same space, and the use of backdrops so that participants can discuss documents, slide shows and movies that appear in the background of the mirror. The work was comprehensive and focused on scenarios of use and the technical implementation of the system which relied on the ISIS programming language that Stephan Agamanolis invented for in his doctorate work.

Figure 15. Refection of Presence and Reflexion, 1997: system configuration, setup and composite view.

TouchFace [Tsukada and Guimbretière, 2010] composites viewers in the same window using background subtraction and supports mouse based interactions with the other person's representation. The researchers recreated the reflexion project as a baseline for evaluating how couples might simulate marks of affection and to simulate where your partner is looking. They report that users preferred their system to the Reflexion setup because of the interactive capabilities they support the the effects they were able to place on the other participant.

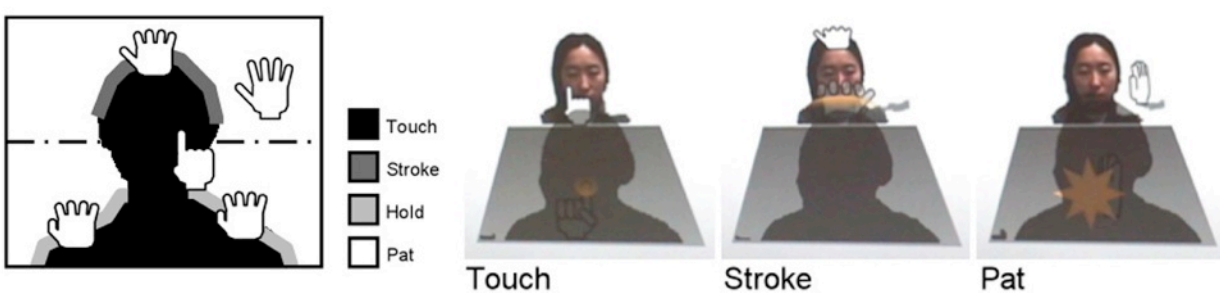

Figure 16. TouchFace, 2010, places viewers in the same space and allows them to touch each other using mouse events.

Maimone et al. [Maimone, Bidwell, Peng, and Fuchs, 2012] present an auto stereoscopic telepresence system that offers fully dynamic, real time 3D scene capture and continuous-viewpoint, head-tracked stereo 3D display. They focus on the technical implementation and describe how to diminish the interference problems you get using multiple Kinect cameras, how to optimize the meshes through hole filling and smoothing and getting rid of discontinuous surfaces on the graphics processing unit (GPU) and merge meshes from different cameras.

Figure 17. MirageTable, 2012; KeckCAVES, 2011

MirageTable [Benko, 2012] creates the illusion of two people working across the table from each other. The mesh of one person is place in a perspective rendered scene on a continuous curved surface. Users can mix real and virtual objects by placing them on the table and face each other as if they were playing a game of chess or having a meal together. *KeckCAVES* [W.M. Keck Center for Active Visualization in the Earth Sciences, 2012] is a high-end 3D system which places people together in a 3D environment and allows them to visualize and interact with geographical data at a distance.

WaaZam builds on the work of these researchers and introduces features intended for creative expression such as dynamic depth thresholding, gestural interaction, puppetry techniques, and a creative studio for creating your own scenes which can be in front or behind the user depending on their position in 3D space. The research examines the social dynamics enabled by such environments as well as examining the types of activities and effects of fantasy and personalized worlds on adult-child relationships within the extended family.

There are several projects that have been developed in collaboration with the *WaaZam* project. I have coordinated and collaborated with these researchers and have shared code libraries we developed in order to enable their research and create a community of researchers interested in shared augmented virtual spaces. The research contributions of each project are distinctly different but contribute to the same space.

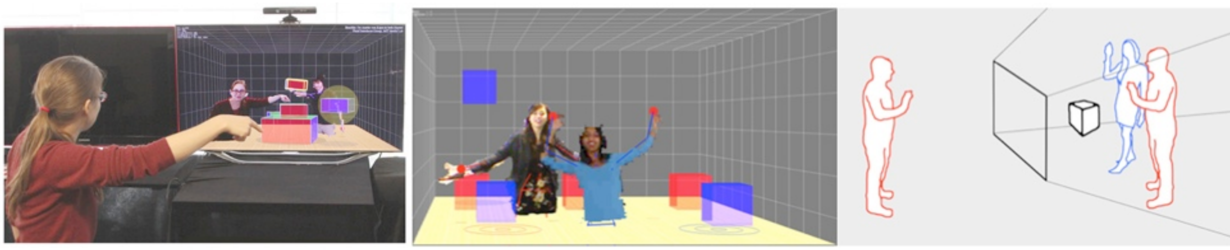

Figure 18. InReach 2013, Von Kapri explores 3D manipulation in a shared remote space

InReach [Von Kapri, 2013] explores how remote collaborators can "reach into" a shared digital workspace and how they can manipulate virtual objects and data. The collaborators see their live three-dimensional (3D) recreated mesh in a shared virtual space and can point at data or 3D models. They can grab digital objects with their bare hands, translate, scale, and rotate them. Von Kapri conducted a user study, which compares face-to-face and side-by-side arrangements for a sorting task. The user study demonstrated that different arrangements of collaborators and different levels of detail for self-representation do not influence the feeling of co-presence significantly. The focus on shared data in a merged space allowed users to explore modes of manipulating data with their bodies that would not have been possible in separate spaces.

Figure 19. OneSpace, University of Calgary (work in progress, 2013) explores representation in merged spaces

OneSpace [Ledo, 2013] is the most closely related project to *WaaZam* with regard to its implementation and goals. In their pilot studies, similar to our results, the researchers found that engagement and enjoyment increased in the space, participants staged visual interactions and appeared to be performing in the environment, and that virtual and physical play did not differ significantly between co-located use of the system and remote use of the system. In discussions with the researchers we distinguished different goals for our platforms and approach. Tony Tang et al is primarily interested in studying how composited environments reduce the reference space and how representation changes between the physical environment, a Skype condition, and the merged environment. We are collaborating and correlating with their findings with a specific focus on customized backgrounds and engagement between parents and children when they personalize their own environments. We are also working closely with these researchers to create a protocol for other people to create shared experiences because we believe that this type of interaction opens up opportunities for shared experiences in online education, family living rooms, creative environments for teenagers, language learning, and distributed theatre, each of which could be research projects on their own.

2.3 Video Mediated Systems for Children

A key benefit of video communication technologies for children is that they can use body language such as gestures, facial expression and voice intonations to convey emotion and reduce confusion [Mehrabian, 1972]. Children are not able to maintain the same level of abstract communication as adults and therefore need to leverage actions and body movement to refer to imagery versus text or voice. Video provides a richer medium [Bruner, 1975] for non-verbal communication with more cues for synchronous communication than other media.

Common ground theory [Clark and Brennan, 1991] suggests that mutual awareness provides awareness that is the foundation for the development of conversation because it provides contextual information. Multiple researchers have found that video mediated communication (VMC) can allow family members to feel more connected and that it enables them to participate in synchronous activities [Kirk, 2010, Ames, 2010, Judge 2011]. Families state that they desire video because they want to be involved in the ongoing lives of other family members, participate in their daily activities, and maintain the feeling of being together.

Gestural gaming interfaces, mixed reality applications, and VMC systems for children have been developed over the last ten years to support shared experiences that begin with shared contextual information and develop into play activities. In the following section I outline the rise in gestural gaming systems, research on mixed reality play and research designed to support shared experiences for families at a distance. This is a area that is rapidly growing within the HCI community as indicated by a workshop this year at Computer Human Interaction with 40 research groups attending entitled "Moving from Talking Heads to Shared Experiences" [Neustaedeter, 2013]. In this workshop attendants explored ways to support social connectedness at a distance and various video transport methods for synchronous and asynchronous awareness and rich social media. Many of the researchers who attended are referenced in the following sections.

2.3.1 Gestural interaction and gaming applications for families

The introduction of the *Wii* to the game console market changed how families interacted with gaming systems. Nintendo, Sony, and Microsoft have all released systems that allow people to interact with screen based systems using their bodies to control characters and views through gestural interaction.

Figure 20. Sesame street for Kinect, developed by the Sesame street workshop.
The sketch on the right is currently just a concept drawing about the experience.

Microsoft Kinect for families is the most closely related commercial product in the space of family connectedness at a distance. Families with Kinect consoles can enroll in the system and interact with avatars together in games. The low bandwidth requirements of passing avatar data such as skeleton positions in 3D space triggering certain motions enables the system to work across a wide spectrum of bandwidths. The designers played with concepts of bringing the child into an imaginary scene using chroma key techniques but this was not implemented in the final version of the product. In discussion with Microsoft, it remains unclear if the upcoming version of the system will include this capability. Core capabilities of Skype will be incorporated into the next version of the XBox to be released during the holiday season in 2013.

The system allows families to play sporting games, interact with Sesame Street characters and content, and supports interaction with Disney content and characters for game scenarios that allow participants to do things like race each other, or influence the unfolding of a narrative structure by triggering actions through gestural interaction. A similar application working on the Kinect platform called *YooStar* allows participants in co-located spaces to play movie backgrounds behind their images and act as if they are in films. The application focuses on role playing and in their commercial, shows college students using the system. It has a community feature that allows people to publish their videos and share them with others on the web. Most of the videos on the site are of parents and children playing together and pretending. The median age of the children on the site is 8-10.

The introduction of the Kinect camera and the speed at which it was hacked by creative technologists also provided numerous opportunities for technologists to build installations for parents and children that are gestural in nature. Chris O'Shea and Theo Watson are artists who have worked during the last six months on applications that are relevant to this thesis.

Figure 21. Weather Worlds, Design-IO, Theo and Emily Watson, 2013

Artists [Watson, 2013] [O'Shea, 2011] have utilized contour analysis systems which detect peaks as arms and hands to track and map graphics to people's bodies in interactive gallery spaces. This work is particularly interesting when viewed in video form because the gestural movements of the participants' bodies influence the environment and the manner in which the participants are rendered through shaders and effects that can be applied relative to bodily movements. Lightning, shooting stars, snowballs and other magical effects are evoked in response to particular gestures. Chris O'Shea employs similar tactics for more theatrical purposes. Children compose a script based on drawings they have made which get rendered to the screen at particular times and respond to the position of their bodies on a live performance stage. During our user studies, our participants expressed a desire to see similar effects in the *WaaZam* environment. These are currently being developed and incorporated into the project.

2.3.2 Mixed reality play at distance

There are several studies that have already explored synchronous mixed reality play for families at a distance with great depth. The three most relevant to this thesis were conducted in a series of collaborations between Microsoft research and Georgia Tech, which I summarize and discuss in this section.

In 2012 Yarosh et al conducted one of the most thorough studies of play at a distance [Yarosh, 2012] by comparing play patterns across four common conditions: (a) Traditional "vanilla" HD videoconferencing (b) using a pan tilt camera operated by a facilitator (c) an augmented floor projection on a rug and (d) a mobile phone with a live video connection to a participant in another room. The children played for 10 minutes in each of the conditions and the videos were coded for play patterns and a peer play ratings.

The researchers had some unique insights. First, the child's model of the other person's view was incorrect or vague at times. The researchers call this the "intersubjective view" of the other person's space. Second, making toys interact a distance was awkward and difficult for the children to coordinate across all conditions. Third, children sometimes had problems managing visibility both when the camera was in their control and when the researcher controlled the camera.

The paper concludes saying: *"Supporting free play across distance may potentially increase the number of opportunities that children have to engage in social free play. However, current technologies for remote free play present some challenges: making sure that your partner can see what you are trying to show (and cannot see what you are trying to hide); understanding when your partner is attending to you and making sure they know when you are attending to them; and, determining how to act towards each other and each other's toys."*

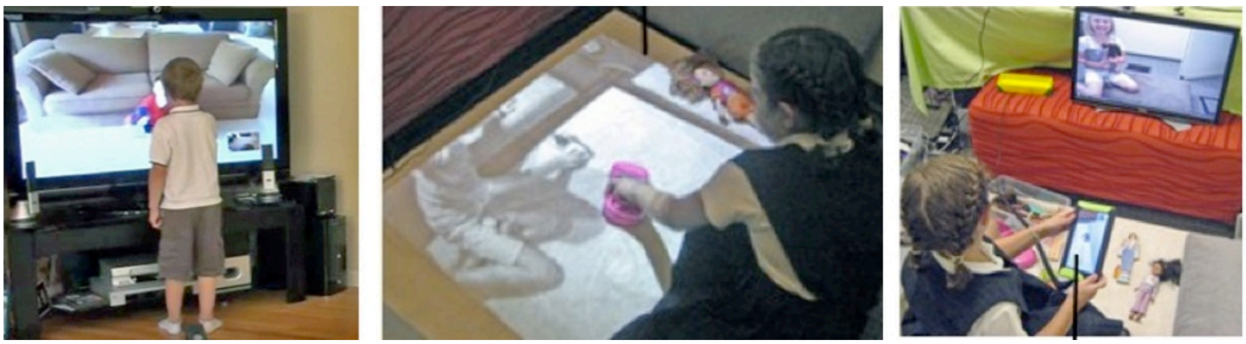

Figure 22. Towards Supporting Free Play at a Distance: pilot study, HDVC condition, floor projection, and mobile conditions. The study was conducted with 26 participants in 13 pairs of play partners.

They continue to postulate possible solutions to these problem by describing a version of the *WaaZam* system:

"With a shared view, it may be easier for the child to attend to the TV while presenting their toy to their partner's side of the screen. Though the children will not be able to feel their toys touch, they will be able to see and gesture with the toys in the same area of the screen. As computer vision researchers develop better algorithms for background subtraction and person-tracking, we may be able to combine the local and remote video streams in more sophisticated ways."

This comment in the discussion prompted me to contact Yarosh for her feedback on our design, and she has continued to provide guidelines for evaluation of our system as well as suggesting possibilities based on her observations of children using the *ShareTable* system and the research conducted in this project.

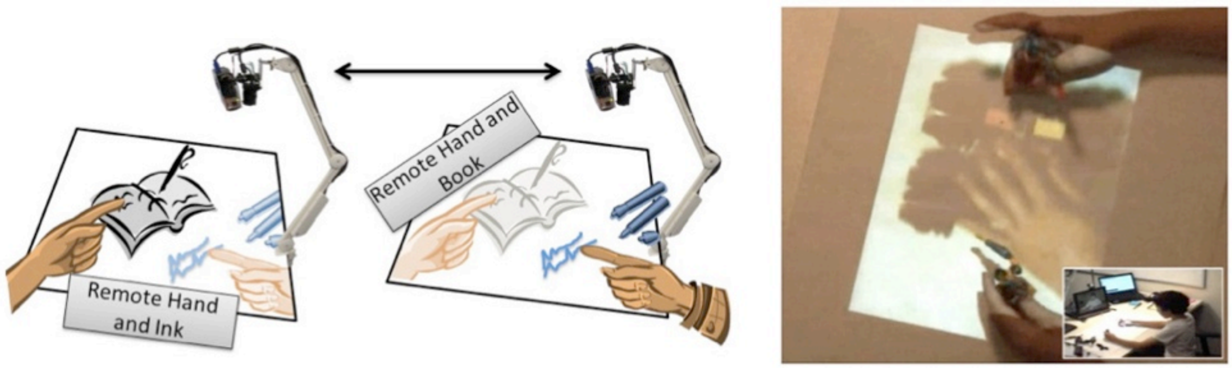

Figure 23. The Illumishare system, 2012 and the evaluative user study with children.

Subsequent to the research on supporting free-play a related system from Microsoft Research entitled *Illumishare* [Junuzovic, 2012] was developed and evaluated. The system allows people to share physical and digital objects on everyday surfaces at a distance using a time differencing between capture and the projection refresh rate. By using a camera that is 120 fps and triggering the camera at times when the projector is refreshing you can effectively remove the feedback problem common in such systems.

Junuzovic found that children preferred having both spaces as opposed to having only one and that combining spaces makes for a natural and seamless interaction between participants. Further the researchers found that removing the task-reference space caused stronger negative disruptions to the play task and engagement level than removing the person space (facial awareness). By merging the the two spaces into one the researchers demonstrated that users find it natural and accommodate for the overlap, even exploiting the freedom of being able to overlap with each other.

Yarosh et al [Yarosh, 2011] proposed design guidelines for supporting family connectedness at a distance as per the following: a) Provide multiple channels for communication (most families complain about audio only communication) b) Make it simple enough to operate without the help of a co-located adult so that children can be free to communicate at any time c) Support diverse play activities and leverage activities that the children have done with the adults before and d) Provide opportunities for care activities on the part of the remote parent (such as help with homework etc.)

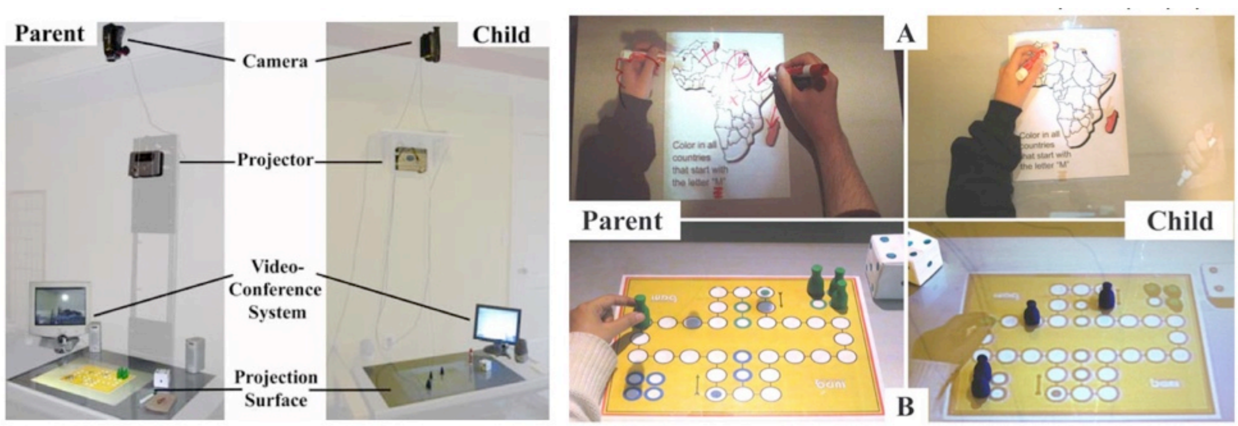

Figure 24. The *ShareTable* system setup, example of shared activity, and the synchronized desk setups

ShareTable is composed of an overhead camera an projector and a diffused horizontal surface where shared activities can be conducted. Like *IllumiShare* the setup created digital overlays of the other participant but did not use an inter-frame tracking technique. The researchers found that session usage during shared activities at the table were longer than a videoconferencing control group. In qualitative interviews of the participants they found that children as young a seven years old were able to use the system without an adult and that interactions were qualitatively different.

Most telling was that may of the participant families wanted to know when such a system would be commercially available for them to purchase. The researchers conclude "Though we cannot directly measure if the *ShareTable* brings two family members closer, we can triangulate self-reported relationship satisfaction with measures of two aspects related to quality parent–child relationships: the amount of interaction and the variety of contexts/topics involved." They measured mutuality and engagement as well as the variety of activities to determine if satisfaction increased by looking more at the actions of the participants than their interview reports.

These studies are closely related to composited video environments because they combine the two views into one frame and attempt to support shared experiences. In evaluations of peer to peer and adult-child play, the studies found that mixed reality scenarios where physical objects could be mixed with digital content were the most seamless and natural for parents and children. Additionally it was found that having a shared activity was more important to children than seeing the face of the adult in terms of engagement and the variety of activities that the child was willing to do with the adult.

2.3.3 Supporting synchronous shared experiences

The final category included in the background section relates specifically to supporting shared experiences and activities at a distance. In our review of the literature and correspondence with other colleagues doing similar work what is really lacking in communication at a distance is a shared context for an engaging activity that makes people feel like they are performing the activity together.

An influential paper in 2012 was published in DIS called '*Focusing on Shared Experiences, Moving Beyond the Camera in Video Communication*'. *This paper* provided some core insights for researchers interested in supporting synchronous family interactions. First, people experience less self-consciousness when using video communication systems as they become more accustomed to the medium. The researchers also claim that video communication constrains physical behavior while exaggerating displays of attention and inattention (in some systems). Most individuals choose to adopt video communication into their personal lives as a result of relocations. Professional adoption is typically imposed as a result of work on projects with distant collaborators. The research notes that video communication helps people tolerate distance in their personal relationships and include remote people in group activities and events to demonstrate membership in that group. Finally the researchers noted that over time individuals were moving beyond conversation only video communications to repurposing video in order to engage in shared activities. Following publication of this paper, a small group of dedicated researchers have studied dedicated shared activities for families at a distance.

Figure 25. The *Family Story Play,* A system for reading with children at a distance.

Nokia research lab played a key role in emphasizing the importance of shared experiences for families through the research of Raffle and Ballagas. In an article on pervasive computing in 2010, the authors published an overarching article outlining a new space for networked family communication that is focused on shared experiences, emphasizing their research in Family Story Play [Raffle, 2010]. StoryPlay consists of networked books in which a grandparent or parent can tell a story with the help of an intermediary character. such as Elmo, which prompts the parent and provides feedback to the child about different aspects of the story.

Figure 26. The *StoryFaces* project in which children can insert themselves into a story

The researchers found that parents and children benefitted more than grandparents and children through shared activities and that the activities increased engagement by up to 400% in trial studies with the system. The researchers advocated for more systems to support long-distance communication with very young children through reading and interaction in a shared context. They emphasize the need for parents to be involved in setting up the calls and maintaining awareness and control throughout the experience.

Raffle and Ballagas have also experimented with software approaches to support story time with grandma. Raffle and Ballagas experimented with ways for children to incorporate their faces in a story and found that both interfaces increased engagement for children by allowing them to participate in the story. This research led to the VideoPlay project [Follmer, 2011] which allows children and parents to insert themselves into a story and attempts to support them by suggesting scaffolding playful activities.

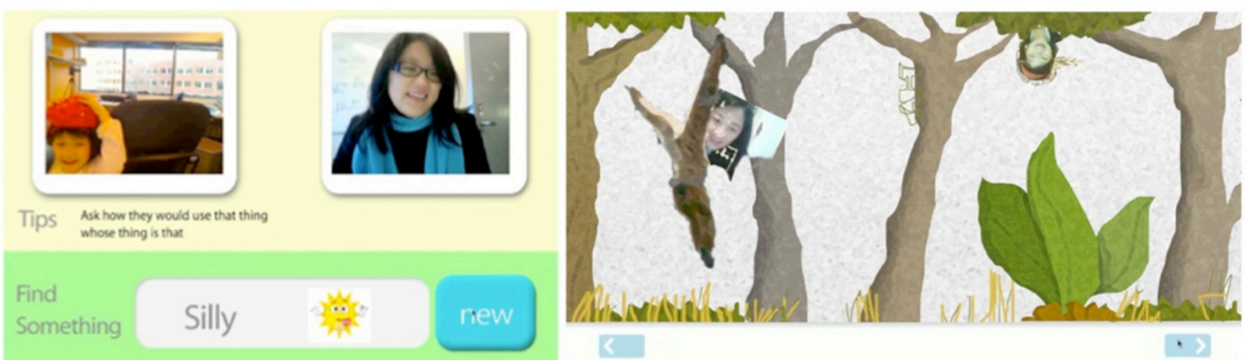

Figure 27. The *VideoPlay,* a system for scaffolding video play at a distance.

Folmer et al found that there is need for content to scaffold the activities of parents and children, that physical objects help to spur conversation and activities, textual prompts were helpful to parents, and that being together within a story increased engagement for both the parent and the child.

Figure 28. The *Family Portals,* 2011 supports shared video activities for up to three families.

A similar system called *Family Portals* [Judge and Neustaedter, 2010] is an always-on media space that provides a continuous video connection between three households. The portal is audio only and allows users to draw in a shared space and message each other through typing and playful interactions such as making faces and costuming the body. The researchers found different patterns of behavior emerged dependent on the relationship of the dyads to each other and the multiparty capabilities were largely unused in the families they studied. They noted that the most popular activities were shared augmentation of the space and asynchronous messaging.

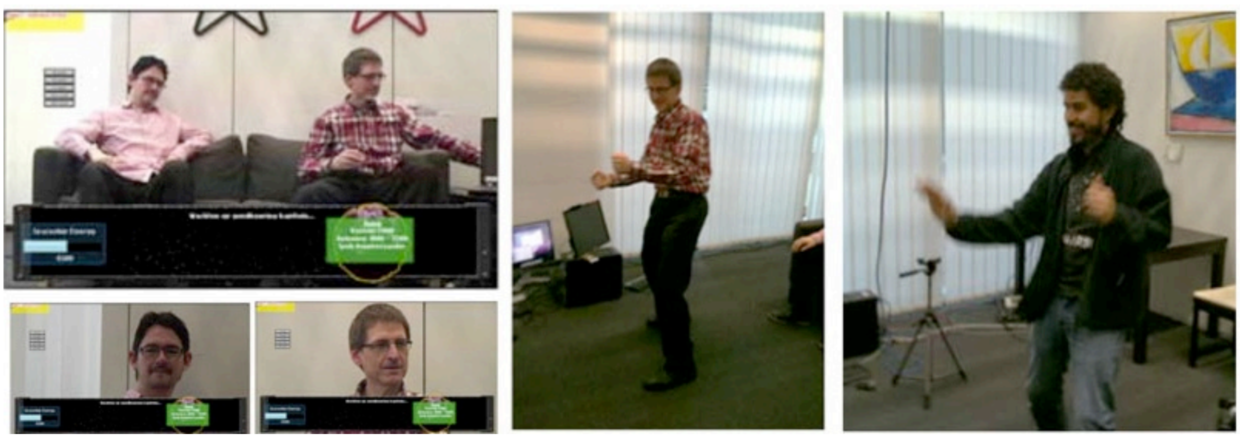
Figure 29. Remote interaction as envisioned by Villemard and the *TA2 System,* 2012

The TA2 project from BT Research [Kegal, 2012] is a high quality audio visual system that supports interpersonal interaction between families through playing mini-games together. The major contribution of the research involves supporting social video conferencing around a shared activity and providing an interaction architecture with the ability to dynamically modify the audiovisual streams so that the movement of participants, verbal and non-verbal interactions and changes in shared activity can be presented effectively and in high resolution. Finally they provide a meta-channel that supports application data for integrating content streams between participants. The project is the most mature to date and has been used in long term studies with families and provides extensive data regarding the social and technical requirements to support a more intimate feeling of being together at a distance.

CHAPTER 3. DESIGN AND FUNCTIONALITY

" The creation of interactive interfaces carries a social responsibility. Whether we intend it or not, we're redesigning the ways that we experience the world and each other." –David Rokeby

This chapter provides an overview of the motivation, design guidelines and functionality of the *WaaZam* platform. The first section outlines the motivation for the design and principles that guided our implementation, followed by a general overview of all the features in the software. The second and third sections focus on describing the functionality of the software interface. These include the set design software (the scene maker) and the networked play environment (the scene player) in which people can be in the same space at a distance.

Figure 30. An illustration of the concept of a networked play environment

Our overall goal in the *WaaZam* project is to allow parents and children to build worlds and play together at a distance. In order to facilitate this we first designed a compositing environment and a network protocol. Then we added a scene renderer with a network syncing mechanism in order to keep each side synchronized. After our pilot study we built a scene maker in order to allow users to create scenes and remix the scenes of others. Finally we added a gestural menu to allow people to access all the features supported by the platform with a full body interface. Please see Chapter 4 for details on the technical implementation of these features.

3.1 Design Principles and General Functionality

3.1.1 Motivation and Design Principles

The design goals that guide this research resulted from reflecting on previous research projects and observing how people engage with each other in interactive systems. My motivation as a designer is to enable people to create and participate in magical digital experiences with each other. I believe it should be easier for people to do creative things together at a distance. Everyday people should be able to create and participate in interaction design and form communities around these activities.

With these goals in mind, I have tried to identify aspects of interface design that enable people to interact in more imaginative and transformative ways with new software algorithms, sensors, and displays. What new experiences will be possible when our phones, computers, and tablets allow us to do things together in a mixed reality? Each point below guided the design and serves as a provocation for designers interested in designing more engaging experiences.

1) Customization: Can I make my own and bring it to life?
Online massive multiplayer designers who make environments like Disney's *Pixie Hollow* and *Club Penguin* understand that every child wants to make something that is uniquely theirs. Children go through an extensive process of choosing the attributes of their character. They modify its skin, hair, eyes, wings, costume and accessories before entering the world. This is in part to motivate the child to feel some ownership of the avatar, and in part to motivate them to spend time in the world playing games in order to earn tokens. The result of this design approach is that the economy of these worlds encourages consumerism.

Many constructivist alternatives exist to online communities that have informed my design approach. *Scratch*, *Alice*, and *Stencyl* provide more open ended environments that allow the child to decide what they want to happen in the environment without the gaming structure that encourages them to level grind. Designing open ended environments that require more creativity can shift the mindset of the child from figuring out the system to discovering what can be generated with the system.

2) Pretending and Fantasy: Can I do impossible things?
One of the most compelling things about playing with children is that their imaginations are not yet limited by empirical knowledge. They love to make pigs fly, let farm animals talk, and give superpowers to their creations. Fantasy is about enacting the impossible because it is exciting! It is also a critical part of the child's development.

Piaget, Erikson, Freud, and Vygostky outline stages of childhood development that transition from egocentric to socially aware. The process begins by acting on the world and discovering object permanence and continues as the child learns to represent things symbolically through pretend play. She will assign agency to her dolls and then fantasize about what they do. Fantasy leads to rule based play, the ability to imagine possibilities, and eventually the perspective of others. This type of socio-dramatic play is more accentuated in girls than boys. Boys tend to focus on amplifying aspects of themselves, often in impossible heroic ways.

Often the stories and metaphors are embedded in media entertainment are appropriated by children across different forms of media. Henry Jenkins [Jenkins, 2006] describes in *Convergence Culture* how many children respond to the Harry Potter fantasy world, joining fan clubs and reinventing aspects of the story to fulfill their desires as they shape their identities. Part of the mass appeal of this epic series is that it satisfies fantasies we have about ourselves by casting them in a world of fictional characters with magical powers that can do impossible things. In our user study many of the children created scenarios where they could fly on a broom with harry potter or stand inside his castle with their parents.

3) Transformation: Can I become something new?

David Rokeby describes media interfaces as “transforming mirrors” in his essays about the metaphors that underly digital experiences [Rokeby, 1991]. He writes that they operate like a “wayward loop of consciousness through which one's image of one's self and one's relationship to the world can be examined, questioned and transformed.”

The virtual space is a liminal space that reflects aspects of our identity and allows us to play with our reflections. Sherry Turkle in *Life on the Screen* says the imperative to self knowledge has always been at the heart of interaction. She portrays online environments as liminal spaces that are between real and fictional, not quite a mirror of the child but reflecting aspects of themselves that they project and transform. Transformation is imbued with surprise and delight because it allows us to experiment with our identities.

One safe way for children to do this is in their homes with their toys. Another is in applications that capture and transform their movements and expressions. Hayes Raffle and Kimiko Ryokai illustrate this idea their work [Raffle, 2010]. The authors conclude: "Our preliminary results suggest that digital authoring can give young children an opportunity to play and reflect on their pretend emotions, and that when these emotions are cast into the context of a traditional narrative, children can engage with and meaningfully transform elements of those stories."

4) Interactivity: Can I make it respond to me?

The speed and automated behavior of computer animations give us the impression of something that is alive and capable of responding to our actions. This capability can be used as a learning tool if we allow children to define relationships between inputs in the world and outputs on the screen.

A child might inquire about what is possible. Can I make it do something when I speak? Can I control it with my body? Interactive designers ask themselves similar questions when they are trying to envision how people will use their software. Why not let children also make choices that inform designers about possible uses?

5) Time-Based Storytelling and Playback: Can I tell a story?

Programs like *Toontastic*, *iStopMotion*, and *Reel Director* are designed to help children author time based media. Narrative is how culture creates and explains life to children and how children construct explanations of life. These programs scaffold the process of creating stories and encourage reflection by allowing the child to replay and modify portions of the timeline. The capability of recording and manipulating video make graphic user interfaces ideal for the creation of time based stories. The linear nature of time can be navigated with tool sets that provide high level representations of time.

6) Social Play at a Distance: Can I play with my friends?

Digital representation allows information to pass so quickly over networks that interaction at a distance feels instantaneous. The richness of these experiences is increasingly compelling in applications like Skype, Google+, and iChat.

Research at Arizona State University [Freed, 2010] has found that children understand the idea of remote connection and can relate through the perspective of dolls. The dolls act as bodily extensions of the child through which they can enact their own concerns. The dolls can also scaffold interactions between children and provide a context for an ongoing dialogue between the characters.

I envision a set of applications that would allow children to invite friends to play in environments they create instead of those designed by companies like Disney. In such a world, children would share assets they have created: character costumes, effects, and even behavior scripts. They might even collaborate to record stories of their characters and share them with each other.

Conclusion

In conclusion, digital representation has the potential to allow children to create their own worlds and make things respond to them in those worlds. Fantasy and storytelling are critical aspects of the their development. Virtual play environments can provide a safe place to enact and explore their ideas about the world.

As children interact with media, they project and transform representations of themselves – enacting possible futures and shaping their identities. The computer becomes a transforming mirror of the child. Interfaces have the potential to empower children when they choose the mappings between the physical and virtual world. Results may take the form of movies, animations, interactive stories, and even games that they can share with other people. *WaaZam* is intended as an example of how this could happen in a networked environment where families can simultaneously create things together.

3.1.2 General Overview of the Functionality

WaaZam is an environment that allows people to make video calls in creative environments that they customize and share with others. In its current form users are paired with each other through a server which allows the application to stream realtime video, audio, and metadata of the scenes and the users.

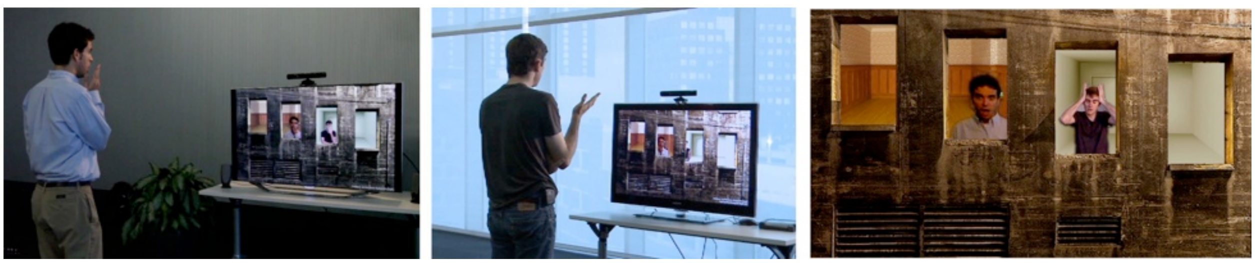

Figure 31. Two participants in a networked video call in WaaZam pretending they are in an apartment together.

It is designed to support shared activities by allowing people to make networked video calls where both participants are in a shared space and can do things together. The software supports three configurations: a more traditional videoconferencing "Skype" configuration where each user see the other's space, a "mirror" mode where one user appears in the other's space, and a "digital" mode where both users appear in a shared virtual space.

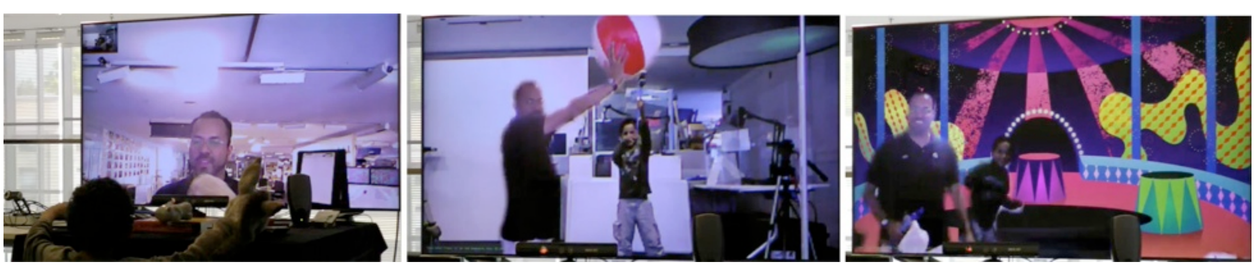

Figure 32. The three primary configurations: "Skype", "mirror" and "digital".

In the "digital" configuration users can transform themselves in the environment and place objects in a layered depth space that is mapped to the 3D space in their physical environment. The ability to hide behind objects and transform themselves creates the illusion of inhabiting the virtual space.

Figure 33. Users interacting in a customized scene together, hiding behind layers in scenes, and using transformation.

Users can design their own sets by either capturing physical objects using a web camera and placing them in the scene-maker, or by searching the web for assets related to the theme they are imagining. Sets are composed of a background image and various foreground images that are placed in the layered depth space.

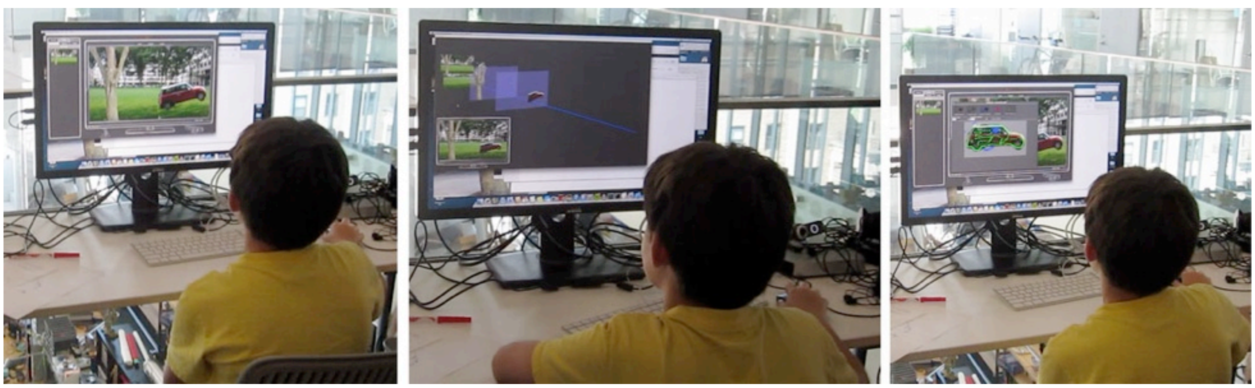

Figure 34. Users creating a scene, ordering foreground images as depth layers, removing a background.

The environment supports embedding animated png elements for sprites with transparent backgrounds and embedding movie files for shared activities such as dancing to a youtube video. The sets are sharable with other users of the *WaaZam* software through the web syncing function when users click the "publish" button inside the sceneMaker.

3.2 WaaZam Environment Design

The *WaaZam* environment is intentionally open ended and simple in terms of the UI and interaction features so that users can focus on being socially connected and immersed in the environment. Our design process was based primarily on watching people interact with the system and making changes based on our observations and their feedback. What follows are general design guidelines for merged compositing environments, our approach to tracking and rendering people, the gestural menu and the design of the scene maker environment.

3.2.1 General implementation guidelines

During the initial phase of our development we reviewed design guidelines for natural user interfaces and the work of artists and designers who implemented full body interfaces in their work. Using an AGILE design process we outlined a set of design suggestions that guided our implementation process. We present these not as comprehensive guidelines but as practical considerations that resulted from our process and that may be helpful to others.

1) Design play scenarios without dependencies on traditional GUI interfaces.

Using the mouse and keyboard requires a table and close proximity to the interface. This interrupts the flow and movement for participants. We noticed that when interacting with the GUI the flow of play and pretending stops. The most seamless manner of allowing users to change activities is a gestural interface similar to XBox Kinect applications. With this, users can access a menu via an unlock gesture which reveals modes, content, help, and hints.

2) Mirror the physical space to the digital environment in an intuitive manner.

Mirroring the physical and the digital space is the most natural way of interacting with screen based content from the perspective of most users. The metaphor of looking into a magic mirror seems fairly embedded into our cultural understanding and provides a starting point for guessing how to interact with the system that translates for children as young as four years old in our studies. In our application we map the available screen space to the 640x480 capture of the user by scaling the video image to the extents of the screen.

3) Allow users to design and build their own environments.

Our application encourages creative forms of play and customization. In order to facilitate this we designed the system to be as flexible as possible, allowing people to choose what is tracked and design the environments in which they play together. We found that this approach gave users the most freedom to improvise and adapt the system to their needs.

4) Define as minimal a gesture set as possible to reduce confusion.

In our application most users were new to gestural interfaces and to date did not have an intuition for gestures that would allow them to transform themselves, choose scenes, and use the available functions in the system. We chose fairly distinct gestures for our menu that were based on normalized joint angles of the participants and followed the nearest hand to control a selection cursor.

5) Support and track multiple users on each client.

Our system supports up to seven users on each side; however, this slows down the frame-rate of the system by 8% for each user added. We optimized the software for 2-3 people on each side and found that this supports most use cases outside of demos for large groups.

The system needs to adapt to the height of the user and adjust the size and position of the user accordingly. One of the advantages of the digital modes is that you can position people relative to the environment. We save camera angle and render settings of enrolled users who start and stop the system.

The ideal position of the camera for children is below the TV whereas for adults it is above the TV. In each case it feels as if the users are standing face to face. If a TV were to be designed for this application it would need to have sensors on all sides of the screen and correlate a perspective from multiple depth images.

6) Design a syncing protocol that applies to objects, users, and the user interface.

Our interactive menu and the shared objects in a scene can be controlled by both clients. To facilitate awareness and negotiate conflicts between users we employ a lock on content. Objects are selected after we check to see if they have not been locked for a few frames. This is partially a technical constraint to prevent a race condition where two users select the same content within the same 100 ms window, generating a loop of lock/unlock messages between clients.

7) Design for the resolution of both the screen and the sensor.

The size, resolution, and aspect ratio all effect the immersive quality of the experience during a video play session. For smaller screens, the range of the depth image should be moved towards the screen (both near and far thresholds) to accommodate for the viewing distance and close proximity of users to each other.

To resize the image for different screens, we used openGL viewports that scale depending on the screen resolution. The quality of the image capture is low on HD screens. In the future we expect time of flight cameras to support high resolution capture and a greater depth range. The shared activities of participants will vary depending on the capabilities of the sensor array and the size of the screen.

8) Include hints which guide users to features and settings.

Although we borrowed many gestures from the XBox, NUI Group, and OpenNI development processes, most users are not familiar with them. In order to inform everyone of the systems capabilities we employed a system of timed hints linked to a state machine. When the user first enters the application, they see a calibration hint, followed by an "open menu" hint and a "control cursor" hint. Other options such as transform and record gestures are displayed dependent on if the user has used them previously.

9) Provide a wide variety of example scenes.

In order to encourage a variety of creative play activities we have included 20 example scenes in the library of the *WaaZam* application. These examples scenes made by previous users of the system are intended to inspire users to try new activities together, remix scenes, and create scenes.

10) Prioritize frame-rate over image quality for shared activities.

The quality of the image allows you to see the expression of the other person more clearly and detect subtle emotions conveyed by the face. In *WaaZam* our focus was on body play, object play, pretend play and narrative play which utilize the entire body in context with the virtual environment. Because of this users were very tolerant of poor image quality due to resolution, compression and segmentation. We expect sensor capabilities to improve in the near future, which will solve many of these resolution and segmentation issues. What is more important for shared activities is the response time when users are overlapping and playing. If the delay is too great, they resort to turn taking which detracts from dynamic capacity of the environment.

11) Design for users that exploit the limits of the sensor and tracking system.

Throughout our user studies children played hide and seek by running in and out of the camera view and tested the capabilities of the sensor. When a user steps off screen we hold state information about them for 30 seconds. Many children were puzzled when they stepped of screen and did not appear immediately or if they were scaled and appeared resized when reentering the screen. We were able to solve many of these problems by adding a mode in which the raw images are merged together into a single image based on the information in each pixel of the depth images.

3.2.2 People, objects, and gestures

The user tracking module described in Chapter 4 outlines how we track the skeleton, contours, and users in our system. During our development process we utilized OpenNI to track and map users. We also process the raw images in openCV to allow users to experiment with different near and far thresholds.

Figure 36. The raw depth image and resulting mixing modes during initial development phases.

We referred to the raw compositing modes as the “puppet mode” and the “ghost mode” to users based on the most common activities we observed during the pilot studies. We referred to the advanced tracking as “everywhere mode”. In this mode users can transform the scale and position of themselves or others.

Figure 37. The three rending modes: ghost (composited), puppet (near range), and everywhere (allows transformation)

By allowing users to choose modes they can decide what they want to be tracked and how it is composited into the system. One of the most important things we wanted to support is allowing people to customize their environments and how they are rendered in it. The more freedom we extended to users, the more they were able to pivot if their attention waned or someone wanted to try something new.

Figure 38. A user masking contour, skeleton and hand tracking, and gesture recognition based on skeleton data.

The tracking system also uses Nick Gillian’s gesture recognition toolkit [Gillian, 2013] to define a set of gestures that unlock a control menu and allow users to interact with the menu.

We are also currently working to support effects, costuming, and interaction with objects in the environment based on a specific set of gestures and mapping objects to the user skeleton when these gestures are enabled. Although these are experimental features, they were the ones most requested in the formal user study.

3.2.3 Set rendering and composite characters

Set rendering is handled by a module in *WaaZam* called the scene player. The job of the scene player is recognize the current scene and display users and objects at a predefined scale, rotation, and Z order in the render mode. We use terminology from theatre where each scene can has a cast, cues, and actions.

Figure 39. Scene assets as separate PNG files, the resulting scene, and people playing in the scene.

Each scene has assets which were designed in the scene maker and saved to XML. The scene-player manages loading scene data, switching between scenes, rendering assets in the correct Z buffer, and applying effects and animations. Each scene has a cast with a list of all the assets. Each scene also has multiple cues, actions, and keyframes. The purpose of this structure is to allow users to include animated sprites and switch between multiple cues within the application.

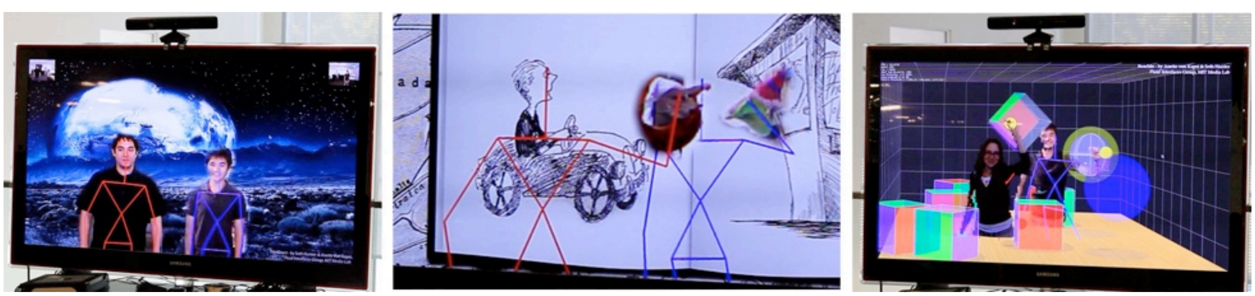

Figure 40. Skelton data rendered in 2D mode, for puppet tracking, and rendered in a 3D environment.

The original design included both 2.5D rendering and 3D rendering depending on the application and scene data. Although our system supports 3D interaction and rendering we did not introduce this capability during our user study. We were focused primarily on research questions regarding engagement in merged environments and the value of personalization for adult-child pairs. In our previous tests we found that in 3D mode the users focused on the interaction UI of the system more than the social interaction between each other. For other applications such as collaborative gaming we think the 3D capabilities would increase engagement by immersing users in a shared task.

figure 41. Example scenes from the user study demonstrating scene making and scene rendering.

The scene-player is built on top of OpenGL and is responsible for rendering the realtime view. Although it is capable of displaying anything one might imagine from the web or photoshop, the rendering itself is similar to OpenGL where you have a list of assets, positions, scales, rotations, and alpha properties. Unlike OpenGL the 2.5D view is limited to a single camera perspective. This provides a constrained design space which works well with the mirror metaphor.

3.3 WaaZam Features

We have developed a set of modules that extend the capabilities of the system. In order to support feature requests and make the UI accessible to all users we added a gestural menu, transformation gestures, effects and costuming, support for object puppetry, and video recording capability. In this section we describe the features and document their implementations.

3.3.1 Gestural Menu

The gestural UI allows users to switch scenes, change modes, and manage recording. Users access the menu by holding their arms in the unlock menu pose. The menu appears with four options available: scenes, modes, record, and close. The user holds out their left or right hand to control a menu cursor.

The cursor is locked to a proximal zone that is calculated based on the normalized distance of the hand relative to the torso of the user within a natural range of movement and smoothed by averaging the current position into an array of previous positions weighted towards the current position. When the menu is revealed the user can either go back to the previous state or switch states. The submenu allows the user to navigate a list of scenes and effects by category.

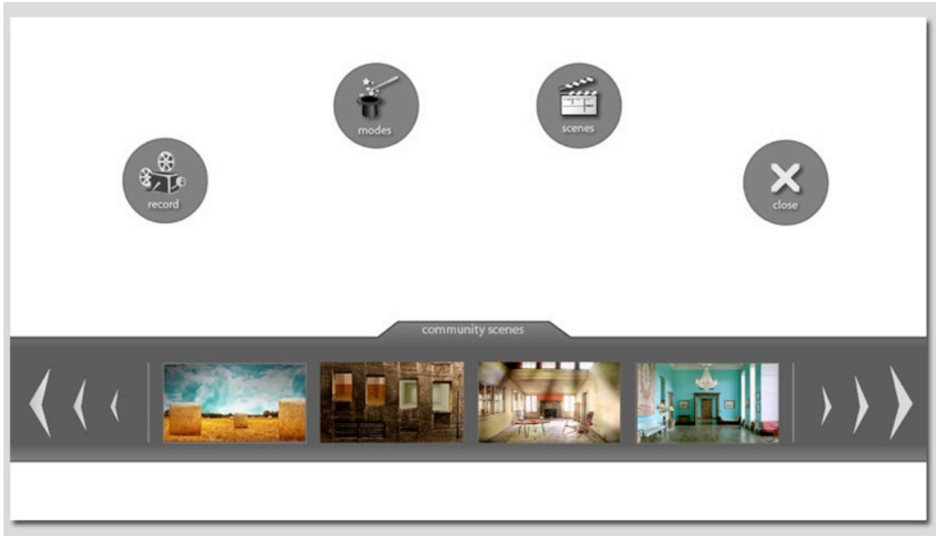

Figure 42. The Layout of the menu and submenu for scene selection.

The submenu has scrolling elements on both sides of the bar which allow the user to control the speed of the scrolling. Content loops infinitely by using a rotating linked list.

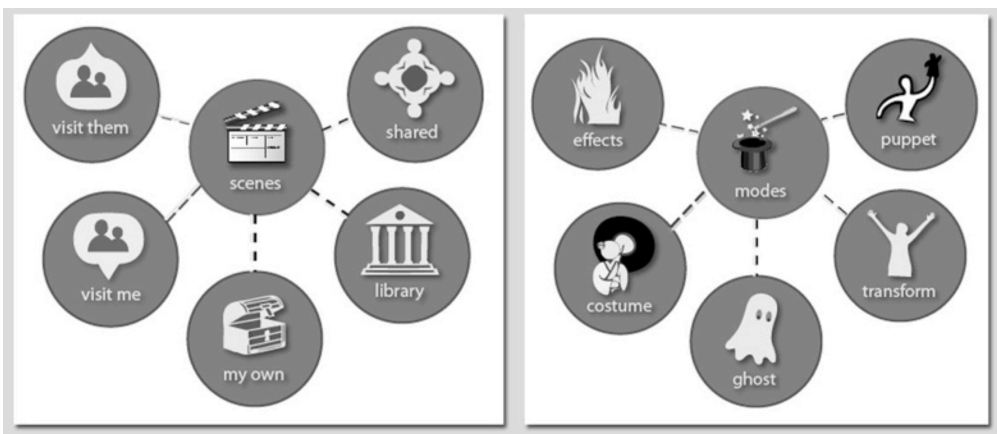

Figure 43. The scenes submenu and modes submenu.

Users can activate two submenus located relative to the scenes node and modes node. The scenes node allows users to choose my scenes, library, web scenes, local, and remote options. These correspond to scenes made in the scene maker, which has scenes designed by artists and developers to support shared activities and scenes shared by other members of the community. Users can also visit each others spaces by choosing either local or remote views, which triggers the system to go into magic mirror mode.

Figure 44. Choosing a menu item with the hand cursor object in the gestural menu.

The modes submenu provides access to many of the features currently under development. Ghost mode, puppet mode, and everywhere mode allow the user choose how they are rendered into the scene and whether or not they want to transform themselves in the scene. Effects allow the user to trigger shaders or sprite animations in relationship to their body poses and overlap with the other person. Costuming is an experimental feature to support Jib-Jab types of interactions and encourage pretend play by adding hats and costumes similar to Google Hangout.

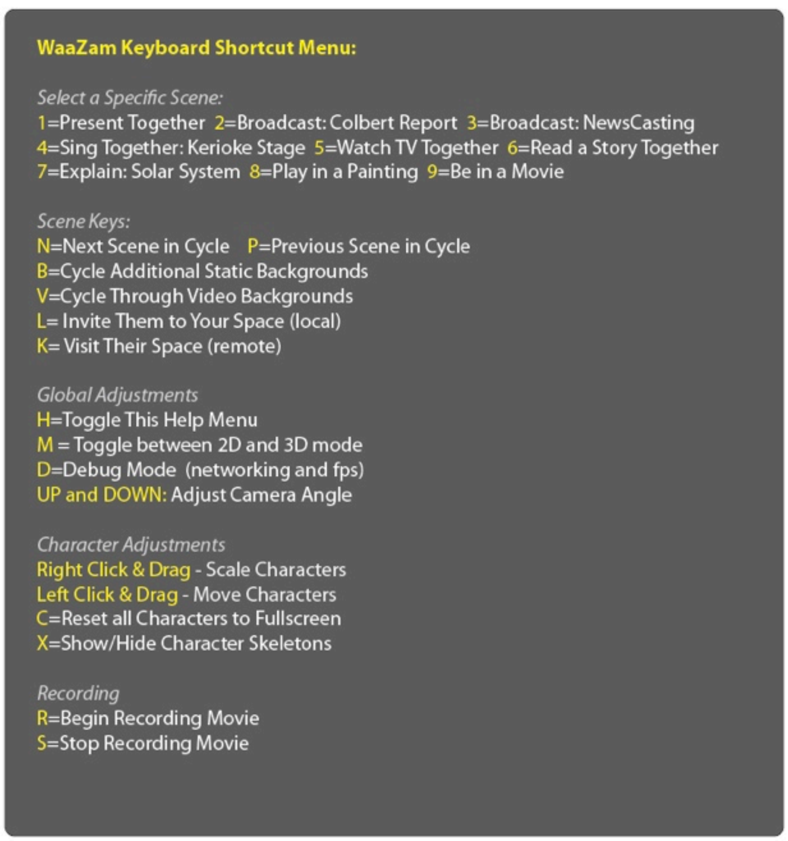

Figure 45. The keyboard shortcuts in *WaaZam* for advanced users.

The gestural menu syncing mechanism supports multiple cursors, one from each side, but the menu is locked for the other user. The menu state is mirrored on both sides to facilitate a discussion about which scenes to choose. The purpose of the menu is to replace many of the hidden functions that were developed on the platform and mapped to keys on a wireless keyboard. A help menu (Figure 45) is available for power users who are interested in using the keyboard to switch modes while recording or control the position of the other at a distance. Many parents and children enjoyed this capability during the user study.

3.3.1 User Interaction Flow Diagram

When the user first starts the application they sign in with their gmail address and we determine what their call status is, who their contacts are and if they are calibrated to use the gestural interface. Figure 46 illustrates a state diagram of the user experience. Hints are displayed in the upper-right hand corner of the screen to help the user determine which pose they should strike to trigger the gestural menu. Incoming calls are automatically accepted if the person is listed as a friend in your Google contacts.

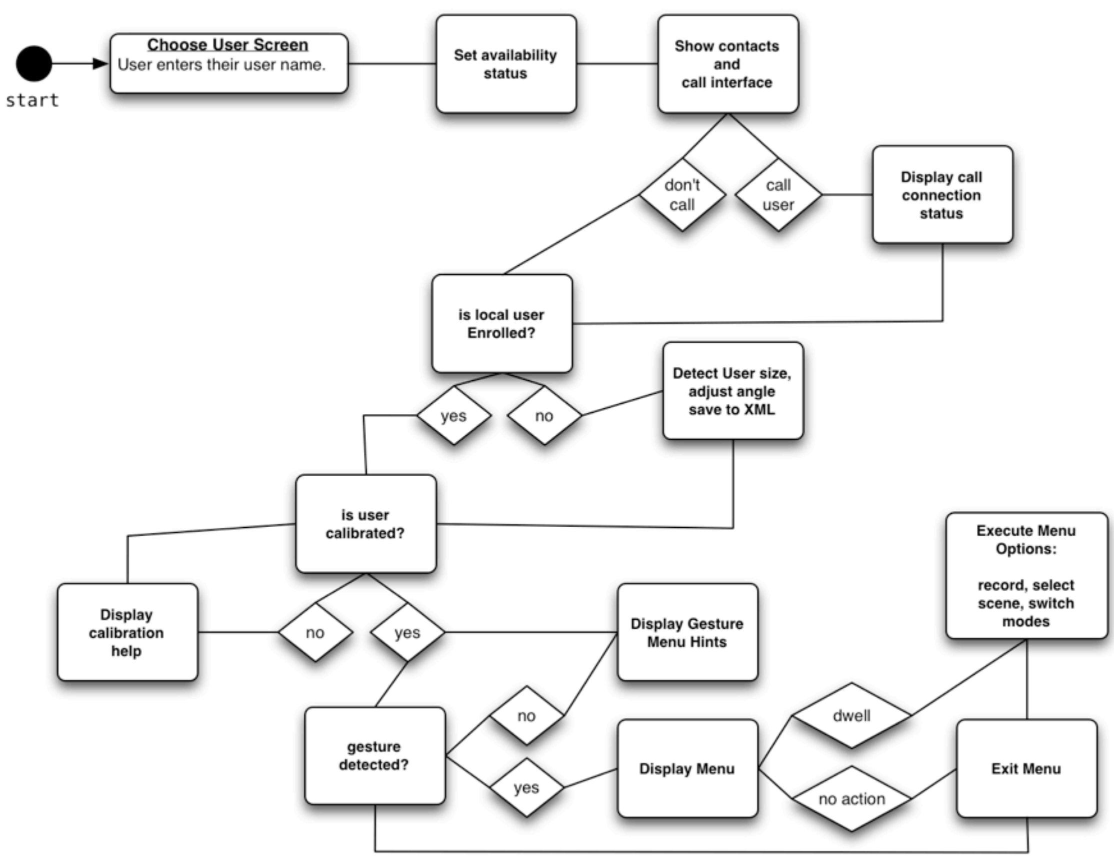

Figure 46. A flow diagram of the states of the user interface in *WaaZam*.

Information about the height of the user and their contacts are stored in an XML file so that the next time the user enters they do not have to complete the camera adjustment or call status options.

3.3.3 Transformation

One of the early components of Myron Krueger's installations was the effect of seeing yourself transformed relative to other people interacting in an artificial reality. The first feature we added to *WaaZam* was a tribute to his work and an experiment to learn how users could transform themselves using gestures within the environment.

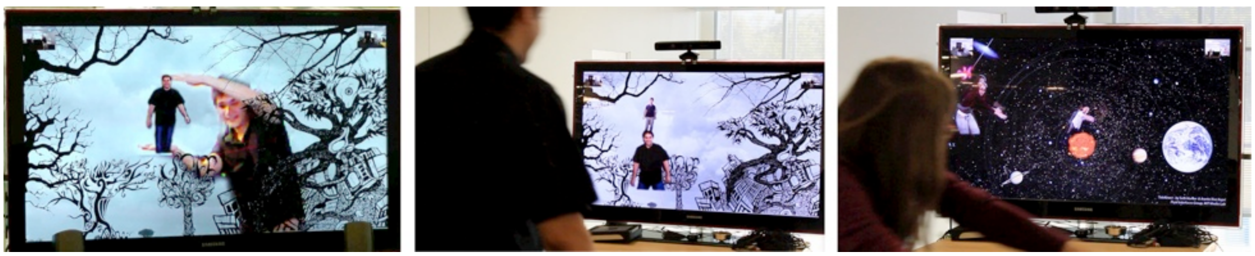

Figure 47. Examples of transformation within the WaaZam environment.

We implemented two ways to transform viewers. First we added a mouse function where when the users clicks on the body of another user, they can scale and move them by switching between the right and left click on the mouse. This worked quite well for testing out the experience with a variety of users and allowing parents and children to manipulate and puppet each other during a live session.

Figure 48. Example uses of transformation with parents and children. The image on the right shows a parent transforming his son using the parental controls.

However, using the mouse interrupts the flow of interactions between people and generally is either a transitional activity or requires coordination and planning between participants. To make transformation more seamless and natural we assigned gestures to "grow", "shrink" and "move" when the user is in "everywhere" mode in the environment. The user holds their hands strait out to grow bigger, crosses their arms to grow smaller and holds their hands above their head to move around. This resulted in many pretend play scenarios where users place themselves in a space and pretend to be in the environment with the other person.

3.3.4 Effects, moving assets, and costuming

The most experimental features in the environment are still under development and were not included in our user study. We have been working on adding them after getting feedback from users in both studies. These include effects that result from gestures, the ability to translate and scale assets on the screen, and designing costumes that are assigned to users in particular scenes.

Figure 49. Examples of users posing for grabbing, costuming, and gesture effects under development

Thus-far we have reserved a gesture for “magic” effects which include fire, ice and electricity dependent on which scene you are in. Many of the kids in our study requested comic book effects and sound effects to accompany them. Although these features are not part of the core research I believe they empower users to do things they cannot do in reality and this adds both enjoyment and entertainment of the session.

The other problem we encountered and are attempting to resolve in our development is the static nature of assets users have created and placed in the environment. Because the users can move around the environment, they expect it to respond to their movements and to have behaviors that respond to them. We are developing a two handed grab and move gestures related to Anette Von Kapri’s research, but with gestures that do not require recognizing hand poses.

Costuming is being implemented in the scene maker. There is a template skeleton to which users can attach assets such as hats, capes, emblems, and armor. These are assigned to the user who chooses a costume from the gestural menu.

3.3.5 Using puppets and toys

Puppetry appeals to children and parents who enjoy pretend play with objects. In childhood development children use puppets as proxies for their own feelings and emotions and externalize their thoughts and feelings by talking through toys and dolls. These pretend play activities are pivotal in helping children develop empathy because it requires imagining how another person might be feeling.

In the toy industry children play with action figures and characters that embody their fantasies and are often represented in accompanying digital animations. During my employment with Hasbro Corporation I developed a series of experiments involving green-screening and chromakeying which influenced my approach to the *WaaZam* design. We built a scene rendering environment for toys that allowed digital assets from TV shows be composited with live video feeds of physical toys.

Figure 50. Experiments in realtime puppetry: floating green-screen, rod-puppets, color tracking and kinect tracking.

For this work we experimented with two primary approaches. First we tracked an object in the environment and mapped digital costumes and behaviors to the object. The second approach takes a live segmented video of the object in the environment and places it in a digital setting. The first approach feels more magical and allows characters to inhabit the digital space and interact with other digital things in the environment. In the second scenario the user has to pretend that such interactions are happening but they can place any object in the environment and improvise with any of the toys they have available in the physical world.

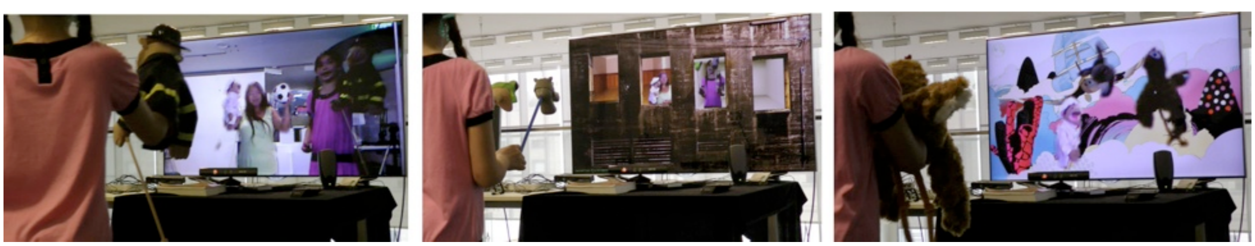

Figure 51. A mother and daughter using puppets together during a play session.

We tried three approaches to tracking puppets in *WaaZam* and compared them in our pilot study. First we tracked the left and right arm of the user, selecting the nearest arm. We then look for a region of interest that is in front of the head and select an approximate threshold range which is dependent on the current depth of the user and an average size of most puppets. We measure the difference between the joint points and segmented the puppet based a histogram of the depth and color values within this region of interest. This approach requires calibration and is problematic for children because they run off the screen and tend to drop to their knees and hold the puppets up in front of the camera, which causes the system to lose the skeleton. The advantage of this approach is that they can move freely in the environment with the puppet and never have to worry about their bodies appearing.

The second approach tracks the object nearest to the camera and sets a dynamic depth threshold. This became problematic when there was more than one puppet on either side of the screen or user's heads were in front of the objects.

Our third and most successful approach to puppet tracking was the most simple to implement. We set a depth threshold at a comfortable fixed distance from the camera for puppets to interact with each other and described it to children as an "invisible puppet curtain". Children immediately understood the idea and quickly found the boundary of the curtain and played within those constraints. The simplicity and predictability of the algorithm made it by far the most successful because no calibration was required and the computer vision system almost never lost track of the objects.

3.3.6 Recording

Recording and sharing were added to enable users to act out scenarios and share them with others. We tried a variety of approaches including: Quicktime, FFMpeg, and GStreamer. We settled on a native OSX QTKit implementation that allowed us to create a frame buffer object and output it in a compressed format as a .mov file with synced audio.

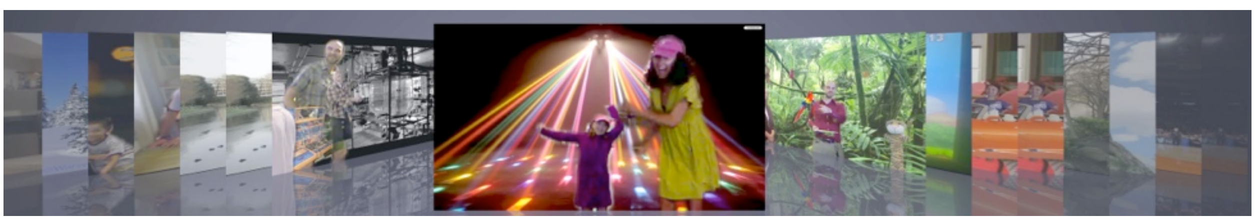

Figure 52. Recordings of play sessions made with the QTKit frame buffer capture module.

To record the user either presses 'r' on the keyboard or opens the menu and triggers the record functionality. Stopping the video occurs when the user opens the menu again and a message is displayed that the video is being saved to the user's desktop. Originally we considered allowing viewers to record, review, and edit files in the environment, but gestural interaction is not suited to collaborative editing and there are many video editing suites available on all operating systems.

One issue which we were not able to resolve was mixing the audio from the remote user and the local user with the video capture file. This can be done by combining the two audio channels with the video channel after the recording has stopped but we found that, in recordings longer than thirty seconds, the audio gets slightly out of sync with the video.

3.4 Scene Maker

The scene maker is an environment where people can quickly search for assets, create backgrounds, add foreground assets, arrange items in the depth space, and share sets with other users in the community. As we outlined the necessary functions required to support the creation of 2.5D environments we observed that all online massive multiplayer games with level design tools use the mouse and keyboard. The precision required to edit images, search, select, and arrange things on the canvas is a natural fit for mouse use and in our study was easy for children as young as 7 to use. This somewhat compromises the collaborative nature of designing a set together; however, we have found that when the audio and video remain on users can work together on a scene and maintain an awareness of each other.

This section describes the UI, 3D mode, layout GUI, and background removal tools in the scene maker. In addition we document image search integration and webcam segmentation for inserting objects from the physical world.

3.4.1 Scene maker User Interface

When users first open the scene maker application they are asked to either create a new user profile,

Figure 53. UI layout: signing in, menubar and design canvas.

or choose their user name from a list. Associating assets and scenes with a user enables us to customize the interface for each user. After identifying themselves, users are presented with two primary elements: an asset management menu bar and a layout canvas. We used the metaphor of the TV and the remote control and cameras below to arrange our interface elements. This is a direct reference to the play environment and provides a more seamless transition between the two programs in the software suite.

The menu bar allows users to choose a scene, add and manage foreground assets (props), and add and manage background assets. Creating a scene typically begins with the user choosing Scenes > New Scene and naming the scene. They can then add a backdrop by choosing from the library, searching the web, or adding a picture from their hard drive. We included the ability to also drag images onto the canvas from the desktop. This allows for easy transitions between a web browser or photo management tool.

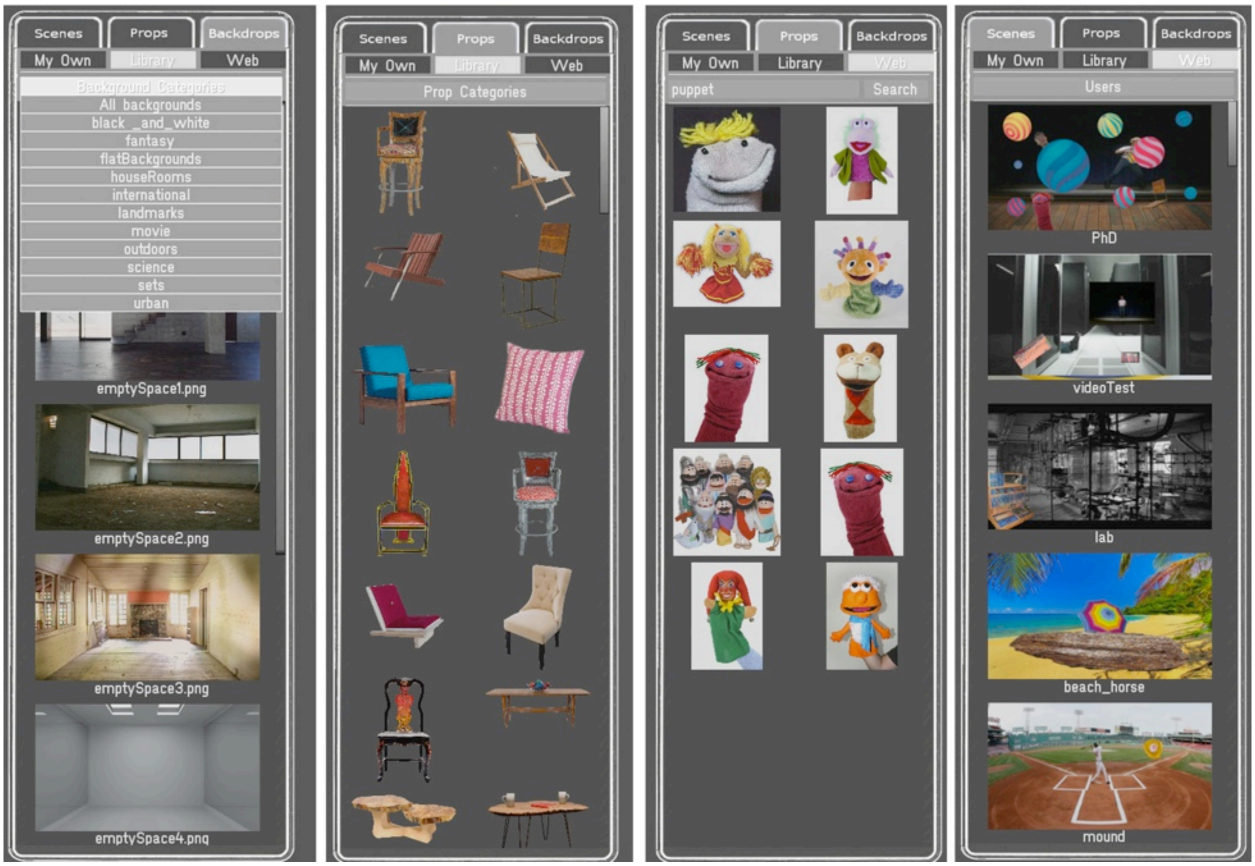

Figure 54. Scenemaker UI: choosing a scene by theme, prop library, searching for props, choosing a shared scene.

When a user drags in an element they are asked if they want to add it as a prop or a background. Background assets are scaled to fit the screen or centered, similar to how people add desktop wallpapers. Props are added at an average size around 500 px. and can be scaled and rotated in either direction. The library provides a set of commonly used props such as furniture organized in categories for users to browse.

In addition, users can search for assets in the menu bar. We utilize a custom search on Google or Bing's search API with a child-safe search filter. One benefit of this search method is its relative safety for children; however, this feature is not scalable beyond 5000 searches per month without the user registering with Microsoft or Google. Another benefit of the custom search is that we can filter out unwanted filetypes and low resolution assets that do not scale well in the environment.

3.4.2 Scene maker user functionality

If a user wants to quickly add a prop they can search within the menubar using a keyword such as "sock puppet" in the example below. Ten results are returned in an asynchronous thread and can be dragged onto the canvas as the thumbnails are loaded into the environment. If a user selects a thumbnail the full size image is requested from the web.

Figure 55. Rotating and scaling a draggable element, searching for props from Bing, and a resulting prop and it's handles.

Once an asset has been loaded into the scene maker the user has five options: move, edit, rotate, scale, and remove. These are shown on the asset when it is clicked by buttons that are center aligned on the corner of each asset. Scaling pivots in any direction from the lower lefthand corner. Rotation is based on the center of the object. The scale and rotation is saved with the prop in XML along with the scene.

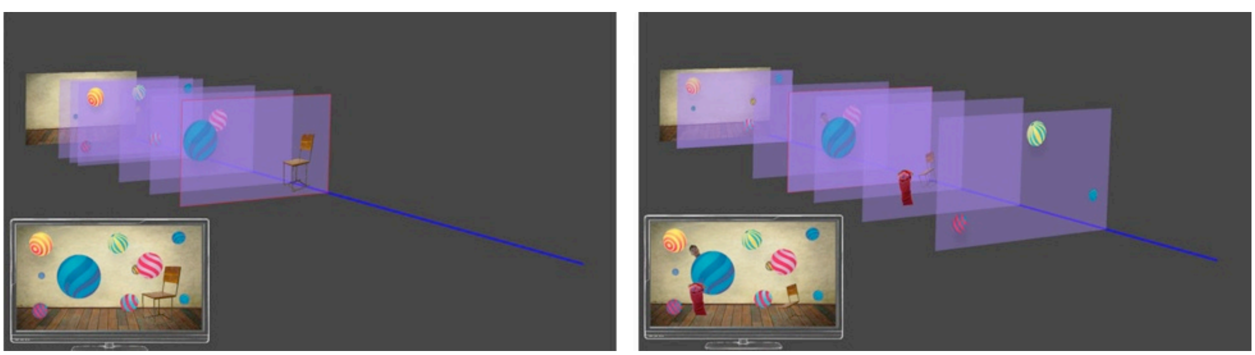

Figure 56. Sorting layers in a 3d view of the image stack according to their depth relative to the viewer.

A depth value is stored for each asset which corresponds to the Z space in an orthographic view. In order to facilitate the depth ordering we added a layering mechanism where users drag layers for each prop along a fixed axis in a perspective view. We found that this worked well for users but in the future a physical distance scale needs to be added as well as some representation of the room and the user in the space. It is unclear whether it is more intuitive to think of the assets as floating in front of the TV or the user rendered behind the TV as either visualization is valid from the point of view of a mirrored environment.

The most advanced feature we included in the environment is a background removal editor which utilizes an implementation of the GrabCut morphological segmentation algorithm in OpenCV 2.2. The function provides a mechanism for classifying pixels as foreground, probably foreground, background or probably background. We provide tools in the editor whereby the user provides hints to the GrabCut algorithm in 5 steps: crop image, provide keep hints, provide remove hints, preview, and save and place the asset.

Figure 57. The scenemaker foreground background editor: selecting foreground and background areas, generating a morphological mask, and placing elements in the scene.

When testing this with children many wanted to outline the image as if they were cutting it out. The algorithm would need to be modified to support this intuition or provided with another mechanism that allows users to dynamically select textured areas. We were surprised how much attention to detail many of the children gave to this editing process.

3.4.3 Adding digital and physical assets

The scene-maker provides tools to add both physical and digital assets to a scene.

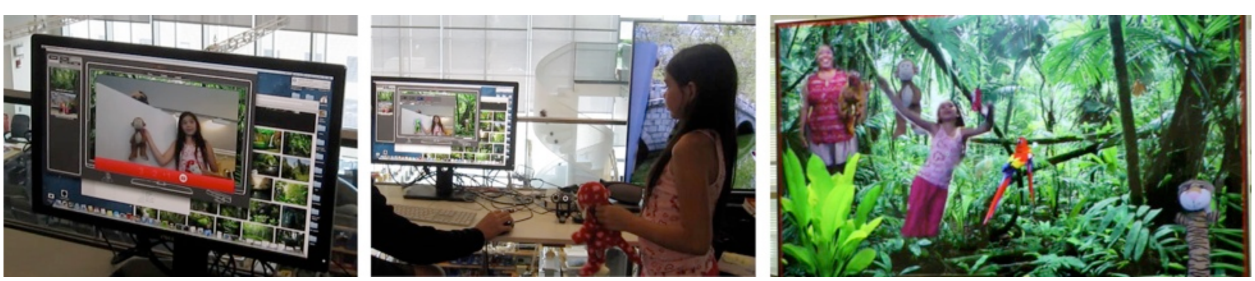

Figure 58. A user capturing her teddy bear, selecting it in the co-design session, and playing in *WaaZam*.

Initially we tried using the Kinect camera to scan objects and allow users to dynamically texture an environment by holding objects up to the screen. In our testing we noticed that it was difficult to find physical assets which one can hold without having some part of the hand occlude the object while using the other hand to “stamp” a segmented version of the object into the environment at a particular depth. The second problem with this approach is that the Kinect sensor does not work at a range closer than 18 in. It also has a large variance at the edges of the depth image due to the geometry computed by a dynamically changing laser projection.

For ease of use, integration, and precision we choose to add a module to support a host of web cameras and allow image captures from the web camera to be imported into the environment as regular props. Advanced users can hold up blank backgrounds while capturing to make editing the background easier later.

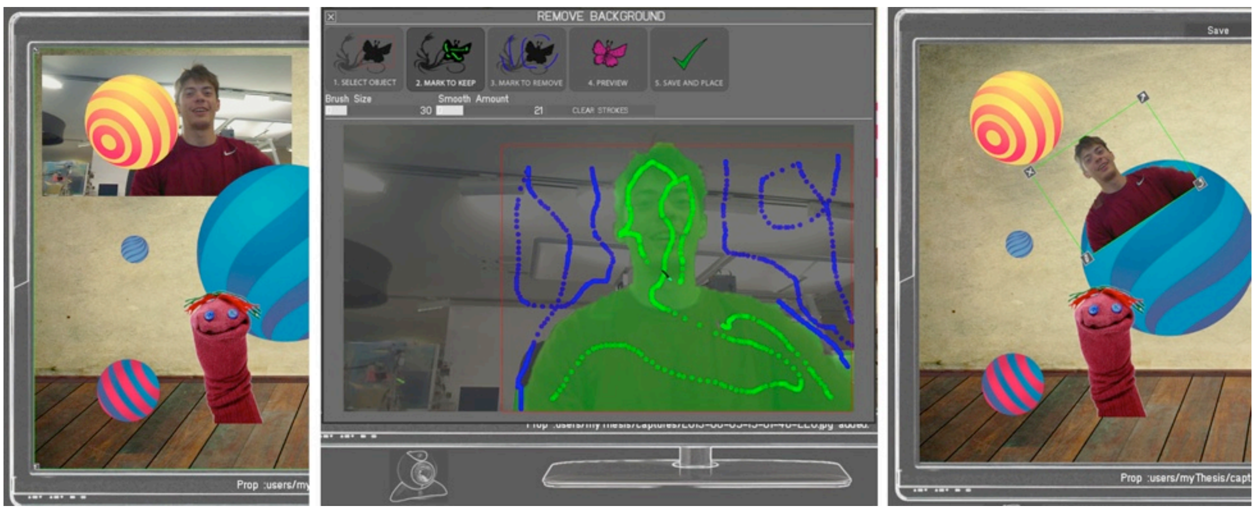

Figure 59. Adding physical assets by choosing the capture picture button and removing the background.

We found that the most common use of this feature was to show LEGOs, stuffed animals, and objects from the environment that users found meaningful, but that a majority of the users preferred to insert objects by searching for them on the web because it allowed for more flexibility and imaginative possibilities.

3.4.4 Publishing and remixing scenes

The example used to demonstrate the features of the UI in Fgure 59 was remixed from a scene in the library. Users can remix an existing scene or a the scene of another user. Research indicates this is a valuable learning tool in the *Scratch* community [Brenan, 2012] . We included remixing to give users more flexibility and encourage them to tinker.

Many users in the user study noted that this was one of their favorite features because they could begin with existing scenes and modify them with their own content. Most of the time users would add themselves, swap the background, or add props.

Figure 60. Choosing a scene to edit and remixing the scene with new elements from the asset library

The scene-maker is intentionally much simpler than Photoshop to enable young children and their parents to quickly and easily create and environment together. It allows users to mock up ideas for sets and incorporate aspects of the physical and digital world. This made it sufficient for us to study the value and meaning of customization and ownership in video mediated communication environments. Our findings from the study and feedback about the environment are presented in Chapter 5.

3.4.5 WaaZam.com: Supporting a Community of Users Online

WaaZam is a made up word which is pronounced "WaaaaaaZam!" - an exclamation that refers to the moment when the other person appears on the screen with you. We wanted to use a name that was fun, unfamiliar, and easy to remember so that people would start to refer to the project as an action. As part of the project we are working on a website is that will be a hub where people can download the software, engage with a community of people creating assets (costumes, background sets and effects) and post movies created in the environment for public viewing.

The goal of the website will be to foster the community, tell a clear and simple story that helps people understand what is currently possible within their homes, and facilitate the technical steps required to setup *WaaZam* in the home. The purpose of sharing the software online is to generate enough interest and value to attract a large base of users when the software is robust enough to support them.

Figure 35. A screenshot of the current implementation of the website

The *WaaZam* application is for anyone interested in being together with friends and families in a shared space that they can design and play in. We plan to provide tutorials for teachers, families, and researchers interested in installing the software modifying it for their own purposes and submitting changes to be added to the repository in some cases. Although it is difficult to maintain software my hope is that releasing it will generate a large user base which will produce content. That would allow us to make a strong case to develop a more robust version of the software that could scale to a larger user base.

CHAPTER 4. IMPLEMENTATION

"As the unity of the modern world becomes increasingly a technological rather than a social affair, the techniques of the arts provide the most valuable means of insight into the real direction of our own collective purposes." – Marshall McLuhan

WaaZam is a built on open source software with the long term goal of releasing a cross platform application that families can use to play with each other at a distance. Our initial goal is to support a small group of users and design the system to be extensible when contact management and NAT traversal solutions are embedded and can managed from a community website.

The current physical setup and hardware are described in the first section of this chapter, with a vision of a more connected system outlined in Chapter 6. The software implementation section is divided into a set of modules and dependencies and relevant algorithms are outlined in section two. Section three describes our current data transfer and management schema which allows users to be enrolled, share scenes and initiate play sessions. The code behind *WaaZam* and the network protocol libraries will be released at *waazam.com* (the telepresence software) and *OFStreamer.com* (the network protocol for researchers and artists).

4.1 Physical Setup

The physical components of the *WaaZam* system are relatively minimal currently, however there will many different sensor options available in the future such as the Kinect 2, the Intel perceptual computing camera system, and sensors from companies like Panasonic, Asus, and Primesense. An overview of the reasons for choosing our current sensor and projections about the future layout are presented here. We built three versions of our system: a LAN configuration, a HDVC piggybacked version, a VPN based network protocol, and an UDP hole punching GStreamer implementation. The hardware components required to support these setups are described here as well.

4.1.1 Hardware configuration

Our current prototype runs on OSX on two mac minis, two Kinect cameras, and two Samsung HD TVs. It is setup in our laboratory at opposite ends of a large room due to space constraints. Users are separated by a temporary partition that divides the room into two spaces. The floor is taped to indicate for new users where in the space the camera can see and where our near and far threshold are located. An ideal setup includes at least a 55 inch TV because makes the user feel as if they are looking into a large mirror and they feel more immersed in the world, as if they were looking into a large mirror.

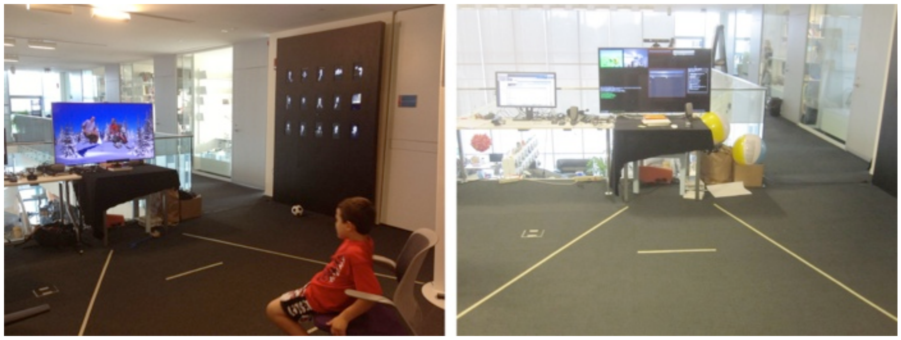

Figure 61. The development and testing area.

The play space extends back approx. 3.5 m in a cone that extends at a 60 degree angle in accordance with the field of view of the camera. In addition it is helpful if the user has 2-3 beyond this to freely move around in the physical environment.

4.1.2 Hardware specifications and projections

The hardware for the platform consists of core elements and optional elements for users of high definition video conferencing systems. The core elements are:

Kinect camera:
The Microsoft Kinect XBox 360 is a sensor that captures an RGB and depth image of 640x480 px in resolution. The capturing rate is 30 frames per second (FPS). The viewing angle is 43° in the vertical axis and 57° in the horizontal. The Kinect has a depth sensing range of approximately 80cm.-3.5m. This supports a large area of interaction. The Kinect has a motor which allows users to shift the viewing direction vertically by ±28°. Additionally, it has a four-microphone array to detect directional sound. The capabilities and affordability of this sensor have made it popular with gamers and hackers who are interested in gestural and 3D interaction.

Mac minis:
We used two mac minis with 2.7 GHz processors and 8 GB of RAM. We have run the software on 10.6, 10.7 and 10.8 on a variety of macs from the late 2008 models to the most current Macs available. The only significant difference between machines is the frame rate which varies from 11 fps to 29 fps depending on the hardware capabilities.

Figure 62. Samsung HDTV, Logitech 9100, Mac mini, Kinect camera, Panasonic HDVC, and Cisco VPN Box

HDTV:
Our displays are 55 inch Samsung LED UN55D7000 Series 7 TVs. They run at 60 Hz and are 3D capable. The TV includes features such as color enhancement, smooth motion performance and a fast refresh rate for moving assets. We found that the image quality of the screen significantly contributes to the immersive quality of the experience and recommend similar specs for others.

The optional elements are:

HDVC Box:
Panasonic gave us a KX-VC300 HD Video Conferencing unit for experimentation. The unit has the benefits of a high end system with HD video compression using the h264 codec MPEG-4 AAC LD / G.722 / G.722.1 / G.722.1 Annex C, 360° Full Duplex Audio, Analog Microphone. It handles all the compression, decompression and sensor interfaces however it is not designed well to support alternative inputs or transport alternative protocols and is not interoperable with other HDVC systems.

BlackMagic Video Transport Card:
Intensity Shuttle for USB 3.0 is a 10 bit HD/SD editing solution for USB 3.0 computers. It supports 10 bit HDMI and HDMI playback. It is capable of repeating displays from different devices realtime to other computers to appear like an input camera. We developed software to import images from the Blackmagic as a go between for HDVC systems but found it very difficult to configure due to differences in standards for display resolutions and refresh rates of camera capture and displays.

VPN Box:
We used a Cisco 9000 series VPN to create a private network that would allow us to connect inside corporate firewalls by tunneling between locations. The cost is prohibitive but the box gives you control over the VPN within your local environment.

4.1.3 WaaZam network configurations

Figure 63 illustrates the hardware required to setup *WaaZam* in any of the primary configurations: Local Area Network, HD VPN, and over a broadband network.

Figure 63. An example of the LAN setup for demo purposes at British Telecommunications in Ipswitch, England.

On our LAN utilized roughly 800 kbps to transfer data without any significant latency other than the time it took to compress and decompress the image on both clients. Although we could pass much higher quality amounts of information over the network we choose to design it to also work on wireless networks so that eventually we can port our compression to a VPN setup for home to home use.

The HD VPN used up to 6 MBPS and allowed us to transfer an HD quality image aligned with the depth image. We had to user a BlackMagic box to route the HD output from the HDVC box to the Mac and special software to simulate a 60 Hz 1080p camera as the input to the box.

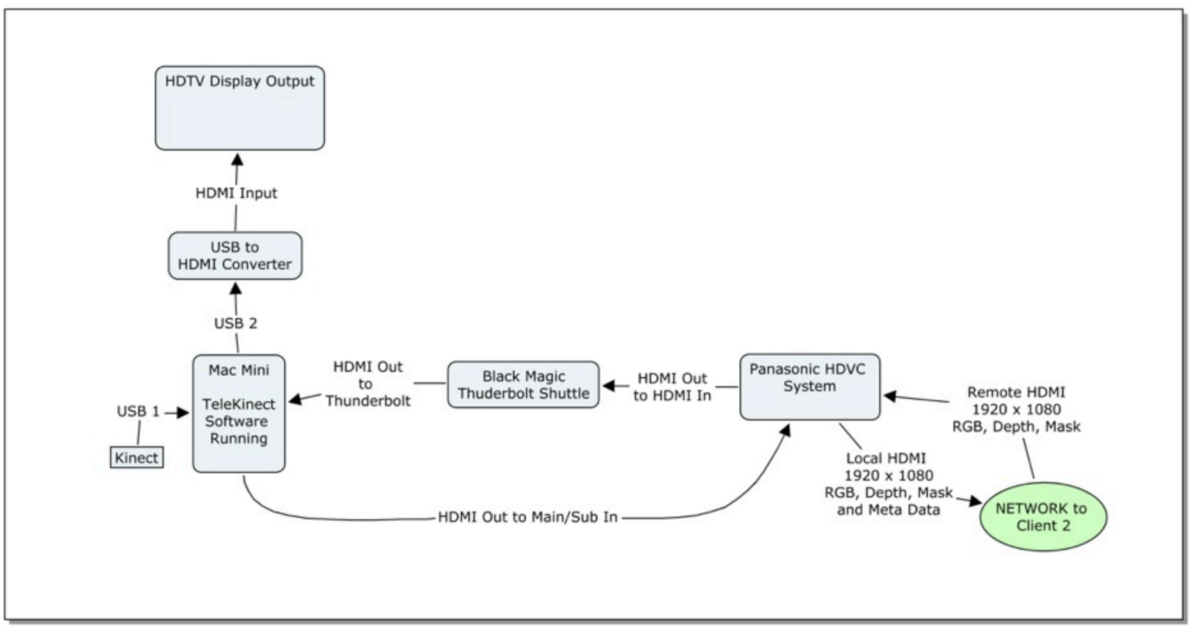

Figure 64. High Definition Video Conferencing pipeline for piggy-backing on any existing HDVC system.

This strategy could also interface with Cisco, Polycom, and Logitech systems as Paul Sermon and the TA2 project did in their research projects.


Figure 65. HDVC setup testing in the lab. The depth image is baked into the 1080p stream and routed back to the PC.

One of the primary issues with this setup was adding a second channel for the depth image and ensuring that it arrives with the current frame at each interval. To accomplish this we add the remote depth image and precomputed mask to the image and send our contour information in the YUUV channels in 3x3 grayscale squares which are resilient to compression. This approach is extremely inflexible and difficult to implement but does work if no other solutions are available.

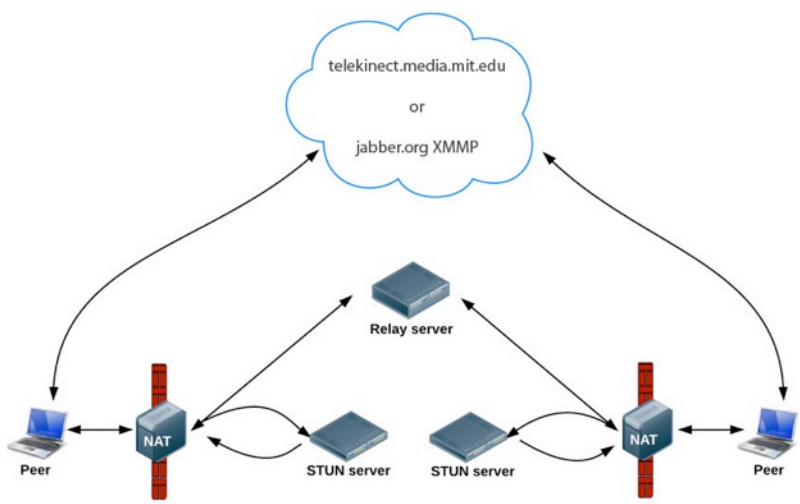

Figure 66. Extensible Network protocol for broadband utilizing ICE Protocol.

Our broadband implementation is the most complex from a software standpoint but requires the least amount of hardware. An XMPP + STUN server is required to pair the users. In 8% of cases the packets get routed via a TURN server but the remaining time we utilize a STUN pairing system facilitated by the ICE protocol standard. We have an initial prototype which demonstrates this capability but the quality is still inconsistent. Users are paired through GMail contacts over GChat based on availability and if they have accepted an invite in the past. We are pursuing this option because it is more extensible, easier to setup and requires less hardware. GStreamer solves a lot of the low level problems not addressed in previous implementations of the network protocol.

4.2 Software

4.2.1 Software development timeline

The inspiration for the *WaaZam* project began by extending experiments with computer vision to work at a distance, so people could make realtime animations with each other when separated. Using the openFrameworks toolkit we can prototype many interactive ideas in an agile fashion, but no libraries exist for developers to connect remote clients. Using the Turbo JPEG libraries and a simple RGB streaming application we added a depth image as a compressed 8 bit grayscale and began to experiment with overlapping the images.

At the time we were using NITE Kinect libraries, so we tried extending the functionality by switching to OpenNI, a set of libraries for prototyping gestural applications. In our experiments within the laboratory I got a lot of good feedback and encouragement to continue development based on the social potential of the experience.

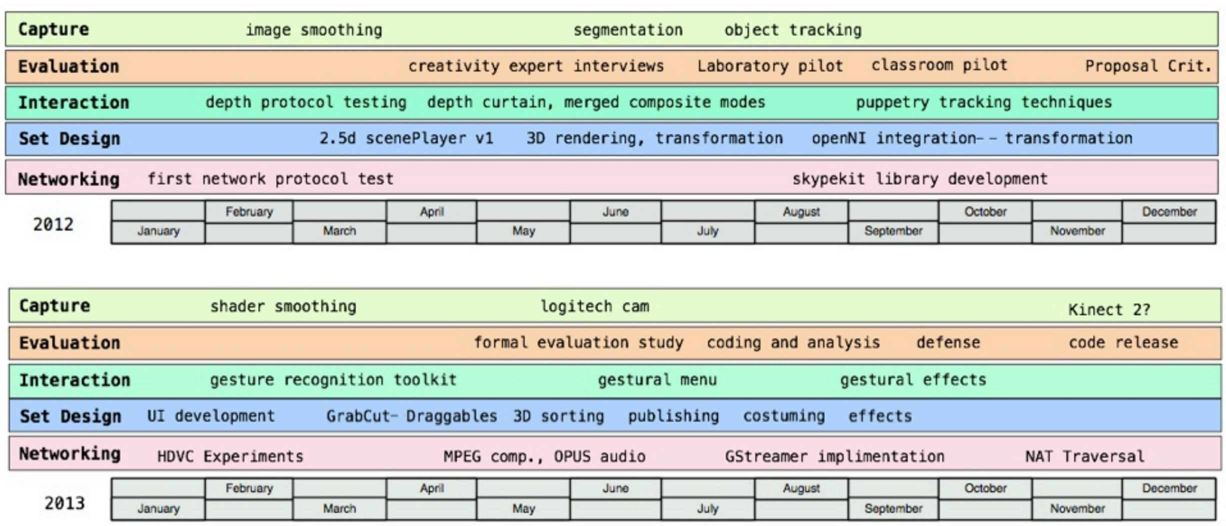

Figure 67. A timeline of the two year development of the *WaaZam* project by category.

As the project progressed I identified areas to do research and development. Kevin Wong, Jon Trevino and John Bowler joined the group for three months and contributed to interaction, segmentation, and set design sections of the timeline from May to August of 2012. We experimented with object tracking, transformation, puppet tracking, gesture recognition, HDVC systems and 3D interaction during the next few months.

Anette Von Kapri joined the project in March of 2012 and decided to focus on 3D interaction at a distance with others. She contributed libraries for tracking the user, integrating with OpenNI, rendering

COLLADA files in a 3D environment and a hand tracking module, some of which are currently used in the *WaaZam* project. At this point we split the technical development into a platform we call *OFStreamer* and the individual applications into *InReach* and *WaaZam*.

After the proposal critique and feedback from a *Computer Human Interaction* 2013 workshop entitled "Moving from Talking Heads to Shared Experiences", we identified creative play and customization and the primary area to develop next. In parallel tracks I applied for funding from Cisco to improve the networking and design an evaluation to study interactions in families. We contracted network engineers, human factors contractors, designers and developers from March to October of 2013 to make the software more robust, cross platform, and solve critical networking issues (such as time syncing, compression, and NAT Traversal) that have prevented us from releasing a version of the software for people to use publicly.

Figure 67 above outlines details regarding the different tracks we developed and many experiments that we researched but did not adopt after a few weeks of development due to CPU usage, insufficient quality, or limitations of the software libraries such as SkypeKit. In total ten developers have worked directly with the code, contributing many of the core and experimental features. I have been a primary developer on all branches of the project, serving as the graphic designer, code reviewer, and debugging features before demos and the user study.

Jason Moran and Noura Howell are currently working with me to refactor the code for open source release and Arturo Castro is developing a more robust network protocol. Documentation of the work in this thesis will be located at *waazam.com* and the network protocol will be located at *OFStreamer.com* for researchers to utilize the libraries we have written to support the research.

4.2.2 WaaZam software architecture

The *WaaZam* codebase was developed for OSX 10.6 10.7 and 10.8 and will be ported to Windows. We used a cross platform codebase called openFrameworks which has historical origins in John Maeda's Aesthetics Computation Group at MIT. It is based on open source C++ libraries that interface with components like OpenCV, audio capture, image reading and saving, OpenGL, and the GLUT windowing engine.

To organize the explanation of the software implementation I have divided the code into key modules shown in Figure 68 below. The functionality for each module is located to the left of the module name and the library dependencies are on the right. I provide an overview of the modules here and explain our approach to each module later in the document.

The user tracking module is necessary to recognize users and track them between frames, implement gestures, and render them. The scene maker module allows users to customize and quickly design new scenes. The segmentation module is responsible for blending, compositing, and cleaning up the edges of people and objects. The scene-player renders layered elements based on their depth, scale, and rotation into a single OpenGL canvas. Scene recording is accomplished with the recorder module. Gestural control and networking run on their own threads and are still under development as this document is being written.

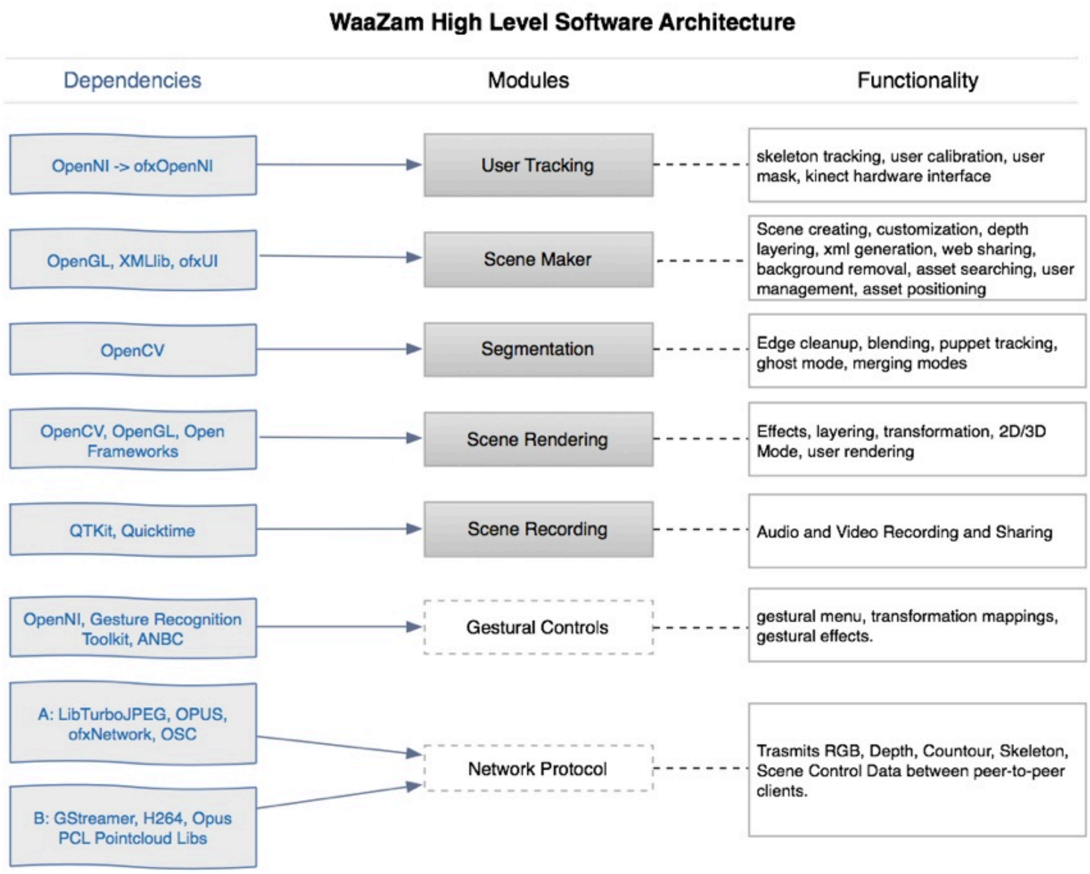

Figure 68. The software architecture of *WaaZam*. The features with dotted lines are still under development.

The application utilizes mostly open source libraries developed for linux and ported to windows and osx by members of the Open Source community interested in creative technologies. A good example is the tracking module because it relies heavily on OpenNI. The skeleton, contours, user mask and various poses can be requested from OpenNI and they use a tree-based machine learning to differentiate users from each other between frames. We developed our own classes to piggy back on Open NI and render the information in OF for debugging. We also added syncing measures for remote and local users, hand tracking, and 2D projections of the elements for both 2D and 3D environments.

Our segmentation module was developed by John Bowler. We tried a number of approaches but ended up choosing simpler methods that don't significantly slow the frame-rate below 18 fps. The methods we used are standard to OpenCV except for a custom edge cleanup operation we designed to address the wobbly edges around each person. This is due to the estimated depth geometries the of the randomized IR laser projection pattern. We looped around the contour and drew overlapping regions of interest 20 pixels wide around the boundaries of the counter. In each ROI we identify three masks: known background, known foreground, and an unknown area.

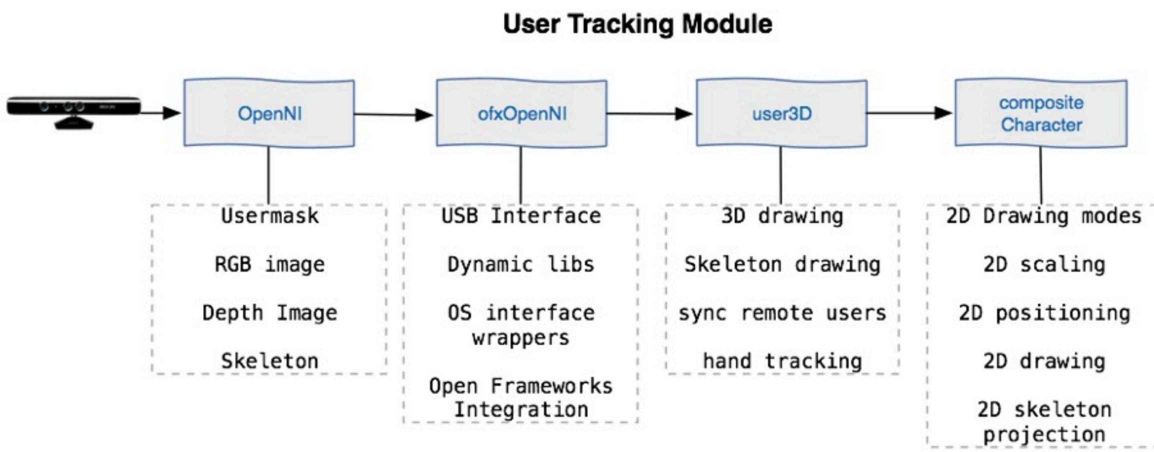

Figure 69. User tracking module for users with properties and functions in the pipeline.

For each pixel in the unknown area we use the expected maximization of the known background and foreground HSV channels to vote if pixel in our unknown mask is foreground or background. This mask is then applied to the image, blurred and thresholded.

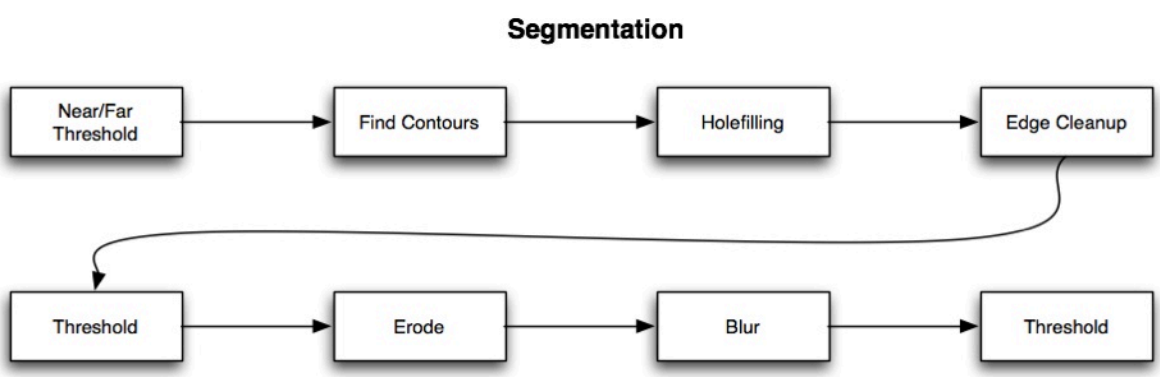

Figure 70. The segmentation pipeline for objects and people.

Movie recording ended up being significantly more difficult to implement than we expected. Cross platform libraries such as FFMpeg and GStreamer did not perform reliably producing results in which the audio was out of sync with the video. We developed a custom solution for OSX with QTKit libraries, saves a frame buffer object to a Quicktime movie. An audio file is recorded and time stamped at the same frequency as the image recording. We have good recordings for up to 4 minute videos on machines with a fast enough CPU but still experience some syncing problems on slower machines due to the cpu slowing down the application during the recording capture.

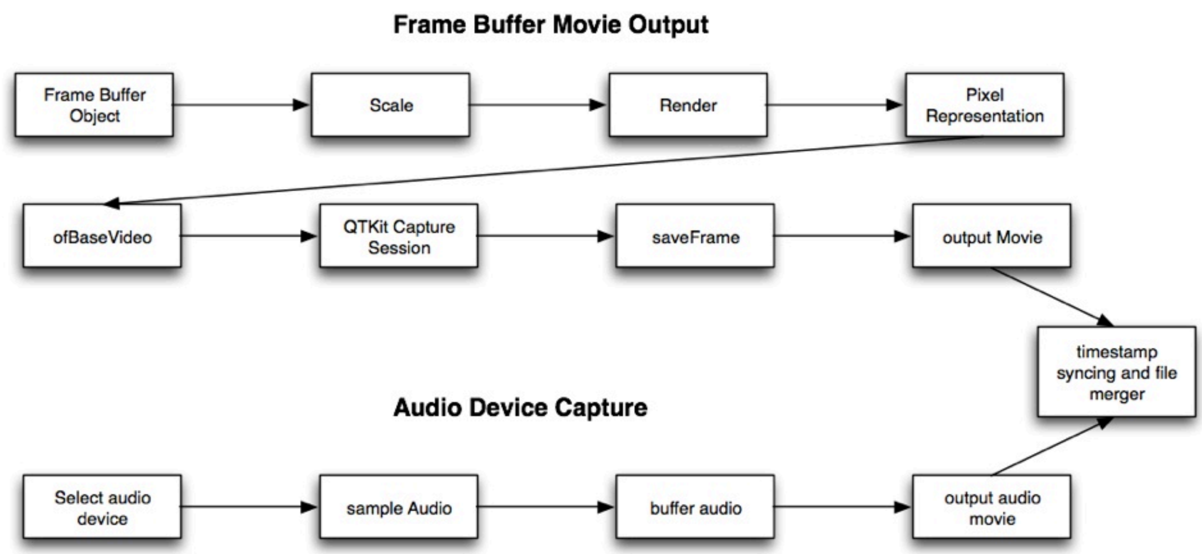

Figure 71. Frame buffer output and audio device syncing pipeline for the video recording.

This plugin has been released and is used by other developers in the OpenFrameworks community. We are still looking for a good solution to record in Windows. The videos are recorded at 720p and compressed using the H.264 codec.

4.2.3 Scene maker software architecture

The scene maker software module is the most complex of the modules due to the extensive UI development needed for asset management, navigation, search, layout, editing and publishing of scenes. The overall concept of the scene maker is fairly simple: it allows users to compose scenes from images and movies in a 2.5D environment in which they can interact together. We designed the software to be flexible, support a wide variety of filetypes and image resolutions, and allow users to make whatever they can imagine.

The software consists of a menubar, a canvas, a web search tool, an editing tool and data management classes. The menubar is managed by the contentManager class which has links to scenes and assets. The menubar relies on a GUI library called ofxUI, which is still in development. We wrote our own scrollbar and dynamic tabbing and a dialog box, but the UI design is still a bit rough compared with a professional tool like Cocoa or QT.

The canvas consists of an instance of the DraggableAssets with CornerButton objects launch editing tools that allow the user to modify the properties of assets. These objects interface with quicktime libraries and openCV 2.2 which has morphological operations that allow the user to quickly select foreground elements. We used the GrabCut implementation written in 2011 for openCV and added a custom GUI to facilitate the use of the algorithm.

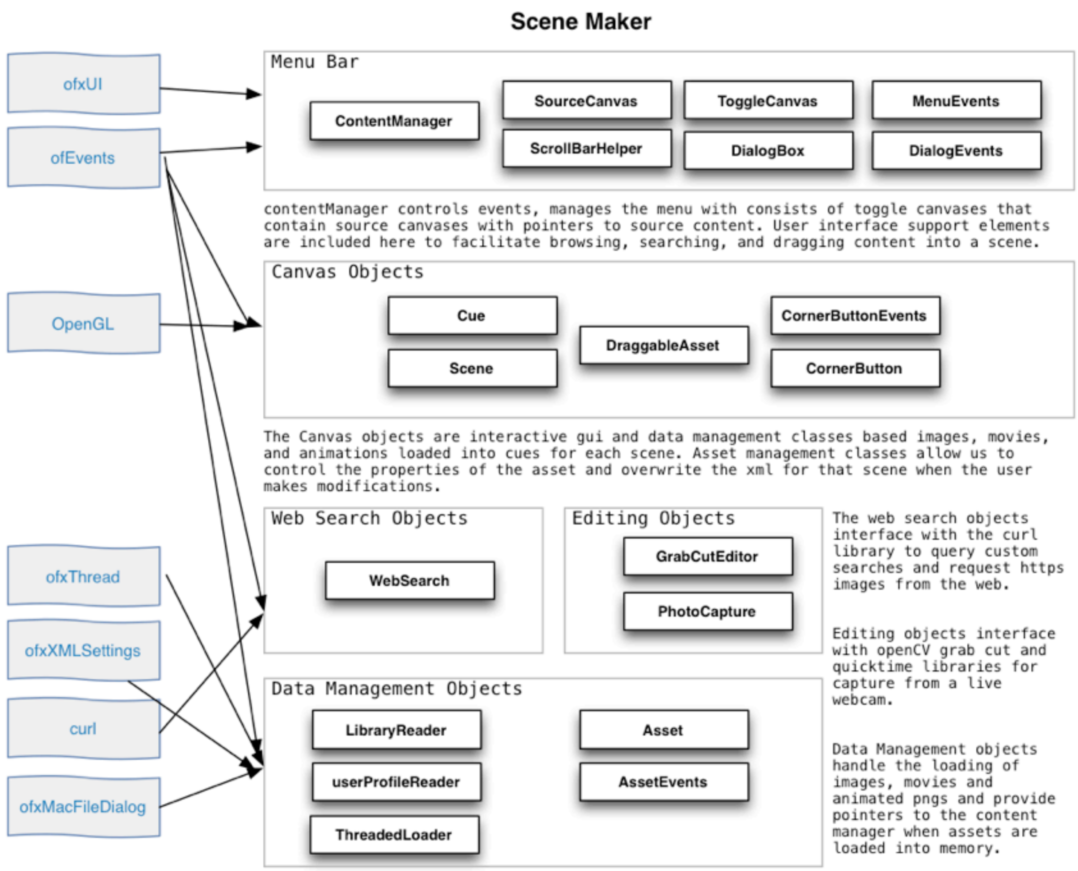

Figure 72. An overview of the classes and supporting libraries in the scene maker.

We are currently working on the transition between scene editing and interacting with customized scenes. The programs began as two separate executables but we are currently incorporating scene editing into the networked environment so people can edit together at a distance. This will require extending our protocol to allow syncing of assets and asset properties through the OSC channel and the rsync implementation currently on the server.

4.3 Data Infrastructure

The data management in *WaaZam* is handled through a variety of protocols that connect to OFStreamer.media.mit.edu, a Linux server on the MIT network. For most of our initial development we pass realtime frame data through a VPN, sync scene data on demand with an Rsync protocol and manage user profiles through a distributed file system on the client side.

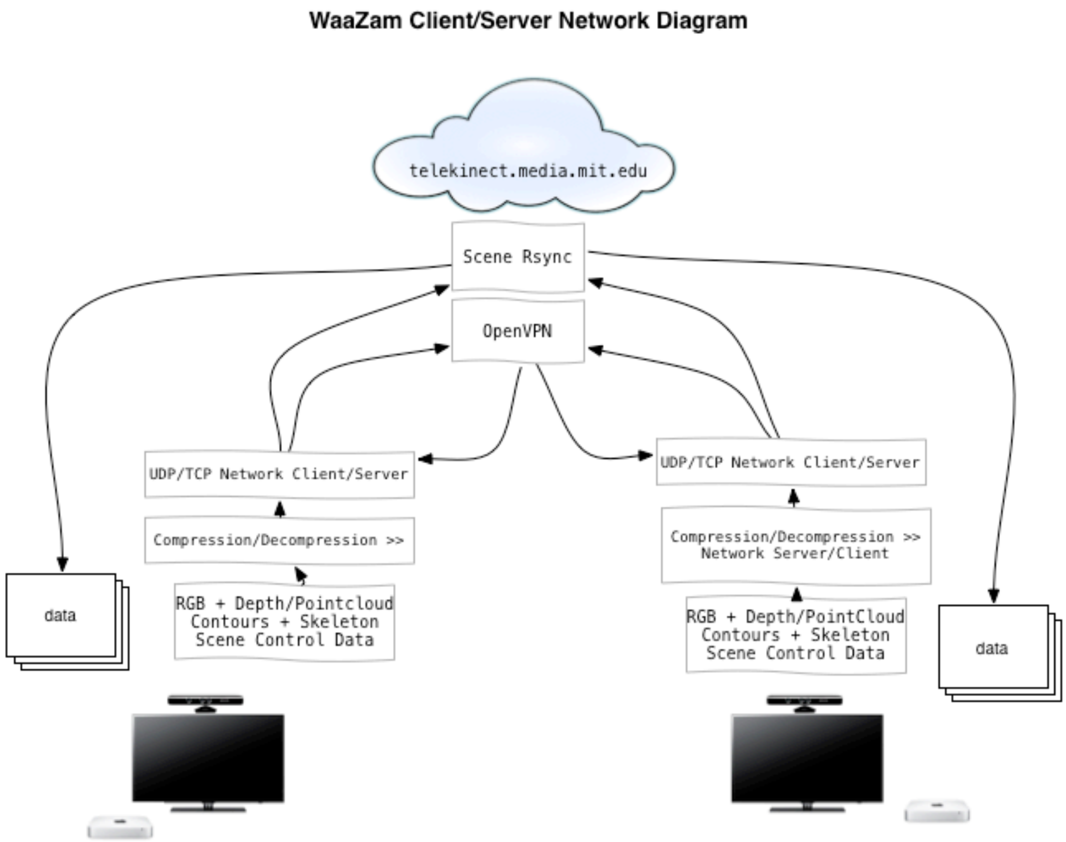

Figure 73. The data exchange between remote participants.

The details of the protocol, scene management, and user information are outlined in the following section.

4.3.1 Network protocol

One of the challenges of creating telepresence applications is streaming the data via a protocol that provides realtime compression and frame synchronization. Many careers are focused entirely on aspects of this pipeline; however, it is possible to cobble together a protocol for most applications using existing open source libraries. Our first attempt is documented in Figure 74, in which we used LibTurboJPEG to do realtime intraframe compression, in which each frame is compressed and decompressed for both the RGB image and depth image on the Kinect camera. We convert the depth image from an 11 bit image to an 8 bit image to use existing grayscale compression algorithms.

We also send floating point (x,y) coordinates of the user contours and an XML format of each of the limbs of the skeleton of users for each frame. This way these features do not need to be computed from a compressed depth and color image and appear the same on both clients. We use a third channel to send control signals when the user changes scenes, opens a menu, or moves an asset to a new position. The fourth channel is reserved for OPUS audio compression and transmission. We used circular buffers to send the required frequency and byte number for each frame to OPUS and set the encoding rate.

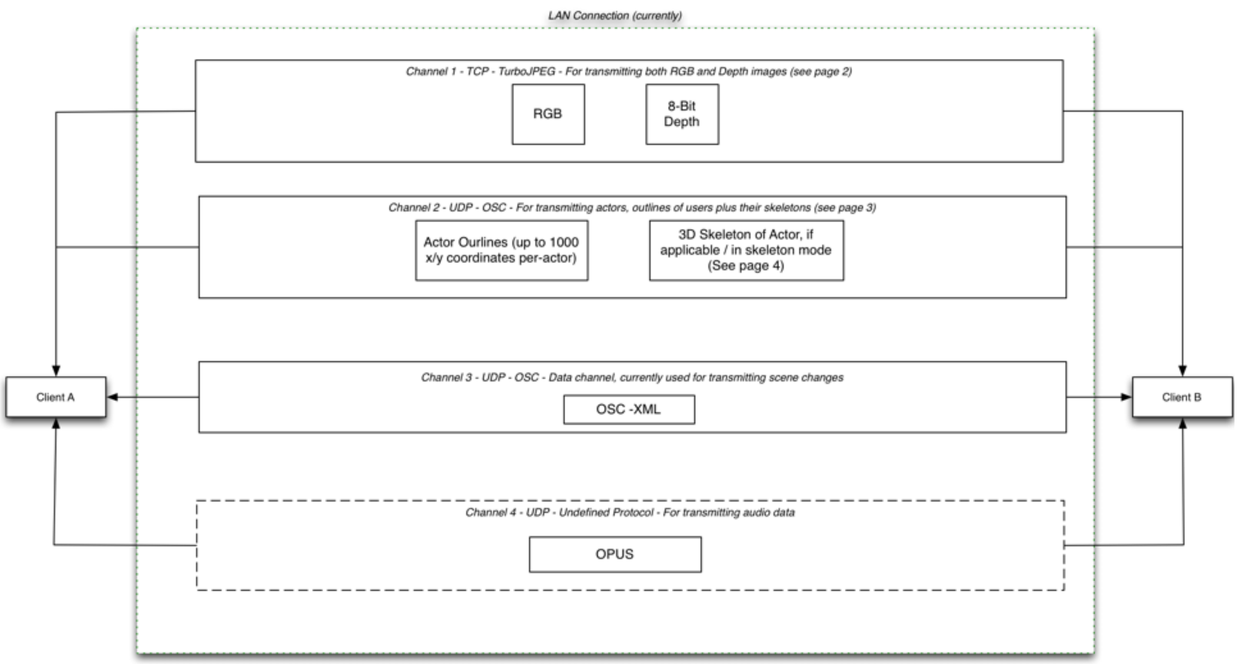

Figure 74. Our initial implementation of the OFStreamer protocol. Channel 1 contains a color and an depth image, the second channel has user data, the third channel contains scene data, and the fourth audio data.

Over LAN this protocol worked perfectly at 700 KBPS, but when we tried to implement it over a VPN the audio channel buffers overflowed and if we dropped a frame in a channel we were not able to re-sync the audio, video, and metadata channels with a time-stamping mechanism because there was no frame count and buffering system built into our implementation. The requirement of sending TCP/IP over networks outside the lab also caused the network to behave erratically when packets that were not received were requested again.

We approached researchers at British Telecommunications, who had implemented a similar protocol for previous research projects, and they recommended two possible approaches to our problem: using the auxiliary video stream descriptor in the MPEG-2 Transport Stream for depth data or using GStreamer (a Linux based multimedia framework) to create separate RTP pipelines for depth, RGB, and metadata.

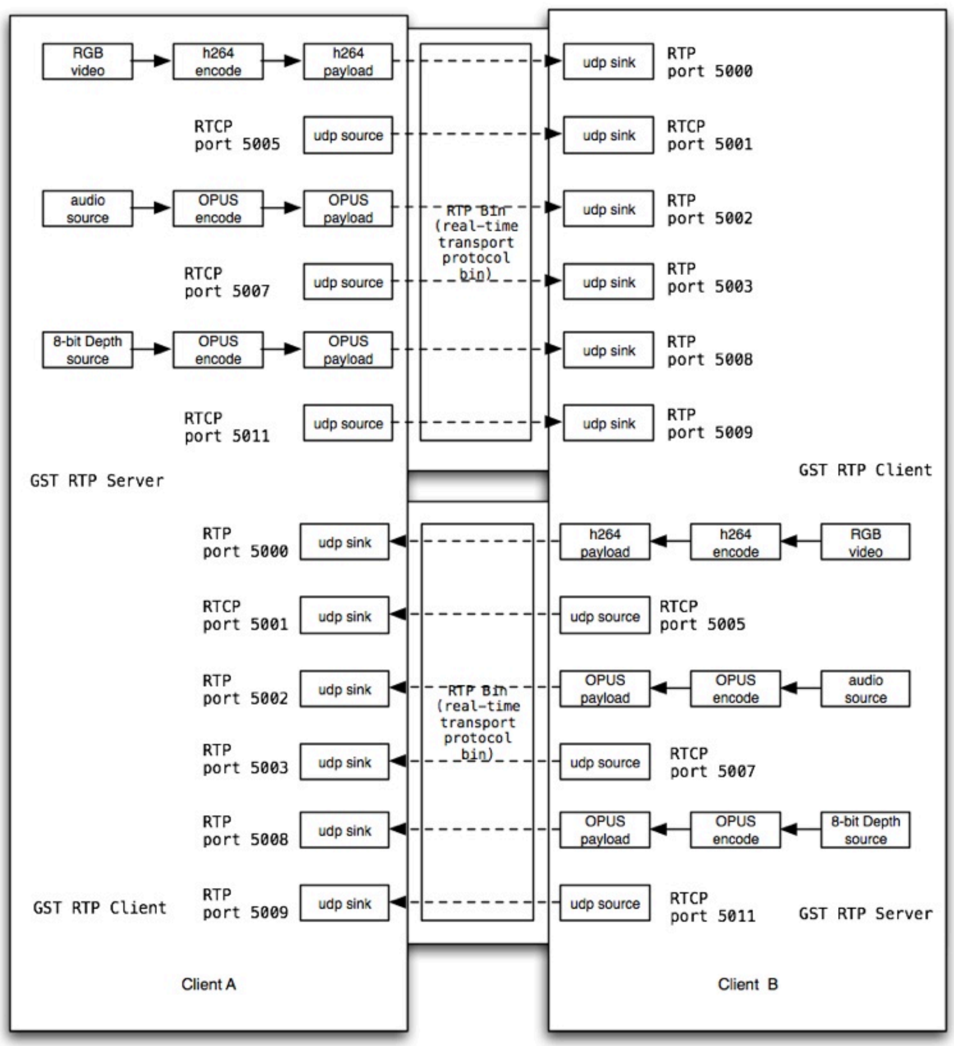

Figure 75. GStreamer breakdown of the UDP Streaming Protocol designed by Arturo Castro for OFStreamer.com

GStreamer wraps FFMpeg and other compression libraries such as OPUS and the H264 codec. The newest version is supported on OSX and Windows. Arturo Castro, an experienced network developer is assisting us in our implementation. GStreamer provides a high level management scheme for complex streaming protocols by using a sourcing and syncing pipeline system which is composed of modules.

We are currently working on using libnice to address NAT traversal issues which are the last remaining hurtle to allowing people to use a peer to peer application that does not require a central server except to initially pair users to each other.

The bitrate required to send audio and video is a third of the previous protocol's at around 300 kbps because the H.264 compression only sends differences between frames. Our initial strategy is to send information in separate channels in order to use robust compression however a lot of the data is lost when assembling a pointcloud or 3D mesh from the depth and color images. We are exploring an alternative approach which sends a point cloud using an octree and entropy compression from the Point Cloud Library point cloud transmission components. This would allow the program to specify a density and size for the points to send, providing more accurate 3D information, and send clusters of data without compressing black space in the grayscale image of the depth channel.

4.3.2 NAT Traversal Strategy

The need for NAT traversal is a legacy of IP4 addresses not having enough numbers to provide unique addresses for all devices. All routers have network address tables (NATs) which map ports and IP addresses from external servers to internal IP addresses on the LAN. A standard called ICE and libraries such as libnice go through a series of steps to identify what kind of NAT the router has and choose a routing method based on the security of the firewall.

We were inspired by Jitsi's source code and the book *The Architecture of Open Source Applications* which accounts lessons and experiences of many developers who have developed alternatives to skype. We researched NAT traversal and creating a video-bridge for multi-party support within the OFStreamer Protocol.

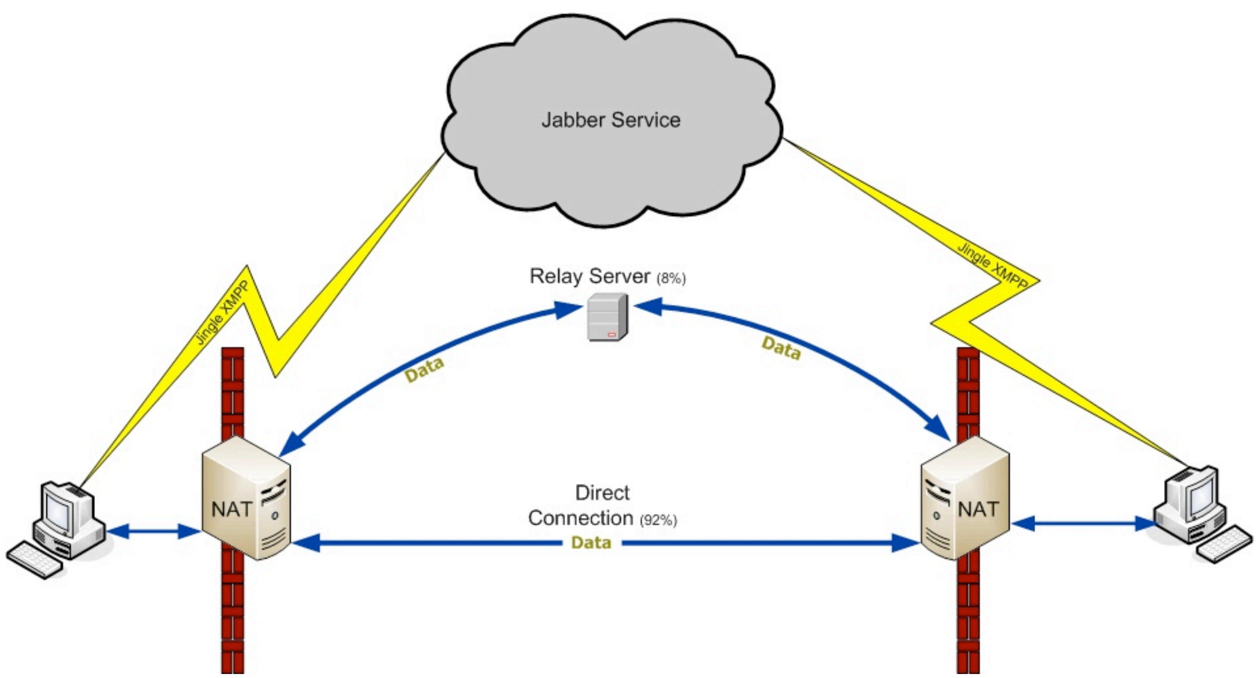

Figure 76. A diagram of Jabber P2P Candidate streaming using ICE and XMPP.

There are four different types of NATs: full cone NATs (once an internal address is mapped to external any host can send to that address on any port), address-restricted-cone NATs (internal host has to send a packet on address to receive packets back and multiple ports can be used), port-restricted cone NATs (internal host has to send to both the external port and the IP address to receive packets back from other servers), and symmetric NATs (only external hosts that directly receive packets on port address can respond). Around 8% of NATs are symmetric. These require a TURN relay server. The rest of the network traffic can be routed using a UDP hole punching method.

Our current approach is to pair users in an initial handshake with each other using an XMPP based Jingle service which uses an XML based protocol to allow users to find each other through google talk or another XMPP server. They exchange IP address over a public port. Each client will use libnice to hole punch the IPs and ports on their routers. This allows the packets to pass to the correct computer on the other side of the router. We are currently do not support TURN relaying due to bandwidth requirements.

4.3.3 OFStreamer: A Research Platform for Creative Telepresence Applications

The idea to make an open source initiative arose from a discussion at a CHI workshop entitled "Moving from Talking Heads to Shared Experiences", in which 50 attendants from all over the world showed up to share their work. Most of the researchers currently use specialized systems over a VPN to avoid the networking infrastructure necessary to deploy the application outside of a research context. We wanted to make an open source standard that would facilitate more research and collaboration at a distance between creative technologists.

To make the protocol open source we are updating our software to be more robust and extensible. We are also documenting our implementation so that other programmers can utilize the examples to develop their own creative telepresence applications. We intend to model the approach to publishing and sharing our code after the community that James George has created around the RGBDToolkit (Figure 77). They have example projects, a concise explanation of the capabilities of the platform and a detailed tutorial for how to get started, calibrate, capture content, visualize it and export it into a video format. The overall scope of their toolkit is very similar to *WaaZam* in size and complexity and shares many code libraries. We will also maintain a Google Group and a Tumblr with news and updates as the project progresses and people generate new content.

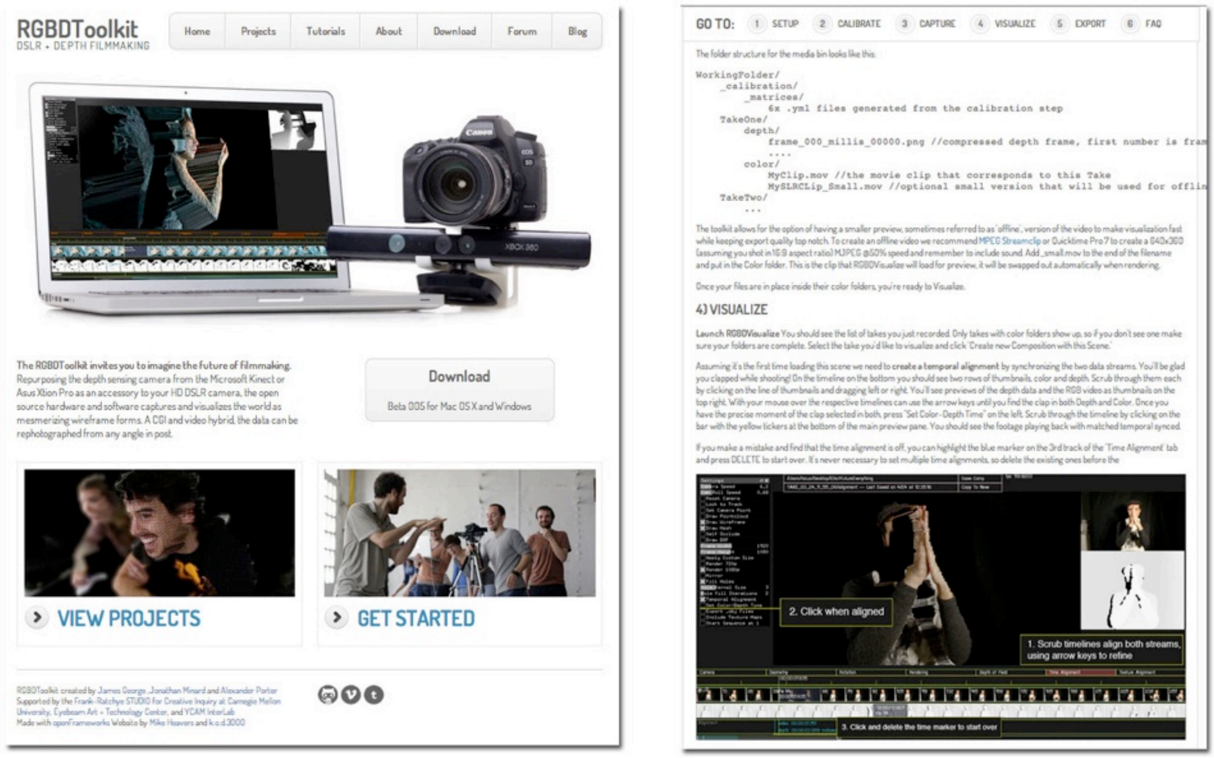

Figure 77. RGBD Toolkit has an online community with tutorials, examples, and a forum.

4.3.4 User profile and scene syncing protocol

Each user has an XML file which contains personal settings. Each user also has as a data folder which includes all the assets you have made with the scene maker. Assets are located in a separate folder from scenes and are synced via a shell script which uses rsync to send the changes in the file system each time a user clicks on the "publish" button.

The OFStreamer server stores all the synced scenes of each user. Synced scenes contain a XML file with a description of the scene, a list of assets and a list of cues with each asset id and the properties of that asset. Properties include the relative file-paths, width, height, x, y, z, transparency, and rate of change for animations.

Each scene folder also contains a file with the string corresponding to the userID of the person who created the scene. When another user wants to see the scenes that are available from the server they request an xml formatted representation of the directory with an HTTPS link to the thumbnail of each of the scenes on the Apache server. These load relatively fast because they are outputted at 320 x 180 px. when the scene is created. When the user clicks on one of the thumbnails the program uses rsync to retrieve the full scene from an identical folder on the clients machine. Once a scene has been synced the first time, only the changes to the scene are sent the next time the user requests the scene.

One of the key components of my vision for *WaaZam* is to build a community of users who make and share assets. When assets are currently published, they are stored in a common folder, but are not tagged according to theme and are not searchable. We can display up to 100 thumbnails in the menubar, after which we need to filter by user, date of creation, and popularity. This will require a database component and server integration with a search component. We are currently considering how to expand the scene-maker web components to include these capabilities.

Figure 78. Developers in the Jitsi community have implimented a cross-platform open source video-bridge.

Building a community of users also requires a contact management system. There are several existing projects such as Tox, Jitsi, and Ekiga that are open source and use these approaches. We have just implemented an XMPP approach using google addresses to make it as easy as possible for people to find each other and see if users are online and available to initiate a peer-to-peer call.

4.3.5 Incorporating multiparty support

The other challenge which we are working to address is having multiple parties in the same call, which Google Hangout and the premium version of Skype support. The most elegant open source solution we have found to date is *Jitsi Videobridge,* an XMPP server component that allows for multiuser video communication.

Unlike the expensive dedicated hardware video-bridges, *Jitsi Videobridge* does not mix the video channels into a composite video stream. It relays the received video channels to all call participants. Therefore, while it does need to run on a server with good network bandwidth, CPU horsepower is not that critical for performance.

The party that sets up a call through a video bridge controls and uses the video bridge with the COLIBRI protocol (COnferences with LIghtweight BRIdging). COLIBRI is an open XMPP extension protocol designed by the Jitsi development team. It allows the conference organizer to allocate channels for everyone add or remove participants and generally remain informed of the call state.

Currently the *Videobridge* interface is Java based and our application is written in C++. We hope to follow a similar approach as our platform matures.

CHAPTER 5. USER STUDIES AND DISCUSSION

"In every real man a child is hidden that wants to play" –Friedrich Nietzsche

This chapter presents initial pilot studies and a formal evaluation of our system. The first section describes our pilot studies with creativity experts, in our laboratory, and in classrooms in Hartford Connecticut. Based on the findings from these pilots and the recommendations of research experts from Microsoft Research, British Telecommunications Research Group, and the University of Calgary HCI Lab we designed a formal user study to evaluate core research questions that emerged during our pilots.

The second section outlines the structure of the formal study in which we examine the differences for families between separate and merged spaces, the effects of digital sets, and the value of customized sets. The final section presents the results of the formal evaluation and a discussion of the implications and key insights.

5.1 Pilot Studies and Expert Evaluations

During initial development of *WaaZam*, we demonstrated it to members of our group as well to visitors of our laboratory. Each group we demonstrated it to suggested a scenario where they could use it in their personal and business lives. We observed that people's behavior when using the system tended to be very playful, exploring the boundaries of their bodies with the other participant and the virtual environment. The inherent playful and creative spirit of people testing the application suggested that we try it with families in the Boston area.

The initial goal of the laboratory pilot study was to observe how children of different ages and their parents interacted with the software in a laboratory setting. We tested a variety of features including resizing users, fantastic backgrounds, mirror configuration, and puppet tracking.

Following the laboratory pilot I invited a puppeteer, two improv actors, and graphic artists to the lab and observed them using the system and interviewed them after play sessions. We used the feedback from these experts to identify features that support creative play in our system and to understand how to scaffold the creative process.

The final stage in our initial evaluation of the platform involved a field study in classrooms in Hartford Connecticut. In this pilot we presented the platform to 4th grade and 7th grade classrooms. Students were given backgrounds related to a historical scenario and acted out scenarios together with puppets in an improvisational manner.

5.1.1 Laboratory Pilot Study Description and Results

We conducted an initial pilot study of the system with children and adults. The study took place in our laboratory with the systems in two different rooms. We invited twelve participants from five different families. The goal of the study was to contrast interactions in person with interactions at a distance and understand the social needs of participants in the system. In addition we hoped to identify technical shortcomings and new use case scenarios. We conducted interviews of parents and teachers, monitored and rated individual sessions, and interviewed the parents and child participants.

Figure 79. Participants in the Laboratory Pilot dancing, using puppets, pretending to give a haircut, and mock fighting

For the duration of the pilot study we recorded both rooms and asked the facilitators to code the videos based on the following activity ratings and attention/engagement ratings. We were interested to see what types of behavior emerged in different environments and how that influenced the play patterns and engagement between participants.

<u>Activity Ratings:</u>
a) Narrative - do they start to tell a story with each other?
Scale: from 1=none to 7=structured storytelling
b) Pretend - how much unstructured improvisation is there?
Scale: from 1=no play to 7=all improvisation
c) Movement - how expressive are they with their bodies?
Scale: from 1=no expressiveness to 7=very expressive
d) Coordination - how do they decide what will happen next?
Scale: from 1=no coordination to 7=coordinated frequently

<u>Attention/Engagement Ratings:</u>
a) Managing Visibility - do they keep themselves and objects visible to the other person?
How? Describe

b) Managing Attention - how often do the children lose track of the other person and engage in solitary play vs. associative or cooperative play?
Scale: 1=self focused to 7=very responsive to other person
c) What did the children do during the session?

The results for the 12 participants were as follows:

Narrative Play: 3.3, STD: 1.03
Pretend Play/Improvisation: 5.5, STD: 1.04
Movement: 5.6, STD: 1.63
Coordination: 4.3, STD: 0.81
Managing Visibility: 4 .0, STD: 1.78
Managing Attention: 4.16, STD: 1.47

Some general trends can be seen in the data. Pretend play and bodily movement were high and indicated excitement and interest in the content across all the sessions. Coordination between participants, visibility and attention suffered at times due to lack of structure, and no suggestion of activities by the interface aside from the theme of the background.

Several participants wanted to make their own scenes and wanted to search for backgrounds on the web. In addition, people requested that they be able to switch scenes through a gestural interface and transform the size and position of their bodies through a gestural interface instead of with the mouse. Children who already owned an Xbox had high expectations about the interactive capabilities.

Post Study Interviews:
Following the study we asked each of the adults in the pilot study the following questions:
1) How can the system be improved?
2) Do children immediately get it, or do they need instruction? What feedback in the system might improve how kids use it? (example: prompting to calibrate, do they understand where the camera can see)
3) What ages are best suited for the system? How do children of different ages describe the experience of "being in the TV" with someone else?
4) Do children who know each other well already enjoy the system more?
5) Which types of scenes do parents and children get engaged in? Does the context change how you interact with each other?
6) If we suggest recording a movie, how can we help them structure the narrative so that they make more than just a silly scene? Should we prompt for "introduction" , "action 1", "action 2", "climax", "conclusion"?
7) What types of scenes/activities would parents like to have when separated with their kids?
8) Do you have any other observations or comments?

A separate interviewer asked each of the children the following questions after play sessions:
1) Do you think the other person sees the same thing as you?
2) What would you like to do if you could make your own scene? Where would you go?
3) Who would you like to do this with?
4) Would you want to do it again?
5) How would you describe this system to a friend?
6) Would you rather use your body or puppets with your friends. What about with a stranger?
7) How can we make it a better experience?

We clustered the comments into the following general observations. All of the participants enjoyed their experience, and wanted to try it with a close friend or relative and had ideas about other things they want to do on the platform.

Children Understand the Intersubjective Viewpoint of the Other Participant:
All children thought the other person sees the same thing on the other side. Most children wanted to do this with their best friend, or with a close relative. Some children wondered why the other person sometimes disappeared when they were outside of the range of the camera. They described the *WaaZam* as:

> *"like we were meeting in real life except we can't touch"*
> *"kind of hard, you can do some stuff, but its different to move around"*
> *"both inside the TV together"*
> *"like I was a ghost and I could haunt the other person"*
> *"like a fight game where I am the fighter but it doesn't hurt"*
> *"a place where people can play together on tv"*

Value of Object Play as a Shared Activity:
Using puppets increased engagement because it introduced a shared activity that involved object play. Children age 6-10 were most engaged in puppetry mode. Kids were more imaginative with puppets than their bodies. They would also revert to arguing or competition when there were more than two were using the system.

Target Age Ranges 7-11:
Children under the age of 7 were more shy and slow to explore. Being with older children decreased their engagement significantly. Encouragement from a parent in one on one scenarios helped increase engagement for the child. Parents seemed to think the system was best for ages 8+. The 6 and 7 year old children struggled some with the puppetry, but could also engage with some help from adults. Children

10+ suggested movie making and sharing on social networks and wanted to upload assets from their mobile devices.

Existing Relationships are Key to Engagement

We found the system was much more engaging for people who already had an existing relationship because they had an existing basis for play and shared activities. Parent child pairs were more imaginative and did more storytelling because the parent was directive and structured the activities through suggestions and prompting.

Support Shared Activities

Scenes that suggested shared activities such as dancing, fighting, acting, singing, etc. increased engagement because they prompted specific activities. Video content at the same scale as the participants on screen was particularly effective at triggering a response from both parents and children. Some parents in the follow up interview suggested "Prompting". (similar to what they do in the Kinect games) to encourage shared activities.

Activities Differed Between Genders

The girls and moms in our study wanted to pose for pictures together and dance together. They pretended to go on an adventure together and generally participated in more role playing scenarios. Boys wanted more action oriented activities such as fighting scenes, fantasy scenes, dragon scenes, and imaginary characters that respond to each other when they overlap. The boys suggested the following themes: Lord of the Rings, monsters, The Little Prince, planets, jumping off of buildings and flying. Boys also suggested sound effects: "POW" and "BOOM" when they overlap with the other person.

Parents were interested in recreating familiar spaces and activities

Two parent participants suggested uploading your photos to share the day's events with each other when separated. They wanted to be able to act out things they had done that day and invite the child to participate in the recreation of the events so that they could be a part of them.

General Improvements to Study Structure

Following this initial trail, we made several key conclusions regarding our evaluation approach. First, its very difficult to evaluate the experience and play patterns of the system when multiple children are using it due to social factors and the limited screen space at the attention of participants. Second, if we want to compare the system to more standard video configurations such as Skype or magic mirror configuration we need to introduce them in a counterbalanced order for different participants because seeing one before the other biases them. Third, making the study more formal will set expectations for the participants before play sessions which will help initiate transitions between activities and allow us to evaluate different conditions within the platform. These insights guided our design of the formal user study in Section 5.2.

5.1.2 Recommendations by creativity experts

I introduced the system to and conducted one hour interviews with several professional improv actors and a professional puppeteer. We asked about technical issues such as: camera position, suggestions for activities, interface suggestions, how to structure storytelling and animation, and limitations they would like to overcome.

Figure 80. Improv actors using the platform to give a mock presentation, a language lesson, and pretending to cook.

We observed improv actors use the system for one hour and brainstormed with the participants after the initial session. They described their experiences and gave suggestions for improving the UI and suggested scenarios for future use.

We invited the actors back two weeks later and filmed them using the system. We constructed many of the scenes they suggested and and asked them to play together as if they were on stage as if they were doing a live internet performance.

Figure 81. Puppeteer Dan Butterworth demonstrating puppet assembly and improv acting with children.

They commented that anything people currently try to do together over Skype, such as a music lesson, a language lesson, a brainstorming session, cooking, etc., would be improved by compositing the environments. A second improv expert was intrigued by the fantastic things you could do that would be impossible in real life. A third said that he would want to host a show making fun of B-rated movies.

We invited puppeteer Dan Butterworth to host a puppetry workshop for eight children ages 6-11. In the workshop children constructed a puppet during a making session and acted out scenarios with the puppeteer in our system. This helped us evaluate our puppet tracking algorithms and learn about the use of objects for play from the perspective of the children. Following this session we interviewed the puppeteer to get his feedback on the design of the system.

The feedback from the puppeteer can be summarized as follows:

-Use felt and hot glue for simple, fast puppet assembly.
-Hats and costuming work really well in the system.
-Simple bold colors work best with the current camera and resolution.
-Hand puppets with rods are best because the camera does not see the rod.
-Positioning the camera at chest level with the participants is the best arrangement for the system.
-Make a story the driving force, such as a journey together, where the participants have to struggle to win.
-Suggested adding effects when kids get near certain objects or their bodies overlap.

Many of the scenes we selected to be in the library of *WaaZam* extend from these discussions. All of the experts suggested scaffolding the experience with examples, adding new contexts of use, and identifying characteristics of the interaction that differ from traditional videoconferencing. The experts also all agreed that being able to load your own content in a simple way would increase engagement and ownership, while providing existing sets will accommodate participants who are less committed and just want to improvise together without making a set first. Many said that they would only make things with a small group of their friends and family, most of the time they would want to just explore, play, and do things together in the environment.

5.1.3 Classroom pilot study and results

The second pilot of the system took place in an intercity classroom in Hartford Connecticut with teachers and 28 students in the 4th grade and 7th grade. Teams improvised together for a live audience, and the audience voted at the end for the best performance.

Figure 82. Participants in Hartford using the system for classroom enactment and improv activities.

Due to a live audience in this study, students were more self conscious and more likely to resort to miming than narrative play. Although the audience increased initial engagement for more outgoing students, it also excluded students who did not want to embarrass themselves. Having an audience can be a motivating factor but the ability to err and try things in a safe and private environment is likely to benefit a broader user base.

The fifth grade class was much more enthusiastic and engaged than the seventh graders because they were less self conscious and inhibited socially in the presence of their peers. We observed that using the system with an adult or a facilitator helped children structure their activities. If the system were used in in a future classroom a template would be needed where users can arrange their sets into a narrative structure. Prompting and scripting would help focus groups on the content of the lessons with a project based curriculum.

5.1.4 Discussion

Sessions involving children and adults had the longest duration and depth. Children seem to understand that what they see is what the other person sees. Boys and girls desired different types of scenarios. Boys requested fight scenes and fantasy scenes, and girls requested mutual activities and role-playing scenarios. We found that children ages 6-11 were most comfortable with the system, and in interviews they expressed a desire to use it with relatives and friends outside of their immediate homes.

During sessions in the first five minutes of sessions most children focused on seeing themselves in the environment. Children then transitioned to improvisation in parallel and associative pretend play. Changing scenes would trigger new types of interactions based on the background content. Narrative play was infrequent and did not develop during our pilot studies, except in the cases of parent/child play. In the post study interviews with the children, more than half the children expressed an interest in choosing their own content, adding personal photos and bringing in digital content.

5.2 Formal User Study

Based on the observations and findings from our pilot studies we developed a software application which allows participants to quickly compose set by importing assets from files or dragging assets from the finder. Our goal in the formal study is to gain insight into differences between configurations, the value of fantastic sets and the value of customization. We are also interested in understanding the types of sets that parents and children will design and if play patterns differ between configurations.

5.2.1 Hypothesis and research questions

We have observed that when children and adults are together in one space it opens up a set of shared experiences not available in Skype. Fantasy sets also seem to increase creative play. We propose that customization tools add depth, richness, and meaning to the experience and can increase engagement between adults and children. To this end, we investigated the following hypothesis:

1) Being together in composite environments enables new activities and increases engagement between children and adults in open play situations.

2) Fantastic backgrounds in video mediated environments increase the frequency of creative play and the duration of engagement between some adults and children.

3) Providing tools that facilitate personalization of background sets increases feelings of ownership and adds to the richness and depth of engagement between adults and children in composited video environments.

These hypotheses can be extended to state our core research questions in three areas:

<u>Seperate vs. Merged Spaces:</u>
What is the difference between separate representations and composite environments for participants? Does being together in one space provide shared experiences not available in conventional VMC environments for families? If so, what are the activities that repeat across participants?

<u>Fantastic vs. Realistic Sets:</u>
Do digital sets increase creative play across participants? What types of digital sets are most effective at suggesting shared activities?

<u>The Value of Personalization:</u>

Does personalization and customization add additional depth, richness and meaningfulness to the experience? How does this impact the play and engagement for the adult/child pairs?

5.2.2 Study procedure

The following section describes the procedure of the study. We recruited 13 families from the Boston/Cambridge area and explained in the recruitment emails that the goal of our project was to make more playful and engaging experiences for families that are geographically separated. We also mentioned that we have designed a study that we thought would be a fun play-date for both adults and children.

The criteria for choosing participants was that the adults be relatives with a child between the age of 6 and 12. We looked particularly for children with family members who are sometimes away or do not live in the same home, due to travel, living arrangements, etc., and selected half our participants from divorced families in the region. We received 22 inquiries about the study and choose 12 core groups (by availability and amount of time separated from child) and 1 pilot group.

The study requested information regarding the age of the child and adult and relationship of the child to the adult. We choose one aunt, one grandmother, five fathers, and five mothers. All of the fathers chosen for the study spend more than half of their time apart from the child due to travel or family separation. Adult participants were between 35-70 and the children ranged from 6-12 years old with an equal distribution across the ages. We also requested information regarding the frequency with which they use video mediated communication. Participants' experience and usage ranged from never to a few times a week with an equal distribution of experience across participants.

The study was conducted in two 1.5 hour sessions on separate days two weeks apart. Having two sessions was recommended by other researchers who have done similar studies because it allows families to accommodate to the system, which helps account for the novelty of the experience. We also wanted to encourage families to reflect their experiences between sessions.

<u>Session 1</u>:
Welcome: *15 minutes*
Greet participants in the lobby. Introduce to our space and both play spaces so they are familiar with the setup. Inside the play spaces physical toys and props will be provided.

To participants: *Before we begin the study we would like to describe what will happen and what will be recorded to insure that you understand the procedures.*

Adults fill out a consent form to participate, be recorded and videotaped, and children filled out a separate assent form.

To participants: *Before we begin we would like to understand a bit more about how you currently play and communicate and how technology is a part of that. We have a five minute online survey for adults, and we would like to ask the child a few questions verbally while the parent is filling out the survey.*

Parents filled out survey regarding experience with technology, gaming systems, mediated communication, and the existing play patters they have with the child. Children answered a few open-ended questions about play patterns in an interview with the facilitators. Details regarding these questions are covered in the section 5.2.3 and the surveys are included in Appendix A, B.

To Participants: *We are studying how families play in video environments in order to improve their experiences. We need your help to discover what the most fun things to do in four different versions of our software.*

Participants were provided with physical props and toys in the room that might initiate object play like puppets, toys, and costumes. Participants were asked to find most fun thing to do together is in each condition. Each session condition was presented in counterbalanced order across participants.

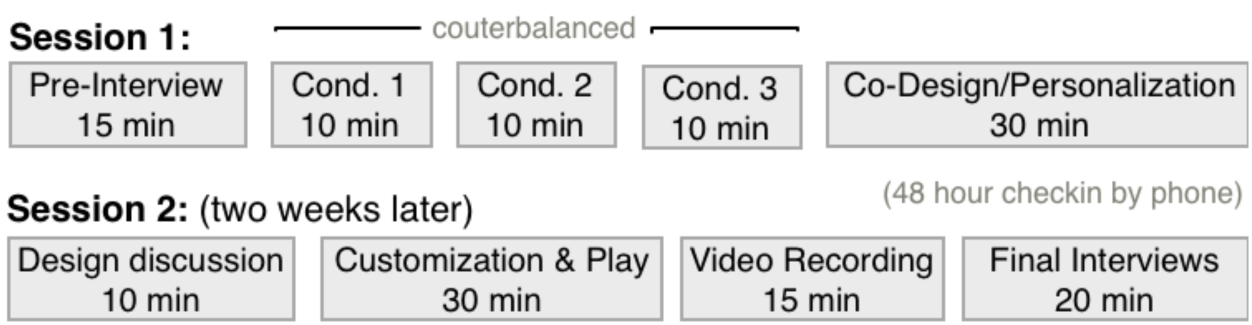

Figure 82a: The study procedure: the first session was focused on understanding the conditions and the second session was focused on design and customization.

To Particpants: *We also need your help finding new things for people to do and discovering what you would do if you could make your own worlds from home. Please design a few personalized scenes with the assistance of the facilitators.*

Co-Design Session and Play Sessions: 30 minutes:
Co design session with the parent, child, and monitor. The task is to come up with 2-3 ideas for customized backgrounds. Once an environment was made parents and children stepped into it and proceeded to play or make modifications.

To Participants: *During the next session we would like to use items from your home and life to make a scene that is especially for you. Please bring physical toys during the next session, or place digital items and photos on this thumb drive.*

We gave each user a thumb drive to save digital assets throughout the week and notified each parent that we would contact them between sessions to hear about any reflections they might have had with the child in the car or at home after the study.

In a mid-point checkin we called participants and emailed them if we were not able to reach them by phone. We asked each participant if they talked about anything on the car ride home, if there had been any subsequent references to their experience in the family, if they had discussed what they would like to do in the environment, and if they had any new ideas about activities themselves. The results from these mid-point checkins are covered in the qualitative discussion in section 5.3.

Session 2:
By the second session users had a chance to see the capabilities of the environment and set maker and gather assets from their personal lives. In this session they designed and played in sets they customized and recorded a short video to take home with them.

To Participants: *We would like to use images and items you brought to make a series of personalized scenes and then record a short video that you can take home with you. We would like for you to imagine if you had this in each others homes the kinds of things you would like to do together.*

30 min: Follow up Co-Design Session
We had short discussions with participants about their ideas in order to establish a baseline and a motivating idea for a co-design session with the facilitators. Using digital assets from the web, or from the users thumb drive we placed assets in customized scenes.

10 min: Record a Short Video
We asked participants to record a short improvised video that they could take home with them. Customization tools were available throughout the session. The facilitator tried to support open ended free play up to the participants.

15 min: Post Study Interviews
We wrapped up the session by interviewing adults and children. We used laddering techniques to try to understand as much as we could from the child about their experiences and ideas. We also interviewed parents, asking exploratory questions regarding how to make it more compelling, the value of

customization, building engagement with the child, and if they would be interested in long term installation.

5.2.3 Coding criteria for play, mutuality, social engagement, and behavior

During play sessions video was recorded of both participants and coded every 15 seconds for the following metrics. We used an inter-rater reliability overlap of two sessions between coders.

Type of Play:
We examine types of activities episodically, coding the video for types of activities and identifying transitions between activities. The National Institute for Play identifies 7 patterns that constitute the elements of play: attunement, body play, object play, social play, pretend play, narrative play, and transformative-integrative play. Types of play are labeled in 15 second intervals during each session condition.

Parent/Child Mutuality:
We examine mutuality using the Piaget/Parten Scale [Rubin et al. 1978], which rates levels of play engagement as solitary, parallel, and associative/cooperative. The work of Parten and Howes, who observed that social play between children is characterized by 5 stages of mutual regard and reciprocity will be used to rate the engagement between the parent and childe. These are: (1) solitary play, (2) parallel play with occasional regard, (3) parallel play with mutual regard, (4) associative play with mutual regard, (5) cooperative and coordinated play on a common activity .

Social Engagement between Adult-Child:
The Adult-Child Rating Scales developed by Crawley and Spiker (1983) were used to code videotaped observations of parent-child play interaction. This scale is divided into two sections: child ratings for play maturity, social initiative, and object initiative, and guidence ratings of directiveness, elaborativeness, and sensitivity. Each characteristic is assessed on a 5-point scale (1 = low, 5 = high). Directiveness is defined as the degree to which the parent attempts to guide the child's behavior. Elaborativeness refers to the extent to which the adult follows and extends the child's self-initiated behaviors. Sensitivity is defined as the adult's awareness of the child's cues and signals and whether the adult's behavior is in accordance with that of the child [Crawley and Spiker, 1983].

Behavioral Characteristics of Participants:
The movement, gestures and speech of the participants were evaluated in the different conditions to understand attention and engagement. The following criteria were coded for and described:

What are people paying attention to in the environment? What is being observed/related to?

How ofter were there non verbal exclamations, laughing, smiling, or other indications of engagement?
Who initiates transitions between episodes?
What activities are common across groups in each condition? How do activities vary across conditions?

The procedure for coding involved a two pass approach. First we went through the videos and identified episodes and labeled episodes with ratings according to the coding criteria for mutuality, types of play, and parent child engagement. We then followed through with a second pass on the video, coding every 15 seconds for level of engagement and attention of the participant. A detailed explanation of the coding criteria is located in Appendix E.

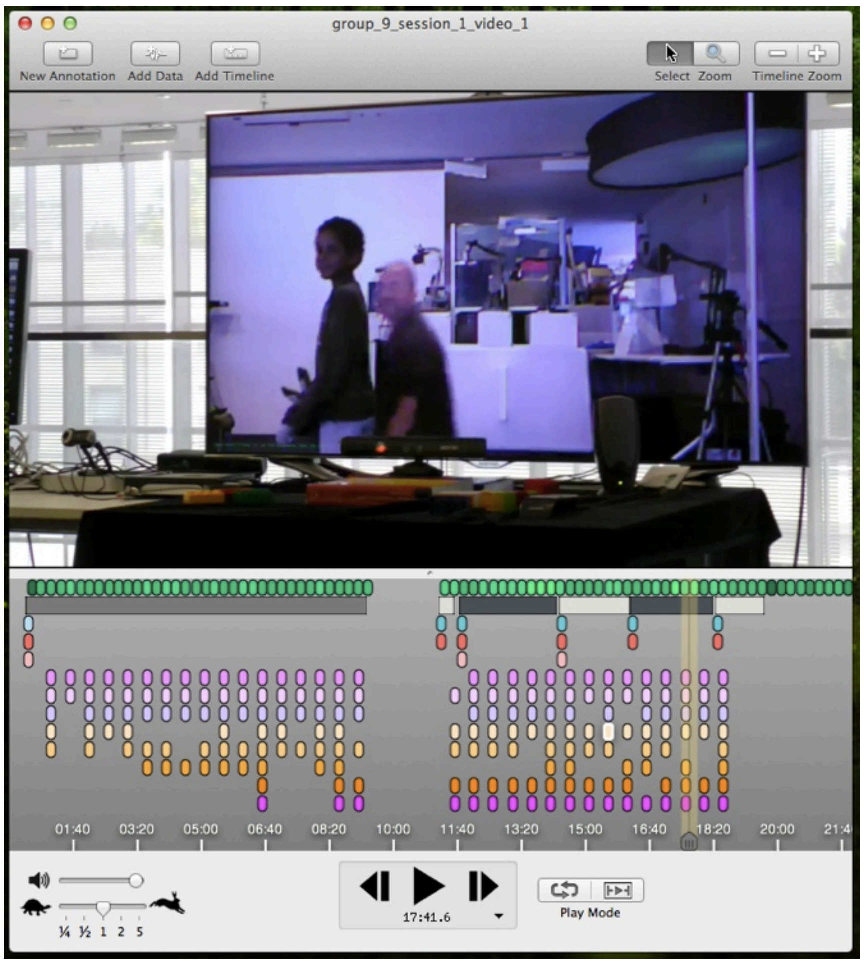

Figure 83. An example video after being coded for engagement, attention, episodes, play types, signals, and mutuality.

We used ChronoViz to code the video because the software allowed us to view multiple views of each session and exported annotations to Excel for further processing and examination. The results and analysis are discussed in Section 5.3.

5.2.4 Post Study Interviews

We asked the adult participants questions 1-5 in Appendix C in the first session and 6-12 in the second session. The goal of the interviews was to understand their general observations, experiences in each condition and the usefulness and relevance of the platform to their current lives.

We asked a similar set of questions to the children in our study but modified the questions for their age and experience level. Children were interviewed using a laddering techniques to encourage them to elaborate. The questions are provided in Appendix D.

5.3 Summary and Discussion of Results

Out of the 23 participant groups who expressed interest we choose 12 groups based on the child's age and the potential need for the system as assessed by their frequency of separation. We wanted to obtain feedback from participants who have a real need to connect with the child and play at a distance due to their separation based on travel, divorce, or work related geographical separation.

Group	Age Adult	Age Child	Relationship	In Same Home	Days Away per month	Use VMC	Use Phone
1	43	10	Mother	no	8	once in a blue moon	more than once a week
2	42	8	Mother	yes	2	once in a blue moon	once in a blue moon
3	36	10	Mother	yes	0	never	never
4	70	8	GrandParent	no	20	once in a blue moon	a few times a month
5	43	10	Mother	yes	0	never	more than once a week
6	38	6	Father	no	10	a few times a month	a few times a month
7	38	8	Father	no	15	a few times a month	Every day
8	43	7	Mother	yes	1	never	once in a blue moon
9	47	12	Father	yes	1	never	Every day
10	48	12	Mother	yes	2	once in a blue moon	Every day
11	49	11	Father	no	15	a few times a month	Every day
12	45	8	Father	yes	18	never	more than once a week

Figure 84. Information about the children and adult pairs participating in the study. Use VMC and use phone columns refer to using these technologies to contact the child when the parent is not in the same home.

Half of our participants were familiar with Skype, Google Hangout, and FaceTime and used one of these media on an occasional basis. The others had never used any of these technologies. We did not see any noticeable differences in play patterns between groups based on familiarity with the technology. We had two groups who were unusually tech savvy and had more interface related suggestions and more meta comments about the overall design of the software. Our participants came from different socio-economic and educational backgrounds. The participants that responded to the study did so because they were interested a system that might let them play at a distance with the child and to have a new activity to do with the child.

In the pre-interview of participants we surveyed them about their play activities together, favorite games and activities, familiarity with video gaming environments and VMC technology, and feelings of connectedness to each other when communication over various mediated technologies at a distance. We used these characterizations of the activities they described to understand how likely participants were to engage in social play, active play, and creative play and see if there were correlations between their existing play patterns and their behavior across conditions.

5.3.1 Observations during the study procedure:

After introducing ourselves to the participants and interviewing them, we showed each participant the area where they would be playing and explained where the camera could see them and that we would switch between conditions at intervals of 10 minutes. Half the participants started with an introduction to the digital condition with scenes made by artists, proceeded to mirror configuration where each participant could visit a representation of the other participants space. We call the traditional VMC configuration "Skype" because most users are familiar with this configuration of interaction. The other half proceeded from Skype to mirror to digital configurations.

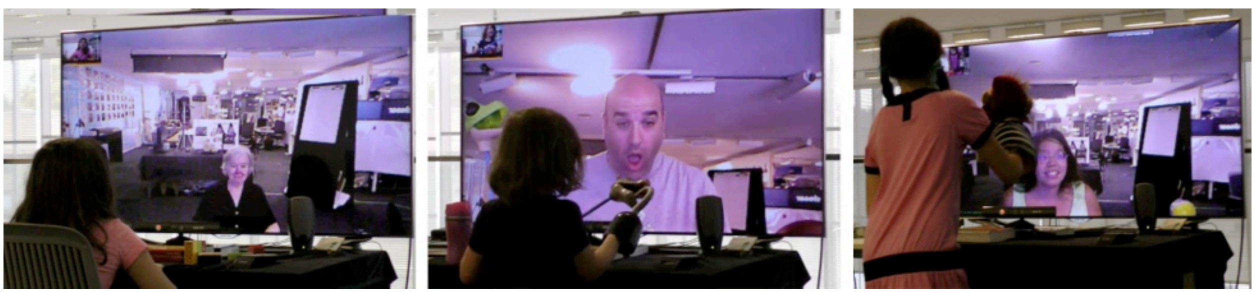

Figure 85. Users interacting in Skype configuration talking, show and tell, and puppetry.

We provided a selection of engaging toys in each space which included puppets, balls, electronic lights, costume props, and LEGOs. Participants choose activities which included charades, puppet show and tell, conversations, talking about the day, playing with individual toys and showing the parent, the child playing with the camera, and individual ball play. In general show and tell and toy play were the most common activities. Participants with children that had high object initiative and social initiative and adults who were directive and sensitive had high engagement in the Skype condition, but some pairs struggled to engage. We had several children with attention difficulties who engaged in more solitary play in the Skype condition than other configurations.

The Skype configuration requires more creativity on the part of parents and children to think of engaging activities. Pairs that found satisfying common activities were exceedingly creative and engaged in the conditions that followed. There were considerably fewer smiles, laughs, exclamations, and gestures in Skype than other conditions. Maintaining an intersubjective representation of the other persons environment requires considerable cognitive load for young children ages 6-8 but seemed easy for the 12 year olds in our study.

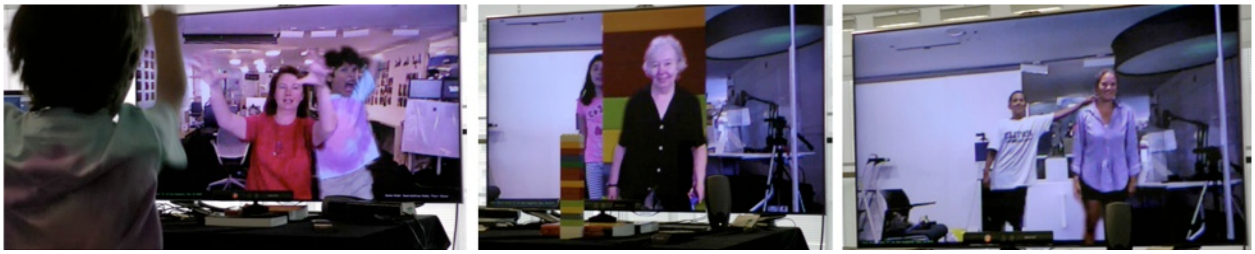

Figure 86. Examples of activities in the mirror configuration: play fighting, using LEGOs to block part of the camera and hide, reaching out to touch each other when first introduced to the composite environment.

When we switched to mirror configuration most participants exclaimed in surprise seeing themselves in the same space. This was usually followed by a playful experimentation of the boundaries of their bodies. Only two groups referred to the elements in the physical room in a direct manner. The most common activities in mirror configuration were play fighting, playing with body boundaries, puppetry, pretending to move around each others spaces, playing with size differences, pretending to eat the other participant and exploring the boundaries of the camera view by hiding and jumping up in surprise.

Although the delight of being together in the same space could be attributed to novelty, families pretending to tickle and play fight with each other engenders familiar feelings of intimacy. Although we are limited by the number of groups in our study, we have seen this behavior with many other people who demo the system. In the mirror configuration, because there are no environmental distractions, the focus is on the participants being in the same space. Many parents told us they would start in the mirror configuration to initially connect with the child and then transitioning to a shared activity in the digital configuration after discussing it with the child.

Figure 87. Examples in the digital configuration: playing hide an see in a forest, pretending to be in separate apartments, and playing with a toy and a puppet with a creepy cemetery background.

When we switched to the digital configuration we spent about 2 minutes verbally introducing participants to the features in the environment from a verbal standpoint. We showed them scenes with layers that they could move in front of and behind, puppetry mode, ghost mode, and transformation capabilities. Participants requested scene changes, mode switches and transformation verbally from the facilitators. In the custom configuration we gave these controls to the parents and observed that the control and coordination increased engagement.

Participants had very distinct preferences about which scenes they liked which usually correlated with the interest of the child. We previewed several scenes for 2-5 seconds each until one of the participants confirmed that they wanted to try that scene. The most popular scenes had many layers and supported hide and seek, or were open ended but imaginative, such as the space scene. Scenes with more constraints, such as the apartment scene, provided clear scenarios for pretend play within a particular imaginary context.

Common activities in the digital configuration included exploring layers, hide and seek, shrinking to the scale of the scene content in perspective, pretending to inhabit the virtual spaces, puppetry, tag, exercising, body play, play acting, and appearing and disappearing from view.

When first entering the digital condition participants generally accommodated to new features during the first few minutes and would settle into an activity together. We tried not to interrupt episodes of play, only responding to requests for changes by participants so that in the 10 minutes that followed transformation only occurred at the participants explicit verbal requests.

5.3.2 Activities and Highlights:

We noted the activities in each play episode and ranked them by the number of occurrences to see what the most common activities were across the groups in all conditions. Eight of the top ten activities are only possible in merged composite modes and the top five activities involve pretend play, which indicates that the environment supports initial forms of creative play.

Overall parents reported the highlights across sessions were: social body play (8) (pretend hugging, tickling, playing with body boundaries, body scaling), scenes that suggested activities (4) (hide and seek, dancing), exploring an environment with the child (3) (circus scene, apartment scene etc), playing with puppets together (3), and being able to create sets that were either familiar or related to the interest of the child (3).

Children reported their favorite activities as customization (7 participants) because it enabled activities specific to the content of the scenes they made, because they like making things, because it was their images and objects, because it was more fun to make their own, and because I could make things that annoyed my mom. They liked being in the environment (4) because it looked like we were in the stars, because you felt like you were there, because it was fun to be there, we could play there. They enjoyed doing things in scenes they made (4) like jumping off a bridge, playing in the park, hiding behind cows, pretending to snowboard with my dad, transformation(3) like being a ghost, flying, shrinking, hide and seek (2), being with the parent (2), and playing with glitches (1) and puppets (1).

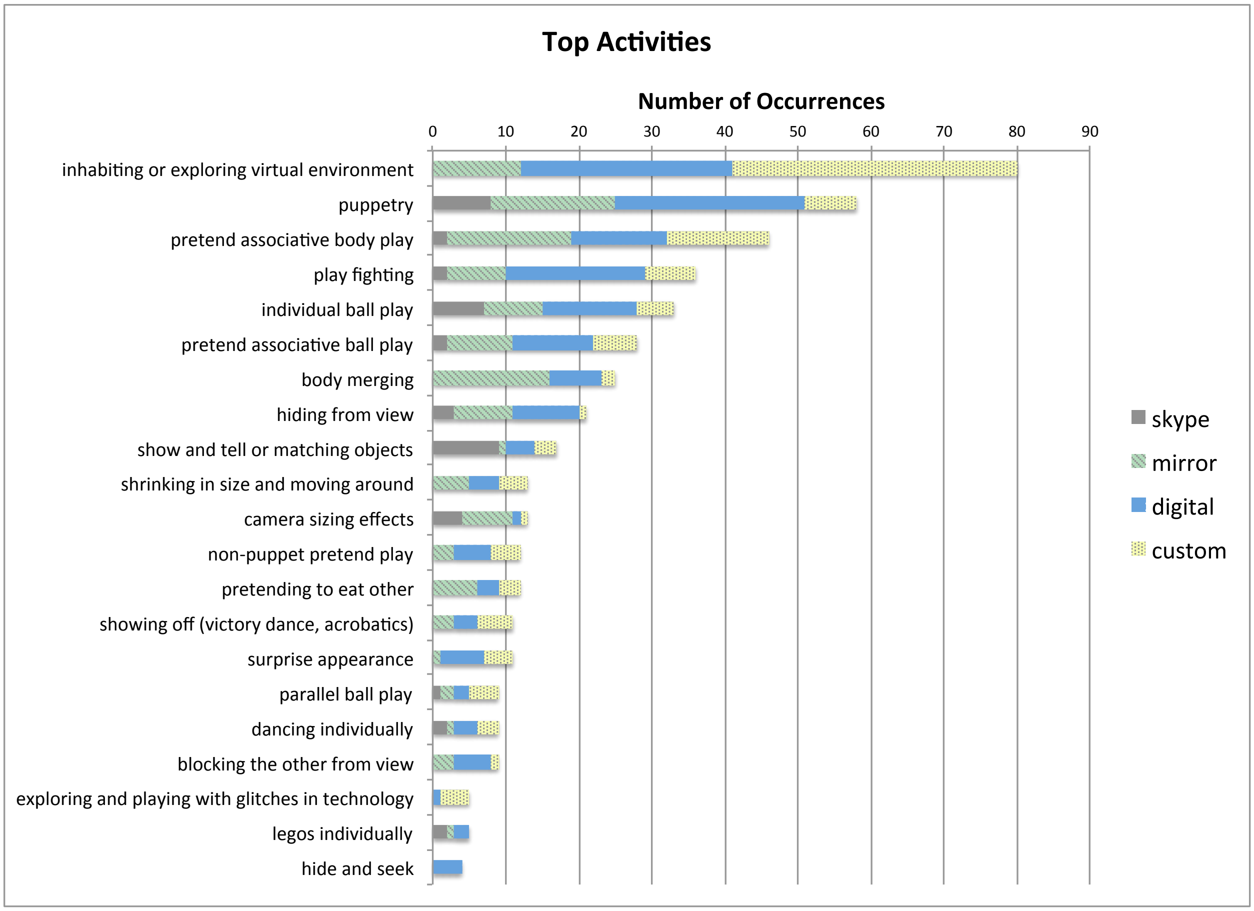

Figure 91. The most common activities across conditions of activities in the study.

Overall, parents seemed to emphasize being with the child, having shared activities to do with the child, and supporting the child's interests through customization. Children were interested in doing things that related to their personal preferences, ideas, and fantasies. They enjoyed feeling like they were immersed in an environment and seeing themselves transformed in surprising ways. They enjoyed shared activities feeling together with the parent.

5.3.3 Differences in Skype Condition between Counterbalanced Groups

Half the participants in the first session went from Skype to Mirror to Digital (SMD) conditions and the other half went in the opposite order (DMS). There were no significant differences in engagement between groups in each condition. DMS groups were rated as not playing 40% of the time, a 30% increase from SMD groups. DMS groups also had shorter sessions in the Skype configuration and some groups requested that we return to the other configruations so they could be together again. There was also a decrease in pretend and social play for groups in the DMS categories.

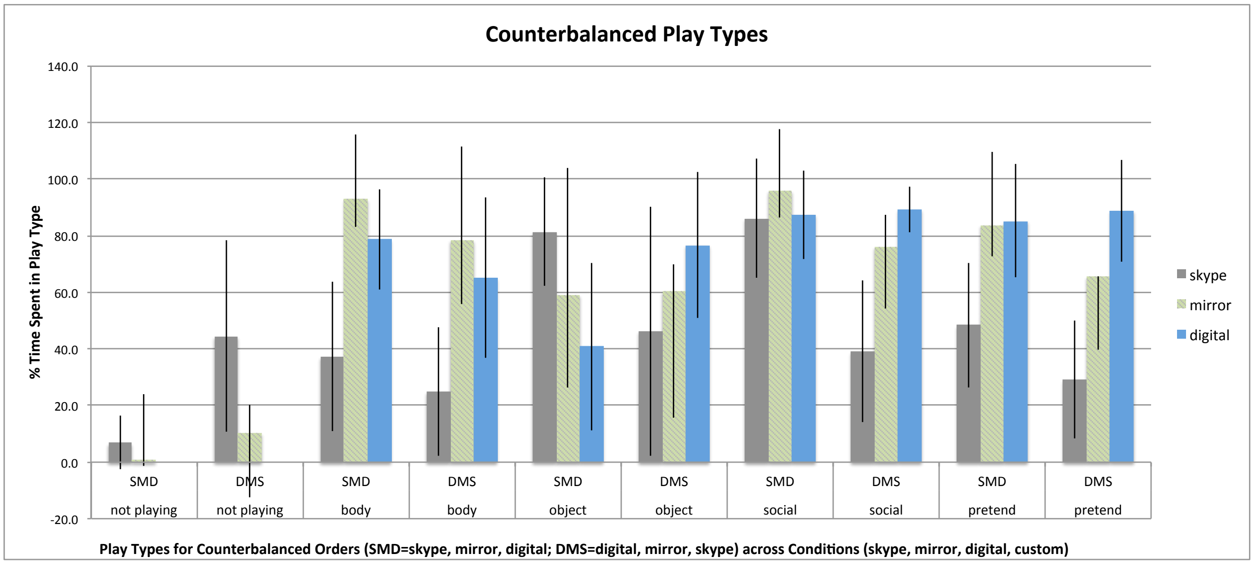

Figure 92. Counterbalanced comparison of play-types across conditions.

We attribute this behavior to the reactions of children who wanted to continue to play with the features that were available in mirror and digital configurations, especially the possibilities that existed for body play with the adult. Many of the participants were also familiar with the Skype configuration and had already established existing modes of behavior such as conversational engagement and show and tell activities.

5.3.4 Mutual Engagement, Creative Play and the Playground Effect:

Adults rated their engagement in the mirror configuration, digital configuration, and custom configuration significantly higher than Skype configuration with only slight shifts between the mirror, digital and custom configurations for both engagement and connectedness. Children rated the Skype configuration an average of 3, the mirror configuration as 3.5 and the digital as 4.5 out of 5. They generally tended to prefer digital configurations over the more static configurations (figure 93).

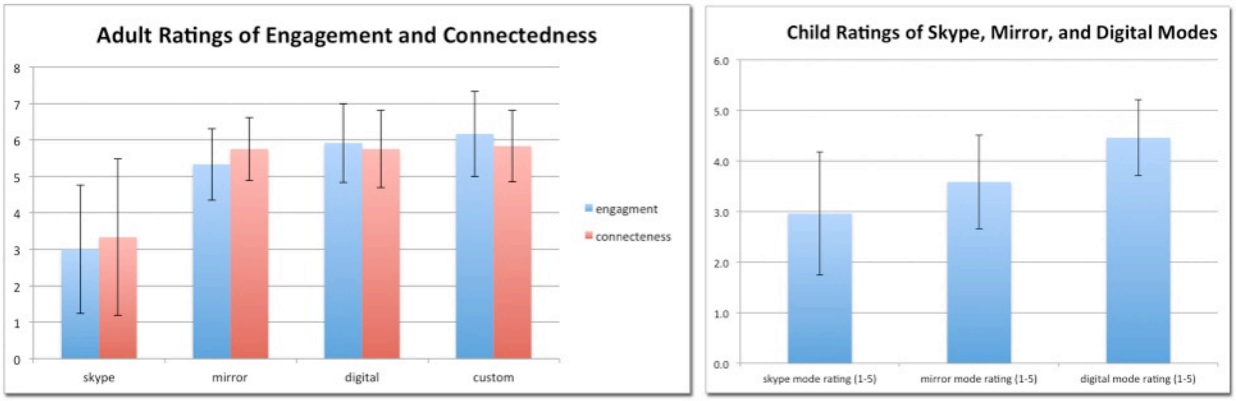

Figure 93. Adult ratings of how engaged and connected they felt with the child, and child ratings of enjoyment.

We measured if the attention of the adults and children was on themselves, the other participant, the mutual activity, the virtual environment or the physical environment at 15 second intervals across all conditions. We noticed a 10-15% shift towards the child in the digital and custom configurations. In custom configurations many of the scenes were designed by groups to reflect the interests of the child. We call this effect the "playground effect" because when observing the videos the children spend more time exploring the environment than in other conditions.

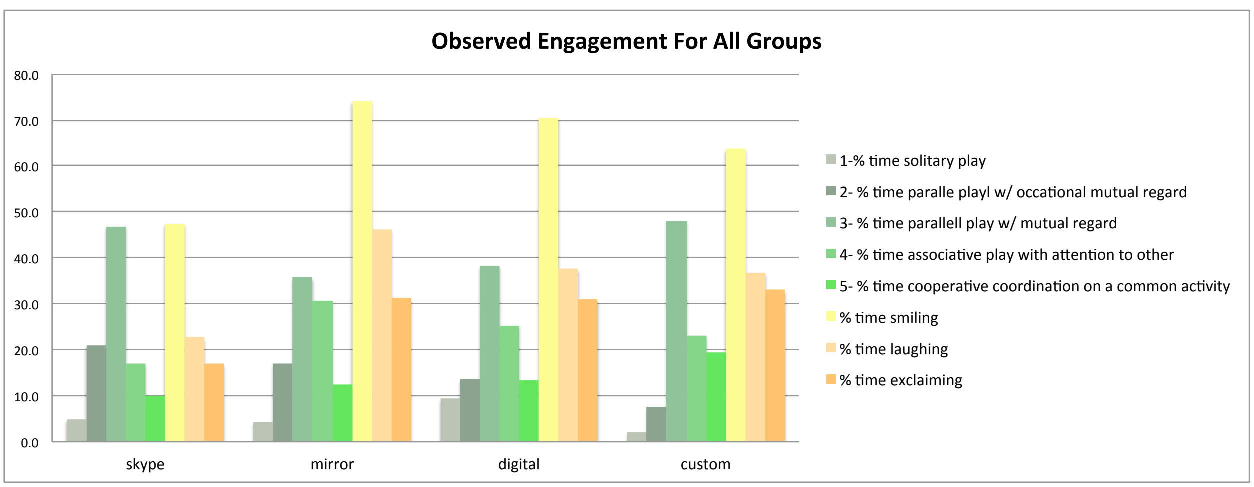

Figure 94. Engagement as measured by Parten's engagement scale 1-5 every 15 seconds and % of time smiling, laughing, and exclaiming in each configuration.

Groups had more exclamations and mutual engagement at level 5 in custom configuration, but spent more time smiling and laughing in mirror configuration. In the post session interviews adult reported that mirror configuration was the most viscerally fun because the focus was on body play more than in other conditions.

This tendency is confirmed in frequency of play types. Active play as indicated by the activity type, play type, and gesturing, was higher in all of the merged conditions with a slight tendency towards more body play in the mirror condition. The types of activities people choose in mirror, digital, and custom configurations were all very active, involving more body movement throughout the physical spaces for both adults and children than in the Skype configuration.

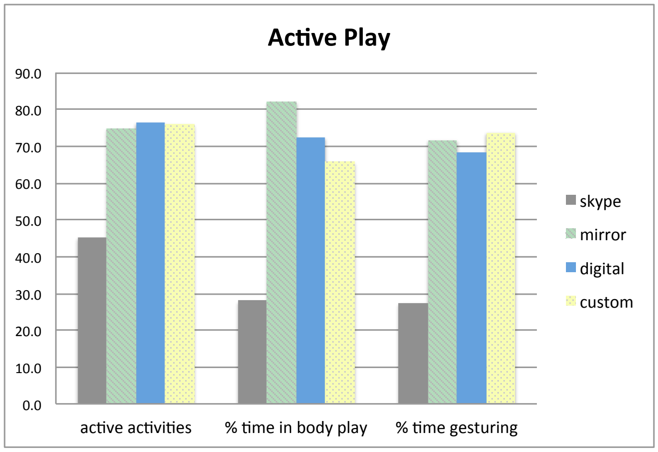

Figure 95. Active activity ratings, the percentage of time in body play and the percentage of time gesturing across conditions.

The percentage of time in creative play activities and engaging in pretend play activities was also significantly higher in the merged configurations, peaking in the custom configuration where participants enacted scenarios related to scenes they designed with an activity in mind. Imagining oneself inhabiting in a virtual space requires some pretend play on the part of both participants and characterizes the nature of interactions in the *WaaZam* environment, which is designed to support pretend play at a distance.

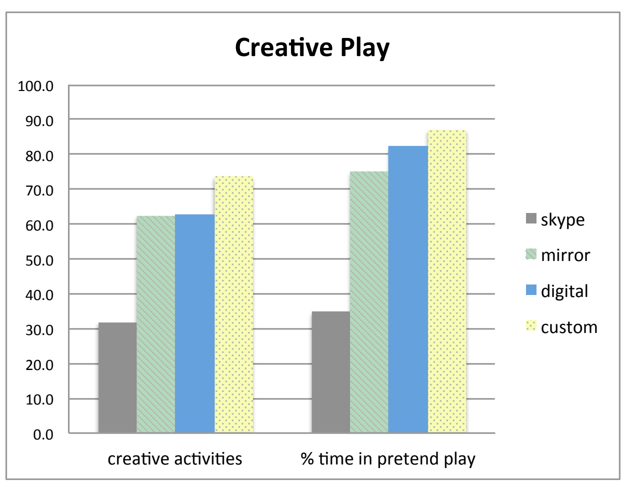

Figure 96. Creative activity ratings and the percentage of time in pretend play.

Social activities, social play frequencies and mutual engagement ratings of 4 or higher provide a more comprehensive view of social engagement between conditions. There was a significant difference between skype and merged configurations but no significant differences between merged configurations that we observed.

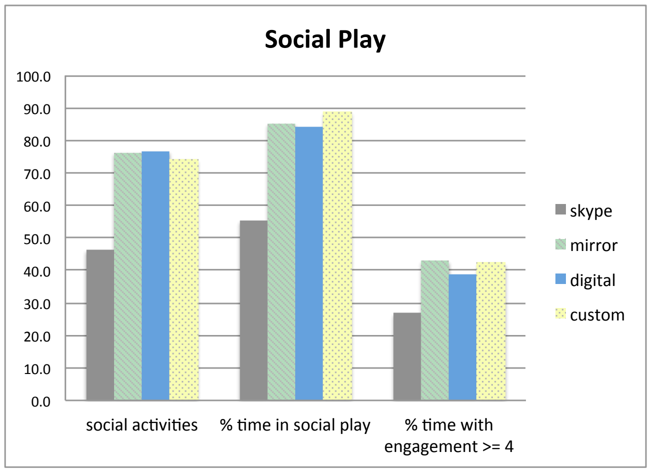

Figure 97. Social activities, percentage of time in social play, and the percentage of time with high engagement.

We noticed that if a group was particularly social their play behavior across conditions was fairly consistent and only varied if new features such as puppetry support or compositing modes were available that allowed them to interact in ways that were not possible in other conditions.

Examining the groups with the highest increase in engagement from digital to customization correlated with the self reports of the most creative activities prior to the study. Groups that engage in pretend play in the digital and mirror configurations were more likely to have higher engagement in the customization configuration. One way of characterizing this tendency would be to say that the features in *WaaZam* provide opportunities for groups to express behaviors that extend their existing play patterns. Customization amplifies existing behavioral tendencies by supporting more creative possibilities.

5.3.5 Customization Feedback

Most of the parents indicated that they would be more inclined to either support the child in making scenes or make scenes together with the child. Rather than making scenes on their own, they would look for example scenes to inspire the child or to find new ideas for things to do together. Most parents also indicated that they would want to be able to import assets from physical and digital assets. The majority said they would be more likely to choose digital assets because there is a larger and more expansive selection.

Customization allowed the software to adapt the diverse needs of our groups which were of different age groups, socio-economic backgrounds, and had different creative interests. Sometimes the environment provided unexpected benefits we did not forecast. One grandparent enjoyed being able to pretend to be more physically active with her grandchild:

"Its been fun for us to play together because I can get on her level and understand her world better. One of the benefits of the environment is that you can do things that are not possible in reality. For example, I am not a very active person but I can run up the walls with her in WaaZam and pretend to go swimming with her in the lake."

Although she was not really more active she felt as if, by participating in the virtual world together they had actually had a shared experience that they could talk about. Another mother from a family with the most detailed and elaborate narrative ideas saw customization as a way of enacting her son's ideas:

"What was interesting for us is that we can act out his ideas and then I can help him develop them into something that we can do together in his worlds. He has lots of ideas that he draws on paper but this would enable us to imagine what it might be like to actually do the things he imagines."

One child said that customization was fun because of it's flexibility:

"You could do anything – you could put a horse on the north pole. Its a lot of fun and pretty easy to try things"

Most of the children found customization fairly easy and rewarding because of all the options available by searching on the web and because they could look for specific content related to their interests:

"It was pretty easy actually, pulling pictures off the internet gives you a lot of choices. I like to find things from my favorite video games and movies (Jurrasic Park) and erase certain things from the pictures".

When we asked children if they preferred making their own sets, using others' or both; half said they liked making them but also didn't mind if other people made scenes that they used, and the other half preferred to only make their own. For example on child said:

"Making my own is good because I can make exactly what I want, like a lake and choosing the angle of a photo to make it look realistic when we are in the background and playing together. I like being able to choose stuff from the internet because there are more choices and I could make a prettier lake".

Remixing seemed to provide examples of what is possible in the environment and some kids were happy to play in other people's environments without modifying them, but others had a distinct preference to have ownership over their sets:

"I like modifying other peoples sets but since mine are better it I could recreate my own using their backgrounds".

We asked children what they thought of customization and we got many interesting responses. Most of them related to doing impossible things, doing things related to their interests, having ownership, and participating with the adult in some capacity. Regarding the creative process one child participant said:

"Its fun to think about doing thing like shaking the president's hand, things you can't do in real life. It cool coming up with ideas and putting them together."

When children and parents were collaborating to make a scene the content was usually oriented towards the child's interest but sometimes the adult was more directive during the creation process. For example one child said:

"Being in a scene is more fun than creating it. I like playing in it. Creativity is a compromise for me because I have to share the place with my grandma and think about what she likes."

5.3.6 How families reported they would use the system at home:

All of the adult participants said that if they were away from home for an extended period of time they would want to use our system as an alternative to Skype, but for different purposes depending on the family. Many parents said they would like to start in mirror configuration so they could see the child at home or show them their space and then transition to digital configuration or a customized scene when they want to do something together. They thought it might be useful to talk about practical things such as homework or what they were eating for dinner before transitioning to imaginary activities.

A few parents said they would like to show the child things they had done that day and invite them to visit the day's events with them. For example, one child really loves trains and planes so the parent said they would take pictures of vehicles while traveling and then pretend to go there with the child. One father was interested in having familiar spaces that he and his daughter had been to previously and pretending to be in them together, such as a bedtime story scene and eating together in the kitchen. Families who were more creatively inclined suggested sharing LEGOs or building scenes together at a distance and then play acting in the scenes. Some parents said they would "cheer the child on" if he or she had a creative idea. Some parents also mentioned that they liked the ability to switch scenes because they could use it to retain the interest of the child when they were no longer engaged.

Many of the children mentioned they would want to use the environment to play with other relatives that do not live near them. Children said that they would use the environment to do imaginary things like being in a cockpit together, going to a playground, playing video games together, playing with their toys, filming what they are doing, making skits, shrinking each other, designing scenes together, and eating together. Most of the activities they suggested were unstructured collaborative activities that involved both digital and physical props.

We also asked the children what they would do if everyone had the system and many of them mentioned playing with their friends in different contexts. One girl said:

"I would play with my friend L who I don't get to see very often. We could go to London together. I would use mirror configuration and digital configuration to see her and we could play board games, draw together, pretend we are roommates, or surfing or making sand castles with the scene maker and we could pretend we are housemates."

Most of the kids mentioned activities that they would do with their friends such as make video games, do movie stunts, hang out together, make videos of themselves, and act out imaginary scenes.

5.3.7 Groups with real need

One third of the participants demonstrated real need for the system because they are separated from the child more than two days a week and said they were interested in a home installation if it were available. All of these families were parents in divorced families.

One girl said that when her father was away she would want to do a mixture of real and imaginary things with him to feel closer to him:

"I would show him my cat snuggling with me, have a picnic with my dad, put a photo of me in for when my brother plays with my dad, have him tell me a story at bedtime, play with toys, pretend to be a toy kitty on a leash so I don't get lost, and make different backgrounds to show him where I've been."

WaaZam maybe serve best as a means of extending a voice call after the father and child have connected initially on an emotional level over the phone. One child's father calls him every night before dinner and said that he would not want to use the system every day even though it is fun. He said that he felt closer to his dad when they talk on the phone because he can tell him whether he had a good day or a bad day.

We tried to find quantitative evidence that the real need groups were more engaged across conditions but because our group sample size is so small, it difficult to claim anything conclusive from the data. Groups appear to be roughly the same across conditions except that the real need groups tended to engage in roughly 10% more play for all major play types.

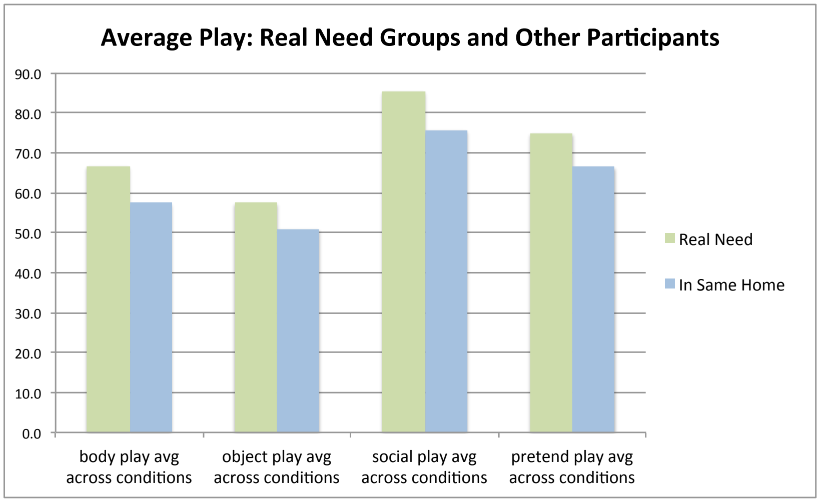

Figure 98. Figure showing slightly higher engagement for real need groups across play types and conditions.

We examined groups from separated families in order to get feedback about the potential impact of the system on families with real need. Although the groups were only with us for around four hours, many spoke frankly about the struggle to maintain the normalcy of family and how important playing is to families in order to maintain relationships.

5.3.8 Requested Features

We asked the parents if there were any things they wanted to do but were unable to in the environment. Parents requested features such as inviting a third person, such as a grandmother, controlling and moving props in the scene while playing for greater interactivity, integrating content from social networks, sound effects, adding a remote control for the parent to have advanced settings, providing more structured activities and physically engaging activities, and adding social games for the child and adult to play together.

Increasing interactivity, adding more dynamic media, cooperative activities, and scaffolding for storytelling would make the environment more engaging and assist parents in suggesting a set of shared activities during sessions. We discuss the features we are currently implementing in the future work section in Chapter 7.

5.3.9 Key Insights

Our study is limited by the number of groups and the highly subjective nature of assessing creative play in open ended VMC environments. Although we cannot make any strong claims which have repeated in multiple studies, we collected extensive information about each group we worked with, spent four hours with each group and spent additional hours coding each groups sessions in two passes for attributes related to mutuality, attention, activities, engagement, and play patterns.

By counterbalancing conditions with different feature sets we were able to identify general trends which correlate with the results of the OneSpace project at the University of Calgary. They have also observed an i**ncrease in body play and pretend play in merged configurations over separated configurations**. We also saw a **significant increase in mutual engagement and many more laughs, smiles and exclamations in the mirror configuration**. The qualitative feedback from both adults and children correlates with our behavioral observations.

Our most convincing results consist of the videos recorded by participants in the second session. The diversity of sets that participants were able to make and the increased amount of pretend play and feelings of ownership, limitless possibilities, and the described feelings of immersion provide evidence that composited VMC environments can be more personally meaningful to participants. This component may contribute to long term engagement by families.

Regarding our three initial hypotheses we had some expected and some surprising results. **Being together in composited environments enables new activities, increases mutual engagement, and increases the diversity of play types**. We cannot confirm that adding digital backgrounds increases the duration of the engagement between parents and children but our results indicate that parents and children would be more likely to play for a greater duration in environments that **suggest an activity, support interactivity, and provide customization tools**.

Digital configurations are like going to the playground and customization is like designing your own playground. Customization can increase creative play and potential peak moments for adult/child pairs. In our post-survey interviews we found that the children overwhelmingly enjoyed having control over content and valued the personally meaningful content more than sets designed by others. **Customization appears to foster feelings of ownership, increase the richness of the scene interactions and potential depth of narrative and pretend play activities.** Most of our parents took on more of a supportive role and were less mutually engaged in digital and customization configuration. Although there was a slight decrease in mutual engagement we attribute the shift to the engaging aspects of the virtual environment. The tradeoff in engagement fosters the creativity of the child and provides motivation to continue to **develop features that allow children to enact and embody their ideas and share them with their parents at a distance**.

We argue that our system supports patterns that are natural in many families. For example, in the evening the father or mother may come home and greet the child in a physical way with a hug or kiss, play for a little while and then find things to do in a common space, some of which are in parallel and some cooperative. **Creative activities seem to result from a safe setting that supports unstructured play where the parent can observe or join the child depending on the context**. We observed many of the same unstructured play patterns in *WaaZam*.

In summary, *WaaZam* has evolved to have a wide variety of features and supports diverse play styles. **By providing support for active, social and creative play, the environment is engaging for many different types of families**. Pretend play is very common and natural in merged configurations and is more personally meaningful to participants when they have customized the content.

In the future we plan to support more active play through shared activities that require rich media such as designed activities and media integration for dancing and exercise. We are interested in scaffolding narrative play by providing interactive features and a story arc that would benefit groups disposed to play acting, skits, and storytelling.

CHAPTER 6. OTHER APPLICATION DOMAINS

"Individuals are moving beyond conversation-only video communication and repurposing video to engage in shared activities." –Jed Brubaker, Gina Venolia, John Tang, 2012

During the course of this research we have had many inquiries about other domains where 3D sensors and innovative software will enable new types of shared experiences at a distance. This chapter is intended to be a launching point for researchers interested in developing in other domains and want consider new and unexplored possibilities. I begin by presenting unexplored tools and techniques and follow with a discussion of application domains that we have discussed but not studied in detail.

6.1 Tools and Techniques for New Shared Experiences at a Distance

Being together extends beyond creative play to many domains. The driving motivation for this research is to support shared experiences at a distance in new and innovative ways no one has explored to date.

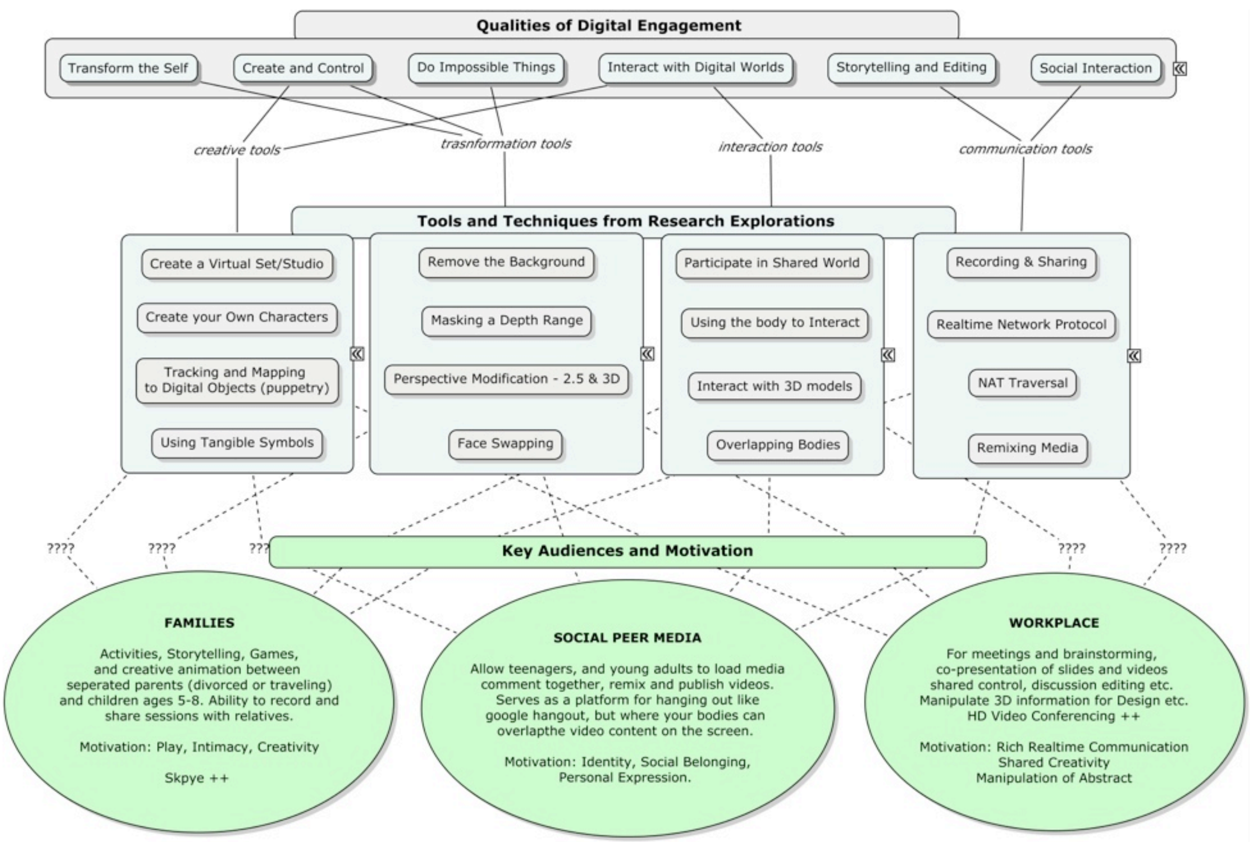

Figure 99: A brainstorm of unexplored qualities of digital engagement for shared VCM systems, the tools that support new experiences, and possible audiences and opportunities.

When we first started brainstorming about this area two years ago I worked with Edith Ackerman to identify qualities of digital engagement that are currently under explored in research systems for shared

VMC. In figure 99 we mapped components of the experience to emerging tools and techniques and postulated potential audiences that could benefit from the application of these tools. To a certain extent the *WaaZam* application contains components of this diagram, however many aspects such as 3D interaction, tracking and mapping objects, and supporting storytelling were not addressed.

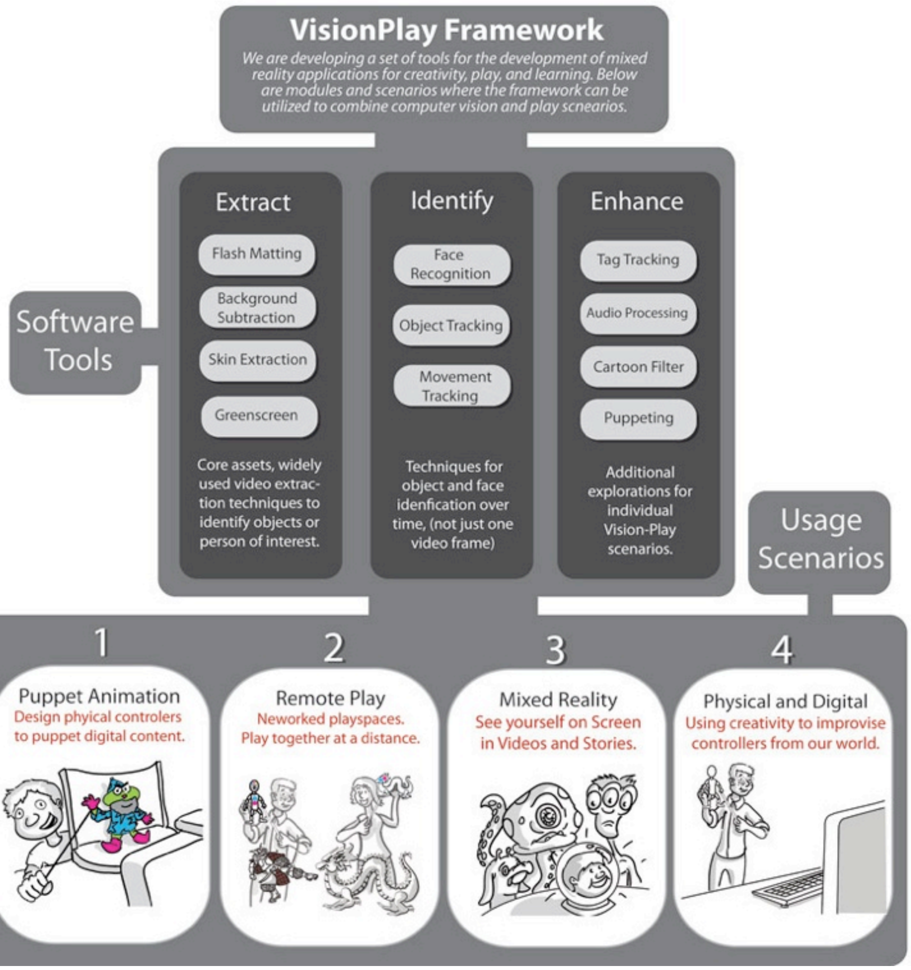

Figure 100. The VisionPlay framework explores scenarios that bridge physical and digital spaces to enable shared play experiences at a distance.

Computer vision techniques are particularly well suited to supporting real time play experiences at a distance. In 2009 I proposed a conceptual framework as a launching point for explorations in remote play at Hasbro Corporation. Many of the software tools are used in *WaaZam* and sensors like the Kinect camera, the LEAP sensor, 3DV systems, and Perceptual Computing time-of-flight cameras have made techniques like green-screening unnecessary but referring to them helps us identify techniques that are tried and true from the past. I am interested in the confluence of improved software algorithms that allow us to extract, identify and enhance information from the real world in order to support augmented virtuality scenarios with others at a distance.

In particular for play scenarios using physical objects and toys to control digital characters affords ways of controlling and interacting that are more natural and expressive. Remote play should support both object play, body play and avatar play. Mixed reality storytelling can use magic tricks to incorporate the viewer into the narrative, using real time capture to make stories personalized. I hope that sharing these diagrams is useful to developers interested in thinking creatively about new types of interactions that bridge physical and digital spaces and potential application domains.

6.2 Application Domain Recommendations

For more generalized uses of creative VMC environments there are many domains we have identified that could benefit from this research. We have identified these by demonstrating our system for hundreds of visitors to the MIT Media Lab and listening to their ideas and feedback. For each new idea we use the scene maker software to design a set to illustrate the idea and understand the needs of that particular domain. To illustrate the variety of possibilities here are some common scenarios that have been suggested:

Figure 101. Example applications illustrated by improv artists: meeting online for a MOOC history class, a language less in French, co-acting an improvisation for an online audience.

Co-broadcasting: teenagers could view YouTube videos, comment together, and post video responses during live video sessions. Aspiring sports fans could co-view and comment on a game for small niche markets of fans who want to tune into their station.

Co-presentations: business partners in two different cities could present to a third party. People in the same class could record a presentation and submit it to the class wiki.

Co-acting and improvisation: real time animation and theatre performance are possible at a distance in constructed environments. Creatively inclined teenagers and adults could produce animations in their own aesthetic and publish the results.

Co-viewing television: while watching a show together family members can step forward to comment, go to the bathroom or suggest alternative content.

Online education: for massive online courses users would be able to collaborate, review, comment and act out content together in a virtual environment while maintaining accountability and authenticity via the live video image.

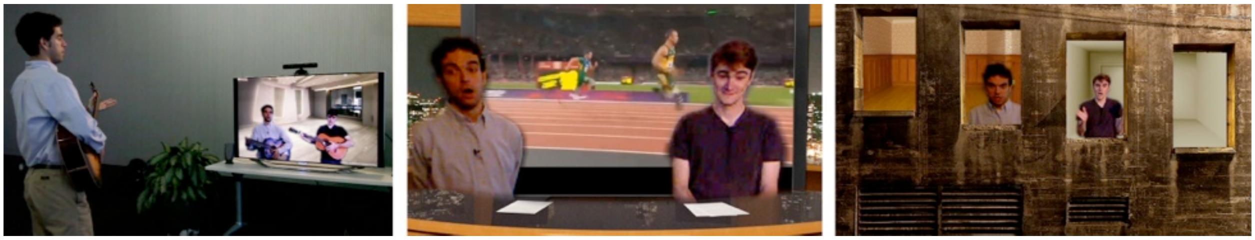

Figure 102. A music lesson, co-broadcasting, co-creating a sitcom

Language learning: many tutors teaching a foreign language use acting as a means of students practicing conversational aspects of the language. If people could teach each other their respective language from their homes a new economy could arise through the exchange of services.

Music lessons: many people learn obscure instruments by contacting musicians online and scheduling Skype lessons. Being together in the same space would allow them to see each other and themselves and better imitate the teacher.

Exercise together: yoga, zumba, cross-fit and dancing are socially motivated activities that people can do together from their individual living rooms. These activities have been popular on the Wii platform.

Family Games: simple games in which the position of your body and limbs influence the environment and in which people have to coordinate together to beat a system.

We also collected feedback from a form at *WaaZam*.com after publishing information about the project and releasing a video. In the survey we received many more suggestions from online users than are summarized here. When preparing to begin a new research project this data provides a reference to tools, techniques and application domains have not yet been explored.

CHAPTER 7. CONCLUSIONS

"Societies have always been shaped more by the nature of the media by which men communicate than by the content of the communication." – Marshall McLuhan

In research the more you know, the more questions you have. Developing the *WaaZam* project has resulted in significant contributions to creative VMC systems for families and we are working to release the software publicly. I am proud of our contributions but there are many unexplored areas as well as limitations to our approach. I present limitations here followed by a proposal for future work that extends the utility, makes the system more interactive and adds support for narrative play. My vision is to enable new, shared experiences for geographically separated people that are engaging, meaningful, and empowering.

7.1 Contribution Summary: Playing Together in Personalized Spaces

In this thesis I presented the design, implementation and evaluation of a uniquely creative system for families to play together at a distance. Our research has resulted in a software application for families and a technical toolkit for researchers. I presented a formal evaluation of our system with key insights related to family play and customization in creative VMC systems. This was followed by recommendations for new systems to support shared activities at a distance. A summary of the contributions include:

1) The novel design and implementation of the WaaZam application
The design of the *WaaZam* system is uniquely focused on families building and playing together in customized worlds at a distance. It includes the tools necessary to compose and share 2.5d sets with other users. The play environment places users in the same virtual space and includes support for puppetry, merged compositing, transformation, a gestural menu, costuming, effects, and set modification. Users can select scenes they design, choose example scenes, or remix scenes made by other users. The system also supports recording short videos for remote theatre purposes. I presented our design guidelines, motivation and implementation in this document.

2) A publicly available application to support creative play at a distance for families
Our team created the *WaaZam* application for families and friends who are geographically separated. We limited the hardware requirements to make it accessible to a broad audience. We are working to release a binary of our software, tutorials, and example videos of people using the system at *WaaZam*.com.

3) An open source toolkit for networking creative applications

The technical overhead required to create an application that stream live video, depth information, audio, and metadata is non-trivial. Once you have data streaming, resolving Network Access Traversal issues adds a layer of complexity that many researchers do not address. We present an open source toolkit which utilizes GStreamer to compress and stream data, XMPP Jabber services for session initiation and NAT Traversal, and OpenFrameworks to interface with the users OS. The examples will be available in OSX, Windows and Linux.

4) A formal evaluation and analysis of our system
After conducting three initial pilot studies we designed a formal evaluation of 24 users of our system that compared conventional VMC setups like *Skype* to magic mirror, digital play sets, and personally customized environments. We found that being together enables new activities, more creative play, and increases engagement for families. Customization appears to foster creative play for children and creates feelings of ownership and personal significance. We demonstrate that our system supports shared activities for families that are active, social and creative by being customizable.

5) Recommendations for further research in other application domains
The possibilities for shared experiences in VMC environments extend to a much broader space than family play. We present an overview of tools and techniques that enable new experiences at a distance, unexplored scenarios, and applications domains that remain unexplored.

7.2 Limitations

Despite the benefits to families and the socially engaging qualities of the *WaaZam* system, there are some distinct limitations to the research. Our user study was conducted with a small set of users and was had a limited feature set. Further studies to repeat our results and test the use of additional features described in the future work section will be needed to provide further evidence that being together in the same virtual space increases engagement and creative play for families. We are confident that the users in our study would benefit from the use of the software and that customization contributes significantly to the experience.

Another limitation of our approach is when using the mirror metaphor you can only see the screen from a front facing or a 3/4 view. If you are interacting with others and you look away from the camera, you no longer feel immersed in the virtual environment. This can be overcome with the right software and hardware pairings. Using a protocol that can send panoramic 3D point clouds you could represent a person at any orientation in space. The challenge is stitching a 3d representation of a person which is complete on all sides so that you can rotate the person and reorient them in space relative to your perspective. This has been accomplished in the literature with 10 Kinect cameras using the PCL libraries but the implementation is quite difficult and has not been streamed over a network.

Another challenge to creating a more realistic interactive experience is tracking the gaze of the user and orienting content to the position and angle of the users head. Even if this is accomplished, things still will not display well for multiple people in the same room.

Our interface did not provide a gestural means of manipulating virtual objects in a natural and intuitive way. Without the use of a wand or controller haptic feedback is also not possible. One solution would be to leverage peoples mobile devices to provide vibration feedback when interacting with the application. We also require users to switch between a gestural and mouse GUI when switching between the scene maker and the scene player. We choose this configuration because we wanted to design a system that early adopters could download and use in their homes without extensive setup requirements.

7.3 Future Work

In order to continue to develop applications that support shared experiences we plan to: a) support the communities that will use the *WaaZam* application and the OFStreamer platform, b) address feature requests that will add more interactivity and immersion to the *WaaZam* application and c) design support tools for storytelling and narrative play.

Develop and Maintain WaaZam.com and OFStreamer.com

Over the course of the first year of developing the network protocol Anette Von Kapri and I began to distinguish the software required to support networked 3D experiences from the client applications because our approaches were different but relied on the same data set. We began to distinguish OFStreamer and *WaaZam* as separate initiatives, the former being the foundational technology and the latter being a domain application.

WaaZam.com will be designed to serve the needs of families who want to play regularly at a distance. The website will provide a demo of the existing features, instructions for downloading the software, a portal for people to share and download assets and a forum for bugs and feature requests. Documentation of the initiative and news about the system will be contained on this website.

OFStreamer is targeted towards developers interested in experimenting with creative telepresence but without deep knowledge of networking libraries, compression codecs, NAT Traversal libraries, and contact management. These are solved problems in other domains and we have been working to slowly assemble and test approaches from Linux libraries currently used for media management, streaming, and NAT traversal. We are currently working with outside developers to implement networking features necessary to scale the application.

Summary of Features to Incorporate in Future Work

During the course of this year we also worked with several artists and designers who made scenes for *WaaZam* and gave us feedback about the types of features they would like *WaaZam* to support. I would like to share these ideas in this document for researchers and designers interested in extending this work and to formalize the ideas which I was not able to explore due to time constraints and the scope of our user study. Birgit Palma and Bruno Selles, designers at *Vasava*, an elite design firm in Barcelona have outlined possible directions for the future of the project in a "blue sky" fashion. This section combines their ideas with feedback from the creativity experts and lessons from our user study. My hope is to illustrate a set of possibilities and suggest possible approaches to improving the environment through interactivity, customization and narrative scaffolding.

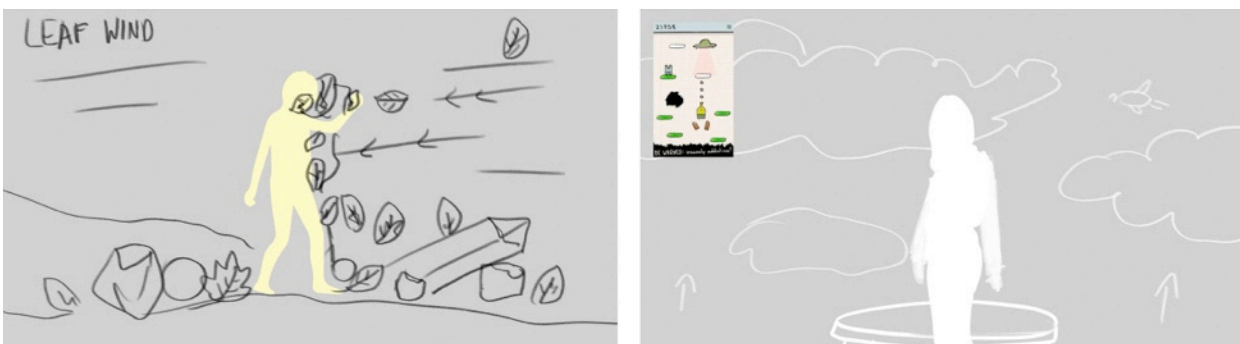

Figure 103. Sketches from Vasava of interactive scenarios: the user would be in space where attributes of the world respond to the user and change the environment, or the user could move left and right to dodge or catch items falling from the sky.

When we combined and categorized the suggestions they cluster into the following categories:

-increase interactivity (control and move props, physics engine, modifying props)
-add more dynamic media (add sound, movies, animation to scenes)
-map gestures to transformation of users (flying, scaling, costuming of self and other)
-create more structured cooperative activities (games, prompted activities, bodily games)
-add magic effects (speech bubbles, gestural effects, triggering scenes and cues)
-add parental controls (meta view, advanced controls, pan tilt camera, controlling rendering modes)
-add storytelling support/scaffolding (narrative structuring, gestural triggers, scene switching)
-explore new artistic experiences (unexpected rendering, surprising overlaps, revealing layers)

We are currently working to add the features that do not require restructuring our application. These include simple effects, gestural transformation, costuming, dynamic media, moving props, and parental controls. The remaining categories require creating separate applications that build on the OFStreamer protocol to support gaming, artistic experiences, and storytelling.

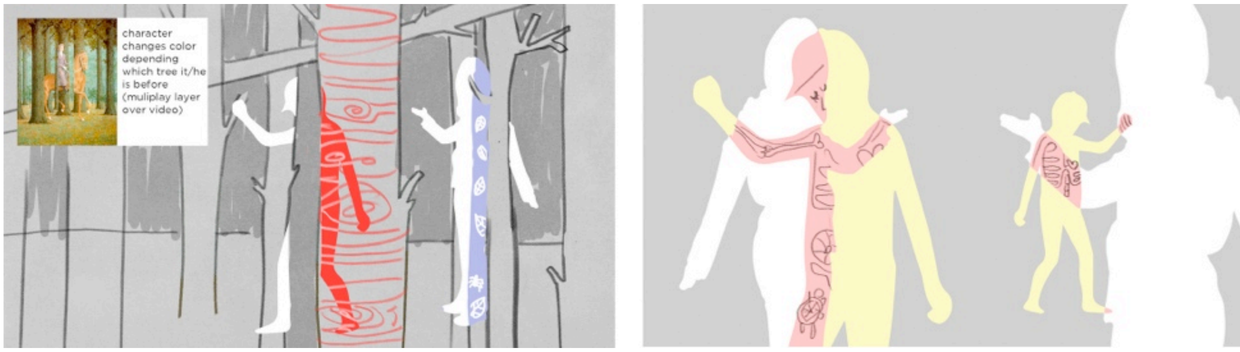

Figure 104. Vasava sketches of interactive experiences that are Magritte-like by revealing hidden layers content in the environment or the figures of other people through the exploration of participants.

Social games (see Figure 103) in which participants work together to accomplish a goal by interacting with elements in the scene similar to many games on the XBox would add a lot of interesting possibilities for structured activities. We have integrated Bullet Physics, a physics simulator used in Blender, Lightwave and many other 3D environments to add interactive capabilities. All elements would need contour polygons which are linked to the Bullet Physics simulator and then these objects would need to be synchronized via a network protocol which assigns one computer as the master and the other as slave.

Figure 105. Vasava sketches of selective rendering based on the location of the user and interaction with specific elements.

Artistic possibilities have been explored in the gallery by artists like Scott Snibbe, Theo Watson, Myron Krueger, and many other early pioneers of interactive art but none of these approaches have been implemented into a telepresence framework that allows people to explore environments at a distance in different artistic environments. Reversing the expectations of how viewers will be rendered often elicits cries of surprise and delight (especially when things happen inside a mirror image of your silhouette).

The best way of accomplishing this effect would be to have settings in the scene maker that allow people to layer blending effects and then apply GPU shaders to selected content. Masking is a fairly simple shader and vertex shaders could be used to designate different render areas.

Support Storytelling and Narrative Play

Supporting storytelling was not the primary focus of our application but is a natural addition from an educational standpoint. This has been done with great success by the developers of *Toontastic* an app for the iPad which helps children animate drawings on their iPad and output animations with very little prior knowledge of animation techniques. The story arc provides a scaffolding structure that helps children understand how to create a story with a setup, conflict, challenge, climax and resolution. The app provides a set of example sets and has a ToonTube where children can watch animations by other children inside the application.

Figure 106. Screenshots from the toontastic app for ipad which allows children to edit scenes within a narrative structure.

Toontastic has been very successful as an app. In *WaaZam*, children and parents could use their bodies to act out different scenarios in magic worlds at a distance. A narrative tutorial and scaffolding technique such as Toontastic's would be very helpful to help them think through the different parts of the story before acting out the scenes together.

7.4 Shared Experiences: How Society will Create and Play in the Future

In 1910 the French artist Villemard created a series of postcards envisioning life in the year 2000. Many of his drawings are cited by historians as some of the most detailed depictions of what Parisian society thought the future would be like. In a picture addressing the future of communication, a gentleman is sitting talking with his wife through a speaker and viewing a live video of her through a projection system. The postcard depicts audio and video components and shows an operator managing the equipment.

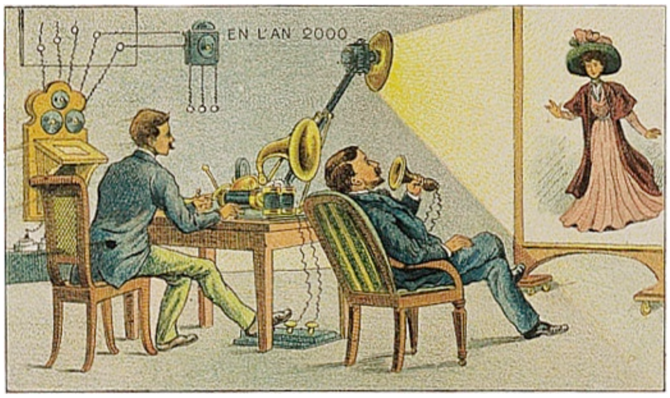

Figure 107. Remote interpersonal communication in 1910 envisioned by the artist Villemard

The role of the assistant in the picture is to facilitate an experience between this man and his wife. I envision myself as that operator when I design computational systems like *WaaZam.* Today artists and researchers design precursors that hint at the future of communication in the next hundred years. During this century VMC systems will support more interactive, personally significant and empowering forms of synchronous remote communication. People will be able to invent their own worlds at a distance. VMC technologies are progressing quickly and present an opportunity for designers to invent a new medium for shared experiences. How will that medium form, what characteristics will it embody, and how will it shape our society?

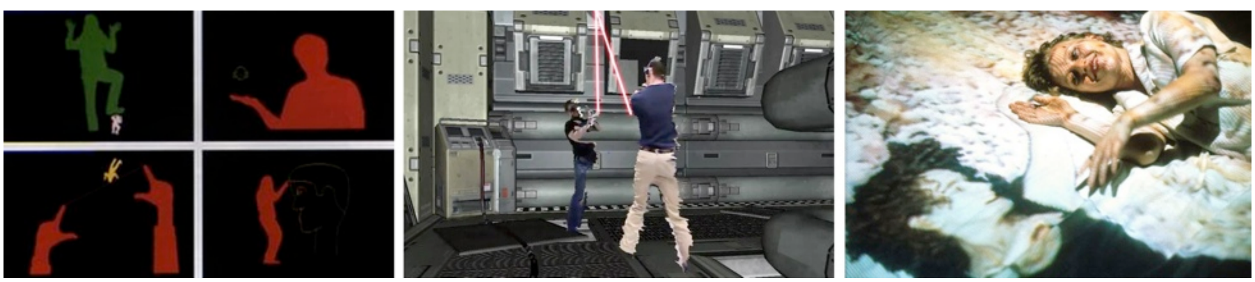

Figure 109. Myron Krueger's original vision, a Kinect based enactment of Star Wars from UC Davis 2011, explorations by the artists Paul Sermon on being together at a distance.

The new shared medium has already been forecasted by artists and hackers. Myron Krueger and Paul Sermon have worked for decades to illustrate a future in which virtual spaces supports shared human activities. Some of the hacks that emerged after the Kinect sensor was released in 2010 illustrated collaborative scenarios in simulated environments. We are replacing the avatars of MMOGs and Second Life with live 3d video images of ourselves.

The medium will blur the boundaries between digital and physical worlds. Doing things together at a distance relies on interaction designers forecasting the capabilities of new sensors and utilizing them in new configurations. Virtual representations will be more accurate and responsive as interaction designers link arrays of sensors together with software libraries that can align them in 3D space. These representations will be combined with digital content and will respond to gestures, orientation, and relationship of participants.

The medium will shape who we are. As VMC systems progress, I advocate that we design systems for the next generation to explore aspects of their identity with friends and family. The users ??will form creative communities that produce media at a distance and share their productions with friends and family online. Instead of merely consuming content, these communities will connect to each other and participate in creating digital culture together.

The medium will be fantastic. What inspires me to design collaborative experiences in virtual spaces is that we can do things in that environment that we could never do together in reality. Participants can become what they envision themselves to be. When an elderly grandmother in my study told me that she enjoyed swimming in the lake with her grandaughter because she could no longer physically run or swim, I realized that she did not distinguish emotionally between pretending and actually doing those activities with her grandaughter.

The medium should be safe place for experimentation. Augmented virtuality is not meant to replace face-to-face interactions. It presents new opportunities for people to maintain active, social and creative relationships at a distance. Our children are growing up in a more connected society. They need virtual playgrounds that are safe places to personify their feelings and fantasies with new media. *WaaZam* is an environment where children can enact imaginary scenarios with their families.

The medium will be invented by its participants. I believe it is the responsibility of designers to invent a media that supports our human purposes and is motivated by democratic values. I envision a future where its users invent the medium in which they interact together. People should have tools that let them appropriate new technologies to serve social and creative needs. This thesis is an initial step in that direction. We should be able design new and delightful experiences with our children when we are apart and in the process build and sustain our relationships with each other.

Appendix A
Pre-Study Questions for Adults in User Study

1) If you do video calls with the child, about how much time do you spend doing this each week?
2) If you do video calls with the child, what kinds of activities do you do during these video calls?
3) What is your level of satisfaction with your experience of these video calls? (1-7)
4) Please check any game consoles you own. (PS, XBOX, WII, Other)
5) If you play video games with the children in your family, about how much time do you spend doing this each week?
6) What is your level of satisfaction with your experience of playing video games together?
7) Please estimate much time per week do you spend on the phone, internet or other technologically facilitated communication with the child?
8) How connected to the child do you feel when communicating at a distance?
9) How connected do you feel right now to the child?
10) Please describe a few play activities that you participated in with the child during the last month:
11) Please indicate what modes of play you engage in with the child, and how much you engage in each mode of play.
    facial expression play (example: peek-a-boo)
    body play (example: swimming, rolling around, wrestling)
    object play (example: LEGOs, marbles)
    social play (example: tag, games, celebratory play)
    pretend play (example: talking with objects, role-playing)
    narrative play (storytelling)
    transformative play (example: playing music, role-playing, becoming a character)

12) Do you and the child ever build things together or engage in fantasy role playing?

Appendix B.
Pre-Interview Questions for Children in User Study

1) What is your favorite game to play with your parents?

2) Do you ever use the computer to talk call people in your family?

3) Do you play any games on TV or the computer?

4) Do you like to make things or build things at home?

5) Do you ever make things on the computer?

Appendix C.
Post-Interview Questions for Children

1) What was your favorite thing to do in the application?
2) Was there anything you did not like?
3) Which scenes were your favorite? Remind them of the scenes.
3a) Did you like visiting the other person's room, using our backgrounds, or using your own designs better? Why?
4) Which one do you think is best to do with (name of other adult)?
5) How easy is it to come up with ideas for your own background sets?
6) If (name of other adult) wanted to do this with you at your house would you want to do this with them? If so what would you guys do together there? How often would you use it?
7) What would make you feel close (connected - like they are there - so you don't miss them as much) to the parent when they are traveling - activities etc. open ended as to what makes them feel close to the parent.
8) What else would you like to make? Was there anything you wanted to make but could not with our system?
9) Are there any toys or objects not here you want to put into the world?
10) Would you rather make your own sets or use the sets of others?
11) Do you have any other comments for us to make this better and more fun for you to do with other people?
12) Initiate an exploratory discussion about if they want in their home for the purpose of being with their parents when they are away - and why?
13) Initiate an exploratory discussion on the meaning of customization to the child? Ownership? Creating their own scenes and recording?
14) Ask about the possibility of sharing scenes/videos with others - who would they share them with?
15) If everyone had this, who would you use it with and what you do?

Appendix D.
Post Interview Questions for Adults

1) Please describe the highlight of your session today.
2) What was the most frustrating experience?
3) Please rate Skype mode, mirror mode, imaginary mode, and customized mode (1-7) in terms of how fun the activities were..
4) Please also rate your how connected (1-7) you felt to the child during each condition.
5) Did you find the making your own scene easy or hard on a scale of 1-7?
6) Imagine you are traveling in the future and you have a version of *WaaZam* that works in your hotel. Would you be inclined to use the system? And if so, how would you use it? How often?
7) Can you describe how you would use the software in this future scenario? Why?
8) What is the likelihood that you would make your own scenes if you had the software at home? Please rate from 1-7 likelihood you do so?
9) Would you be more likely to make a scene by yourself, or collaboratively with the child?
10) What types of props/objects would you want to add in the future? Would they be primarily digital or physical things or both?
11) If you had this environment in your home would you search for sets made by other participants? Would you share your own sets?
12) Are there any activities you wanted to do in *WaaZam* but were unable to? Please describe what you imagine doing.
13)How many days a week do you see the child? *If they are separated more than half the time* do you think you would benefit from using this platform? Would you be interested in a long term home installation if the opportunity arises?

Appendix E.
Video Coding Criteria Categories:

The procedure for coding was as follows: First we went through the videos and identified episodes and labeled episodes with ratings according to the coding criteria for mutuality, types of play, and parent child engagement. We then followed up with a second pass on the video, coding every 15 seconds for level of engagement and attention of the participant.

Episodic:
Label each interval.
At transition points assign episodic characteristics
Characteristics:
1) Who initiated transition
2) Child Maturity 1-5
3) Child Social Initiative 1-5
4) Child Object Initiative 1-5
5) Parent Directiveness 1-5
6) Parent Elaborativeness 1-5
7) Parent Sensitivity 1-5
8) Body Play 1,0
9) Object Play 1,0
10) Social Play 1,0
11) Pretend Play 1,0
12) Narrative Play 1,0
13) Transformative-Integrative Play 1,0
14) Laughed
15) Smiled
16) Had Performative Gestures
17) Exclaimed! 1,0
18) Include in Documentation later 1,0
19) Which condition (digital, mirror, skype)

Noted by coder for each episode:
-- Activity Description (for categories of play)
-- Observations of Interest - was it funny, original, compelling, sweet, what is your qualitative feeling regarding the episode.

Interval: Loop back to 15 sec intervals within the episode and mark:

A) Parten's levels of engagement:
0-unoccupied (no play)
1-solitary
2-parallel with occasional mutual regard
3-parallel play with mutual regard
4-associative play with attention to the other person
5-cooperative/mutual play coordinating with a common task

B) Categories of Attention:
1-Self
2-Other
3-Us
4-Virtual Environment
5-Physical Environment

Bibliography:

Agamanolis, S., & Bove, Jr., V. (1997) Multilevel Scripting for Responsive Multimedia. *MultiMedia , 4* (4), 40-50, 1997

Ames, M. G., Go, J., Kaye, J., and Spasojevic, M., (2010) "Making Love in the Network Closet: The Benefits and Work of Family Videochat," in Proc. of CSCW, pp. 145–154.

Ballagas, R., Kaye, J. J., Ames, M., Go, J., and Raffle, H., (2009) "Family communication: phone conversations with children," in Proc. of IDC, (Como, Italy), pp. 321–324, ACM.

Benko, H., Jota, R., and Wilson, A.D. (2012) MirageTable: freehand interaction on a projected augmented reality desktop. In Proc. of ACM SIGCHI, 199-208.

Brubaker, J. Venolia, G. and Tang. J (2012). Focusing on shared experiences: moving beyond the camera in video communication. In Proceedings of the Designing Interactive Systems Conference (DIS '12). ACM, New York, NY, USA, 96-105.

BruckMan, Amy, Luther, Kurt, Kevin Ziegler, Kelly Caine (2009). "Predicting Success in Online Creative Collaboration." ACM Conference on Creativity and Cognition, Berkeley, California.

Bruner, J. S. (1975). The ontogenesis of speech acts. Journal of Children Language, 2, 1–40.

Bodrov A, E. and Leong, D.J. (1996): *The Vygotskian approach to early childhood.* Ohio, Merrill, Prentice Hall.

Buckingham, D. (2007). Beyond technology: Children's learning in the age of digital culture. Malden, MA: Polity.

Chatting, D. Galpin, J and Donath, J. (2006) Presence and portrayal: video for casual home dialogues. In Proceedings of the 14th annual ACM international conference on Multimedia (MULTIMEDIA '06). ACM, New York, NY, USA, 395-401

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. Perspectives on social shared cognition. Washington, DC: American Psychological Association.

Consolvo, S., Roessler, P., and Shelton, B. E., (2004) "The CareNet Display: Lessons Learned from an In Home Evaluation of an Ambient Display," in Proc. of UbiComp, pp. 1–17.

Dewey, J. (1938). Experience and education. New York: Collier Books.

Duckworth, E. (1987). The Having of Wonderful Ideas and Other Essays on Teaching and Learning. Teachers College, Coloumbia University, New York.

Evjemo, B., Svendsen, G. B., Rinde, E., and Johnsen, J.-A. K., (2004) "Supporting the distributed family: The need for a conversational context," in Proc. of NordiCHI, no. 4, pp. 309–312.

Fish, R. Kraut, R and Chalfonte, B. (1990) The VideoWindow system in informal communication. In Proceedings of the 1990 ACM conference on Computer-supported cooperative work (CSCW '90). ACM, New York, NY, USA, 1-11.

Follmer, S., Raffle, H., Go, J., Ballagas, R., and Ishii, H., (2010) "Video play: playful interactions in video conferencing for long-distance families with young children," in Proc. of IDC, pp. 49– 58, 2010.

Gee, J.P. (2003). What video games have to teach us about learning and literacy. New York: Palgrave.

Hemmeryckx-Deleersnijder B and Thorne. J (2008) Awareness and conversational context- sharing to enrich TV-based communication. Comput. Entertain. 6, 1, Article 7

Isaac, E. A., & Tang, J. (1994). What video can and cannot do for collaboration: a case study. Multimedia Systems, 63-73

Jenkins, Henry (2006). "Convergence Culture: Where Old and New Media Collide". New York: New York University Press. p. 308

Judge, T. K. and Neustaedter, C., (2010) "Sharing conversation and sharing life: video conferencing in the home," in Proc. of CHI, (Atlanta, Georgia, USA), pp. 655–658, ACM, 2010.

Kirk, D. S., Sellen, A., and Cao, X., (2010) "Home video communication: mediating 'closeness,'" in Proc. of CSCW, (Savannah, Georgia, USA), pp. 135–144, ACM.

Ekman, P., & Friesen, W. (1968). Nonverbal behavior in psychotherapy research. Research in Psychotherapy: Proceeding of the Third Conference, 3, 179–183.

Frued, S. (1959): *Beyond the pleasure principle.* (J. Strachey, transl.) New York, Norton.

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., and Kramer, A. D. I., (2004) "Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks," Human-Computer Interaction, vol. 19, no. 3, p. 273.

Garvey, C. (1977): *Play.* Cambridge, Massachusetts, Harvard University Press.

Gaver, W. W., Sellen, A., Heath, C., and Paul Luff, (1993) “One is not enough: multiple views in a media space,” in INTERCHI, (Amsterdam, The Netherlands), pp. 335–341, ACM.

Gillian, Nick, (2013) Gesture Recognition Toolkit: http://www.nickgillian.com/software/grt

Hemmeryckx-Deleersnijder B and Thorne. J, (2008) Awareness and conversational context-sharing to enrich TV-based communication. Comput. Entertain. 6, 1, Article 7.

Inpen, K. and Junuzovic, S., (2012) “PixIO : Sharing Any Surface,” in Proc. of CHI.

Ishii, H., Kobayashi, M., & Arita, K. (1994). Iterative design of seamless collaboration media. *Communications of the ACM , 37* (8), 83-97.

Izumi Taylor, Satomi; Rogers, Cosby Steele; & Kaiser, Javid. (1999). A comparison of playfulness among American and Japanese preschoolers. In Stuart Reifel (Ed.), *Play and culture studies: Vol. 2. Play contexts revisited* (pp. 143-150). Stamford, CT: Ablex.

Johnson, J.E., Christie, J.F., and Yawkey, T.D. (1999). Play and early childhood development. Addison Wesley Longman, Glenview, Illinois.

Judge, T. K. and Neustaedter, C., (2010) “Sharing conversation and sharing life: video conferencing in the home,” in Proc. of CHI, (Atlanta, Georgia, USA), pp. 655–658, ACM.

Krueger, Myron (1991) Artificial Reality, Addison-Westly

Lieberman, Zack: (2009) Reface, http://thesystemis.com/projects/reface/

Ledo, D., Aseniero, B.A. Greenberg S. Boring, S. and Tang, A (2103) OneSpace: Shared Depth-Corrected Video Interaction. In Proc. ACM CHI April 27- May 3. A CHI Work in Progress.

Levin, Golan, (2003) Messa di Voce, http://www.flong.com

Harmon, A. (2007). Grandma’s on the computer screen. The New York Times. November 28, 2007

Irina Verenikina, Pauline Harris, and Pauline Lysaght. (2003). “Child's play: computer games, theories of play and children's development.” (CRPIT '03), June Wright, Anne McDougall, John Murnane, and Jenny Lowe (Eds.), Vol. 34. Australian Computer Society, 99-106.

Junuzovic S., Inkpen K., Blank, T and Gupta, A, (2012) "IllumiShare: Sharing Any Surface", in CHI 2012, ACM Conference on Human Factors in Computing Systems.

Leontiev, A.N. (1981): *Problems of the development of the mind.* Moscow, Progress Publishers.

Maimone, A., Bidwell, J., Peng, K., & Fuchs, H. (2012) "Enhanced personal autostereoscopic telepresence system using commodity depth cameras." *Computers & Graphics , 36* (7), 791-807.

Mehrabian, A. (1972). Nonverbal communication. Chicago: Aldine-Atherton.

Morikawa, O., & Maesako, T. (1998). HyperMirror: toward pleasant-to-use video mediated communication system. *Proceedings of the conference on Computer supported cooperative work* (pp. 149-158). Seattle, Washington, USA: ACM.

Mynatt, E., Rowan, J., Jacobs, A., and Craighill, S., (2001) "Digital family portraits: supporting peace of mind for extended family members," in Proc. of CHI, (Seattle, Washington), pp. 333 – 340, ACM

Nescher, T., & Kunz, A. (2010). An interactive whiteboard for immersive telecollaboration. *The Visual Computer: International Journal of Computer Graphics - Special Issue on CYBERWORLDS 2010 , 27* (4), 311-320.

Neustaedter, Oduor, Venolia, Judge. (2012). Moving beyond talking heads to shared experiences: the future of personal video communication. In Proceedings of the 17th ACM international conference on Supporting group work (GROUP '12). ACM, New York, NY, USA, 327-330.

Nikolopolou, A. (1993): Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. *Human Development* 36:1–23.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas.* New York, Basic Books

Piaget, Jean (1954) *The construction of reality in the child*, New York, Basic Books

Piaget, J. (1962): *Play, dreams and imitation in childhood.* New York, Norton.

Plaisant, C., Clamage, A., Hutchinson, H. B., Bederson, B. B., and Druin, A., (2006) "Shared family calendars: Promoting symmetry and accessibility," ACM ToCHI, vol. 13, no. 3, pp. 313–346.

Rainie, L. and Zickuhr, K., (2010) "Video calling and video chat," in Pew Internet & American Life Project: http://pewinternet.org/Reports/2010/Video–chat.asp

Raffle, H., Revelle, G., Mori, K., Ballagas, R., Buza, K., Horii, H., Kaye, J. J., Cook, K., Freed, N., Go, J., and Spasojevic, M., (2011) "Hello, is Grandma there? Let's Read! StoryVisit : Family Video Chat and Connected E-Books," in Proc. of CHI, pp. 1195–1204.

Raffle, H., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J., Reardon, E., Mori, K., Kaye, J., and Spasojevic, M., (2010) "Family Story Play: Reading with Young Children (and Elmo) Over a Distance," in Proc. of CHI, ACM.

Renger, B. Feng, Dan, Chang, and Barbosa. (2011) VoiSTV: voice-enabled social TV. In Proceedings of the 20th international conference companion on World wide web (WWW '11). ACM, New York, NY, USA, 253-256.

Rokeby, David: Transforming Mirrors, (1996) An essay which appears on his website: http://www.media-studies.ca/articles/rokeby.htm

Romero, N., Markopoulos, P., Baren, J., van, Ruyter, B., de, Jsselsteijn, W., & Frashchian, B. (2009). Connecting the family with awareness systems. Personal and Ubiquitous Computing, 11, 303–329.

Rozin, Daniel: (2003) Mechanical Mirrors: http://www.smoothware.com/danny/

Salen, K. (2007). The ecology of games: Connecting youth, games, and learning. Cambridge, MA: MIT Press.

Sawhney, N. Wheeler, S. and Schmandt. C (2001). Aware Community Portals: Shared Information Appliances for Transitional Spaces. Personal Ubiquitous Comput. 5, 1, 66-70.

Saywer, Keith (2012) "Explaining Creativity: The Science of Human Innovation", Oxford University Press, pp. 389-405

Siio, I., Rowan, J., and Mynatt, E., (2002) "Peek-a-drawer: communication by furni- ture," in Ext. Abst. of CHI, (Minneapolis, USA), pp. 582–583, ACM.

Snibbe, Scott, (2002) Deep Walls: http://www.snibbe.com/projects/

Synovate, "Smart Phone Parenting," Tech. Rep. July, 2011.

Tang, A., Neustaedter, C., & Greenberg, S. (2007). VideoArms: Embodiments for Mixed Presence Groupware. In *People and Computers XX — Engage* (pp. 85-102). London: Springer.

Tang, J. C. and Minneman, S. L., (1991) "Videodraw: a video interface for collaborative drawing," ACM Trans. Inf. Syst., vol. 9, no. 2, pp. 170–184.

Tsukada, Y. and Guimbretière, R. (2010) TouchFace: the interaction between cursors and live video images for casual videoconferencing. In Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10). ACM, New York, NY, USA, 39-42

Utterback, Camille (1999), Text Rain: http://camilleutterback.com/

Von Kapri, A, Wong, K, Hunter, S, Gillian, Maes, P (2013) "InReach: Manipulating 3D Objects Remotely Using your Body" Intl. Conf. Human Factors in Computing (CHI 2013), Workshop paper.

Vetere, F., Davis, H., Gibbs, M., and Howard, S., (2009) "The Magic Box and Collage: Responding to the challenge of distributed intergenerational play," Intl J of Human Computer Studies, vol. 67, pp. 165–178.

Vygotsky, L.S. (1977): *Play and its role in the mental development of the child*. In *Play: Its Role in Development and Evolution.* New York, Basic Books.

W.M. Keck Center for Active Visualization in the Earth Sciences, (2013) *KeckCAVES.* Retrieved from http://keckcaves.org

Watson, Theo (2009), Knee Deep: http://www.theowatson.com/

Wertsch, J. (1985). Vygotsky and the Social Formation of Mind. Cambridge: Harvard Press.

Yarosh, S., Cuzzort, S., Muller, H., and Abowd, G. D., (2009) "Developing a media space for remote synchronous parent-child interaction," in Proc. of IDC, pp. 97–105.

Yarosh, S., Inkpen, K., and Brush, A., (2010) "Video playdate: toward free play across distance," in Proc. of CHI, (Atlanta, GA), pp. 1251–1260, ACM.

Yarosh, Radu, Hunter, and Rosenbaum. (2011) Examining values: an analysis of nine years of IDC research. In Proceedings of the 10th International Conference on Interaction Design and Children (IDC '11). ACM, New York, NY, USA, 136-144.

Zuckerman, O. and Maes, P., (2004) "Awareness system for children in distributed fam- ilies," in Proc. of IDC, (Boulder, USA), ACM.