
Designing Systems that Direct Human Action

by Ana Ramírez Chang

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Marc Davis
Research Advisor

(Date)

* * * * *

Professor Jennifer Mankoff
Research Advisor

(Date)

Designing Systems that Direct Human Action

Ana Ramírez Chang

Electrical Engineering and Computer Science Department
University of California at Berkeley
Berkeley, CA 94720-1770
anar@cs.berkeley.edu

December 11, 2005

Abstract

In Active Capture applications, systems that direct human action, the system works with the user to achieve a common goal, for example, taking her picture and recording her name for inclusion on a department web page. The design of Active Capture applications draws from the areas of direction and cinematography, computer vision and audition, and human-computer interaction and results in a diverse design team. Without a tool to help the design team work together and leverage expertise of each of the members, the design of Active Capture applications is limited to a design team with members who can bridge the gap between the different disciplines Active Capture applications draw from. In this report I present a design process for Active Capture applications, a visual language to help the members of the design team work together and a tool to help the design team integrate the computer vision part of the system and the interaction with the user. These three pieces are work toward an integrated tool to support the design of Active Capture applications. At the end of the report I include recommendations for such an integrated tool.

Contents

1	Introduction	1
2	Active Capture Applications	3
3	Active Capture Application Components	6
3.1	Interaction Script	7
3.2	Action Recognizers	9
4	Challenge	9
5	Methodology	12
5.1	Ethnographic Study	13
5.1.1	Design Strategies	15
5.1.2	Direction and Feedback Strategies	16
5.1.3	Mediation Strategies	17
6	Active Capture Design Process	20
6.1	Bodystorming	20
6.2	Wizard-of-Oz User Study	22
6.3	Designing Action Recognizers	24
7	ACAL	26
7.1	State Transition Diagrams	26
7.2	Statecharts	29
7.3	Hypermedia Authoring Systems	31
7.4	ACAL Design History	31
8	Related Work	40
8.1	Active Capture	40
8.2	Active Capture Design Process	45
8.3	ACAL	47
8.4	SIMS Faces system	49

9	Evaluation	50
10	Conclusion and Future Work	52
A	SIMS Faces System Interaction Script	61
B	Wizard-of-Oz User Study Script	67
C	Wizard-of-Oz Study Consent Form	69
D	Wizard-of-Oz Study Media Release Form	70
E	Wizard-of-Oz Study Follow-up Questionnaire	71
F	Wizard-of-Oz Study Questionnaire Data	73
G	SIMS Faces System User Study Consent Form	86
H	SIMS Faces System User Study Media Release Form	87
I	SIMS Faces System User Study Follow-up Questionnaire	88
J	SIMS Faces System User Study Questionnaire Data	90

List of Figures

1	Active Capture	2
2	Headturn and Terminator II Trailer Images	4
3	Picture Module from SIMS Faces Application	10
4	ACAL Representation of Headturn Active Capture Module	14
5	Design Process	21
6	Smile Picture and Head Turn Action Recognizers	24
7	Smile Recognizer	25
8	State Transition Diagram of Headturn Module	28
9	Example of hierarchy in statecharts.	29

10	Example of orthogonal combinations in statecharts.	30
11	Headturn Active Capture Module as a Statechart.	32
12	Path of least mediation in headturn module in Authorware.	33
13	First Step in Headturn Authorware Implementation	33
14	Step Chart Example	36
15	Path of Least Mediation in the SIMS Faces Application	38
16	ACAL representation of the SIMS Faces Application	39
17	Path of Least Mediation in the Picture Active Capture Module	40
18	Picture Module from SIMS Faces Application	41
19	Lab setup for Wizard-of-Oz user study	51

List of Tables

1	SIMS Faces Application Abbreviated Interaction Script	8
---	---	---

Acknowledgments

Sincerest thanks to my advisers, Professors Marc Davis and Jennifer Mankoff. Thanks to Arian Saleh, Brett Fallentine, Rita Chu, My Huynh, Pauline Jojo Chang, Leo Choi, William Tran and Madhu Prabaker for their work on the SIMS Faces application. Thanks to Ka-Ping Yee for his help with ACAL. I also thank the participants of both studies. The author is supported by an NSF fellowship.

1 Introduction

In Active Capture applications [Dav03b], systems that direct human action, the system works with the user to achieve a common goal, for example, taking her picture and recording her name for inclusion on a department web page. The design of Active Capture applications draws from the areas of direction and cinematography, computer vision and audition, and human-computer interaction and results in a diverse design team. Without a tool to help the design team work together and leverage expertise of each of the members, the design of Active Capture applications is limited to a design team with members who can bridge the gap between the different disciplines Active Capture applications draw from. In this report I present a design process for Active Capture applications and a visual language to help the members of the design team work together. These two pieces are work toward an integrated tool to support the design of Active Capture applications. At the end, I include recommendations for such an integrated tool.

The design process follows the design-test-analysis methodology used in SUEDE [SKL02] with the use of bodystorming [OKK03], wizard-of-Oz studies [DJ93] and traditional user studies. It allows for rapid iteration during the early phases, and reduces the cost of iteration throughout the design process. The visual language is based on a survey of languages that support the passage of time, or control flow, and supports both the passage of time and control flow. It allows all members of the team to understand the design, allowing the team to work on the design together without treating some members of the team as consultants.

The design process and visual language are based on experience designing and

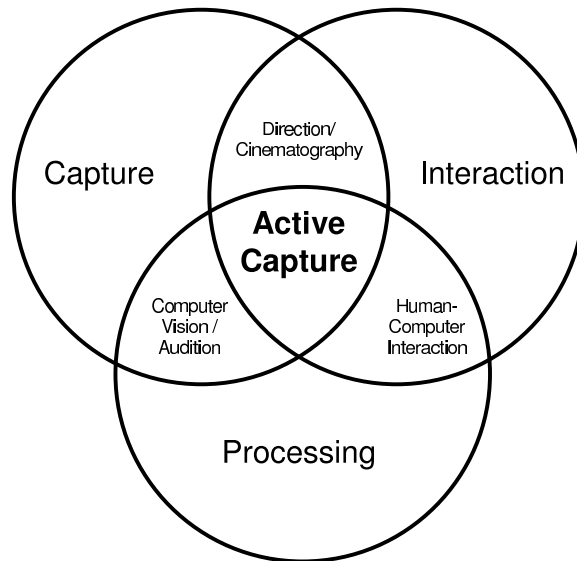


Figure 1: Active Capture applications bring together capture, interaction and processing, enabling system direction of human action.

developing the SIMS Faces application [Cha05]. This Active Capture application works with the user to record her saying her name, and take her picture for inclusion on the department web page. The application will allow the department staff to easily capture a picture of each new student and an audio clip of each student pronouncing her name without having to take the pictures and record the names manually. The new students can use the application as many times as they would like without placing a burden on the department staff.

In the following sections, I will describe Active Capture applications in detail and the challenges they present for the design team and my methodology for the project. I will present the user-centered design process and the visual language followed by a set of recommendations for an integrated Active Capture design tool.

2 Active Capture Applications

“Active Capture” is a new paradigm in multimedia computing and applications that brings together capture, interaction, and processing and exists in the intersection of these three capabilities (See [Figure 1](#)). Most current human-computer interfaces largely exclude media capture and exist at the intersection of interaction and processing. In order to incorporate media capture into an interaction without requiring signal processing that would be beyond current capabilities, or bridging the *semantic gap* [DV01], the interaction must be designed to leverage context from the interaction. For example, if the system wants to take a picture of the user smiling, it can interact with the user to get them to face the camera and smile and use simple, robust parsers (such as an eye finder and mouth motion detector) to aid in the contextualized capture, interaction, and processing. From the computer vision and audition side, Active Capture applications are *high context multimedia recognizers*. They augment computer vision and audition parsers and recognizers with context from the interaction with the user. Active Capture applications bring together three disciplines (direction / cinematography, human-computer interaction, and computer vision / audition), giving them more power and making them harder to design and implement than applications that draw from just one or two of the disciplines alone. Not only are the applications interdisciplinary, but the design team and design tools must also draw from and support each of the disciplines.

In this report, I will refer to two Active Capture applications, the SIMS Faces application (which my team designed and implemented) and the Kiosk Demo [Dav03a].



Figure 2: Pictures of an Active Capture participant performing a head turn. The figure shows both the original captured footage and corresponding images from an automatically generated Terminator II trailer.

The SIMS Faces application works with the user to achieve two goals, take her picture, and record her saying her name. The Kiosk Demo is similar to a photo kiosk in the mall, but instead of taking the user's picture, it takes a few videos of the user, and automatically creates a personalized commercial or movie trailer starring the user. There are two parts in the Kiosk Demo, the Active Capture part works with the user to capture a shot of her looking at the camera and screaming, and a shot of her turning her head to look at the camera. The second part of the Kiosk Demo uses Adaptive Media technology described in [Dav03c, DL96]. The shots of the user screaming and turning her head are automatically edited into a variety of commercials and movie trailers including a 7up commercial, an MCI commercial, and the Terminator II movie trailer. Figure 2 shows how the head turn is parsed into the Terminator II movie trailer.

Aside from the two Active Capture applications we have built, many other examples exist and are possible in a variety of domains, such as sports instruction, health, and job training.

When studying a sport, there are many repetitive actions that need to be mas-

tered. A personal instructor is often complemented by a system that monitors the learner as she practices repetitive movements and provides her with feedback. For example an automated golf instructor monitors your stance and your golf swing and provides feedback on the speed and projection of your swing as well as the angle of your shoulders and your posture. Such a system would also be useful for other patterned physical activities such as Aikido.

In the health area, Active Capture applications can be used with telemedicine [RD04] and health monitors among other areas to collect patient data for diagnosis after capture. In telemedicine, doctors in bigger cities use video and images sent in from smaller villages to help screen for common illnesses. This allows the doctors to reach people who are too far to come to them and who are too few and/or too expensive to be deployed to all remote village locations. At home, Active Capture applications can help the patient monitor her health. For example, a posture monitor can help the patient remember what the doctor told her, and train her to have better posture. On the preventative side, a keep awake system [Ayo03] helps the person prevent injuries while driving due to extreme drowsiness.

In the job training area, consider a building guard who must be trained to stand still and face forward no matter what is happening around him (such as at Buckingham Palace). An Active Capture application would help the guard learn to not react to stressful stimuli which could be varied and presented to the guard in an Active Capture training application.

In the examples above, the application of the Active Capture paradigm increases the symmetry between the creation and consumption of multimedia. By using human-computer interaction design to interactively simplify the context of

capture and thereby enabling computer vision and audition algorithms to work robustly, the Active Capture applications radically reduces the skill and effort needed to create high quality reusable media assets. In addition to creating high quality media assets, applications in the Active Capture paradigm can focus on the action the guest is performing as an output. For example, a sports instruction system that works with the guest and directs her to improve her skills in that sport is also in the Active Capture paradigm. In an automated golf instructor, the system would work with the guest to help her improve her golf swing. It is important to note that a sports instruction system that does not *direct* the user is not a system in the Active Capture paradigm.

3 Active Capture Application Components

An Active Capture application is made of a set of Active Capture modules, each one with an action it tries to elicit from the user. An Active Capture module is made of two interdependent components: the interaction script and the action recognizers. The interaction script together with input from the action recognizers allows the computer and user to work together to achieve the desired action. The SIMS Faces application has three Active Capture modules, in the first the user and computer work together to get the user to stand in front of the camera and look at the camera. In the second they work together to get the user to say her name, the third they work together to get the user to look at the camera and smile.

3.1 Interaction Script

The interaction script describes how to work with the user to achieve the common goal. It is a flow chart that describes what the system says or does to elicit the desired action from the user and what to do when something goes wrong. There is much that can go wrong in the interaction that the interaction script must be able to handle. Not only might the system's recognizers make incorrect inferences, the participant may misunderstand the system's instruction or give performances that do not meet the programmed requirements. As a result, the system must adopt strategies for directing the user and provide appropriate feedback to shape the desired performance. For example, when the user is getting her picture taken, she may be partially out of frame. The system asks her to move so she is in the frame: "I don't entirely see you, perhaps sitting down or standing on a stool might help. Now smile." This corrective interaction technique is known as *mediation* [MHA00]. Mediation techniques are used to resolve ambiguity in systems where there exists an apparent discrepancy between the system's model of the state of the world, in this case, the physical position of the user, and the state of the world. Table 1 lays out an abbreviated version of the interaction script for the SIMS Faces application. The *path of least mediation* through the interaction script is made up of the prompts and closing statements and describes how to get the user to perform the common goal. The mediation instructions describe what to do when something goes wrong. The action recognizer analyzes and recognizes human actions using simple multimedia parsers (e.g. eye finding, gross motion analysis, sound detection) and the context of interaction including expectations about what the user will do.

3.1 Interaction Script 3 ACTIVE CAPTURE APPLICATION COMPONENTS

Welcome

Prompt *Welcome to SIMS and the SIMS Faces application developed by Garage Cinema Research.*

Position

Prompt *Please stand on the white marks on the floor and look at the camera.*

Mediation

While ... Say ...

Can't see guest *Hello? I can't see you. Please make sure you are standing on the white marks on the floor and that you are looking at the camera.*

Guest not framed *Hmm, I can't see all of you. Please be sure you are standing on the white marks on the floor and look at the camera.*

Can't find eyes *I can't see your eyes. Please make sure you are facing the camera so that your name will be recorded clearly.*

Closing *That's great. Next we are going to record your name so people will know how to pronounce it.*

Name

Prompt *Please look at the camera and state your full name now.*

Mediation

While ... Say ...

Didn't hear anything *I didn't hear anything. I'd like you say your first name and your family name. Now please state your name.*

Utterance too short *Wow, that's pretty short for a name. Just in case, let's rerecord your name. Please be sure to state your first and last name.*

Utterance too long *I heard you say something, but it sounded too long for a name. Let's try again. Please say your full name, that is your first and last name now.*

Closing *Thanks for saying your name. Now we are going to take your picture.*

Picture

Prompt *Please stand on the white marks on the floor and look at the camera. Smile.*

Mediation

While ... Say ...

Not standing still *Please stand still while I take our picture. OK, smile.*

Can't see eyes *I can't see your eyes. Perhaps you are wearing glasses or a hat. Please remove them and look at the camera. Smile.*

Can't find smile *Your picture will be nicer if you smile. Please look at the camera and smile.*

Closing *That was really great.*

Thanks

Prompt *Thank you for using the SIMS Faces application developed by Garage Cinema Research.*

Table 1: SIMS Faces application abbreviated interaction script. See Appendix A for the complete interaction script

3.2 Action Recognizers

The action recognizer is driven and shaped by the interaction script. In the context of the interaction, the input from the parsers can be interpreted much differently than if used alone. For example, the likelihood that the motion detected in the user's mouth after she is asked to smile is a smile is very high. An action recognizer for the desired action, in this case the user smiling, can be defined as the automatic recognition of a mouth motion by the multimedia parser after a trigger in an instruction (e.g. "smile"). [Figure 3](#) shows the relationship between the interaction script and the action recognizer in a simplified version of the picture-taking module part of the SIMS Faces application.

4 Challenge

The challenges in designing and implementing Active Capture applications come from the diverse team, the complex system, the interaction with the user, and the available multimedia parsers.

The design of Active Capture applications draws from the areas of direction and cinematography, computer vision and audition, and human-computer interaction and results in a diverse design team. The diverse team presents extra challenges in communication among the team members. Computer scientists may be able to work on the application together using control flow diagrams or finite state machine diagrams, but if the only artifacts, representations, and discourses used in the design process are those of computer science then the film and theater team members will not be able to contribute to the design. The team needs a descrip-

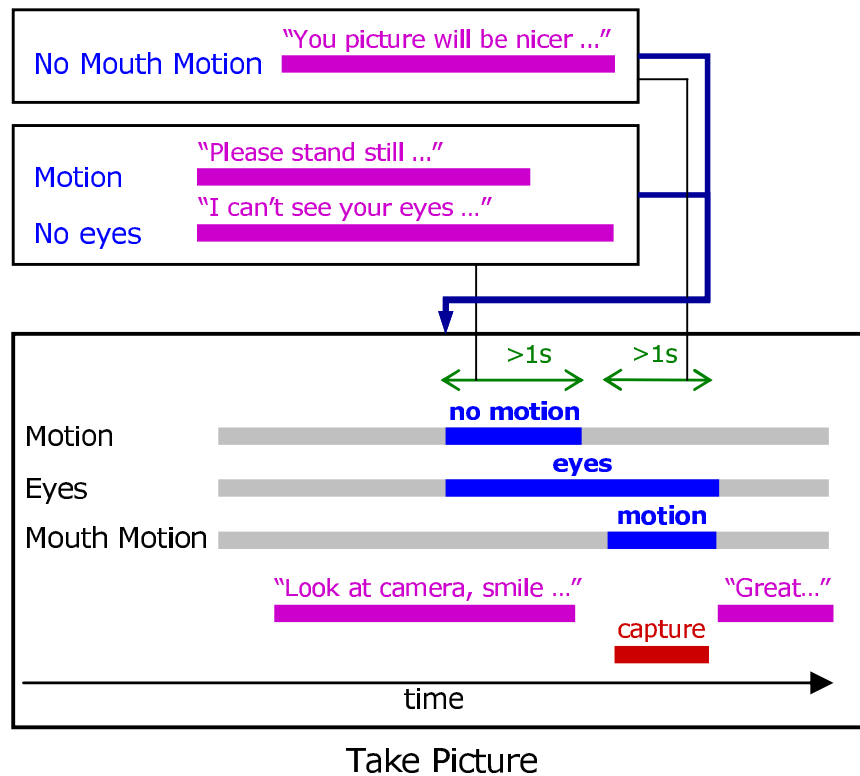


Figure 3: The Picture module from the SIMS Faces Application with mediation. The small timestrips above the main strip describe what to do if something goes wrong, if one of the time constraints are violated.

tion of the application that all members of the team understand and can contribute to. The representation must encapsulate the control flow of the interaction script along with the description of the action recognizer and the relation between the two. The interdependence between the interaction script and action recognizer is what allows Active Capture applications to bring together human-computer interaction, direction / cinematography and computer vision / audition. Since this interdependence is so important in Active Capture applications, it is very important the whole team understand how they interact in the application.

The interaction script provides context for the action recognizer, and depends on structure of the action recognizer. This interdependence between the design of the interaction script and the action recognizer means the team needs to be able to iterate many times on the design of the application, requiring a rapid iteration cycle.

The interaction with the user presents its own set of challenges. If the user is going to work with the system, and be directed by the system, the interaction with the system must be of high enough quality for the user not to mind using the system. The interaction script must describe how to work with the user, most importantly, what to do in the cases where mediation is necessary.

The parsers that are applicable to a given Active Capture application can be limited by the constraint to run in real-time, or by the sensors available to provide data to the parsers. In the SIMS Faces system, the sensors include a microphone and video camera. The simple multimedia parsers include a gross motion detector, an eye finder and a mouth motion detector. The action recognizer uses context from the interaction script and data from the multimedia parsers to recognize the

desired action. The team must figure out what context from the interaction script would be feasible and useful in the design of the action recognizer and design the interaction script accordingly. Based on the context from the interaction script and the data from the multimedia parsers, the team can design the action recognizer.

5 Methodology

In this report I present an Active Capture design process, a visual language for Active Capture applications and a tool prototype for designing the action recognizer part of Active Capture applications. The three contributions are based on related work and experience implementing a working Active Capture application, the SIMS Faces application.

The design process and visual language are work toward an integrated tool to support the Active Capture design team. Before we began work to understand how to support the design team, we did an ethnographic study to understand how to design the interaction with the user [HGR⁺04]. While this work is not presented as part of this report, I will summarize the resulting design space and design strategies at the end of this section.

In order to understand how to support an Active Capture design team, we first looked at existing Active Capture applications, and then implemented the SIMS Faces application following the design strategies from the ethnographic study.

We looked at the head turn and scream Active Capture modules in the Kiosk Demo. The scream module asks the user to look at the camera and SCREAM! It ensures the scream is long enough and loud enough. The head turn module asks

the user to look away from the camera and then to turn her head to look at the camera. It uses eye detection, gross motion, and head motion and ensures the turn is slow enough, begins with the user looking away from the camera, and ends with the user looking at the camera. [Figure 4](#) shows the action recognizer and the interaction script for the head turn Active Capture module.

5.1 Ethnographic Study

Before we began designing the SIMS Faces application, we did an ethnographic study to better understand the design space of the interaction with the user in Active Capture applications. In an effort to inform the design of Active Capture applications and design patterns for use in computer-human interaction, we conducted a series of contextual interviews with human-human interaction experts [[HGR⁺04](#)].

The people interviewed included two film and theater directors, a children's portrait photographer, a golf instructor, an aikido instructor, a 911 emergency operator, and a telephone triage nurse. These interviews revealed successful direction and mediation techniques used by experts in human-human interaction under different circumstances. For example, the 911 operator is an expert in communicating and getting feedback over a limited sensory channel connection (the phone). Our interviews uncovered numerous strategies employed by experts to guide specific human actions including different design strategies, direction and feedback strategies, and mediation strategies. We used many of these strategies in the design of the SIMS Faces application which influenced the design process and ACAL. In the

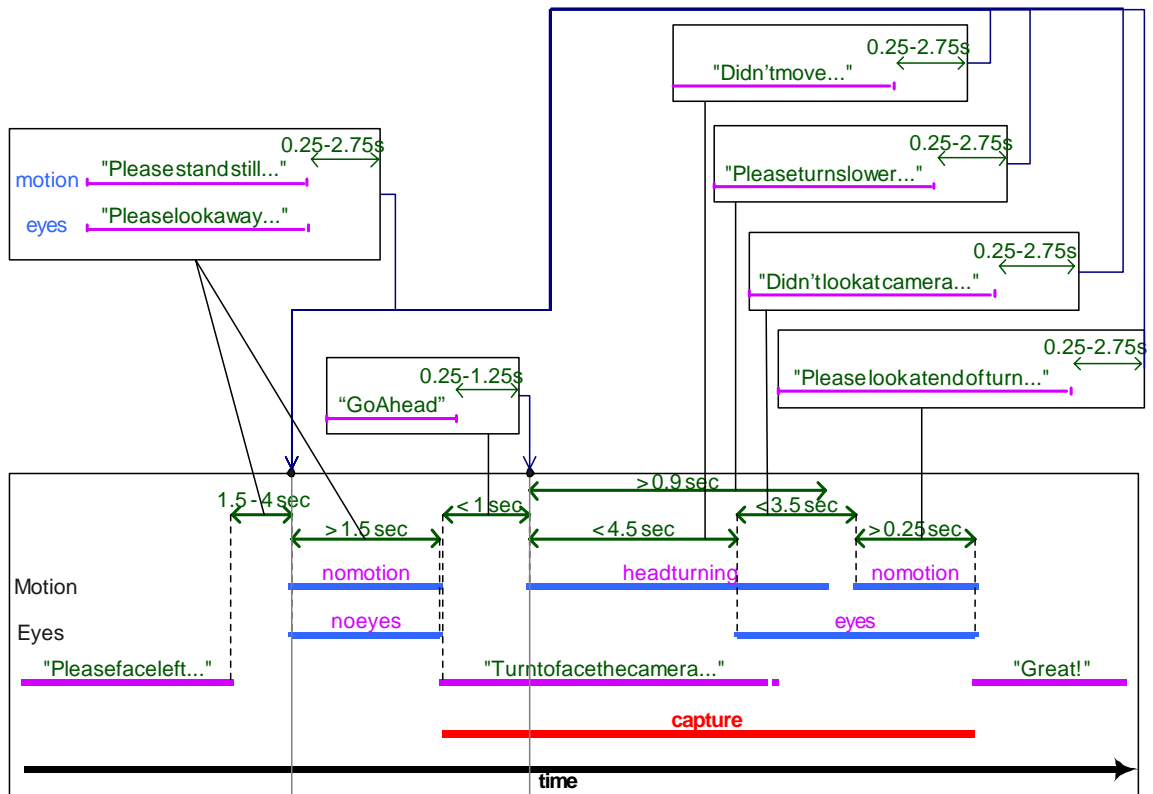


Figure 4: The head turn Active Capture module. In the head turn module, the user is asked to look away from the camera and stand still. Next she is asked to turn her head to look at the camera. The system ensures the head turn is long enough and not too fast, and ends with the user looking at the camera and not moving. See Figure 2 for images of a user at the beginning and end of the head turn. The interaction script (the pink text in quotes) is abbreviated to fit in the figure.

design process, the bodystorming session focuses on figuring out how to direct the user, what could go wrong, and what to do when something goes wrong. The design team uses the strategies summarized in this section as they design the interaction script. ACAL provides a structure for some of the mediation strategies summarized below. Currently the design process and ACAL only support a few of the strategies. In the future we plan to develop an Active Capture design tool that will support more of these strategies, if not all of them.

5.1.1 Design Strategies

Strategy	Use in SIMS Faces Application
<i>Anticipation:</i> Anticipate common errors before they happen. Actively seek out these problems and address them before they disrupt the interaction.	We position the user in front of the camera (the microphone is on top of the camera) before asking her to say her name to make sure she is a good distance from the microphone and is facing it.
<i>Appropriate System Impression:</i> Adopt the appropriate tone and role for the context of the interaction. Urgent tasks may necessitate a curt and insistent style, while recreational contexts afford a more relaxed and flexible tone.	The application uses a professional tone as it is designed to be used with new students in the department.

5.1.2 Direction and Feedback Strategies

Strategy	Use in SIMS Faces Application
<p><i>Decomposition:</i> Break down complex actions into a series of simpler sub-actions. Decompose troublesome actions into sub-parts, identify those parts that are causing difficulties, and address those explicitly before re-attempting the larger, composite action.</p>	<p>First we ask the user to get in front of the camera, once we know she is in front of the camera and facing the camera, we ask her to say her name instead of asking her to stand in front of the camera, face the camera and say her name in one instruction.</p>
<p><i>Imaginative Engagement:</i> Immerse the subject in the experience by engaging their emotions or imagination. Internal motivation is one avenue for accomplishing this.</p>	<p>The system does not make use of this strategy.</p>
<p><i>External Aids:</i> Use physical props or other external aids to guide human actions and provide implicit feedback. Examples include lines on a golf putter to assist aiming and footmarks on a floor indicating where to stand.</p>	<p>If the user is not framed, we suggest she try sitting on a chair if she is too tall to be entirely in the frame.</p>

Confirmation: Explicitly query the subject to ensure they are in the expected state. For complex tasks it may be desirable to have the subject verbalize their understanding of the task. Even if a system cannot parse these utterances, the practice can still aid the subjects learning and memory.

We did not make use of this strategy.

Consequences: Explain the consequences, both positive and negative, of particular actions. This provides more concrete incentives to maintain beneficial actions and correct detrimental behavior.

The system explains why it wants to record the user saying her name and explains if she smiles for her picture, her picture will look nicer on the department web page.

5.1.3 Mediation Strategies

Strategy

Use in SIMS Faces Application

<i>Freshness:</i> Avoid repeating utterances, even when giving an instruction nearly identical to a previous one. Maintain consistent vocabulary, but don't repeat items verbatim. At a bare minimum, designers should make multiple versions of instructions that may be repeated.	The system has a few variations on instructions that are used more than once.
<i>Progressive Assistance:</i> Address repeated problems with increasingly targeted feedback. Provide "successively more informative error messages which consider the probable context of the misunderstanding".	The system makes use of this strategy in many cases. If it cannot find the user's eyes, it suggests the user take off her glasses or her hat.
<i>Method Shifts:</i> When one form of instruction fails, try another. Direction can vary between telling the subject what to do, showing them how to do it, or making them do it.	When the system cannot find the user's smile after asking her to smile, the system shifts methods and tried to make her smile by telling her a joke.

Modality Shifts: When a particular direction approach repeatedly fails, switch or augment the modalities of communication, e.g., use visual rather than auditory cues. Changing or using multiple modalities may prove more effective to a larger audience due to the different aptitudes of auditory, visual, and kinesthetic learners.

Our implementation only makes use of auditory cues, so it does not make use of this strategy.

Level of Discourse: Simplify the vocabulary and language structure when people are having difficulty understanding. Conversely, be concise once grounding is established.

The system does not make use of this strategy.

Backtracking: When grounding is lost, backtrack to the last state of mutual understanding. By returning to a state of common ground, the system and user can then again proceed towards the goal.

The system does not make use of this strategy.

Graceful Failure: When all else fails, provide the subject natural exits from the interaction. Recognize over-repetition and respond by pursuing an alternate goal or allowing the interaction to proceed to completion, despite the error.

After a few attempts in each of the modules, the system fails gracefully and moves to the next module, letting the user know it is not necessarily her fault and she can come back and try again another time.

6 Active Capture Design Process

The Active Capture design process follows the design-test-analysis methodology from SUEDE [SKL02] with the use of bodystorming [OKK03], wizard-of-Oz user studies, traditional user studies, and an iterative design cycle (See Figure 5). The design of the interaction script and the action recognizers are interleaved in the design process as they are interdependent in the application.

6.1 Bodystorming

The design process begins with the desired action in mind and a bodystorming session to inform the first draft of the interaction script. Bodystorming is the technique of acting out full body contextual interactions. It is similar to paper prototyping as it allows the designers to rapidly debug a full body interaction using a low fidelity medium. It allows for rapid low cost iteration at the beginning of the design process. With the desired scenario in mind, the design team acts out different variations of the interaction script and various reactions to the script (what could go wrong in the interaction). In addition, the sample interactions are recorded,

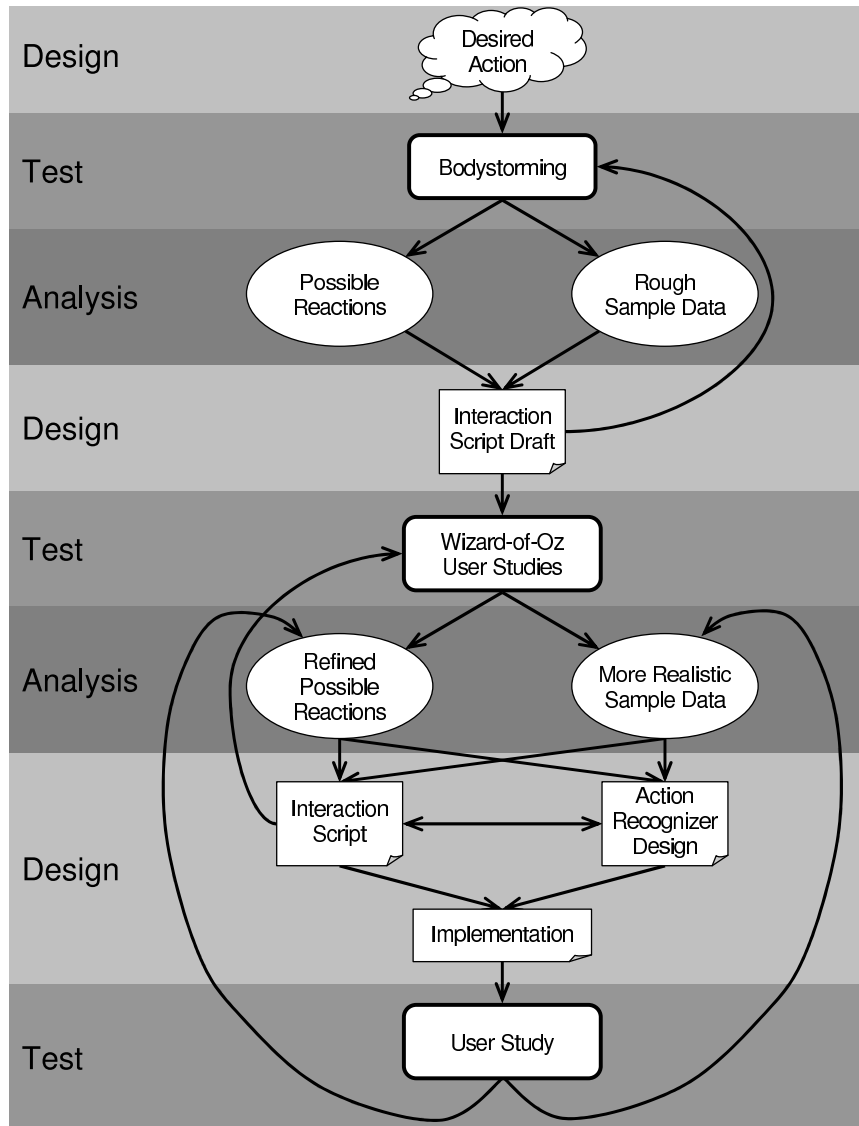


Figure 5: Design Process

providing sample data for use in the design of the parsers and interaction script. In the case of the SIMS Faces application, the bodystorming session raised and attempted to answer the following questions: Suppose we want to take the user's picture, how will we get her to stand in front of the camera? What if she is moving too much to take her picture? What if she is not framed properly? What if her eyes are closed? What if she doesn't smile? These questions show where mediation is needed in the interaction script and should be answered by the interaction script.

6.2 Wizard-of-Oz User Study

Once the design team has a draft of the interaction script, the team can run a wizard-of-Oz study. In order to run the wizard-of-Oz study, the team must digitize the instructions and triggers described in the interaction script. In the wizard-of-Oz study, the computer plays the clips and records the data, but the human wizard decides when to play each clip. Since humans react differently to computers than they do to other humans, the wizard-of-Oz study is important to test user's reactions to the interaction script when it is coming from the computer. It simulates the human-computer interaction bodystorming cannot simulate because in bodystorming the interaction is between humans. For example, during the bodystorming sessions for the SIMS Faces application, many of the people getting their picture taken were smiling through the whole process, so asking them to smile was not necessary. During the wizard-of-Oz study, none of the people getting their pictures taken were smiling through the whole process, so it was necessary to ask the user to smile and to check for it.

In addition to testing the flow of the interaction, the wizard-of-Oz study tests and reveals the triggers in the interaction script (the words or phrases that make the user react). The interaction script is designed with triggers, some may work well, others may not result in the desired reaction and there may be still others that weren't intended as triggers. For example, in the SIMS Faces application, the system offers to tell the user a joke to get her to smile after a few failed attempts to smile. "Let me tell you a joke. A guy walked into a bar, ow!" We expected the "ow" to be a trigger for a smile, but it turns out "Let me tell you a joke" also turned out to trigger a smile.

The interaction script can *tell* the user what to do, *show* the user what to do or *make* the user perform the desired action. In any case, the interaction script must include triggers so the user knows when to perform the action. In a *tell* instruction, the trigger might have emphasis on it, to indicate the instruction is over and the user should perform the action. In the SIMS Faces application, the system asks the user to smile: "Please look at the camera and *smile*." In this example, the word "smile" is emphasized. When the system can not see the user smiling, it tries to *make* the user smile by telling a joke. Here, we expected the trigger to be the punchline in the joke, which was the case, but there turned out to be two jokes, "Let me tell you joke" and "A guy walked into a bar, ow!"

The data collected in the wizard-of-Oz study provides realistic examples with close to realistic timing details of the interaction and resulting actions. This data is crucial for the design of the action recognizer. The refined set of possible actions and realistic sample data allow the designers to iterate on the script and design the recognizer for the desired action, the action recognizer. Currently, the data is

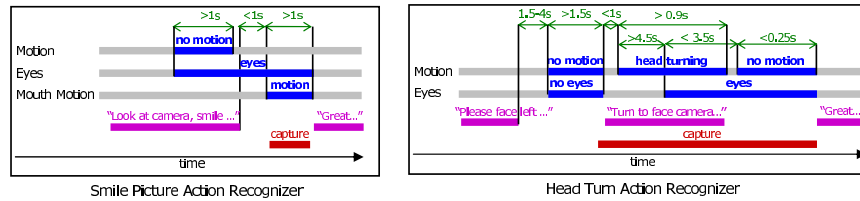


Figure 6: The smile picture and head turn action recognizers.

collected manually, and analyzed with the help of the action recognizer tool.

6.3 Designing Action Recognizers

An action recognizer defines the desired human response in terms of the multimedia parsers in the context of the interaction script. The action recognizer is the high-context computer vision recognizer. The multimedia parsers alone could be used to make a computer vision recognizer, but with the context from the interaction script, a high-context computer vision recognizer can be created. A computer vision algorithm for a smile recognizer might look at mouth motion or mouth shape. A high-context computer vision recognizer on the other hand, not only uses the mouth motion multimedia parser, but also the context from the interaction, when the guest is asked to smile. With context from the interaction, the computer vision recognizer is more sophisticated, or can use a simpler multimedia parser. Figure 6 shows the action recognizer for the smile recognizer and the action recognizer for the head turn.

At this point in the design process, the team has the interaction script and data from the wizard-of-Oz study. The wizard-of-Oz data contains useful examples of the action in terms of the multimedia parsers and their relation to the triggers in

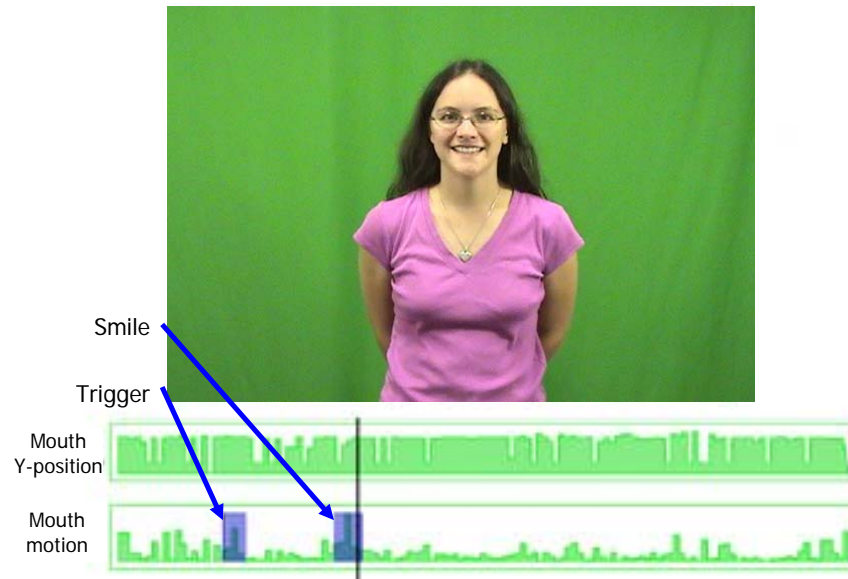


Figure 7: The user’s smile corresponds with a peak in mouth motion after the trigger to smile.

the interaction script. The design team uses this data to figure out how to make use of context from the interaction script, and the multimedia parsers to design the action recognizer. We developed a tool to help us sort through the wizard-of-Oz data and design the smile action recognizer, which is described in Chapter ???. We noticed if we just looked at the mouth motion data, the peaks in mouth motion did not correspond to the user smiling. But a peak in the mouth motion after one of the triggers to smile did correspond with the user smiling. [Figure 7](#) shows an example of how a peak in mouth motion after the user is asked to smile corresponds to her smiling.

7 ACAL – Active Capture Authoring Language

In order to make use of the varied experience on our design team, we needed a way to describe the application so the whole team could understand it. To this end, we developed a visual authoring language. The visual authoring language had to help all members of the design team understand and iterate on the interaction script, the action recognizer and the interdependence between the two. It had to be able to express the control flow details from the interaction script and the timing details from the action recognizer together. Active Capture interactions appear natural and intuitive to the user, but involve considerable complexity in the system's program for dealing with the wide variety of possible states and transitions in the interaction script. As such, the Active Capture design process requires representations that can help Active Capture designers manage the complexity of the interaction script and the action recognizer in the design process, especially on multidisciplinary design teams. ACAL does not have an authoring tool yet, but this is part of the future work. We (Ka-Ping Yee and I) began by looking at existing visual scripting languages, the three most relevant languages are state transition diagrams, statecharts [Har87] and hypermedia authoring systems. We designed ACAL based on the strengths of these languages and feedback from the SIMS Faces design team.

7.1 State Transition Diagrams

State transition diagrams are made up of states connected by transitions. They have the advantage that they are standard and many people understand how they

work and how to use them, but they are:

- bad at handling time
- have no way of expressing concurrent actions
- are a flat description
- in practice need extra variables to keep track of some state.

[Figure 8](#) shows the head turn module as a state transition diagram. Many of the self-looping edges in the graph are used to keep track of the passage of time. There are quite a few extra variables that keep track of state related to mediation. For example, one variable keeps track of how many times the actor has attempted the performance, in order to prevent the actor from having to repeat the loop forever without ever succeeding. Two observations cannot be easily monitored at the same time: first, the system checks to make sure the actor is standing still, then it checks to make sure the actor is looking at the camera. While it is checking to make sure the actor is looking at the camera, the program assumes that the actor is still standing still, but it has no way of monitoring these observations concurrently. While state transition diagrams are standard and many people already understand how to read them, they are very tedious to create and once created are difficult to reason about. When looking at a state transition diagram one of the most difficult things to reason about is the passage of time.

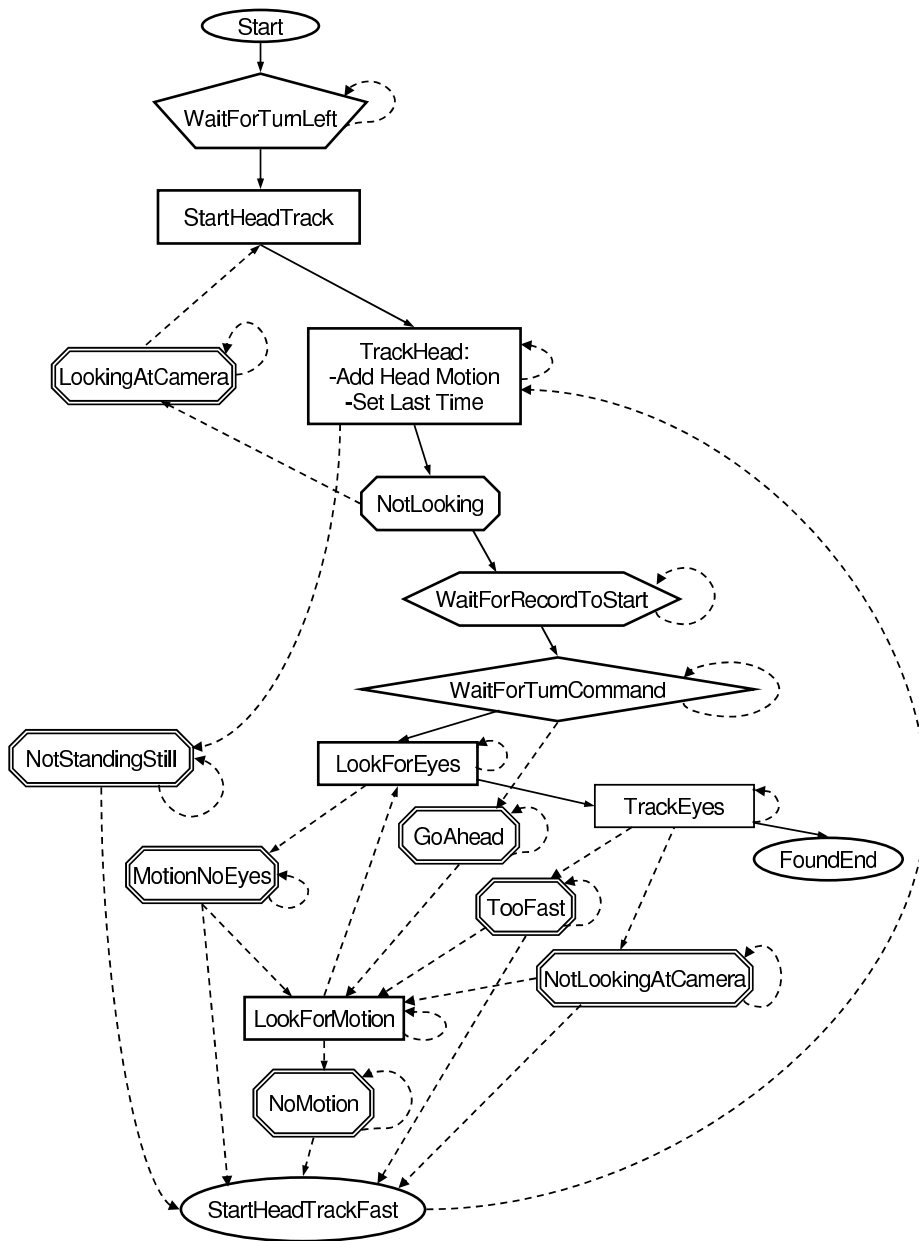


Figure 8: Headturn state transition diagram. pentagon : Wait for a command to play / Play commands in sequence hexagon : Waiting for capture to start hexagon : Wait for command to play and mediate diamond : Wait for participant square : Wait for and observe participant circle : Get input from participant (observer participant) *Solid Edge*: Interaction path with no mediation (no errors). *Dashed Edge*: Interaction path with mediation.

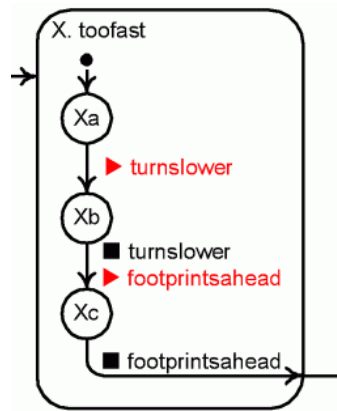


Figure 9: Example of hierarchy in statecharts.

7.2 Statecharts

Statecharts [Har87] are similar to state transition diagrams but they include various additional notational features to express hierarchy, orthogonal combinations, and timeouts. Hierarchy is expressed by drawing statecharts within individual states. In Figure 9 the steps to take when the actor turned her head too fast are encapsulated in one statechart node with more specific statechart nodes inside. This helps in reasoning about a statechart.

The ability to express orthogonal combinations allows for simpler diagrams as compared to state transition diagrams. Orthogonal combination constructs support concurrent events. In Figure 10, the duration of the turn and whether the eyes can be seen are monitored at the same time. The transitions out of the state “R.turning” on the bottom right part depend on which state you are in on the left side of the dotted line and the right side.

The timeouts in statecharts allow some timing details to be represented in the diagram, eliminating the necessity for the self loops that were necessary in state

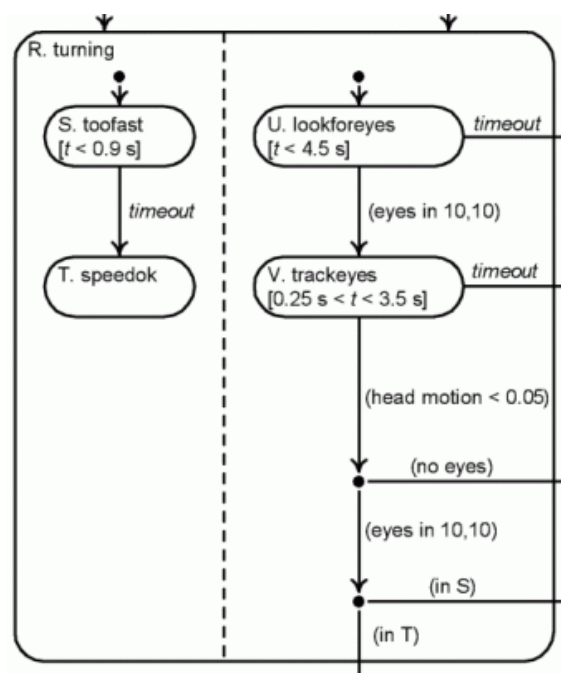


Figure 10: Example of orthogonal combinations in statecharts.

transition diagrams. See [Figure 11](#) for a statechart of the whole head turn module.

7.3 Hypermedia Authoring Systems

Hypermedia authoring systems make it easy to create an interactive program using a GUI. We looked at an example of such a system called Authorware [Mac] from Macromedia. It facilitates making a simple program, but does not support representing and manipulating temporal constraints or variables. We implemented a simplified version of the head turn example in Authorware. It was relatively easy to use to create the example and provides constructs to organize programs in a hierarchical structure similar to statecharts, but the hierarchy is not always optional. This causes an explosion of windows while editing or trying to debug an application making reasoning about an existing application or debugging an application very tedious. Figure 14 shows the path of least mediation. In order to see what happens in the first step, three windows must be opened (Figure 14). Authorware provided a good example for how a debugger for an ACAL program might look, except for the lack of the ability to express the flow of time.

7.4 ACAL Design History

As we looked at the head turn example and explored its representation in state transition diagrams, statecharts, Authorware, as well as visual step charts, and timeliness, we determined that a timestrip-based visual language would offer the best mix of expressibility and simplicity for representing timing, control flow, and hierarchy in the design of Active Capture interaction scripts. We began by creating

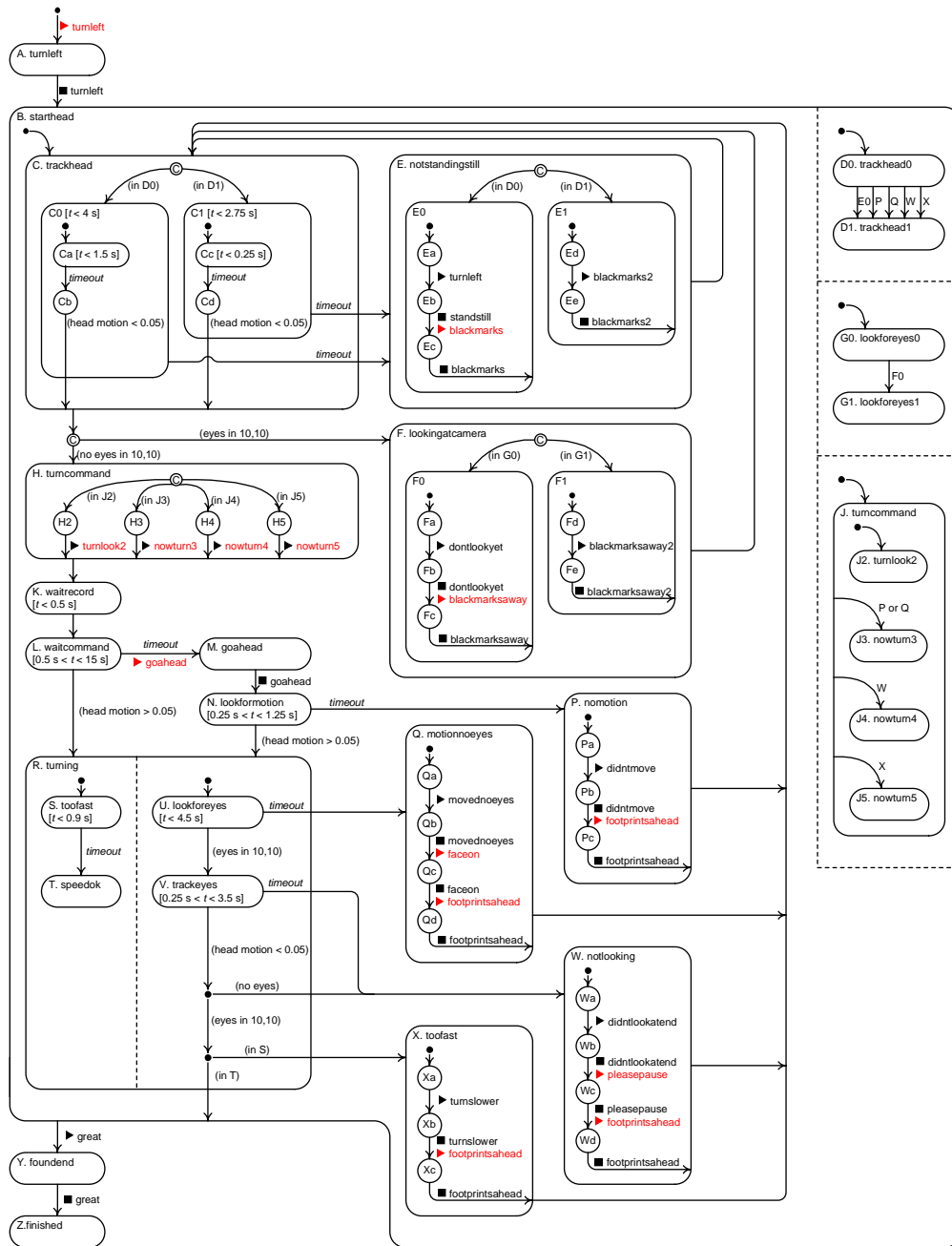


Figure 11: Headturn Active Capture module as a statechart.

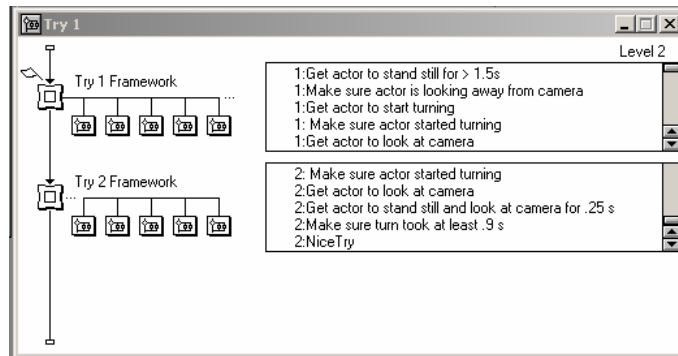


Figure 12: Path of least mediation in headturn module in Authorware.

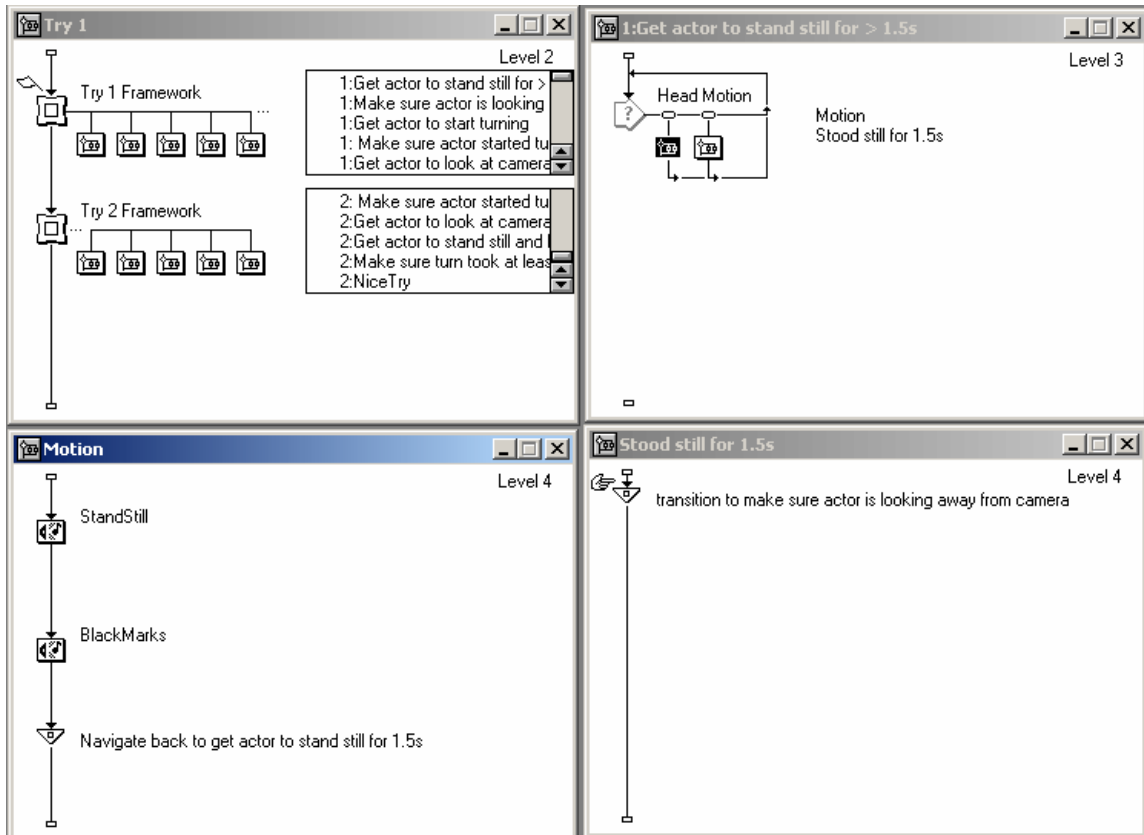


Figure 13: Windows necessary to see what happens in the first step in our Authorware implementation of the head turn example.

a state transition diagram for the head turn module. The state transition diagram was extremely complicated and motivated us to design a simpler, better suited language for the task of describing Active Capture applications. After the state transition diagram, we tried statecharts, devised visual step charts, time lines, and finally arrived at a hybrid visual design combining time lines with control-flow arrows.

In an effort to tame state transition diagrams and statecharts, we devised step charts, which serialize the set of conditions necessary to achieve the desired result. We decided to try a more constrained, structured representation in the hope that it would simplify writing and visualizing Active Capture interaction scripts. Step charts express the path of least mediation as a sequence of steps, where each step has four parts:

1. The stimulus to motivate the desired result for the step
2. The desired result for the step
3. The success condition
4. The failure condition (possibly a timeout condition)

In the headturn module, the first step has the following parts:

1. Stimulus: "Please turn to your left, stand on the black marks on the floor and look straight ahead."
2. Desired Result: Get user to stand still for at least 1.5 seconds.
3. Success Condition: No head motion for at least 1.5 seconds.

4. Failure Condition: Time out and use mediation techniques after waiting 4.5 seconds.

See [Figure 14](#) for the headturn module as a step chart.

After doing an informal evaluation of step charts with one of the theater majors on the team, we discovered that the level of abstraction in step charts was not high enough for him to be able to design an Active Capture module. The level of abstraction is higher than that in state machines, but does not provide enough structure to minimize simple errors. In our informal interview, we asked the directing major to describe the conditions he would need to check in order to see if an actor was running in place. He came up with the following steps:

Make sure the actor is in the frame.

Make sure the actor is standing still.

Make sure the actor is looking at the camera.

Ask the actor to start running in place. (Get the actor to run in place).

He forgot to make sure the actor was looking at the camera after running in place (to make sure she did not run in place and turn at the same time). He also forgot to make sure the actor was still in the frame. These mistakes are simple enough that the language should be able to catch them, or simply not allow them to be made. This observation led us to our timestrip design. The timestrip representation allows the ACAL designer to express the different observations that must be true in order for the actor to perform the desired action while minimizing simple mistakes such as remembering to check to make sure the actor is still looking at the camera at the end of an action if she was supposed to be looking at the

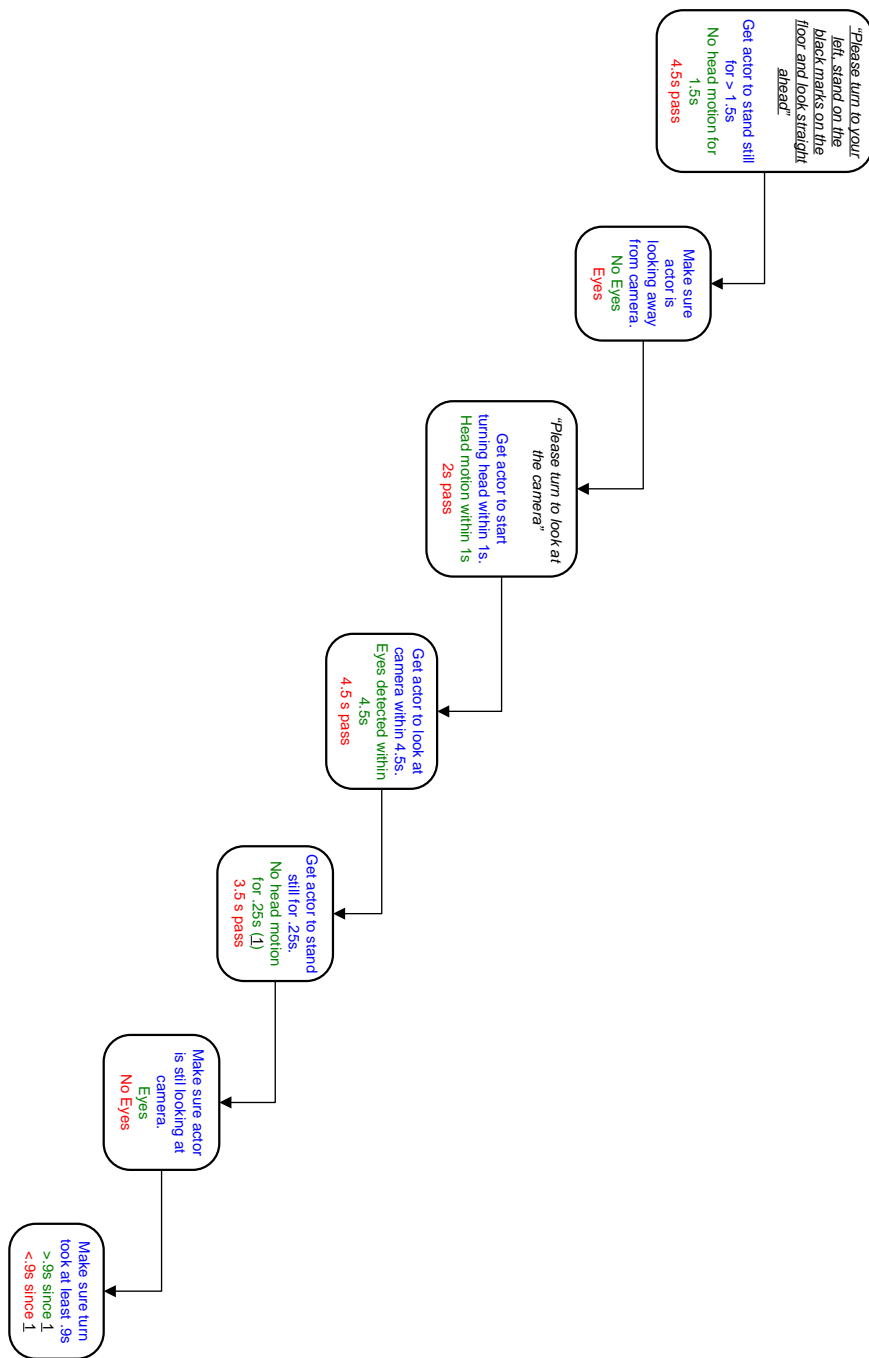


Figure 14: Step chart for head turn example.

camera during the whole action. Timestrips provide a natural way of expressing concurrent events, temporal relationships, and time constraints. Control flow on the other hand is not as naturally expressed. This drawback lead us to a hybrid design between timestrips and control flow graphs.

Finally, we tried to devise a visual notation that would combine the most useful properties of the languages we surveyed and attempted to design. Our hybrid timestrip design incorporates the concept of multiple levels of detail, as inspired by statecharts, the control flow notation from state transition diagrams, the clear path of least mediation from the simple timestrip and from step charts, and concurrency, temporal relationships, and time constraints from timestrips. The hybrid timestrip provides support for mediation and the design strategies for the interaction script from the contextual interviews.

ACAL depicts an Active Capture module as a set of rectangular timestrips connected by arrows representing jumps. The primary timestrips describes the path of least mediation and the action recognizers. The simpler time strips express the mediation steps. There is one primary timestrip for each Active Capture module. In the SIMS Faces application, there are three primary timestrips, one for the position module, one for the picture module and one for the name module. [Figure 15](#) shows the three primary timestrips for the SIMS Faces application. [Figure 16](#) shows the three modules in the SIMS Faces application with some mediation cases.

A timestrip contains consists of several tracks: one for each observable feature or multimedia parser, one for stimuli, and one for capture control. Segments reside on feature tracks indicating a requirement that the feature be true or false. The arrangement of segments on a time line expresses temporal ordering among the

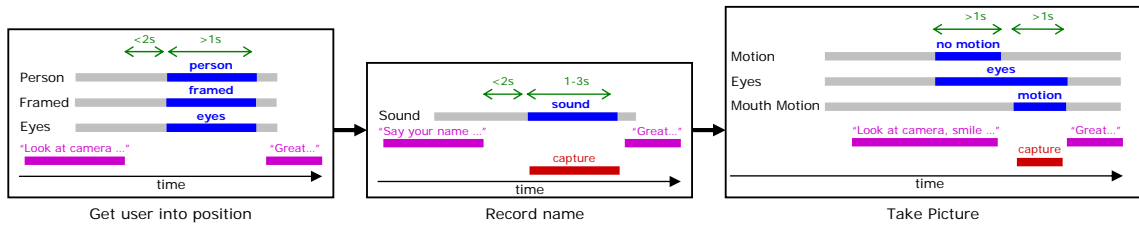


Figure 15: Path of least mediation in the SIMS Faces application. The SIMS Faces application has three Active Capture modules, one to get the user into position in front of the camera, one to record the user’s name and one to take the user’s picture.

segments (in any of Allen’s 13 temporal relationships [All83]); segment triggers are then inferred from the ordering. Time constraints may be added among the anchors on the observation segment to indicate how long to wait before mediating or how long a condition is required to remain true before mediating. The segments on the stimulus track and the capture track describe when to play a stimulus and which parts should be captured. Figure 17 shows the path of least mediation in the picture module of the SIMS Faces application with the different parts from ACAL marked.

The simpler timestrips hanging off of the primary timestrip on the time constraints in Figure 18 express the mediation steps that take place when a time constraint is violated. Figure 16 shows the three primary timestrips in the SIMS Faces application with some of the mediation cases hanging off. The figure is simplified to fit on the page.

As we designed the SIMS Faces application we used ACAL to describe the evolving design. First we laid out the path of least mediation with primary time strips so the whole team would understand what we were trying to get the user

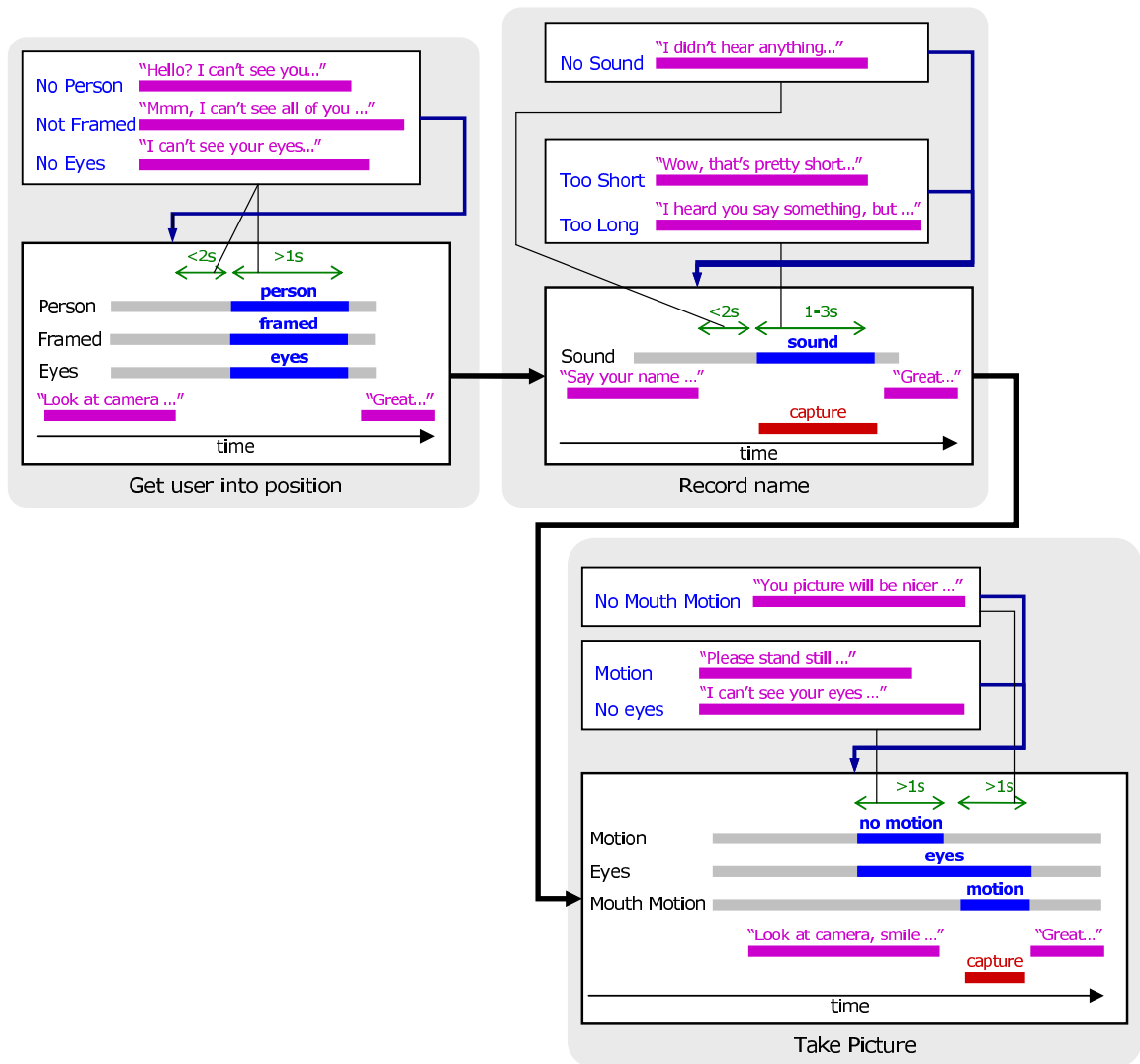


Figure 16: A simplified version of the SIMS Faces application. The text in quotes describes the interaction script (above pink strips). The set of (blue) segments on the (grey) parser strips describe the smile recognizer.

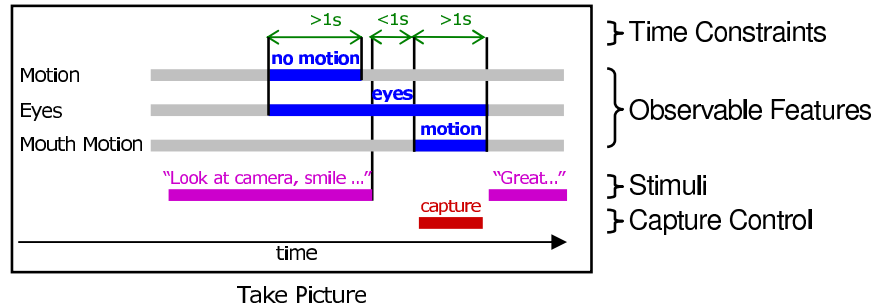


Figure 17: Path of least mediation in the picture Active Capture module in the SIMS Faces application.

to do. As we did bodystorming sessions, we added mediation cases. They were tied to time constraints, but the numbers on the time constraints weren't fill in yet. After the wizard-of-Oz study, we were able to fill in the timing details based on the wizard-of-Oz timing data. Note, the wizard-of-Oz part of the design process is not supported with a tool yet. The data is collected from the data streams manually.

8 Related Work

In this section, I divide the related work into the following topics of this work: Active Capture, the Active Capture design process, ACAL – the visual language for Active Capture applications, and the SIMS Faces system – the Active Capture case study presented in this report.

8.1 Active Capture

Active Capture brings together three actions – capture, interaction and processing. Each of the three pairwise combinations of these actions make up existing and

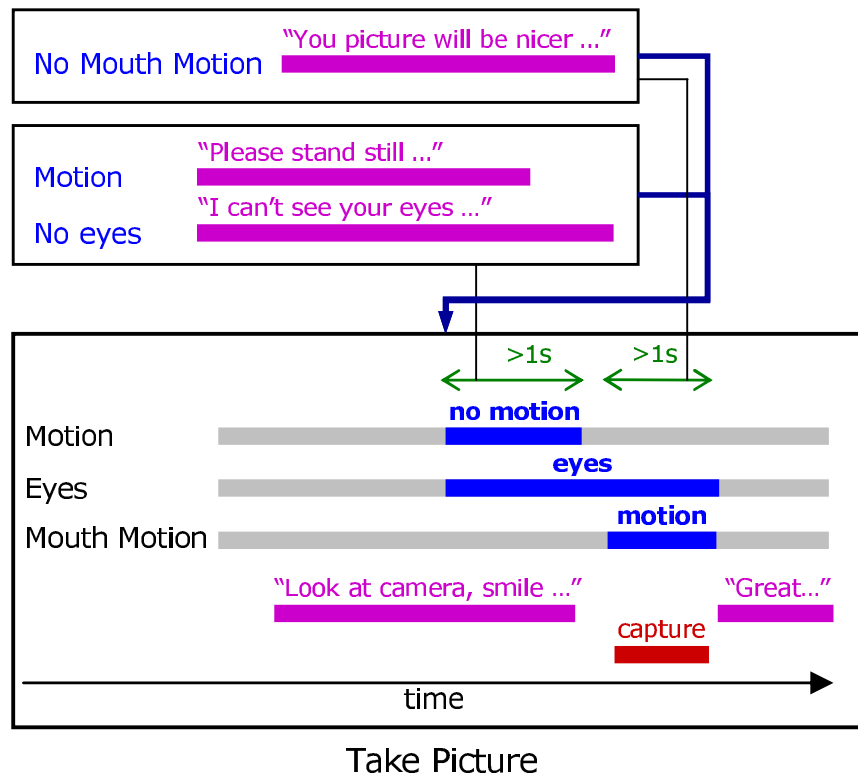


Figure 18: The Picture module from the SIMS Faces Application with mediation. The small timestrips above the main strip describe what to do if something goes wrong, if one of the time constraints are violated.

well-established fields of study. Capture and processing come together in computer vision and audition, capture and interaction come together in the field of direction and cinematography, while interaction and processing come together in human-computer interaction. Each of these three fields have a long and independent history, each with their own challenges and limitations.

Computer Vision and Audition. In computer vision and audition, one of the main limitations is called the *semantic gap* [DV01]. The semantic gap describes the gap between the shallow content descriptions computer vision algorithms can infer from a picture or video, and the rich meaning users want. Humans want rich semantic descriptions, for example, they want to know a picture has a car in it, not the colors or simple shapes that are in the picture. Computer vision algorithms are generally limited to identifying syntactic descriptions, such as the colors in the picture, or the shapes in the picture using edge or blob detection/segmentation. In order to bridge the semantic gap in computer vision, the algorithm has to leverage context from somewhere. Image search on web pages uses text surrounding images to infer a semantic meaning of the images [FSA96]. Active Capture bridges the semantic gap by using context from the interaction with the user [Dav03b].

Direction and Cinematography. In direction and cinematography, one of the main challenges for the director is to get the actor to perform a scene the way she pictures the scene. Degree programs and a vast array of books are dedicated to this very subject. In Active Capture, the director is distilled into a computer program, and the computer directs the human to perform an action. Such a *director in*

a box is considerably limited compared to a human director. A human director is an expert in communicating with other humans and is excellent at making use of visual and auditory feedback from the actor. A *director in a box* is limited by our ability to understand the psychology of conversation and conversation management [CB91, BH95] (both in general and computationally.) It is also limited by the information about the user the computer can get from real-time computer vision and audition algorithms. This limitation is analogous to removing some of the director's feedback from the user. Consider a telephone triage nurse, she must direct a caller to perform actions such as CPR over the telephone. The only feedback she has about the caller's reactions to her directions is whatever she can hear on the phone. She cannot see how the caller is treating the patient. The challenges in designing an Active Capture application are similar to those a telephone triage nurse faces. In an Active Capture application, the system has limited information about the state of the user using the system, some of the interaction strategies described in Section 5.1 are based in part on interviews with professionals who direct humans to perform actions over the telephone.

Human-Computer Interaction. In traditional interfaces, the human tells the computer what to do by using direct manipulation or interface agents [SM97] or with a mixed initiative interface [Hor99]. Active Capture turns the interaction around and uses expertise from direction and cinematography to have the computer direct the user. The system monitors the user's actions with data from computer vision and audition algorithms which often result in a low recognition rate of the users actions. In human-computer interaction, interfaces that use input devices with low

recognition rates such as speech recognition and handwriting recognition use *mediation* techniques that allow the user to correct recognition errors. Ainsworth and Pratt [AP92] and Baber and Hone [BH93] identify speech recognition errors and introduce and evaluate mediation strategies for resolving them. Yankelovich *et al.* [YLM95] present an advanced speech system with error-correction support. In multimodal interfaces, modality shifts are used to disambiguate recognition [Ovi00, SMW01]. Mankoff *et al.* review past work on mediation techniques [MA99] and provide support for mediation in a GUI toolkit [MHA00] that focuses on interaction techniques supporting choice mediation. For Active Capture applications, mediation strategies are necessary to resolve recognition errors from the use of computer vision and audition algorithms, as well as to resolve ambiguity in the instructions or commands used to direct the user. The results from the ethnographic study of expert in human-human interaction reviewed in Section 5.1 lay out a set of mediation strategies specific for Active Capture applications.

Tellme Networks Inc. [Tel] designs interfaces for 411 phone calls, voice mail and many other phone-based applications. The design of voice interfaces presents similar challenges to the design of Active Capture applications, namely they include mediation cases for when something goes wrong. Just as Active Capture applications use context from the interaction with the user and input from simple multimedia parsers to create more accurate action recognizers, the voice interfaces use the context of the interaction to limit the dictionary used for the speech recognizer, creating a more accurate “response recognizer.” In contrast, however, the Tellme applications do not direct the user, but the user directs the system. This difference eliminates the interdependence between the interaction script and the

action recognizer in Active Capture applications, making the design and development less complex. The interdependence between the interaction script and the action recognizer is one of the main challenges that ACAL and the design process for Active Capture applications address.

8.2 Active Capture Design Process

The Active Capture design process is an iterative process with three main stages – design, test and analysis. It uses bodystorming [OKK03] to test early stages in the design, followed by wizard-of-Oz [DJ93] studies to test intermediate stages, and finally user studies with implemented prototypes to test the final iterations in the design. The Active Capture design process draws from previous design processes, such as the design process for designing speech interfaces supported in Suede [SKL02] and the hypertext design process [NN95]. It also draws from traditional interface design principles. Such as iteration [GL85], prototypes and early sketches [Ret94, BB94], and wizard-of-Oz studies [DJ93], and addresses the challenges unique to directive technologies.

It is well understood that iteration is key to any design process. The faster a design team can iterate on a design, the more iterations it will have the resources to complete. In order to reduce the cost of each iteration in a design process, the team uses prototypes with a level of fidelity that matches the maturity of the design. At the beginning of the design process, the iterations are shortest and the prototypes are lowest in fidelity. As the design process progresses, the length of each iteration increases as does the fidelity of the prototypes. **Similar to the design**

process for hypertext [NN95] and for voice interfaces [SKL02], the Active Capture design process focuses on reducing the cost of iteration at each stage of the design. In the early stages of design, it uses bodystorming, while in intermediate stages of design, it uses wizard-of-Oz studies before implementing the application in the final stages.

Traditional early stage prototyping methods, such as paper prototyping [Ret94], do not directly apply to Active Capture applications because the interaction medium is considerably different, but the same prototyping concepts apply. In Active Capture applications, the medium of interaction is auditory or visual and involves the whole body. Thus, the design process uses bodystorming [OKK03] to prototype early stages in the design. Bodystorming is the technique of acting out full body contextual interactions. It has similar benefits as paper prototypes, as it allows the designers to rapidly debug a full body interaction using a low fidelity medium.

A wizard-of-Oz study allows a design team to test a prototype before it is functional. The wizard simulates the parts of the prototype that are not yet implemented. For voice interfaces, the wizard often reads out clips for parts of the application that are not yet functional. In the wizard-of-Oz studies for user-friendly natural language office information applications [Kel84], the wizard intervenes as necessary to keep the dialog flowing, so only the clips that have already been implemented in the system come from the computer, and those that are still necessary are uttered by the wizard. For Active Capture applications, it is important for the directive clips to be prerecorded because users react differently in front of a computer than they do with another human. For example, we observed this phenomenon first hand. We incorporated in our system a joke to get the user to smile

(“let me tell you a joke, a guy walked into a bar, owe”) expecting the user to smile after the punch line as if a human told it, but surprisingly almost all users reacted immediately upon hearing the computer say “let me tell you a joke.”

8.3 ACAL

ACAL draws from visual languages used to represent time flow, and those used to represent control flow. Time flow refers to the timing details of an application, for example, how long to wait before prompting the user again to please say her name. Control flow is the flow between events or modules in an application based on a set of decision points. More concretely, an informal description of some control flow in the SIMS Faces system is: if the user says her name, tell her “Thanks for saying your name,” otherwise ask her to say her name again.

For describing time flow, time lines or time-series plots are the most frequently used visualization technique, which have been used since the tenth- or possibly eleventh-century [Tuf83]. They have been used in many fields and for many purposes, including visualizing debugging information [Kar94]. The main difference between traditional time lines, and the time-strips in ACAL, is that ACAL uses a non-absolute representation of time in the time-strips. Traditional time lines use an absolute representation for time along one of the axis. The rate of time passing along the axis does not have to be constant, but it is described in absolute terms. For example, the TimeSlider [KSK97] visualization of a large time line uses a non-linear time line to provide context when focusing in on a specific time on the time line. In ACAL, the time-strips organize a sequence of events along a time

line, but each event has a minimum and maximum length instead of an absolute length. This non-absolute representation of time is similar to the use of time lines in Flex [KL91], a language for building flexible real-time systems.

As described in the section on ACAL, we looked at a variety of visualizations of control flow that also support time-flow (state charts [Har87] and Authorware [Mac]), but the balance between the support for time-flow and control-flow did not meet our needs. The resulting visual language ACAL places more of an emphasis on the time-flow, using little time lines or timestrips, and linking the timestrips together to support control flow. Edward *et al.* present visualizations using time lines for multilevel undo and redo [EILM00], and to support autonomous collaboration [EM97]. Similar to ACAL, the time line visualizations use branches in the time line to represent control flow. The data shown on the time line are events, similar to ACAL, however the undo/redo events and document states do not have any time length associated with them. In ACAL, the events need to have minimum and maximum lengths associate with them.

ACAL draws from the visualizations used in some existing applications. The *a CAPpella* application [DHB⁺04] supports the design of recognizers for use in context-aware systems. The system allows the user to collect and annotate data of the situation she wants the system to recognize. The user is presented with the data in streams, one stream for each type of data. The user then selects or annotates which sections of the different data streams are important. The system uses machine learning techniques to build a recognizer based on the data collected. The *a CAPpella* application does not require the user to find the pattern in the data, but it does not allow the designer to see the resulting pattern or modify it. While ma-

chine learning may be very useful in extracting a pattern from the data, an Active Capture interaction designer may want a pattern that is more flexible and interpretable, or allows a larger space of variations on the desired action than demonstrated in the wizard-of-Oz data.

MediaCalc [DL96] allows the user to experiment with different parsers as well as composition of parsers on rich input streams such as video or audio. While its interface allows both visualization of the data flow of the inputs and outputs of media analysis and synthesis algorithms linked to a time line visualizing each step of the processing results, the challenge in ACAL [RD04] is to also visualize and interactively edit the control flow of the computer-human interaction.

Interval Scripts [PB03] present a possible implementation for applications described in ACAL. Interval Scripts were developed to describe interactive environments and computer characters. While the domain is not exactly the same as that of Active Capture applications, the language addresses similar challenges, including temporal events as opposed to instantaneous events. Interval Scripts use a strong temporal algebra, which is used to describe the temporal relationship between events based on Allen's set of temporal relationships [All83]. I plan to adapt Interval Scripts to implement Active Capture applications described using ACAL in an integrated development tool for Active Capture applications.

8.4 SIMS Faces system

The SIMS Faces system follows from the Video Guest Book [THB⁺04] developed at the Fuji-Xerox Palo Alto Laboratory. The Video Guest Book takes the guest's

picture and scans in her business card. This system focuses on data collection but does not focus on the interaction with the user or the quality of the data. The SIMS Faces system takes the Video Guest Book to the next level and focuses on the interaction with the guest and the quality of the data.

The media equation [RN96] describes how humans have a strong tendency to treat computers as humans. They only treat them as machines when they are focusing hard on the fact that it is a machine, but easily fall back to treating them as humans. The SIMS Faces system uses pre-recorded human voice utterances, and as such does not work against the guest's tendency to treat the computer as a human.

9 Evaluation

The Active Capture design process and ACAL are based on the related work described above and our experience in designing the SIMS Faces application. In order to be able to do that, we needed to make sure the SIMS Faces application is a successful Active Capture application. We ran a wizard-of-Oz user study with 10 participants and a user study with the implemented SIMS Faces application with 7 participants.

In our wizard-of-Oz study, participants were lead into a room divided by a green curtain. The mock application was set up on one side and the "wizard" on the other. The participant was lead to believe the room was divided so the study would not disturb other people working in the room. The "wizard" monitored the participant's actions via a wireless camera and selected the clips to play on

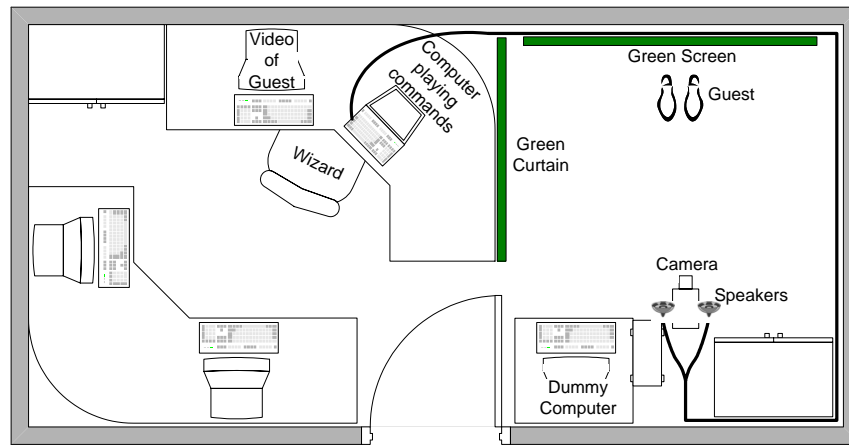


Figure 19: Lab setup for Wizard-of-Oz user study. The participant was asked to stand on the right hand side of the lab and was lead to believe the computer on the right hand side of the lab was connected to the speakers, camera and microphone. The wizard sat behind the green curtain, on the left hand side of the lab and selected which instructions to play.

a computer behind the curtain. The computer played the clips through speakers situated next to the computer believed to be running the SIMS Faces system. (See [Figure 19](#)). The participants were asked to fill out a followup questionnaire (See [Appendix E](#)). The quantitative and qualitative data informed the next iteration of the system.

In the study with our implemented SIMS Faces system, we asked 7 students to compare their picture taken by the SIMS Faces system with their picture on the department web page. Each of the 7 participants in our traditional user study has pictures posted on a department web page taken by a human (the system administrator). Both pictures were the same resolution and cropped similarly, although unlike the photos on the departmental web page, the SIMS Faces photos were cropped automatically by the system. Six of the students preferred the picture

taken by the SIMS Faces system and one student said both pictures were about the same. In addition to a picture, each student successfully recorded her name and selected to keep her recorded name after hearing it. These results demonstrate the SIMS Faces system successfully records the participant saying her name and takes a portrait photo she is happy with. These are the desired actions we set out to elicit and record with the design and implementation of the SIMS Faces system. The system lowers the cost of iteration in taking the student's picture and recording her name while ensuring the quality of interaction. This allows the student to select from multiple images, resulting in a better picture, and the students don't mind iterating because of the quality of the interaction.

While a working Active Capture system is not an evaluation of the proposed design process, it does serve as a proof of concept. The proposed design process has led to a working Active Capture system and encapsulates the lessons learned from the process of designing the SIMS Faces System.

Although the design process and ACAL have not been formally evaluated, they are based on the design of a successful Active Capture application and we iterated the design of and used ACAL as we designed the SIMS Faces system.

10 Conclusion and Future Work

The diverse design team and complexity of Active Capture applications present unique challenges. I presented a design process and visual language for Active Capture applications and a tool to help the design team integrate the interaction with the user and the computer vision parsers. The work is toward an integrated

design tool for Active Capture applications.

The interdependence between the interaction script and the action recognizers present unique challenges in designing Active Capture applications. I presented a user-centered iterative design process and a visual language for Active Capture applications and a prototype of a tool for designing the action recognizers. The three systems are based on experience from the design of a working Active Capture application, the SIMS Faces application. In the future, I plan to combine the design process, visual language and action recognizer design tool in an integrated design tool for Active Capture applications.

The design process follows the design-test-analysis methodology and leverages the benefits of bodystorming, wizard-of-Oz user studies and traditional user studies. Each step in the process provides important data and examples for the next step of the interaction script and action recognizers. As the interaction script unfolds, the action recognizers start to take form. As we describe the action in terms of the multimedia parsers, we understand how to work with the user to reach our desired action.

The bodystorming session and the strategies and design space lay the foundation for the interaction script. The wizard-of-Oz study refines the interaction script and provides valuable sample data for the design of the action recognizer. In addition, the wizard-of-Oz study is important in the design process for two main reasons. 1) Humans act differently around other humans than they do with computers. The wizard-of-Oz study provides realistic data for how the user will react to the script when it is coming from the computer. 2) It is important to test the triggers in the commands. Some triggers may not work as expected and others may

appear in unexpected places.

The visual language allows the diverse design team to work together on the design of Active Capture applications. It supports the timing details from the action recognizer and the control flow details from the interaction script and allows the design team to iterate on the two interdependent components, the interaction script and action recognizer, more easily.

The Active Capture design process reduces the cost of iteration and manages the complexity inherent in these types of systems. The visual language helps the design team work together and keep track of all the details. A design tool for Active Capture applications should further reduce the cost of iteration, better manage the inherent complexity and support the diverse design team. In particular, it should support the design-test-analysis methodology, help the design team manage the complexity by allowing them to work at varying levels of detail, support the design strategies presented by Heer *et al.*, and help the team keep track of the context of the interaction.

By supporting the Active Capture design process presented, the future design tool will reduce the cost of iteration and help the team better manage the details and complexity in Active Capture applications. As we did manually in designing the SIMS Faces application, each iteration of the design-test-analysis cycle in the design tool should allow the team to add more detail to the design of the Active Capture system. When they begin to design a new system, they should be able to describe the system at a high level, specifying the instruction or trigger for the requirements for the desired action. Once the team has the high level ideas in the design tool, they should be able to start filling in the instructions for the cases when

something goes wrong. We did this on paper using ACAL, but the tool should provide an authoring environment for ACAL. Once they have a rough draft of the interaction script, they should be able to follow the design process and record all of the clips and run a wizard-of-Oz study. The tool should provide the interface for the human wizard and collect the sample data. The sample data should be presented back to the team so they can iterate on the design. Next the team should be able to add the timing details to the design based on the data from the wizard-of-Oz studies. This part was done manually, but a tool could help associate data from the wizard-of-Oz study with parts of the design in ACAL. At this point they may have several modules laid out, but not connected together in a coherent system. Now they will need to think about the pre and post conditions of each module in the application to make sure the flow from one to the next makes sense and will work. The wizard-of-Oz studies will help them get a feel for what might not work, and thinking about the pre and post condition details will also help them here. All of these details would be too overwhelming at the beginning of the design process, as we experienced in designing the SIMS Faces application, but are necessary to work through. The design tool should support this progression and support the designer as they are ready to add more detail without requiring all the details to begin with.

As the design team fills in the interaction script for what to do when something goes wrong, they should be guided to follow the interaction script design strategies laid out in Heer *et al.* As with any design tool, the path of least resistance should lead the team to a well designed system. In this case, the path of least resistance in the design of the interaction script should lead the team to an interaction script

that supports the design strategies such as progressive assistance, freshness, and graceful failure. How to support the design strategies is not clear yet and left for future investigation.

The use of context from the interaction in the action recognizer is what makes Active Capture systems powerful and one of the things that presents challenges in the design of these systems. The design tool should help the designers keep track of the context from the interaction script both by helping them visualize it (possibly with key frames from the video of each action) as well as allowing them to make annotations about the context from the interaction script.

By supporting the design team as they work together and manage the details inherent in Active Capture applications, such a tool would make the design of these systems more feasible.

References

- [All83] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [AP92] W. A. Ainsworth and S. R. Pratt. Feedback strategies for error correction in speech recognition systems. *International Journal of Man-Machine Studies*, 36(6):833–842, 1992.
- [Ayo03] Richard Grace Aaron Steinfeld Ayoob, Ellen M. A user-centered drowsy-driver detection and warning system. In *Proceedings of Designing for User eXperience*, San Francisco, California, USA, 2003. ACM Press.
- [BB94] Daniel Boyarski and Richard Buchanan. Computers and communication design: exploring the rhetoric of HCI. *Interactions*, 1(2):25–35, April 1994.

- [BH93] C. Baber and K. S. Hone. Modeling error recovery and repair in automatic speech recognition. *International Journal of Man-Machine Studies*, 39(3):495–515, 1993.
- [BH95] Susan E. Brennan and Eric A. Hulteen. Interaction and feedback in a spoken language system: a theoretical framework. *Knowledge-Based Systems*, 8(2):143–151, April-June 1995.
- [CB91] H.H. Clark and Brennan. Grounding in communication. In Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley, editors, *Perspectives on Socially Shared Cognition*. APA Books, 1991.
- [Cha05] Ana Ramírez Chang. Designing systems that direct human action. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, Portland, OR, USA, 2005. ACM Press.
- [Dav03a] Marc Davis. Active capture: Automatic direction for automatic movies (video). In *Video Proceedings of 11th Annual ACM International Conference on Multimedia*, Berkeley, CA, USA, 2003. ACM Press.
- [Dav03b] Marc Davis. Active capture: Integrating human-computer interaction and computer vision/audition to automate media capture. In *IEEE International Conference on Multimedia and Expo (ICME 2003) Special Session on Moving from Features to Semantics using Computational Media Aesthetics*, volume II, pages 185–188. IEEE Computer Society Press, 2003.
- [Dav03c] Marc Davis. Editing out video editing. *IEEE MultiMedia*, 10(2):54–64, April - June 2003.
- [DHB⁺04] Anind K. Dey, Raffay Hamid, Chris Beckmann, Ian Li, and Daniel Hsu. a cappella: programming by demonstration of context-aware applications. In *Proceedings of the 2004 conference on Human factors in computing systems*, pages 33–40. ACM Press, 2004.
- [DJ93] Nils Dahlbäck and Arne Jönsson. Wizard of oz studies – why and how. In *Proceedings of International Workshop on Intelligent User Interfaces*, pages 193–200, 1993.
- [DL96] Marc Davis and David Lezitt. Time-based media processing system. US Patent 5,969,716. Continuation of US Patent 5,969,716. Filed: August 6, 1996. Issued: October 19, 1999, 1996.

- [DV01] Chitra Dorai and Svetha Venkatesh. Computational media aesthetics: Finding meaning beautiful. *IEEE MultiMedia*, 8(4):10–12, October 2001.
- [EILM00] W. Keith Edwards, Takeo Igarashi, Anthony LaMarca, and Elizabeth D. Mynatt. A temporal model for multi-level undo and redo. In *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 31–40, New York, NY, USA, 2000. ACM Press.
- [EM97] W. Keith Edwards and Elizabeth D. Mynatt. Timewarp: techniques for autonomous collaboration. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 218–225, New York, NY, USA, 1997. ACM Press.
- [FSA96] Charles Frankel, Michael J. Swain, and Vassilis Athitsos. Webseer: An image search engine for the world wide web. Technical report, Chicago, IL, USA, 1996.
- [GL85] John D. Gould and Clayton Lewis. Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311, March 1985.
- [Har87] David Harel. Statecharts: A visual formalism for complex systems. science of computer programming. In *Science of Computer Programming*, volume 8, pages 231–274, 1987.
- [HGR⁺04] Jeffrey Heer, Nathaniel S. Good, Ana Ramírez, Marc Davis, and Jennifer Mankoff. Presiding over accidents: system direction of human action. In *Proceedings of the 2004 conference on Human factors in computing systems*, pages 463–470. ACM Press, 2004.
- [Hor99] Eric Horvitz. Principles of mixed-initiative user interfaces. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 159–166, New York, NY, USA, 1999. ACM Press.
- [Kar94] Gerald M. Karam. Visualization using timelines. In *ISSTA '94: Proceedings of the 1994 ACM SIGSOFT international symposium on Software testing and analysis*, pages 125–137, New York, NY, USA, 1994. ACM Press.
- [Kel84] J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, 2(1):26–41, January 1984.

- [KL91] Kevin B. Kenny and Kwei-Jay Lin. Building flexible real-time systems using the flex language. *Computer*, 24(5):70–78, May 1991.
- [KSK97] Yuichi Koike, Atsushi Sugiura, and Yoshiyuki Koseki. Timeslider: an interface to specify time point. In *UIST '97: Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 43–44, New York, NY, USA, 1997. ACM Press.
- [MA99] Jennifer Mankoff and Gregory D. Abowd. Error correction techniques for handwriting, speech, and other ambiguous or error prone systems. Technical Report TechReport GIT-GVU-99-18, Georgia Institute of Technology, Atlanta, GA, USA, June 1999.
- [Mac] Macromedia. Authorware, version 6. <http://www.macromedia.com/software/authorware>.
- [MHA00] Jennifer Mankoff, Scott E. Hudson, and Gregory D. Abowd. Interaction techniques for ambiguity resolution in recognition-based interfaces. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 11–20. ACM Press, 2000.
- [NN95] Jocelyne Nanard and Marc Nanard. Hypertext design environments and the hypertext design process. *Communications of the ACM*, 38(8):49–56, August 1995.
- [OKK03] Antti Oulasvirta, Esko Kurvinen, and Tomi Kankainen. Understanding contexts by being there: case studies in bodystorming. *Personal and Ubiquitous Computing*, 7(2):125–134, July 2003.
- [Ovi00] Sharon Oviatt. Taming recognition errors with a multimodal interface. *Communications of the ACM*, 43(9):45–51, 2000.
- [PB03] Claudio S. Pinhanez and Aaron F. Bobick. Interval scripts: a programming paradigm for interactive environments and agents. *Personal Ubiquitous Computing*, 7(1):1–21, 2003.
- [RD04] Ana Ramírez and Marc Davis. Active capture and folk computing. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2004) Special Session on Folk Information Access Through Media*, Taipei, Taiwan, 2004. IEEE Computer Society Press.
- [Ret94] Marc Rettig. Prototyping for tiny fingers. *Communications of the ACM*, 37(4):21–27, April 1994.

- [RN96] Byron Reeves and Clifford Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.
- [SKL02] Anoop K. Sinha, Scott R. Klemmer, and James A. Landay. Embarking on spoken-language nl interface design. *The International Journal of Speech Technology*, 5(2):159–169, May 2002.
- [SM97] Ben Shneiderman and Pattie Maes. Direct manipulation vs. interface agents. *Interactions*, 4(6):42–61, 1997.
- [SMW01] Bernhard Suhm, Brad Myers, and Alex Waibel. Multimodal error correction for speech user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(1):60–98, 2001.
- [Tel] Tellme Networks. <http://www.tellme.com>.
- [THB⁺04] Jonathan Trevor, David M. Hilbert, Daniel Billsus, Jim Vaughan, and Quan T. Tran. Contextual contact retrieval. In *Proceedings of International Conference on Intelligent User Interfaces (IUI 2004)*, 2004.
- [Tuf83] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, 1983. "The time-series plot is the most frequently used form of graphic design."
- [YLM95] Nicole Yankelovich, Gina-Anne Levow, and Matt Marx. Designing speechacts: issues in speech user interfaces. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 369–376, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

A SIMS Faces System Interaction Script

Welcome

Prompt *Welcome to SIMS and the SIMS Faces application developed by Garage Cinema Research. The SIMS Faces application will automatically record you saying your name and take your picture for inclusion on the SIMS web page on the SIMS internal web site. The SIMS web page can only be seen by members of the SIMS community and helps people get to know what you look like and how to pronounce your name. OK, here we go.*

Position

Prompt *Please stand on the white marks on the floor and look at the camera.*

Mediation	While ...	Say ...
Can't see guest		<p><i>Hello? I can't see you. Please make sure you are standing on the white marks on the floor and that you are looking at the camera.</i></p> <p><i>Hmm, I still can't see you, please stand on the white marks on the floor, they're the ones that look like shoe prints and look at the camera.</i></p> <p><i>(Graceful Failure) Although I can't see all of you, let's move on. If you are not happy with your results, you can come back and try again.</i></p>
Guest not framed		<p><i>Hmm, I can't see all of you. Please be sure you are standing on the white marks on the floor and look at the camera.</i></p> <p><i>Your picture would look a lot better if I could see all of you. Please stand on the white marks on the floor and look at the camera.</i></p>

(Graceful Failure) *Although I can't see all of you, let's move on. If you are not happy with your results, you can come back and try again.*

Can't find eyes

I can't see your eyes. Please make sure you are facing the camera so that your name will be recorded clearly.

I still can't see your eyes, please face the camera and I'll do a much better job of recording your name.

(Graceful Failure) *Although I can't see your eyes, let's move on. If you are not happy with your results, you can come back and try again.*

Closing

If ...

Say ...

Success

That's great. Next we are going to record your name so people will know how to pronounce it.

Failure

Now you are going to tell us your full name so people will know how to pronounce it. Please look at the camera and state your full name.

Name

Prompt

Please look at the camera and state your full name now.

Mediation

While ...

Say ...

Didn't hear anything

I didn't hear anything. I'd like you say your first name and your family name. Now please state your name.

I still didn't hear your name. But perhaps an example will clear this up. My first name, the name that people call me by is Harry, my family name is Potter. I'd say to the camera, Harry Potter. Now it's your turn. Go ahead.

I'm having some trouble hearing your name. Maybe I have too much wax in my audio input. Is this thing on?? Please face the camera and clearly say your full name.

Perhaps you go by another name. Consider using another name or just clearly restate your full name.

(Graceful Failure) I think something might be wrong with the microphone because it isn't picking up your name. That's OK, you can come back and try again later. Now let's move on and take your picture.

Utterance too soft *That was great, but I need you to speak louder so we can clearly record your name. Go ahead.*

I'm sorry but I didn't catch that either. Please be sure to speak up. Go ahead.

(Graceful Failure) I wasn't able to clearly hear your name, but I did hear something. This is what I heard ;user's name;. Would you like to rerecord your name?

Utterance too short *Wow, that's pretty short for a name. Just in case, let's rerecord your name. Please be sure to state your first and last name.*

I didn't quite catch that. Please be sure to state your full name. For example Daniel Boone. Your turn.

		<i>(Graceful Failure) I'm pretty sure your full name was longer than that. However, just to make sure, I'll play back what was recorded. ;User's name;. Would you like to rerecord your name?</i>
	Utterance too long	<i>I heard you say something, but it sounded too long for a name. Let's try again. Please say your full name, that is your first and last name now.</i> <i>Hmm, that still sounded too long to be a full name. An example would be Harry Potter. Your turn.</i> <i>(Graceful Failure) I'm still not sure that was your full name, so next we'll play back your name for verification. ;User's name;. Would you like to rerecord your name?</i>
Closing	Success	<i>Thanks for saying your name. Now we are going to take your picture.</i>

Picture

Prompt	<i>Please stand on the white marks on the floor and look at the camera. Smile.</i>	
Mediation	While ...	Say ...
	Guest not framed	<i>You don't seem to be framed. Please be sure you are standing on the white marks on the floor. Now smile.</i> <i>I don't entirely see you. Perhaps sitting down or standing on a stool might help. Now smile.</i> <i>(Graceful Failure) Although I had some trouble seeing all of you I do have a picture of you.</i>

Guest not standing still *Please stand still while I take our picture. OK, smile.*

You will get better results if you are standing still for the picture. Please stand the white marks on the floor and stand still. Now smile.

(Graceful Failure) Although you seemed to be moving while I was taking your picture, I do have a picture of you.

Can't find eyes *I can't see your eyes. Perhaps you are wearing glasses or a hat. Please remove them and look at the camera. Smile.*

Hmm, no, can't see your eyes. Please be sure to look at the camera while I take your picture. Now smile.

I still can't see your eyes. Try opening our eyes wider. Let's try again. Smile.

Maybe you aren't framed properly. Perhaps sitting down or standing on a stool will put you in the frame. Now smile.

I'm sorry but I'm having a problem finding your eyes. Let's try one last time. Smile.

(Graceful Failure) I couldn't find your eyes. But I did get a picture of you.

Can't find smile *Your picture will be nicer if you smile. Please look at the camera and smile.*

Hmm, I still didn't see you smile. Here, let me tell you a joke. A guy walks into a bar. Owe.

(Graceful Failure) I didn't catch your smile. But I do have a picture of you.

Closing If ... Say ...

Success *That was really great.*

Failure

*Now you will select the best picture for the
SIMS Faces page.*

B Wizard-of-Oz User Study Script

15 minutes before subject arrives:

[*Wizard*] Prepare room for study: This entails ensuring lab space is available for the study and allowing users of the space to know that a study will commence shortly. Make sure the camera works and all the connections are in place (especially the RCA video feed to the “wizard’s monitor”). Be sure to position the equipment in the room to minimize the possibility of the subject “looking behind the curtain”. In addition, open the Director application and test that the sound works properly (adjust the sound levels). Run through a few scenarios if time permits to ensure everything is functioning properly.

[*Proctor*] Prepare the materials for the study: Prepare the relevant papers for the “expert” or “experienced” subject, including the consent form, the instructions, the post-experiment questionnaire, and the explanation paper. Get the computer adjacent to the camcorder up and running with the mock active capture software. Cue up the tape and insert into the dv camcorder.

When subject arrives:

[*Wizard*] Be in the room and act occupied. Be sure to turn off the monitor displaying the camcorder output. [*Proctor*] Greeting and introduction outside the room. Knock on the door and allow the Wizard to let you in to ensure that nothing is exposed to compromise the illusion of a fully functioning device. Lead the participant into the lab and place a “study in progress” sign on the lab door to prevent interruptions.

Administering consent forms and explaining study:

[*Proctor*] After you administer the documents, ask the Wizard if it’s okay if you conduct a study (Wizard can ask how long it will take or something of that nature). After the participant signs the consent forms administer the study explanation page. After they read it, emphasize that you are studying how the portrait camera performs and not the participant. Also mention that they helping you by allowing us to get feedback on the device’s performance in the early stages of production. Lastly, mention again the expected time and that it is okay to quit at any

time. Be sure to hit record on the camcorder to document the study.

Initiating the study:

[*Wizard*] After the Proctor positions the subject in place, turn on the monitor and ready Director to begin. Be ready to hit the “welcome” sound file when the Proctor hits the space bar on the dummy computer.

[*Proctor*] Position the subject in place and ask if the subject is ready to begin and if they have any questions before they begin. When ready, announce that you will begin the experiment and hit the space bar on the dummy computer audibly so the Wizard may begin the “welcome” sound file on cue.

After the study:

[*Wizard*] Be sure to shut off the monitor before the study commences (perhaps when the “thank you” sound clip is playing). Resume acting busy and unconcerned with the study in progress.

[*Proctor*] Present the follow-up questionnaire to the subject. After they complete it, present the explanation paper as well as answer any questions that the participant may have.

After the participant leaves:

[*Wizard and Proctor*] Remove the “study in progress” sign from the door and store it. As you are organizing the participant’s files discuss any interesting behavior. Before you leave the room, be sure to place any necessary cables or papers in a locatable place for the next study. Be sure to remove the dv tape from the camcorder and label it with the participant’s id number and store it in the participant file.

C Wizard-of-Oz Study Consent Form

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

SCHOOL OF INFORMATION MANAGEMENT AND SYSTEMS
BERKELEY, CALIFORNIA 94720

Consent Form for User Study

Thank you for your interest in our user study. If you agree to take part by completing this consent form, you will participate in an approximately 15 minute in-laboratory experiment, in which you will be directed to perform simple actions by the computer, followed by a 15 minute exit questionnaire. The experiment will take place in 110a South Hall on the campus of the University of California, Berkeley and will be videotaped.

There are no known risks to you from taking part in this research, and no substantial benefit to you either. However, it is hoped that this research will provide insights into the design and usability of multimedia based interfaces.

All of the information we obtain from you during the research will be kept confidential. Your name and other identifying information about you will not be used in any reports of the research. After this research is completed, we may save our notes for use in our future research. If you sign the corresponding sections in the release form, we may also save the audio and video records. However, the same confidentiality guarantees given here will apply to future storage and use of the materials. Although we will keep your name confidential, you may still be identifiable to others on the audio, video, and/or photographic records.

Your participation in this research is voluntary. You are free to refuse to take part. You may refuse to answer any questions and may stop taking part in the study at any time.

If you have any questions about the research, you may contact Ana Ramirez (phone: 510-847-1985, email: anar@cs.berkeley.edu). If you agree to take part in the research, please sign the form below. Please keep the other copy of this agreement for your future reference.

If you have any question regarding your treatment or rights as a participant in this research project, please contact Professor Marc Davis at (510) 643-2253 or marc@sims.berkeley.edu.

I have read this consent form and I agree to take part in this research study.

Signature

Date

(revised 03/04)

D Wizard-of-Oz Study Media Release Form

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

SCHOOL OF INFORMATION MANAGEMENT AND SYSTEMS
BERKELEY, CALIFORNIA 94720

AUDIO AND VIDEO RECORDS RELEASE CONSENT FORM

As part of this project we have made audio and video recording of you while you participated in the research. We would like you to indicate below what uses of these records you are willing to consent to. This is completely up to you. We will only use the records in ways that you agree to. In any use of these records, your name will not be identified.

All: The records can be used in all of the contexts described below.

(initials) Audio & Video _____ Audio _____ Video _____

1. The records can be studied by the research team for use in the research project.

(initials) Audio & Video _____ Audio _____ Video _____

2. The records can be shown to subjects in other experiments.

(initials) Audio & Video _____ Audio _____ Video _____

3. The records can be used for scientific publications.

(initials) Audio & Video _____ Audio _____ Video _____

4. The records can be shown at meetings of scientists interested in the study of tangible user interfaces and/or ubiquitous computing.

(initials) Audio & Video _____ Audio _____ Video _____

5. The records can be shown in classrooms to students.

(initials) Audio & Video _____ Audio _____ Video _____

6. The records can be shown in public presentations to nonscientific groups.

(initials) Audio & Video _____ Audio _____ Video _____

7. The records can be used on television and radio.

(initials) Audio & Video _____ Audio _____ Video _____

I have read the above description and give my consent for the use of the records as indicated above.

Signature _____ Date _____

(revised 03/04)

E Wizard-of-Oz Study Follow-up Questionnaire

I.

a) Age: _____

b) Gender: MALE FEMALE

c) What is your primary language: _____

II.

a) Please rate your familiarity with portrait photography using the following scale:

Very Unfamiliar (1 2 3 4 5 6 7) Extremely Familiar

b) Have you ever directed others or been directed? YES NO

If "YES", please elaborate below:

c) Do you currently own a computer? _____

If "YES", please circle the approximate frequency of use:

Less than 5 hours/week 5 hours/week 3 hours/day 8 hours/day More than 8 hours/day

d) Do you currently own a film camera? _____

If "YES", please circle the approximate number of photos taken per month:

1 or less photos/month 10 photos/month 30 photos/month 40 photos/month 50+ photos/month

e) Do you currently own a digital camera? _____

If "YES", please circle the approximate number of photos taken per month:

1 or less photos/month 10 photos/month 30 photos/month 40 photos/month 50+ photos/month

III.

a) Please rate the overall clarity of instructions offered by the portrait camera using the following scale:

Very Unclear (1 2 3 4 5 6 7) Extremely Clear

b) Please comment on any instructions that worked well for you?

c) Please comment on any instructions that did not work well for you?

d) How could we improve the interaction?

e) What did you like about the portrait camera experience?

f) What did you dislike about the portrait camera experience?

g) Is there anything that you would add to the portrait camera?

IV.

a) What do you think about using the portrait camera for taking pictures for the SIMS Faces page?

b) What would work well from the portrait camera for this application?

c) What would need to be changed from the portrait camera to better fit the needs of this application?

F Wizard-of-Oz Study Questionnaire Data

Cumulative Data

I. (a) *Age:*

- **Average: 24.875**
- **Range: 18 - 30**
- **Median: 25**

(b) *Gender: 5 Female, 3 Male*

(c) *What is your primary language? 5 English, 1 Hebrew, 1 Italian, 1 German*

II. (a) *Please rate your familiarity with portrait photography using the following scale:*

1 (Very Unfamiliar)	1
2	0
3	0
4	2
5	2
6	2
7 (Extremely Familiar)	1

(b) *Have you ever directed others or been directed? 7 Yes, 1 No*

(c) *If you have directed others, please elaborate below:*

- **I wrote a CHI paper about direction, I've also been in numerous plays.**
- **Have directed short videos.**
- **Photo shoots in a studio with professional advice. Selection of pictures on LCD screen and feedback and more pictures afterwards... et ad infinitum**
- **Family-friends pictures and videos.**
- **Directed a play. Been in several plays. Directed films and shot films.**
- **For school photographs; we would be directed to position ourselves a certain way or to do different things.**
- **Taken part in media photo shoots.**

(d) *How many hours do you spend using a computer?*

<i>Less than 5 hours/week</i>	0
<i>5 hours/week</i>	2
<i>3 hours/day</i>	2
<i>8 hours/day</i>	3
<i>More than 8 hours/day</i>	1

(e) *How many pictures do you take with a film camera per month?*

<i>1 or less photo/month</i>	3
<i>10 photos/month</i>	3
<i>30 photos/month</i>	0
<i>40 photos/month</i>	0
<i>50+ photos/month</i>	0
<i>Does not have a film camera</i>	2

(f) *How many pictures do you take with a digital camera per month?*

<i>1 or less photo/month</i>	0
<i>10 photos/month</i>	3
<i>30 photos/month</i>	0
<i>40 photos/month</i>	0
<i>50+ photos/month</i>	2
<i>Does not have a digital camera</i>	3

III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:*

1 (Very Unclear)	0
2	0
3	0
4	1
5	2
6	2
7 (Extremely Clear)	3

(b) *Please comment on any instructions that worked well for you.*

- **The joke was great. I couldn't help but smile.**
- **Where to stand, where to look, what to say.**
- **All instructions were understandable**
- **It explains well what is going on and why.**
- **Instructions clear and easy to follow.**

(c) *Please comment on any instructions that did not work well for you.*

- I didn't know exactly when it was going to take my picture.
- With the instructions to say my name, I had the impression I was supposed to speak before I was actually supposed.
- Timing for saying your name perhaps a bit unclear.

(d) *How could we improve the interaction?*

- Equalize the audio. Use a more natural timing with the voice commands. Let me see the result right away.
- Less silly of an operator voice.
- Asking participants if they are happy with the results, and offer them another chance if not.
- Bigger feedback screen (was too far away and too small). Also, zooming: heads have different sizes and an automatic adjustment could help to make oneself feel better about the picture quality.
- Show picture after it's taken. Play back name and ask if you want to record it again. Unclear when image is being taken and when to speak. Batch mode. Don't repeat intro so many times. Better lighting; softer. Turn white marks so that are turned.
- Simplify; you don't need to tell a person you're going to ask for their name before asking for their name.
- Re-record instructions with proper pauses.

(e) *What did you like about the portrait camera experience?*

- Quick, painless. Fun trying to subvert the instructions.
- Looking at myself in the camera.
- I like the idea of pronouncing the name, since most people mispronounce my name.
- Simple.
- Very cheerful!
- The experience is not overly overbearing and does not cause nervousness.

(f) *What did you dislike about the portrait camera experience?*

- Not being able to see the result.
- Coldness of no human interaction.
- No options, no selection of pictures.
- Longer than normal, no feedback.

- **Would prefer more flattering lighting.**
 - **It seemed to take longer than needed for just a picture to be taken and a name to be taken down.**
 - **None**
- (g) *Is there anything that you would add to the portrait camera?*
- **See above.**
 - **No**
 - **In addition to (f), bigger screen, count down switch-on screen last two seconds to avoid distraction.**
 - **I'd like to be able to choose the picture, or take another one.**
 - **Perhaps a special indicator - a tone, beep, or ding that indicates exactly when you need to speak and when the camera takes a picture.**
 - **No**
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?*
- **Seems like a good idea, but have to see what happens in real world use.**
 - **Cool and novel idea.**
 - **I would prefer personal interaction**
 - **ok**
 - **It's nice that it is automated - reduces costs too.**
 - **I think it is a good idea and perhaps should be implemented widely to increase name/face recognition.**
- (b) *What would work well from the portrait camera for this application?*
- **Both images and pronunciation would work well in theory... but its hard to say how algorithms will hold up and it would have been nice to review the results.**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?*
- **Show the result! There also nothing to prevent me from making up a crazy name.**
 - **Like (d) in previous section.**
 - **see (g)**
 - **None.**

Participant 1

- I. (a) Age: **24**
(b) Gender: **Male**
(c) What is your primary language? **English**
- II. (a) Please rate your familiarity with portrait photography using the following scale:
5
(b) Have you ever directed others or been directed? **Yes**
(c) If you have directed others, please elaborate below: **I wrote a CHI paper about direction, I've also been in numerous plays.**
(d) How many hours do you spend using a computer? **More than 8 hours/day**
(e) How many pictures do you take with a film camera per month? **Does not own a film camera.**
(f) How many pictures do you take with a digital camera per month? **10 photos/month**
- III. (a) Please rate the overall clarity of instructions offered by the portrait camera using the following scale: **6**
(b) Please comment on any instructions that worked well for you. **The joke was great. I couldn't help but smile.**
(c) Please comment on any instructions that did not work well for you. -
(d) How could we improve the interaction? **Equalize the audio. Use a more natural timing with the voice commands. Let me see the result right away.**
(e) What did you like about the portrait camera experience? **Quick, painless. Fun trying to subvert the instructions.**
(f) What did you dislike about the portrait camera experience? **Not being able to see the result.**
(g) Is there anything that you would add to the portrait camera? **See above.**
- IV. (a) What do you think about using the portrait camera for taking pictures for the SIMS Faces page? **Seems like a good idea, but have to see what happens in real world use.**

- (b) *What would work well from the portrait camera for this application?* **Both images and pronunciation would work well in theory... but its hard to say how algorithms will hold up and it would have been nice to review the results.**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **Show the result! There also nothing to prevent me from making up a crazy name.**

Participant 2

- I. (a) *Age:* **18**
- (b) *Gender:* **Male**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
7
- (b) *Have you ever directed others or been directed?* **Yes**
- (c) *If you have directed others, please elaborate below:* **Have directed short videos.**
- (d) *How many hours do you spend using a computer?* **5 hours/week**
- (e) *How many pictures do you take with a film camera per month?* **Does not own a film camera.**
- (f) *How many pictures do you take with a digital camera per month?* **Does not own a digital camera.**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **7**
- (b) *Please comment on any instructions that worked well for you.* **Where to stand, where to look, what to say.**
- (c) *Please comment on any instructions that did not work well for you.* **-**
- (d) *How could we improve the interaction?* **Less silly of an operator voice.**
- (e) *What did you like about the portrait camera experience?* **Looking at myself in the camera.**
- (f) *What did you dislike about the portrait camera experience?* **Coldness of no human interaction.**
- (g) *Is there anything that you would add to the portrait camera?* **No**

- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **Cool and novel idea.**
- (b) *What would work well from the portrait camera for this application?* -
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* -

Participant 3

- I. (a) *Age:* **28**
- (b) *Gender:* **Female**
- (c) *What is your primary language?* **Hebrew**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
1
- (b) *Have you ever directed others or been directed?* **No**
- (c) *If you have directed others, please elaborate below:*
- (d) *How many hours do you spend using a computer?* **8 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **10 photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **50+ photos/month**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **7**
- (b) *Please comment on any instructions that worked well for you.* -
- (c) *Please comment on any instructions that did not work well for you.* -
- (d) *How could we improve the interaction?* **Asking participants if they are happy with the results, and offer them another chance if not.**
- (e) *What did you like about the portrait camera experience?* **I like the idea of pronouncing the name, since most people mispronounce my name.**
- (f) *What did you dislike about the portrait camera experience?* -
- (g) *Is there anything that you would add to the portrait camera?* -
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **I would prefer personal interaction**

- (b) *What would work well from the portrait camera for this application? -*
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application? Like (d) in previous section.*

Participant 4

- I. (a) *Age: 26*
 - (b) *Gender: Female*
 - (c) *What is your primary language? German*
- II. (a) *Please rate your familiarity with portrait photography using the following scale: 6*
 - (b) *Have you ever directed others or been directed? Yes*
 - (c) *If you have directed others, please elaborate below: Photo shoots in a studio with professional advice. Selection of pictures on LCD screen and feedback and more pictures afterwards... et ad infinitum*
 - (d) *How many hours do you spend using a computer? 8 hours/day*
 - (e) *How many pictures do you take with a film camera per month? 10 photos/month*
 - (f) *How many pictures do you take with a digital camera per month? 10 photos/month*
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale: 4*
 - (b) *Please comment on any instructions that worked well for you. All instructions were understandable*
 - (c) *Please comment on any instructions that did not work well for you. -*
 - (d) *How could we improve the interaction? Bigger feedback screen (was too far away and too small). Also, zooming: heads have different sizes and an automatic adjustment could help to make oneself feel better about the picture quality.*
 - (e) *What did you like about the portrait camera experience? Simple.*
 - (f) *What did you dislike about the portrait camera experience? No options, no selection of pictures.*

- (g) *Is there anything that you would add to the portrait camera? In addition to (f), bigger screen, count down switch-on screen last two seconds to avoid distraction.*
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page? ok*
- (b) *What would work well from the portrait camera for this application? -*
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application? see (g)*

Participant 5

- I. (a) *Age: 34*
- (b) *Gender: Female*
- (c) *What is your primary language? Italian*
- II. (a) *Please rate your familiarity with portrait photography using the following scale: 6*
- (b) *Have you ever directed others or been directed? Yes*
- (c) *If you have directed others, please elaborate below: Family-friends pictures and videos.*
- (d) *How many hours do you spend using a computer? 8 hours/day*
- (e) *How many pictures do you take with a film camera per month? 1 or less photos/month*
- (f) *How many pictures do you take with a digital camera per month? 50+ photos/month*
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale: 6*
- (b) *Please comment on any instructions that worked well for you. It explains well what is going on and why.*
- (c) *Please comment on any instructions that did not work well for you. I didn't know exactly when it was going to take my picture.*
- (d) *How could we improve the interaction? -*
- (e) *What did you like about the portrait camera experience? -*

- (f) *What did you dislike about the portrait camera experience?* **Longer than normal, no feedback.**
- (g) *Is there anything that you would add to the portrait camera?* **I'd like to be able to choose the picture, or take another one.**
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* -
- (b) *What would work well from the portrait camera for this application?* -
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* -

Participant 6

- I. (a) Age: **30**
- (b) Gender: **Female**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
4
- (b) *Have you ever directed others or been directed?* **Yes**
- (c) *If you have directed others, please elaborate below:* **Directed a play. Been in several plays. Directed films and shot films.**
- (d) *How many hours do you spend using a computer?* **5.5 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **510 photos/month**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **5**
- (b) *Please comment on any instructions that worked well for you.* -
- (c) *Please comment on any instructions that did not work well for you.* -
- (d) *How could we improve the interaction?* **Show picture after it's taken. Play back name and ask if you want to record it again. Unclear when image is being taken and when to speak. Batch mode. Don't repeat intro so many times. Better lighting; softer. Turn white marks so that are turned.**

- (e) *What did you like about the portrait camera experience? -*
- (f) *What did you dislike about the portrait camera experience? **Would prefer more flattering lighting.***
- (g) *Is there anything that you would add to the portrait camera? -*
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page? -*
- (b) *What would work well from the portrait camera for this application? -*
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application? -*

Participant 7

- I. (a) *Age: **19***
- (b) *Gender: **Female***
- (c) *What is your primary language? **English***
- II. (a) *Please rate your familiarity with portrait photography using the following scale: **4***
- (b) *Have you ever directed others or been directed? **Yes***
- (c) *If you have directed others, please elaborate below: **For school photographs; we would be directed to position ourselves a certain way or to do different things.***
- (d) *How many hours do you spend using a computer? **3 hours/day***
- (e) *How many pictures do you take with a film camera per month? **10 photos/month***
- (f) *How many pictures do you take with a digital camera per month? **Does not have a digital camera***
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale: **5***
- (b) *Please comment on any instructions that worked well for you. -*
- (c) *Please comment on any instructions that did not work well for you. **With the instructions to say my name, I had the impression I was supposed to speak before I was actually supposed.***

- (d) *How could we improve the interaction?* **Simplify; you don't need to tell a person you're going to ask for their name before asking for their name.**
 - (e) *What did you like about the portrait camera experience?* **Very cheerful!**
 - (f) *What did you dislike about the portrait camera experience?* **It seemed to take longer than needed for just a picture to be taken and a name to be taken down.**
 - (g) *Is there anything that you would add to the portrait camera?* **Perhaps a special indicator - a tone, beep, or ding that indicates exactly when you need to speak and when the camera takes a picture.**
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **It's nice that it is automated - reduces costs too.**
- (b) *What would work well from the portrait camera for this application?* -
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* -

Participant 8

- I. (a) *Age:* **20**
- (b) *Gender:* **Male**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
5
- (b) *Have you ever directed others or been directed?* **Yes**
- (c) *If you have directed others, please elaborate below:* **Taken part in media photo shoots.**
- (d) *How many hours do you spend using a computer?* **3 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **Does not have a digital camera**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **7**

- (b) *Please comment on any instructions that worked well for you. **Instructions clear and easy to follow.***
 - (c) *Please comment on any instructions that did not work well for you. **Timing for saying your name perhaps a bit unclear.***
 - (d) *How could we improve the interaction? **Re-record instructions with proper pauses.***
 - (e) *What did you like about the portrait camera experience? **The experience is not overly overbearing and does not cause nervousness.***
 - (f) *What did you dislike about the portrait camera experience? **None***
 - (g) *Is there anything that you would add to the portrait camera? **No***
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page? **I think it is a good idea and perhaps should be implemented widely to increase name/face recognition.***
- (b) *What would work well from the portrait camera for this application? -*
 - (c) *What would need to be changed from the portrait camera to better fit the needs of this application? **None.***

G SIMS Faces System User Study Consent Form

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

SCHOOL OF INFORMATION MANAGEMENT AND SYSTEMS
BERKELEY, CALIFORNIA 94720

Consent Form for User Study

Thank you for your interest in our user study. If you agree to take part by completing this consent form, you will participate in an approximately 5 minute in-laboratory experiment, in which you will be directed to perform simple actions by the computer, followed by a 10 minute exit questionnaire. The experiment will take place in 110a South Hall on the campus of the University of California, Berkeley and will be videotaped.

There are no known risks to you from taking part in this research, and no substantial benefit to you either. However, it is hoped that this research will provide insights into the design and usability of multimedia based interfaces.

All of the information we obtain from you during the research will be kept confidential. Your name and other identifying information about you will not be used in any reports of the research. After this research is completed, we may save our notes for use in our future research. If you sign the corresponding sections in the release form, we may also save the audio and video records. However, the same confidentiality guarantees given here will apply to future storage and use of the materials. Although we will keep your name confidential, you may still be identifiable to others on the audio, video, and/or photographic records.

Your participation in this research is voluntary. You are free to refuse to take part. You may refuse to answer any questions and may stop taking part in the study at any time.

If you have any questions about the research, you may contact Ana Ramirez (phone: 510-847-1985, email: anar@cs.berkeley.edu). If you agree to take part in the research, please sign the form below. Please keep the other copy of this agreement for your future reference.

If you have any question regarding your treatment or rights as a participant in this research project, please contact Professor Marc Davis at (510) 643-2253 or marc@sims.berkeley.edu.

I have read this consent form and I agree to take part in this research study.

Signature

Date

(revised 12/2004)

H SIMS Faces System User Study Media Release Form

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

SCHOOL OF INFORMATION MANAGEMENT AND SYSTEMS
BERKELEY, CALIFORNIA 94720

AUDIO AND VIDEO RECORDS RELEASE CONSENT FORM

As part of this project we have made audio and video recording of you while you participated in the research. We would like you to indicate below what uses of these records you are willing to consent to. This is completely up to you. We will only use the records in ways that you agree to. In any use of these records, your name will not be identified.

All: The records can be used in all of the contexts described below.

(initials) **Audio & Video** _____ Audio _____ Video _____

1. The records can be studied by the research team for use in the research project.

(initials) **Audio & Video** _____ Audio _____ Video _____

2. The records can be shown to subjects in other experiments.

(initials) **Audio & Video** _____ Audio _____ Video _____

3. The records can be used for scientific publications.

(initials) **Audio & Video** _____ Audio _____ Video _____

4. The records can be shown at meetings of scientists interested in the study of tangible user interfaces and/or ubiquitous computing.

(initials) **Audio & Video** _____ Audio _____ Video _____

5. The records can be shown in classrooms to students.

(initials) **Audio & Video** _____ Audio _____ Video _____

6. The records can be shown in public presentations to nonscientific groups.

(initials) **Audio & Video** _____ Audio _____ Video _____

7. The records can be used on television and radio.

(initials) **Audio & Video** _____ Audio _____ Video _____

I have read the above description and give my consent for the use of the records as indicated above.

Signature _____ Date _____

(revised 12/2004)

I SIMS Faces System User Study Follow-up Questionnaire

I.

a) Age: _____

b) Gender: MALE FEMALE

c) What is your primary language: _____

II.

a) Please rate your familiarity with portrait photography using the following scale:

Very Unfamiliar (1 2 3 4 5 6 7) Extremely Familiar

b) Have you ever directed others or been directed? YES NO

If "YES", please elaborate below:

c) Do you currently own a computer? _____

If "YES", please circle the approximate frequency of use:

Less than 5 hours/week 5 hours/week 3 hours/day 8 hours/day More than 8 hours/day

d) Do you currently own a film camera? _____

If "YES", please circle the approximate number of photos taken per month:

1 or less photos/month 10 photos/month 30 photos/month 40 photos/month 50+ photos/month

e) Do you currently own a digital camera? _____

If "YES", please circle the approximate number of photos taken per month:

1 or less photos/month 10 photos/month 30 photos/month 40 photos/month 50+ photos/month

I SIMS FACES SYSTEM USER STUDY FOLLOW-UP QUESTIONNAIRE

III.

a) Please rate the overall clarity of instructions offered by the portrait camera using the following scale:

Very Unclear (1 2 3 4 5 6 7) Extremely Clear

b) Please comment on any instructions that worked well for you?

c) Please comment on any instructions that did not work well for you?

d) How could we improve the interaction?

e) What did you like about the portrait camera experience?

f) What did you dislike about the portrait camera experience?

g) Is there anything that you would add to the portrait camera?

IV.

a) What do you think about using the portrait camera for taking pictures for the SIMS Faces page?

b) What would work well from the portrait camera for this application?

c) What would need to be changed from the portrait camera to better fit the needs of this application?

V. Which picture do you prefer, the one on the SIMS Faces page or the one taken by the SIMS Faces system?

J SIMS Faces System User Study Questionnaire Data

Participant 2 Answers

Cumulative Data

I. (a) Age:

- Average: 27.86
- Range: 23 - 42
- Median: 25

(b) Gender: 2 Female, 5 Male

(c) What is your primary language? 7 English

II. (a) Please rate your familiarity with portrait photography using the following scale:

1 (Very Unfamiliar)	0
2	0
3	0
4	4
5	2
6	0
7 (Extremely Familiar)	1

(b) Have you ever directed others or been directed? 4 Yes, 3 No

(c) If you have directed others, please elaborate below:

- I have done some amateur portrait photography.
- I teach dance classes and direct others in how to do dance steps pretty often.
- Often direct others, especially when strangers want their pictures taken in San Francisco.
- I have done some amateur portrait photography.

(d) How many hours do you spend using a computer?

Less than 5 hours/week	0
5 hours/week	0
3 hours/day	1
8 hours/day	5
More than 8 hours/day	1

(e) *How many pictures do you take with a film camera per month?*

<i>1 or less photo/month</i>	7
<i>10 photos/month</i>	0
<i>30 photos/month</i>	0
<i>40 photos/month</i>	0
<i>50+ photos/month</i>	0

(f) *How many pictures do you take with a digital camera per month?*

<i>1 or less photo/month</i>	0
<i>10 photos/month</i>	3
<i>30 photos/month</i>	1
<i>40 photos/month</i>	0
<i>50+ photos/month</i>	3

III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:*

<i>1 (Very Unclear)</i>	0
<i>2</i>	0
<i>3</i>	0
<i>4</i>	0
<i>5</i>	4
<i>6</i>	3
<i>7 (Extremely Clear)</i>	0

(b) *Please comment on any instructions that worked well for you.*

- **It asked me position myself properly and say my name loud enough.**
- **The instructions were very clear.**
- **I liked the enthusiasm in the voice, and the fact that it was a natural voice rather than a generated one - it's easier to understand.**
- **Instructions were pretty clear and explicit.**
- **smile worked ok**
- **First and Last Name**

(c) *Please comment on any instructions that did not work well for you.*

- **I didn't know exactly when to say my name because the system said something that seemed like I should have said my name but then it followed up by really asking for my name.**
- **Some of the instructions were a bit condescending, even maybe for people who didn't understand the directions.**

- I was confused when it asked for my name, then kept talking when I gave my name. I was also confused at when, exactly they took the picture. Also, after two tries the application exited.
- The camera does not understand if there is any fault with the equipment rather than the user's position
- Name repetition was tedious.
- standing on white footprints made my image unentered in the camera, but the image came out ok.
- "Family Name" was strange to hear.

(d) *How could we improve the interaction?*

- I didn't know how loud to say my name. Maybe a display of the microphone detection/volume display could help.
- Turn up the volume on the mic, or put it closer so you don't have to yell. Also, I am tall and the system had a hard time recognizing that I was in the frame.
- Clarify the beginning of the name dialog - ask for the name once, then pause before talking again so that people will know when to say their names. Play a camera shutter sound when the pictures are taken (like camera phones do). Let the user select whether to try again after a few tries, rather than exiting.
- More sensitive equipment
- Indicate when photos are captured. There was a long time during which photo might have been taken. A prompt like "taking photo NOW" or a shutter click would be good.
- Seems like the interaction could have been tighter (faster). The audio was clear.
- More immediate feedback would be good.

(e) *What did you like about the portrait camera experience?*

- Simple and straightforward
- I liked being given the option of several portraits. An idea would be to give me a chance to rearrange my hair (or whatever) and then take another one, allowing me to chose from the two sets.
- I liked that I could listen and re-record anything I didn't like. I also liked the perky feedback.
- Its interactive and gives u a chance to improve.
- funny picture.

- Kinda funny having a machine talk to you.
- I liked how it walked me through both the name and photo without much interaction on my part.

(f) *What did you dislike about the portrait camera experience?*

- Nothing
- I had to yell my name and still it didn't quite hear me. The whole recording of my name was a bit irritating.
- I was frustrated with the lag, as well as with the issues I mentioned above.
- Sometimes is not as intuitive as a human.
- Name recording was tedious – it didn't hear me for a long time. Too slow to take photo, unclear when the photo was taken.
- Sometimes when someone is taking your portrait, they gesture or give off non verbal communication. Perhaps this is what is lost without a human "in the loop", as they say.
- Nothing!

(g) *Is there anything that you would add to the portrait camera?*

- Nope.
- stated above.
- maybe shutter noises when capturing photo, or some other audio prompt. Maybe allow user to see their video on screen while photo is being captured.
- Sometimes it's a good idea to goof with the subject, even to go as far as taking a goof picture. It loosens things up and the subject feels more at ease. Seems like subjects often are waiting for something to happen and folks get a strained look on their face when they are waiting and trying to understand what is going on (what is taking time). Indeed, the sense of time a subject experiences is always more extreme when they are in front of a camera waiting for something to happen.
- An indication of when it was taking the individual photos. It was hard to know if it was done taking all the pictures, since it took quite a while and I didn't hear any feedback.

IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?*

- Seems like a good idea. Can be used for other photo taking procedures (ie. DMV, etc.)
- Good idea. Especially saying your name...
- Yes, definitely (though it may get backed up if everyone takes as long as I did).
- Good idea.
- It needs to get faster.
- I think it's a good introduction to SIMS. It shows students that we are doing new things right off the bat on the first day.
- I think it would be a good idea.

(b) *What would work well from the portrait camera for this application?*

- (Didn't understand this question)
- Everything, I think, as long as the audio portion is improved.
- I'm not sure the name part would be as important for me because my name isn't hard for most English-speakers to pronounce, but it'd be useful for others. I liked the fact that I could repeat it as many times as I wanted, and I didn't feel like I was wasting the photographer's time (though I was sort of wasting the experimenter's time this time ...)
- Photo taking
- Users get to pick from a set of photos.
- I don't really understand the question. I'm guessing you are asking what text would work well, script that is. Seems like it could ask what you are interested in. Or where you are from.
- Immediate feedback, people could pick the picture they liked, and bad pictures wouldn't be posted.

(c) *What would need to be changed from the portrait camera to better fit the needs of this application?*

- Nothing
- But you would still need someone to adjust the height of the camera for tall people. Or you could have people sit on a stool to minimize the range of height differences.
- Maybe the option to skip the name part, if they want.
- More better voice recording
- Speed needs to improve.

- Seems like it could go faster.
- Speedier, and maybe a quieter place for the name recording.

- V. (a) Which picture would you prefer on the SIMS Faces Page?
- | | |
|--|---|
| The picture currently on the SIMS Faces page | 0 |
| The picture taken during this study | 6 |
| Both would be okay | 1 |

Participant 1

- I. (a) Age: **24**
(b) Gender: **Male**
(c) What is your primary language? **English**
- II. (a) Please rate your familiarity with portrait photography using the following scale: **5**
(b) Have you ever directed others or been directed? **Yes**
(c) If you have directed others, please elaborate below: **I have done some amateur portrait photography.**
(d) How many hours do you spend using a computer? **3 hours/day**
(e) How many pictures do you take with a film camera per month? **1 or less photos/month**
(f) How many pictures do you take with a digital camera per month? **10 photos/month**
- III. (a) Please rate the overall clarity of instructions offered by the portrait camera using the following scale: **5**
(b) Please comment on any instructions that worked well for you. **It asked me position myself properly and say my name loud enough.**
(c) Please comment on any instructions that did not work well for you. **I didn't know exactly when to say my name because the system said something that seemed like I should have said my name but then it followed up by really asking for my name.**
(d) How could we improve the interaction? **I didn't know how loud to say my name. Maybe a display of the microphone detection/volume display could help.**

- (e) *What did you like about the portrait camera experience?* **Simple and straightforward**
- (f) *What did you dislike about the portrait camera experience?* **Nothing**
- (g) *Is there anything that you would add to the portrait camera?*
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **Seems like a good idea. Can be used for other photo taking procedures (ie. DMV, etc.)**
- (b) *What would work well from the portrait camera for this application?* **(Didn't understand this question)**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **Nothing**
- V. (a) *Which picture would you prefer on the SIMS Faces Page?* **The picture taken during this study**

Participant 2

- I. (a) *Age:* **25**
- (b) *Gender:* **Male**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
4
- (b) *Have you ever directed others or been directed?* **No**
- (c) *If you have directed others, please elaborate below:*
- (d) *How many hours do you spend using a computer?* **8 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **10 photos/month**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **5**
- (b) *Please comment on any instructions that worked well for you.* **The instructions were very clear.**

- (c) *Please comment on any instructions that did not work well for you. Some of the instructions were a bit condescending, even maybe for people who didn't understand the directions.*
 - (d) *How could we improve the interaction? Turn up the volume on the mic, or put it closer so you don't have to yell. Also, I am tall and the system had a hard time recognizing that I was in the frame.*
 - (e) *What did you like about the portrait camera experience? I liked being given the option of several portraits. An idea would be to give me a chance to rearrange my hair (or whatever) and then take another one, allowing me to chose from the two sets.*
 - (f) *What did you dislike about the portrait camera experience? I had to yell my name and still it didn't quite hear me. The whole recording of my name was a bit irritating.*
 - (g) *Is there anything that you would add to the portrait camera? Nope.*
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page? Good idea. Especially saying your name...*
- (b) *What would work well from the portrait camera for this application? Everything, I think, as long as the audio portion is improved.*
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application? But you would still need someone to adjust the height of the camera for tall people. Or you could have people sit on a stool to minimize the range of height differences.*
- V. (a) *Which picture would you prefer on the SIMS Faces Page? The picture taken during this study.*

Participant 3

- I. (a) *Age: 23*
 - (b) *Gender: Female*
 - (c) *What is your primary language? English*
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
4
- (b) *Have you ever directed others or been directed? Yes*

- (c) *If you have directed others, please elaborate below:* **I teach dance classes and direct others in how to do dance steps pretty often.**
 - (d) *How many hours do you spend using a computer?* **More than 8 hours/day**
 - (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
 - (f) *How many pictures do you take with a digital camera per month?* **50+ photos/month**
- III.
- (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **5**
 - (b) *Please comment on any instructions that worked well for you.* **I liked the enthusiasm in the voice, and the fact that it was a natural voice rather than a generated one - it's easier to understand.**
 - (c) *Please comment on any instructions that did not work well for you.* **I was confused when it asked for my name, then kept talking when I gave my name. I was also confused at when, exactly they took the picture. Also, after two tries the application exited.**
 - (d) *How could we improve the interaction?* **Clarify the beginning of the name dialog - ask for the name once, then pause before talking again so that people will know when to say their names. Play a camera shutter sound when the pictures are taken (like camera phones do). Let the user select whether to try again after a few tries, rather than exiting.**
 - (e) *What did you like about the portrait camera experience?* **I liked that I could listen and re-record anything I didn't like. I also liked the perky feedback.**
 - (f) *What did you dislike about the portrait camera experience?* **I was frustrated with the lag, as well as with the issues I mentioned above.**
 - (g) *Is there anything that you would add to the portrait camera?* **stated above.**
- IV.
- (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **Yes, definitely (though it may get backed up if everyone takes as long as I did).**
 - (b) *What would work well from the portrait camera for this application?* **I'm not sure the name part would be as important for me because my name isn't hard for most English-speakers to pronounce, but it'd be useful for others. I liked the fact that I could repeat it as many times as I**

wanted, and I didn't feel like I was wasting the photographer's time (though I was sort of wasting the experimenter's time this time ...)

- (c) *What would need to be changed from the portrait camera to better fit the needs of this application? Maybe the option to skip the name part, if they want.*
- V. (a) *Which picture would you prefer on the SIMS Faces Page? The picture taken during this study.*

Participant 4

- I. (a) Age: **27**
(b) Gender: **Female**
(c) *What is your primary language? English*
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
4
(b) *Have you ever directed others or been directed? No*
(c) *If you have directed others, please elaborate below:*
(d) *How many hours do you spend using a computer? 8 hours/day*
(e) *How many pictures do you take with a film camera per month? 1 or less photos/month*
(f) *How many pictures do you take with a digital camera per month? 10 photos/month*
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale: 6*
(b) *Please comment on any instructions that worked well for you.*
(c) *Please comment on any instructions that did not work well for you. The camera does not understand if there is any fault with the equipment rather than the user's position*
(d) *How could we improve the interaction? More sensitive equipment*
(e) *What did you like about the portrait camera experience? Its interactive and gives u a chance to improve.*
(f) *What did you dislike about the portrait camera experience? Sometimes is not as intuitive as a human.*

- (g) *Is there anything that you would add to the portrait camera?*
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **Good idea.**
- (b) *What would work well from the portrait camera for this application?* **Photo taking**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **More better voice recording**
- V. (a) *Which picture would you prefer on the SIMS Faces Page?* **The picture taken during this study.**

Participant 5

- I. (a) *Age:* **31**
- (b) *Gender:* **Male**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
4
- (b) *Have you ever directed others or been directed?* **No**
- (c) *If you have directed others, please elaborate below:*
- (d) *How many hours do you spend using a computer?* **8 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **50+ photos/month**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **6**
- (b) *Please comment on any instructions that worked well for you.* **Instructions were pretty clear and explicit.**
- (c) *Please comment on any instructions that did not work well for you.* **Name repetition was tedious.**
- (d) *How could we improve the interaction?* **Indicate when photos are captured. There was a long time during which photo might have been taken. A prompt like "taking photo NOW" or a shutter click would be good.**

- (e) *What did you like about the portrait camera experience?* **funny picture.**
 - (f) *What did you dislike about the portrait camera experience?* **Name recording was tedious – it didn't hear me for a long time. Too slow to take photo, unclear when the photo was taken.**
 - (g) *Is there anything that you would add to the portrait camera?* **maybe shutter noises when capturing photo, or some other audio prompt. Maybe allow user to see their video on screen while photo is being captured.**
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **It needs to get faster.**
- (b) *What would work well from the portrait camera for this application?* **Users get to pick from a set of photos.**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **Speed needs to improve.**
- V. (a) *Which picture would you prefer on the SIMS Faces Page?* **Both would be okay.**

Participant 6

- I. (a) *Age:* **23**
- (b) *Gender:* **Male**
- (c) *What is your primary language?* **English**
- II. (a) *Please rate your familiarity with portrait photography using the following scale:*
5
- (b) *Have you ever directed others or been directed?* **Yes**
- (c) *If you have directed others, please elaborate below:* **Often direct others, especially when strangers want their pictures taken in San Francisco.**
- (d) *How many hours do you spend using a computer?* **8 hours/day**
- (e) *How many pictures do you take with a film camera per month?* **1 or less photos/month**
- (f) *How many pictures do you take with a digital camera per month?* **50+ photos/month**
- III. (a) *Please rate the overall clarity of instructions offered by the portrait camera using the following scale:* **5**

- (b) *Please comment on any instructions that worked well for you.* **smile worked ok**
 - (c) *Please comment on any instructions that did not work well for you.* **standing on white footprints made my image unentered in the camera, but the image came out ok.**
 - (d) *How could we improve the interaction?* **Seems like the interaction could have been tighter (faster). The audio was clear.**
 - (e) *What did you like about the portrait camera experience?* **Kinda funny having a machine talk to you.**
 - (f) *What did you dislike about the portrait camera experience?* **Sometimes when someone is taking your portrait, they gesture or give off non verbal communication. Perhaps this is what is lost without a human "in the loop", as they say.**
 - (g) *Is there anything that you would add to the portrait camera?* **Sometimes it's a good idea to goof with the subject, even to go as far as taking a goof picture. It loosens things up and the subject feels more at ease. Seems like subjects often are waiting for something to happen and folks get a strained look on their face when they are waiting and trying to understand what is going on (what is taking time). Indeed, the sense of time a subject experiences is always more extreme when they are in front of a camera waiting for something to happen.**
- IV. (a) *What do you think about using the portrait camera for taking pictures for the SIMS Faces page?* **I think it's a good introduction to SIMS. It shows students that we are doing new things right off the bat on the first day.**
- (b) *What would work well from the portrait camera for this application?* **I don't really understand the question. I'm guessing you are asking what text would work well, script that is. Seems like it could ask what you are interested in. Or where you are from.**
- (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **Seems like it could go faster.**
- V. (a) *Which picture would you prefer on the SIMS Faces Page?* **Both would be okay.**

Participant 7

- I.
 - (a) Age: **42**
 - (b) Gender: **Male**
 - (c) What is your primary language? **English**

- II.
 - (a) Please rate your familiarity with portrait photography using the following scale:
7
 - (b) Have you ever directed others or been directed? **Yes**
 - (c) If you have directed others, please elaborate below: **I have done some amateur portrait photography.**
 - (d) How many hours do you spend using a computer? **8 hours/day**
 - (e) How many pictures do you take with a film camera per month? **1 or less photos/month**
 - (f) How many pictures do you take with a digital camera per month? **30 photos/month**

- III.
 - (a) Please rate the overall clarity of instructions offered by the portrait camera using the following scale: **6**
 - (b) Please comment on any instructions that worked well for you. **First and Last Name**
 - (c) Please comment on any instructions that did not work well for you. **"Family Name" was strange to hear.**
 - (d) How could we improve the interaction? **More immediate feedback would be good.**
 - (e) What did you like about the portrait camera experience? **I liked how it walked me through both the name and photo without much interaction on my part.**
 - (f) What did you dislike about the portrait camera experience? **Nothing!**
 - (g) Is there anything that you would add to the portrait camera? **An indication of when it was taking the individual photos. It was hard to know if it was done taking all the pictures, since it took quite a while and I didn't hear any feedback.**

- IV.
 - (a) What do you think about using the portrait camera for taking pictures for the SIMS Faces page? **I think it would be a good idea.**

- (b) *What would work well from the portrait camera for this application?* **Immediate feedback, people could pick the picture they liked, and bad pictures wouldn't be posted.**
 - (c) *What would need to be changed from the portrait camera to better fit the needs of this application?* **Speedier, and maybe a quieter place for the name recording.**
- V. (a) *Which picture would you prefer on the SIMS Faces Page?* **Both would be okay.**