

SaySee

An Audio-Video 2-way Communication Tool for Communities in Developing Nations

Research Proposal by Carla Gomez Monroy
Advising Professors: *Walter Bender & Michael Bove*

I. Real Life Problem

Rural communities in Mexico lack audio-visual communication tools to:

- Discover markets,
- Access markets,
- Reduce emigration by having properly paid jobs (suited to local knowledge and skills),
- Control their production which is dependent on or affected by natural phenomena,
- Reduce intermediaries who pay little, at times, less than production cost,
- Improve the educational system (distribution systems, collaboration, resources),
- Improve the health care system,
- Improve women's productivity and participation,
- Express needs and feelings,
- Share knowledge and resources,
- Organize collaborative projects,
- Obtain job opportunities
- Carry out e-commerce
- Have updated market prices in other places (prices of their products or needed raw materials).

Matters are worse for communities with poor transportation infrastructure.

Key issues to approach the problem

Improve **organization** and **collaboration** between communities, their emigrants, researchers, and governmental and institutional entities.

Key issues to solve the problem

Provide the necessary tools to enable communication and organization between the villagers and their exterior.

II. Research Problem

Can an audio-video communication tool empower rural communities?

III. Research Objective

The main purpose of the tool is to be a relatively low-cost and efficient means of communication that will contribute to empowering developing communities by enabling them to better express their needs and feelings, share knowledge and resources, and organize collaborative projects in spite of time differences and distance. This audio-video tool is expected to help deepen and broaden socio-cultural identities and business practices as well as to help maintain family ties between community members who live in their village and those who are away temporarily or permanently, e.g., migrant workers abroad.

Salient features are its relative low cost and its *Novice Ready* capability, that is, no computer skills are required to learn how to handle the tool nor to access the Internet because of its specific-user friendly interface that is tailored to suit the face-to-face communication preferences and scarce reading and writing skills of the target population.

Once having a system that bridges the communication gap in rural communities, it can be applied to many different purposes by adapting the system to each of them, for example, e-commerce and to obtain governmental and other institutional support.

Other features are of importance if we want to adequately integrate the digital environment into a certain community's reality through communication.

- Software or hardware features are tailor made for a specific community or purpose.
- Video images of people or goods (fixed shots, skippy video motion, smooth video motion).
- Direct and automatic access to the Internet and to specific Internet service providers and to the services themselves, for example, direct access to a chat room by just pushing one or two controls.
- Easy introduction to computers. For those that want it or need it, SaySee can lead them into ever more sophisticated computer and www knowledge and skills.
- Low cost long distance calls from developing communities to developed communities (the Mexico-USA call is much more expensive than the USA-Mexico call).
- Up-to-date prices and availability of their products and raw materials at different locations can be displayed visually and heard in their own language by just touching an icon.
- Advertising one's own goods, raw materials, expertise in similar ways to the above mentioned.
- Life-like online negotiating and bargaining capabilities because of the following feature.
- Voice and video (facial and hand gestures) and touchscreen/tablet scribbling and drawing capabilities.
- TV / radio online capabilities.
- Sharing local social, cultural, artistic, musical manifestations, e.g., paintings, singing, dancing, ceremonies, lifestyles.

Stages and Objectives: The tool will be developed by stages.

Our end users have:

- Little or null reading and writing skills,
- No familiarity with computers,
- Strong family connections (the stronger the feeling of togetherness the better), and a
- Wide range of ages.

The One Step Away Feel means that users have to carry out a minimum number of actions to achieve desired ends. I propose to design a tool that enables users to reach their contacts with just a few actions, something like a telephone where you just dial in a number after picking it up (or press one single memory key) to reach your contact. And when several actions are inevitable to achieve a certain end, I propose to design the tool in such a way that novice users feel that they are one step away from their desired end and never get lost or forget what to do next.

At present, to reach your contact you have to carry out a number of actions:

- | | |
|---|--|
| 1· Turn on the computer, | 8· Reach Chat service provider icon with mouse, |
| 2· Reach Internet Service provider icon with mouse, | 9· Click mouse, |
| 3· Click mouse, | 10· Log in, |
| 4· Log in (user name), | 11· Switch to password section via mouse or keys), |
| 5· Switch to password section via mouse or keys), | 12· Provide Password, |
| 6 Provide Password, | 13· Hit enter |
| 7· Hit enter, | |

These are, approximately, the number of actions. But most important are the added complexities of remembering or knowing what one must do every step of the way, e.g., in one or two of the previous actions, you must double click.

The Tool Must be:

- User-friendly
- Easy Plug and play installation
- Adapt to the culture (like not making them use another language)
- Cheap
- Reliable

IV. Some Why's and How's

A. Long distance communication is 300% cheaper.

Benefits:

- Lower telephone bill anxiety, which benefits communication.
- More independence to initiate communication because regular outgoing telephone calls are too expensive in developing countries.
- Neither party has the economic pressure to hang up quickly → in business this means better chances to negotiate or bargain → providing more eCommerce opportunities (getting more quotations on raw materials, transport, and distribution).

B. Audio, Video, Graphic, Textual Publishing: News, Advertising, Informative, Educational, Cultural Pieces

B.1. Incoming pieces can be heard and/or seen on regular screen, big screen, wall (projection). If only some get to see it, word gets round anyway.

B.2. Outgoing Audio, Graphic, Video, Textual pieces can be recorded locally and “published,”* that is, uploaded to an *ePool* [Appendix 3], the net, or broadcast by local, regional, national and international radio and TV stations.

{*References missing}

B.3. Access, Storage, and Updating: Both incoming and outgoing Audio, Video, Graphic, Textual pieces can be accessed, stored and retrieved later in order to benefit from them or to update or discard them. Individual filters and organizers channel the pieces of interest to personal mailboxes or archives.* Individual users or sets of users, e.g., community A, goat breeders, relatives of migrants receive in their personal mailboxes what is of most relevance to them. Some pieces can be stored in a database to be repeated for a certain period of time or at certain intervals.

B.4. Examples of Audio, Graphic, Video, Textual Pieces:

- Workshops in nearby communities (about how to graft a more resistant species of avocado into the local avocado trees or products).
- Price lists of local productions in the closest cities or communities, allowing for well-informed negotiation between producers and intermediaries.
- To post inquiries –or give advice– concerning productivity or diseases of animals or plants.

B.5. Examples of How the Tool can be Used:

- **On Screen Advertisement Pages** contain images (photos or icons) of what is on sale (a donkey, handcrafts, a plow). The user selects an image and the full audio-video advertisement takes place. Through the same means the user can immediately contact the vendor to initiate negotiations.
- **Visually Enhanced Negotiating and Bargaining** is possible by directly finger drawing, sketching, and scribbling on the touch screen-pad. For example, someone interested in buying the donkey on the screen can reinforce verbal argumentation by finger drawing, directly on the pad, a circle around the donkey's defective limbs and by crossing out the advertised price and scribbling –or selecting– a much lower price as an offer. Crossing out or erasing each other's prices can go on until they reach an agreement. Another possibility is offering and showing through a link or the video camera an item or merchandize for swap.

- **Collaboration with personalities, institutions, and other media** can augment the efficacy of the impact. Getting the collaboration of the local priest, politicians, educational institutions, radio, or TV stations could be a way to introduce the technology and/or to boost its use and acceptance. For example a politician or the priest can do the community a service by plugging the tool into their megaphones or PA systems so more people can hear and learn of the device or see it at work as it delivers broadcast music, stories, and news. If backed up by a large screen or wall projection, all the better because video can accompany audio and even TV shows can be delivered.
- **Radio and TV through the Internet** can be brought in tailored modes to meet specific likes and needs. For more on this, see Walter Bender's ideas at <https://www.media.mit.edu/research/group.php?type=researchGroup&id=10> under HDL.

V. Background

In this section, I briefly describe my motivation, preliminary research, beneficial effects of technology in developing countries, and related audio/video tools.

SaySee is an audio-video user-friendly interface for people with no or very little reading and writing skills and with no computer skills that live in a remote rural setting but have electric power and at least one telephone line.

My motivation began when I got to know the people of a community in a remote rural location. They proved to be very willing to try out ideas, participate in programs aimed at improving their development, and, most importantly, they have achieved a minimum infrastructure and level of empowerment. Nevertheless, they are still in dire need of so much that is factually inaccessible yet actually at hand, given certain preliminaries.

The preliminaries are exhaustively discussed in “The Pool Effect: More Access, Sharing, & Resources” [[Appendix 3](#)]. In that paper I describe the background and motivating factors that led me to realize that there must be many communities that have the necessary willingness and infrastructure and are just one step away from the benefits of eDevelopment. Nevertheless, this step has not been taken due to so many complexities inherent to the Internet, its vastness is a case in point, and to so many software and hardware complexities for getting on the Internet, specifically the written-medium, keyboard-dependent interfaces, and to so many logistic complexities of what to do with the internet and of how to do it.

The more general issues can be taken care of by *the ePool* and *the Pool Effect* [[Appendix 3](#)] –or some equivalent– and the more specific software and hardware issues can be taken care of by the tool I have named SaySee.

A preliminary study was done to see the feasibility of my ideas. My first ideas and findings are presented in [Appendix 2](#), which is a summary of interviews done, articles and theses read, and applications tested.

The World Bank Group and some other researchers stress the success of technology in several rural communities or organizations around the world to improve and expand their access to markets and opportunities [4].

Much of the work on audio/video compression and transmission has already proved successful. Little of this work has been done for networks of low bandwidth. However, the most pressing issues are logistic.

Examples are

1. How to contact some one in another village?
2. How to know when somebody is trying to contact you?
3. How to set up multi-conference appointments with people at different places?
4. What kind of interface would be most user-friendly, most intuitive for people with reading, writing, and computer skills that range from null to beginning?
5. What to do with the Internet? How to benefit from the Internet?, And

6. How to do it?

After having done this research, my main concern is that people with little or no reading and writing skills living in developing countries are, in fact, being left out from the available benefits of eDevelopment.

Therefore my main purpose is to develop an interface for people living in remote rural communities in developing countries while keeping in mind the importance of ‘introducing cold technologies with warm process.’[Appendix 3: References [15]]

The interface will focus on the necessities and capacities of a given target population.

VI. Hypothesis

An audio-video communication tool designed to meet the characteristics, abilities and necessities of rural people will trigger their economic, social, educational, and technological development.

G. Research Questions	H. Possible Solutions
<ul style="list-style-type: none"> • How can people in different locations set an appointment to talk to each other remotely? 	<ul style="list-style-type: none"> • Sending list of possible times to meet / reply –send until agree in certain day and time. • Timer-satellite (send time and am or pm) → pager • Caller Identifier of the person who called (will call back in x minutes)
<ul style="list-style-type: none"> • How to register for the first time? 	<ul style="list-style-type: none"> • Name • Save their picture / say password
<ul style="list-style-type: none"> • How to log in keeping preferences? 	<ul style="list-style-type: none"> •
<ul style="list-style-type: none"> • How to add contacts? 	<ul style="list-style-type: none"> • Phone number • Create Identity • Directory / approval
<ul style="list-style-type: none"> • How to know they are connected? 	<ul style="list-style-type: none"> • Ping / reply • Picture in a menu
<ul style="list-style-type: none"> • How to dial? 	<ul style="list-style-type: none"> • Multiple cards/ one card for each person you usually talk to. The cards hold all your information and the information of the contact, so as soon as you slip it in, the system takes charge of making the connection.
<ul style="list-style-type: none"> • How to choose whom to talk to? 	<ul style="list-style-type: none"> • Selecting a picture
<ul style="list-style-type: none"> • How is the user going to interact with the tool? 	<ul style="list-style-type: none"> • Audio Receptor Transmitter Handheld container • A container with speaker and buttons (controls)
<ul style="list-style-type: none"> • How many buttons? 	<ul style="list-style-type: none"> • Ideally, not more than 3 kinds •
<ul style="list-style-type: none"> • How to select options? 	<ul style="list-style-type: none"> • Voice, pressing buttons, touching the screen
<ul style="list-style-type: none"> • Buttons? 	<ul style="list-style-type: none"> •
<ul style="list-style-type: none"> • Figures/colors/voice? 	<ul style="list-style-type: none"> • Big buttons, different shapes and colors
<ul style="list-style-type: none"> • How to make it very easy to use and learn? 	<ul style="list-style-type: none"> • Few controls, few menus, intuitive
<ul style="list-style-type: none"> • How is handling know-how to be transmitted? 	<ul style="list-style-type: none"> • By watching how someone operates it
<ul style="list-style-type: none"> • How to input images? • How to register for the first time? • How to add contacts? • How does the tool work? (functionality) • What is it made of? (embodiment) • How does it communicate? (Interaction) • Does it learn? What does it learn? (learning) → how to transmit? how to use it? • What is it going to do? 	<ul style="list-style-type: none"> •

- | | |
|---|--|
| <ul style="list-style-type: none"> • How is it going to do it? • What does the user have to do? • How is the user going to interact? • What does the system do? • How does the user make selections? | |
|---|--|

By solving some of these problems of logistics and of functionality the system can do everything automatically for the following sessions.

VII. Delimitations and limitations

- Low resources
 - Low bandwidth (Telephone line) → not convenient for audio transmission, more complex with the video transmission.
 - Low costs of implementation, maintenance,
- Little or none writing and reading skills.

VIII. Assumptions

End Users:

- Little or null reading-writing skills.
- Family and friends connections are very important (the stronger the feeling of togetherness the better).
- Not used to computers
- Variety of ages

Infrastructure:

- At least one telephone line per town or Internet connection by other means (even though there are communities without access to communications.)

Tool:

- User-friendly
- Cheap
- Reliable
- Easy Plug and play installation
- Must adapt to culture (like not making them use another language)
- Something like the phone that it just rings and you answer it.

Other issues:

- Time zone (Different times in different parts of the world at the same moment.)

IX. Tool Design

Community requirements:

Communities should have:

- An available telephone line or some wireless Internet link (at least one or get one or be provided with one)
- Willingness to develop and to try out new things (devices and processes)

Physical Installation of the Tool

1. The location of the tool depends on number and location of available telephone lines or Internet connection by other means (*e.g.* satellite).
2. The tool is plugged into available telephone lines (houses and establishments) or Internet connection. If the community only has one single line, the tool could be placed in a public booth, and the local system of letting people know about their calls is used. (A simple local network of ringing devices could reach each participating home so villagers can run to the booth to answer their call.)

Physical device:

In *Appendix 1*, I outline some of the interfaces I considered in order to design a tool with the above mentioned features.

The interface of my choice is a large flat touch screen with all the physical computational devices (See: *A Single Device* in *Appendix 1*). All the area of the screen is touch sensitive and can be used for choosing options. The device itself includes a camera, two speakers, a microphone, and inlets for many other devices that can be attached to it, such as, a bigger screen, projector, audio system, telephone line, electric power (to keep it on stand-by mode and to charge battery), a CPU.

Connection

To make long distance calls cheaper the most economical medium is Internet-only, that is, parties at both ends are on the Internet. Other less economical possibilities are Internet-telephone and telephone-Internet-telephone.

Functionality

The tool will serve two kinds of purposes: private communications and public communications.

- **Public** communications can be delivered to the whole community.
 - Advertising
 - Price lists
 - Products and services
 - TeleMeetings (with people from other places to organize an event or solve a problem jointly)
 - Cultural, educational, informative, fun stuff
- **Private** communications are just seen or heard by one or a few people.
 - Long distance call communication (relatives, friends, business, inquiries)
 - Other Internet possibilities
- Make the software and hardware do almost everything
- Communication could be any: 1-1, 1- M, or M-M
- Transmission of frames every time there is space on the connection.
- Transmission of silhouettes, using the idea of Myron Krueger and Telemurals of Karrie Karahalios of the Sociable Media Group.
- Through the tool, send a text message to a pager, or an audio or text message to a cell phone.
- The tool should have a minimum of buttons. For further or more sophisticated functions there is the possibility of setting options, personalization, and menus. There also are inlets for keyboard and other devices.

Interaction:

In this section, the functionality of the device depends on the implementation stage of the tool. Implementation will go from simple or single functionality to complex or multi-functionality. By complex or multi-functionality, I mean that the novice user can evolve and become a sophisticated user of a sophisticated tool and use the tool for a wide variety of tasks. Improvements of the tool will aim at automating it as much as possible to make it as simple to use as possible.

i. Log in Possibilities

In this first stage the goal will be to start the 'Reaching Your Contact' application as soon as the device is turned on.

The application shows the picture of several possible contacts. By just clicking on the picture of your choice you reach your contact, that is the tool does all the work of going through the necessary steps to log in, etc. Alternatively, instead of seeing and choosing pictures, a one-word spoken command (the name of your contact) automatically makes the tool go through the necessary routine so you reach your contact with only one action.

I have considered implementing in a later phase, some kind of recognition system or some other system to add some security to the system, avoiding usernames and passwords because our target user may well find it difficult to adapt to that kind of procedures. Even frequent computer users, at times, forget their user names or passwords or some necessary step or sequence. At present, security per se is not the issue since I am assuming there would be no interest in invading other people's directories, but very easy automated access could lead to accidental or naïve entering other people's domains.

ii. Contacting someone

Logging In & Seeking My Contact

After selecting the picture of himself in the 'Community Picture Album,' that is, after logging in, the user will see his personal picture album, that is, 'My Picture Album,' containing the pictures of his contacts. While going through the pictures, the picture being pointed at, at any given moment, is amplified. Once a contact is selected the tool automatically seeks him or her out.

Contact Options

For those familiarized with regular telephones, there will be the option of dialing a given person's phone number directly on the tool. This option may come in handy if something goes wrong with the video, the screen, or the system.

Contact On Line

Once a picture in 'My Picture Album' is selected, the system will make the connection. The user can recognize the contacts that are on-line because they will be flashing quickly and they will be calling out, "Hi, Mike. I'm here."

Contact Off Line: Voice Mail

If the contact is not on line a new screen for recording an audio/video message will automatically pop up and say, "Mike's not here. You wanna leave a message? Yes? (Flashing icon and letters standing for 'Yes' etc.)" The screen will have iconic options for recording, stopping, replaying, canceling, and sending, plus audio cues that lead the user, step by step, into recording, etc. his own voice mail.

But before leading the users into recording their own messages, the system will offer a few personalized prerecorded messages that can be customized right their and then, e.g. "Hi, *Ann*. This is *Mike*. See you tomorrow night at 9." (*Italics* indicate personalization done by the tool automatically. Underlined indicates customizing, that is, options user selects.) This message can be saved for later use.

You Got A Message

If they are not online but have left a message, they will be flashing slowly and saying, "Hi, Mike. I left you a message." The audio is optional and it can be turned on and off, or a given picture can be selected to sound louder than all the others.

Making New Contacts

If a picture from the 'Community Picture Album' is selected, it will be saved in the user's 'My Picture Album' and the tool will initiate contact procedures.

The following features will also be included in later stages.

- Latest contact is First in 'My Picture Album' the next time I open it.
- Most frequent contacts are also nearer to the beginning of 'My Picture Album.'
- Search by image, audio recognition of the names, or text
- My contacts can be reached to their regular telephones through my machine to make the call cheaper.
- They, in turn, can dial back from their phones to my machine.
- Messages can be posted on the Community Bulletin Board by clicking the corresponding icon.

System Assumptions

Audio/video messaging (either for private or public transmission) can be recorded in real time and transmitted by chunks or entirely (if possible) later when there is space on the network.

On the screen, I see all the pictures or video transmission of all the participants in the conversation.

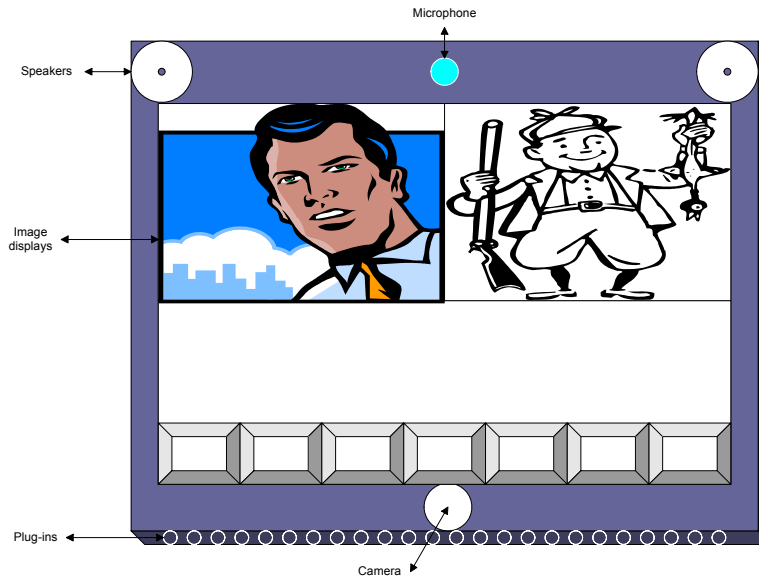
GENERAL DIRECTIONS: If nothing is going on, on the screen and the user doesn't know what to do, then the screen should be touched to see if it changes, does something, or gives some advice.

????Pattern Recognition????

How feasible is it for the system to take a picture of you and compare it with the last one taken –or a set– to see if you are the same person? How hard is it to determine whether a new picture matches a given person? Will the patterns distort the distinguishing features of a person? Is it good or not since I was told that sensors and pattern recognition are not very reliable yet? And what about fingerprint, iris and voice recognition????

Appendix 1: Interfaces that I have thought over

A. A SINGLE DEVICE



Visual Language

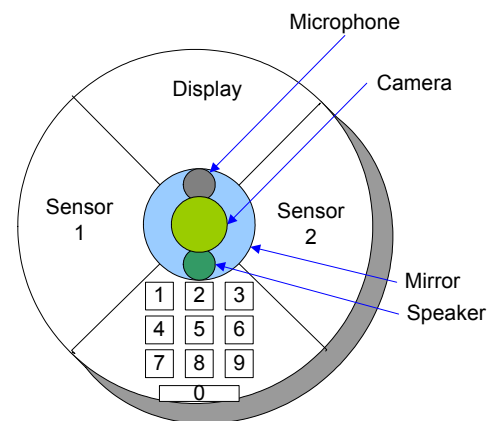
The Single Device can also be used with a visual language. At the beginning the system can have a few important words, say 30. The keys on the touch screen can be incremented if the addition of new vocabulary is required; or reduced if some of the keys are not being used. The keys-section of the screen can be divided into the number of concepts that are most used. The concepts can be grouped by people, things, verbs, and modifiers. This could be used to speak across languages since the vocabulary is reduced and people could agree on common meanings reinforced by pictures or icons.

Each machine is personalized locally so the system can convert the common visual language contained in the keys into different local spoken languages.

B. CONTROLLERS:

a. Like a CD

- Size: CD diameter
- Thickness: at most 5 cm.
- Microphone and camera covered by the polarized Mirror (view from the inside to the outside)
- Middle part fixed
- Sensors move to the sides and can be pressed down
- The numbers are to dial the phone
- Sensor 1 & 2 are directional and to select inputs
- Mini display? Willy's? It can work without it → if the displayer is not available, the camera still transmits my video, and others can see me even though I can't see them.



b. JOY STICK

i. Physical:

The tool will be one Box, one Control, and the camera and microphone could be attached to the Box and to the Control respectively.

- **Display:** the tool can be plugged into any displaying device (TV, computer screen, projector)
- **Speakers:** the tool can be plugged into any audio device with a line-in, audio-in, CD-in, or mic-in jack (PA system, TV, VCR, DVD player, passive, active or stand-alone speakers) or, via a cassette-like adapter, into any cassette player (walkman, stereo, car stereo).
- **Box:** where the processes run.
- **Control:** tool (Squeezable Joy-stick: for selection of images. Squeeze will work like clicking a mouse.)

ii. Interaction: (One Scenario)

Directions: “If you do not know what to do or what is going on, squeeze or move the stick until you see what is going on“

Person	Tool
<ul style="list-style-type: none"> ▪ He* Looks at the camera lens. 	<ul style="list-style-type: none"> ▪ Displays video
<ul style="list-style-type: none"> ▪ He Squeezes the stick 	<ul style="list-style-type: none"> ▪ Freezes image ▪ Asks, “What is your name?”
<ul style="list-style-type: none"> ▪ He Says his name 	<ul style="list-style-type: none"> ▪ Asks “What is your Secret Word (password)?” (not sure how to contextualize it, and assure the security of the system, yet) → a secret word, game of colors? ▪ Changes display to “Contacts” ▪ “Dials the phone number” (first time no contacts)
<ul style="list-style-type: none"> ▪ He Dials (directly in the phone) 	<ul style="list-style-type: none"> ▪
<ul style="list-style-type: none"> ▪ She answers 	<ul style="list-style-type: none"> ▪ Video transmission ▪ If she is wasn’t on line, then, while talking, she has to go through the same process as He. When everything is ready, ▪ Her image freezes at both sites. (Settings are saved in both machines) And his image also freezes at her site. ▪ Video transmission starts. (Voice was always being transmitted.)

* He is in one town\\She is in another town

c. Jog/wheel Knob

Button projecting the image (photo / video) of the remote person

Scroll the button to look for the image of the person you want to talk to / choose by stopping and pushing to dial,

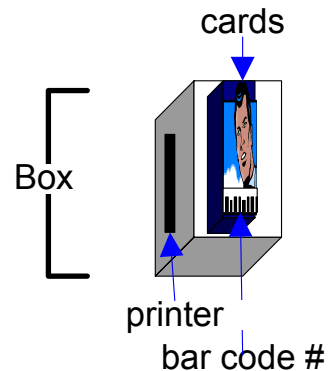
B. UNIQUE INSERTABLE CARDS

It generates a card per phone number / person (including yours)

Using Bar code / scanner.

Box: (like Nintendo) where the processes run. Input and output devices can be attached to the Box.

- **Display:** the tool can be plugged into any displaying device (TV, computer screen, projector)
- **Speakers:** the tool can be plugged into any speaker device (not exactly any: see Joystick section above)
- **Mini printer?** To generate a card with a picture of the contact and his bar code number (the picture is for the human to identify his contacts, and the bar code number for the machine). The bar code number contains the telephone number and picture of the contact (probably the spoken name)
- **Scanner** to read the bar codes of the cards.



Alternative ‘Insertables’: Magnetic-Strip cards, Chip cards, Diskettes: On each ‘Insertable’ is the picture of each one of my contacts. Each Insertable is like a unique key that gives me access (logs me in) and opens the way to one specific contact.

FIRST TIME USER (One scenario)

Person	Tool
<ul style="list-style-type: none"> ▪ He* looks at camera lens ▪ He places an empty card in the slot. 	<ul style="list-style-type: none"> ▪ Displays video (himself)
<ul style="list-style-type: none"> ▪ He presses Sensor2 ▪ Or if Sensor2 is not pressed, after 5 seconds 	<ul style="list-style-type: none"> ▪ Freezes image ▪ Prints card ▪ Commands “Dial the phone number”
<ul style="list-style-type: none"> ▪ He dials his contacts phone number 	<ul style="list-style-type: none"> ▪ Dials the internet service provider
<ul style="list-style-type: none"> ▪ She answers 	<ul style="list-style-type: none"> ▪ Video transmission ▪ If she is connecting for the first time, then, while talking she has to go through the same process as he. ▪ When everything is ready ▪ Her image freezes at both sites. (Settings are saved in both machines) And his image also freezes at her site. ▪ Video transmission starts. (Voice was always being transmitted.)

* He is in one town\\She is in another town

CONTINUOUS USER (One Scenario)

Person	Tool
<ul style="list-style-type: none"> ▪ He looks at camera lens ▪ He places the card of the contact on the slot. 	<ul style="list-style-type: none"> ▪ Displays video (himself until the contact is transmitting video) ▪ Machine logs him in and seeks contact. ▪ If she is available, the connection is established.
<ul style="list-style-type: none"> ▪ He or she extracts the card to finish the conversation. 	

Appendix 2: Preliminary Research

Preliminary Ideas:

My initial idea was regular audio/video transmission, considering that rural connections are through telephone lines.

Michael Bove's idea, regarding the constraint of the bandwidth, was to do an audio/video messaging system. I started thinking on how it could be. This idea looked to me like the best general approach and it could be fun to send audio/video messages in an asynchronous way.

But, I still wanted to do something with video transmission in real time.

Probably making the audio part very reliable and then have some kind of video. Both, audio and video should be compressed as much as possible.

I tried to think in many different ways of getting some kind of video transmission, such as, the transmission of only a physical pattern of the person and the parts that are being moved, and then reanimate the image in the receiving machine. The reanimation could be that movement would be added onto the real image, maybe with animated patterns or avatars that maintain the main characteristics of the person.

Some of these ideas could be either not reliable or too hard to implement. Since the problem is transmission, I considered sending a minimum of video data.

The following chart shows some of my findings.

Preliminary research:

Name	Topic / doubts	Research and Things Learned	Contact
ISIS / ICom	Tool to develop multimedia applications.	I read Stefan's theses. I tried Isis and learned its basics. I also got in touch by mail with him a couple of times. Stefan told me that Isis only handles raw audio, but there is a high probability that I can use Isis in addition to some other compression libraries. But, definitely not to try to re-invent the bandwidth conservation strategies that already exist for online audio/video conferencing.	Human Connectedness Stefan Agamanolis stefan@media.mit.edu
Convivo	Communication tool (voice) for rural communities	I read his thesis. His thesis is about Reliable Voice Communication in poor network performance and about a multi-literate graphical user Interface for users with limited literacy and computer skills.	eDevelopment Marco Escobedo
Audio – Video	Audio-video transmission Video abstraction	Karrie is working on the Telemurals project, it is done in Isis and C, running in Linux and uses UDP. She told me that just using video (gif) compressed the band width. It would be too little. → Compression lowers quality. {we would not use her text, nor her sound parts}. Silhouette movement (back - front), change colors with the voice volume → to give life to the characters locally. The two people share the same digital space (at each end there could be more than one person, but it is a one to one connection) Talk to Stefan about audio – audio/video compression http://web.media.mit.edu/~kkarahal/thesis/projects/telemurals.html	Sociable Media Karrie Karahalios kkarahal@media.mit.edu

Pattern Recognition	Eye movement recognition (send only the movement)	<p>Ashish Kapoor recommends:</p> <ul style="list-style-type: none"> • Not to use sensors because they are not reliable • Nor speech recognition processing • Applications that use text-image (to animate them) → trying to convert them into speech-image will make it monolingual. [people working on that Tommy Ezzat (his advisor is Tommy Poggio) at MIT] <p>His recommendation is to send less frames per minute or second, reduce their resolution, jpeg files, compress them. Make voice very reliable.</p>	<p>Affective Computing Ashish Kapoor ash@media.mit.edu</p>
Artificial Reality	Silhouette transmission	<p>Reviewed : http://bubblegum.parsons.edu/~praveen/thesis/html/wk05_1.html</p>	<p>Myron Krueger praveen@parsons.edu</p>
Networks	Rural networks	<p>I talked to Alex Pentland. His recommendations: It depends on the place where the tool is going to be applied. Recommended use</p> <ul style="list-style-type: none"> • For communication → messaging • For e-commerce → voice • For education → picture & voice <p>Video is not needed, not useful, it does not work, even with the best resources available. No animated photos. Rural networks are pretty bad. General suggestions for my system:</p> <ol style="list-style-type: none"> 1. Asynchronous systems 2. Conferencing (more than 2 people) [one at a time] like a T.V. show, and each participant with its own microphone to keep track of who's talking 3. Make a Web Site to post video clips 	<p>Human Design Alex Pentland sandy@media.mit.edu</p>
Animation	Video local reanimation of video (sending data of the parts that the person moved or animate the pattern of the person) Reanimate photos	<p>I talked to Hannes Hogni I read about these tools and tried some of them:</p> <p>Animated faces: http://emote.3dme.com/emote/ http://www.lifefx.com/main.html</p> <p>Avatars: http://avatars.www.media.mit.edu/avatars/ http://www.ccon.org/ http://www.digitalspace.com/traveler/index.html http://www.smartvr.com/ http://lindenlab.com/ http://www.prod.there.com/</p>	<p>Gesture and Narrative Language: Hannes Vilhjalmsson hannes@media.mit.edu</p>
Speech	Speech-text-speech Speech recognition (to train it through repetitions) Speech menus	<p>Gerardo told me that:</p> <ul style="list-style-type: none"> • Speech recognition for regular conversation is not recommendable. • Speech recognition with a limited vocabulary (menus) works pretty well. • Speech-text-speech not reliable and very limited (not Multilanguage) <p>Recommended I look for:</p> <ul style="list-style-type: none"> • LPC (Linear Prediction Coding) • APCM • Speech coder-decoder (sampling is implicit) <p>For voice → periodic pulses: vibration of the vocal cords in the presence of airflow.</p>	<p>Speech Interfaces Gerardo Vallejo gvallejo@media.mit.edu</p>

Some REFERENCES of articles and theses I read:

1. Agamanolis, Stefan. High-Level Scripting Environments for Interactive Multimedia Systems. S.M. Thesis, MIT, February 1996.
2. Agamanolis, Stefan. **Isis, Cabbage, and Viper**. Ph.D. Thesis, MIT, June 2001.
3. Escobedo, Marco. Convivo Communicator: an Interface-Adaptative Voice over IP System for Poor Quality Networks. S.M. Thesis, MIT, June 2002.
4. Jain, Ramesh. "Folk Computing". Communication of the ACM. April 2003/ Vol. 46, No. 4, Pages 27-29.
5. Narayan, Deepa and Talat Shah. "Connecting the Local to the Global: Voices of the Poor". December 11, 2000. Local to Global Connectivity for Voices of the Poor Workshop, Washington, DC.
6. Gould, John and et. al. "Making Usable, Useful, Productivity: Enhancing Computer Applications." Communication of the ACM. January 1991/ Vol. 34, No.1, Pages 74-85.
7. The World Bank Group. "Information and Communication Technologies." Washington, D.C., 2002
8. Goetze, Marcel and Thomas Strothotte. "An Approach to Help Functionally Illiterate People with Graphical Reading Aids." University of Magdeburg, Germany.
9. Rondot, Pierre and et. al. "Empowerment Tools and Practices." The World Bank Group.
<http://www.worldbank.org/poverty/empowerment/toolsprac/tool13.htm>
10. Kenny, Charles. "The Internet and Economic Growth in Least Develop Countries." World Institute for Development Economics Research. August 2002. <http://www.developmentgateway.org/node/133831/sdm/blob?pid=3951>
11. Joshi, Anirudh and Shama Rasal. "Interaction Paradigm for Semi-literates and illiterates." December 2002, http://thinkcycle.mle.ie/thinkcycle/main/development_by_design_2002/publication_interaction_paradigm_for_semi_literates_and_illiterates/final.pdf
12. Humphrey, John and et. al. "The Reality of E-commerce with developing countries." March 2003. <http://www.gapresearch.org/production/Report.pdf>

Appendix 3: The Pool Effect: More Access, Sharing, & Resources
<http://web.media.mit.edu/~carlagm/papers/PoolEffect.pdf>