Muhammad Obaidullah

Research Papers Summaries

CSE B10

Mnemonic Techniques for Self Identification for Illiterates

By Dinesh S Katre:

This technique was tested on villagers & 2000 were the subjects. Subjects took pictures according to their fields and knowledge. 6 pictures and 1 color scheme was taken as identification. 9 picture per screen were used. After selection of pictures, make them remember to the user by telling him/her the sequence. Specific location and size does matter for selection of pictures. Pictures with simple backgrounds were preferred. The gender base choice of pictures need to be addressed.

Exploring the Feasibility of Video Mail for Illiterate Users

By Archana Prasad, Indrani Medhi,

Kentaro Toyama & Ravin Balakrishnan:

User needs to set up his account. In future, he needs to identify himself uniquely and remember his identification. Graphics were preferred over text for identification. Consistent help for setting up the account and for future use is required. Welcome message has got a good effect on users. Users were divided into certain categories like literate, semi-literate etc. numeration idea for illiterates was also proposed. Users preferred video mailing instead of others. Abrupt page changes made the user fed up. Instead it should be animated/slow. Picture quality should be high and size and position on the screen does matter. Many users remembers digits by their specific shape and constant position. Precise instructions were preferred instead of a length one.

A solution for a mobile computing device along with supporting infrastructure for the needs of illiterate users in rural areas

By Joy Bose, Dipin K. P

Focus on mobile computing technology for illiterates. Proposed a way for modifying mobile application for illiterates. Applicable in the field of health care, education, agriculture, NGOs projects, facilities like maps, public transport and government registration. The proposed solutions are:

* Image/animation driven interfaces; Subject + verb + object. Images are stored against these actions in the database. System will construct the sentence by the combination of relevant images (Artificial Intelligent System).
* Speech/Voice Driven Interface. The speech recognition software will hear the voice, parse it and construct the sentence based on what it has heard.(NLP system)
* Hybrid system (image + oral + animation + query based). On the user selection, the system will ask some relevant queries which can be answered in a single word. On the basis of the answer it will construct the sentence.

These could be applied to the following application in mobiles.

Telephony: the caller can speak out the name of the person to call, and the software on the phone can either match the voice with the pre-recorded name of the contact when it was first spoken and stored by the user, or perform speech recognition to identify the name of the contact to call. In the case where a user has just received a new call and wishes to add the caller to the contacts list, the user can speak the name and assorted details (male or female, organization etc) of the person. In this way no text needs to be input.

SMS: By constructing a sentence based on user selection of image or animation icons representing names, places, actions etc. The phone will have a dictionary of words along with image icons, which can be a subset of the entire language dictionary based on the profile of the user. The vocabulary subset for a farmer, for instance, would be different from that of a weaver.

By constructing a sentence based purely on speech recognition. The user speaks the message to send and the speech recognition system can convert the speech to words. The advantage of this approach is the widespread availability of speech recognition solutions. However, their accuracy may vary especially for unfamiliar accents and languages.

Browser: One way is by introducing a reader for the webpage. On selection of a particular portion on the page with gestures, the device should start reading that out. Whenever the reader encounters any hyperlinks it should stop reading and prompt the user with voice for getting the user input on whether the browser should load the hyperlink or skip(image blinking).

Another method can be using a custom browser to render only the relevant information in audio or video format for use by those who are unable to read the text. Hence we propose a new browser solution which is limited in capacity that can render only a few relevant tags including audio and multimedia as compared to the full featured browser. The audio/video tag will have a proposed new attribute called text free browser whose value can be either yes or no. This new attribute will be ignored by normal browsers whereas the custom browsers in our phone will only render this and

nothing else.

Theses can also be applied on video chat and radio. They have considered some browser based modifications to render web pages in the form of speech or images, text free voice or image based solutions to construct text messages to send as SMS, a telephony solution without text, and also looked at what kind of infrastructure might be needed to support such a system in villages, with one or more centralized WiFi hotspots serving a variety of rural users.

Optimal Audio-Visual Representations

for Illiterate Users of Computers By Indrani Medhi, Archana Prasad, Kentaro Toyama

This presents research leading toward an understanding of the optimal audio-visual representation for illustrating concepts for illiterate and semi-literate users of computers. 200 illiterate subjects each of 13 different health symptoms in one representation randomly selected among the following ten: text, static drawings, static photographs, hand-drawn animations, and video, each with and without voice annotation. (1) Voice annotation generally helps in speed of comprehension, but bimodal audio-visual information can be confusing for the target population; (2) richer information is not necessarily better understood overall; (3) the relative value of dynamic imagery versus static imagery depends on various factors

1) DESIGN ITERATIONS: It is good to try to compare videos against photographs, but what should the videos be of, that our subjects would respond meaningfully?

1.1) Choice of Domain: Health Symptoms; 1)Admit to visual representations, 2) Consist of information which they are not exposed to on a daily basis, 3) Have universally recognized meaning, 4) Offer a range of cognitive complexity, and 5) Offer a range of visual complexity.

1.2) Culturally Appropriate Representations: To accomplish this, we involved a sample design group of subjects, who were part of the same community from which we drew our test group, in our design process

1.2.1) Generating Gestures for Graphics: Each participant performed the exercise in isolation, without interaction with other participants. Two techniques were used to elicit the appropriate gestures from our subjects: (1)2D paper cut-out dolls: The participants of the exercise were asked to depict a given symptom using these paper dolls…(2)Enacting: The participants were asked to enact the given symptom without using verbal communication.

1.3) Consistency across Representations: To arrive at some consistency, they undertook the following process: Video, Photograph, Animation, Static Drawing.

1.4) Designer Biases: he designers felt that a dynamic mode of representation (animation) to depict swollen feet was better because a time-based progression from a normal foot to its extreme swollen state would be far more informative than a static visual. Results were different for different symptoms.

2) Quantitative Results: Based on the data we collected from user studies, they have the following findings supported by the standard Students t-test of statistical significance.

2.1) Overall, the five representation types with audio were much better understood than the representation types without audio. All visual representations with voice annotation were better recognized than the corresponding representations without. With audio, results show that accuracy is greater in all versions.

2.2) They found that among the cases with audio, hand-drawn representations accompanied by voice annotation was the most accurately recognized. They suspect that one reason why hand-drawn static images and animations fared well was because semi-abstracted drawings in which only the essential information is depicted is better grasped than photorealistic imagery that contains extraneous visual features.

2.3) The cases without audio, animation was the most accurately identified among the five representations without audio. It resulted in 7.35 correct identifications out of 13.

3) Qualitative Results: During their tests they made a number of qualitative observations, as well. These results were not statistically significant, but they present them here as qualitative data.

3.1) Richer information is not necessarily better understood overall: static drawing proved to be better as the audience were confused by the additional rich information (of intense physical activity) before the primary action (principal actor feeling weak) and thus drew inaccurate conclusions.

3.2) Unrelated context laying activity changes value of static versus dynamic imagery: The pattern observed, with the visual representations without audio assist, was that placing a contextlaying activity that was unrelated to the principal action cue caused confusion amongst the subjects. For example, while describing the symptom of mild fever the video used the context laying activity of an attendant wringing the cloth and then placing it as a cold compress on the fevered forehead of a patient. The photo representation of the same symptom showed only the action cue of the principal actor (the patient) lying down with a cold compress being placed on forehead. The latter version was more clearly interpreted by the subjects

3.3) Socio-economic conditions correlated with cognition: They noticed that subjects from the economically better-off slums who typically had more formal education performed better overall throughout the experiment, generally responding more quickly, accurately, and with greater confidence.

3.4) For some, audio-visual cognitive coordination was difficult: In the representation types withvoice annotation, we saw that some of our subjects were unable to fully process the bimodal stimulus. They inevitably used a strategy of focusing either on the graphic or the audio, but not both. When they tested the text+voice representation with totally illiterate subjects, they listened intently to the audio cue, and ignored the on-screen text completely

4) Experimental Issues: A number of issues came up during the experiments which could have biased the results in one way or another

4.1) Short-term category conditioning: if the first three symptoms presented in were headache, backache, sore throat, one after another, then the user was more likely to perceive the next symptom presented as an ache or pain of some sort

4.2) Test/text anxiety of subjects: Subjects who had some basic level of schooling exhibited significant nervousness when they were shown the text based version, as they seemed to feel the same pressures they may have faced in school.

SpeechRecognition for Illiterate Access to

Information and T echnology By Madelaine Plauché, Udhyakumar Nallasamy, Joyojeet Pal, Chuck Wooters, Divya Ramachandran.

Speech recognition has often been suggested as a key to universal access, but success stories of speech-driven interfaces for illiterate end users are few and far between. The challenges of dialectal variation, multilingualism, cultural barriers, choice of appropriate content, and, most importantly, the prohibitive expense of creating the necessary linguistic resources for effective speech recognition are intractable using traditional techniques.This paper presents an inexpensive approach for gathering the linguistic resources needed to power a simple spoken dialog system. In their approach, data collection is integrated into dialog design: Users of a given village are recorded during interactions, and their speech semi-automatically integrated into the acoustic models for that village, thus generating the linguistic resources needed for automatic recognition of their speech.

A spoken dialog system (SDS), which would allow users to access information by voice, either over the phone lines or at a kiosk, could play a role in overcoming current barriers of cost and literacy. Speech-driven user interfaces (UI) are cheaper than display-based UI solutions and more accessible than text based UI solutions. peech technologies, such as automatic speech recognition(ASR), require the collection and hand annotation of a large corpus of speech and a dictionary of all possible words in the language with all possible pronunciations for each word. The availability of linguistic resources, such as training data and pronunciation dictionaries, which are arguably the most costly part of development. In this paper, they investigated the possibilities for a speech driven UI design for users in developing regions who may have never been exposed to formal education. They proposed a multi-modal interface that requires no training and provides users with a variety of visual and audio cues and input options, allowing each user the flexibility to interact with the application to the best of their ability and the system's ability. Participants valued accuracy of information and time of interaction, which averaged about three minutes. Participants expressed pride and excitement at hearing a computer speak their language.

Recording illiterate speakers saying the words for digits 0-10 took approximately six times as long as recording the same data from literate speakers. This discrepancy was due to difficulties explaining the task, limitations in the protocol (no reading aloud), inflexible and demanding occupations of participants, and apprehension involving participation, resulting in many missed appointments. They sought to explore with our field recordings was whether illiterate speech varied substantially from literate speech. Unfortunately, in our data, the differences between literate and illiterate speakers were due to differences in elicitation protocols and environmental factors rather than linguistic characteristics of speech production. We attempted an adaptive experimental protocol but failed to anticipate the cognitive difference between the task of determining a number of displayed fingers and the task of reading numbers aloud. Touch screen combined with digital photographs or graphics produced by a local artist is an ideal input for users of all literacy levels.

The Local Language Speech Technology Initiative– localisation of TTS for voice access to information Roger Tucker,

Ksenia Shalonova

Phone-based services in local languages would allow widespread access to medical, market pricing, weather and other timely information, as well as email and on-line transactions. Such services are dependent on two types of language technology – textto speech (TTS) to deliver information, and automatic speech recognition (ASR) to access it and control its delivery. The Local Language Speech Technology Initiative (LLSTI) is solving this problem by bringing together motivated groups in developing countries, providing tools, expertise, support and training to enable first TTS and then ASR to be developed in their own local languages. For command-and-control ASR, it is mainly the very time-consuming task of collecting and annotating a lot of data. But a good quality TTS system requires deep knowledge of the morphology, syntax and phonology of the language as well as the technical skill to build the TTS system. A single TTS system can cover quite a large region (using a neutral dialect, as is used for broadcasting), but ASR requires extensive data collection throughout the region it is intended for.

E-Inclusion Innovation for Rural India:

Mobile Voice and Tablet Based Educational Services

By Mikko Ruohonen, Markku Turunen, Juhani Linna, Jaakko Hakulinen,

Amit A. Nanavati, Nitendra Rajput

In this paper, they study the deployment of voice-based mobile educational services for developing countries, especially in India. Our research is mainly based on a Spoken Web technology developed by IBM Research Labs. In this paper we focus on educational services.

Spoken services have been used successfully in many areas in many countries. Typical examples of speech applications include transport information services, such as automated train and bus timetable services. The first speech applications were telephone-based interactive voice response (IVR) systems, which used speech outputs and telephone keys for interaction. These applications are still popular and possibly the most important examples of widely used commercial speech applications. Typically, these applications are designed to replace human operators. Problems with these systems include that the DTMF (colloq. touch-tone) interface may be awkward and user satisfaction poor, if the service does not match the quality offered by human operators.

The current trend is that DTMF inputs are replaced by speech inputs. IBM Research Labs have developed the Spoken Web to deliver data and information to illiterate people. In a nutshell, Spoken Web content is stored in the form of voice-sites instead of text based web-sites. The content is in local dialects, making it much easier for illiterate people to access this information. It has been observed that IT systems with a voice-based feedback have much more appeal for the illiterate and semiliterate population of these regions as compared to graphical user interface (GUI)-based systems such as Internet Kiosks.Voice-links allow for navigation between voice-sites using voice commands from a limited permissible vocabulary set. The interconnection of voice-sites leads to a WWTW or World Wide Telecom Web on the lines of the well-known World Wide Web. Voice-sites can be accessed by dialing a phone number. The much higher access to mobile phone connections, therefore, makes access to voice sites much easier.

Content Management and User Interface

for Uneducated People

By Zainab Mahmood, Syeda Sana Shahzadi and Sahar Tariq

From literature review we found that many research have been conducted to provide interactive interfaces for Un-Educated users. Un-educated people are certainly part of the group which has been referred to as the “Information Poor”. With weak reading skills they cannot navigate, explore, and use the web effectively and understand information presented there. A detailed review of the accessible literature leads us to the question “How to enhance the information gathering (from web) method for un-educated user?” Visually with pictures,

video or text, acoustically with sound or spoken Language considering regional languages (Voice Directions gives you turn by turn navigation instructions in voice), icons and menus.

• To examine the rural community’s need of information gathering

• Provide interaction styles for information presentation

Text-Free User Interfaces for Illiterate and

Semi-Literate Users

By Indrani Medhi, Aman Sagar,

and Kentaro Toyama

They describe work toward the goal of a user interface designed such that even novice, illiterate users require absolutely no intervention from anyone at all to use. They built two applications using these principles, one for job search for domestic labourers, and another for a generic map that could be used for navigating a city. Results also show that the text-free designs are strongly preferred over standard text-based interfaces by the communities which we address, and that they are potentially able to bring even complex computer functions within the reach of users who are unable to read. The work presented in this paper was motivated by a single goal: to provide useful applications to communities of illiterate users, with a user interface designed such that even novice, illiterate users required absolutely no intervention from anyone at all to use. t the goal does not seem that far out of reach.This paper presents two applications which were designed to ferret out the requirements for a text-free user interface. In the first, an employment-search application, the intent is to provide job information to a group of domestic labourers. In the second, we explore a text-free UI for maps that should allow illiterate users to answer questions having a geographic dimension.

1. Avoid text (but using numbers may be okay). b) Use semi-abstracted graphics, photorealism with deeper interaction. c) Pay attention to subtle graphical cues. User response may depend on psychological, cultural, or religious biases. d) Provide voice feedback for all functional units. e) Provide "help" on all screens.

Subjects recognized semi abstract cartoons and more photo-realistic graphics much better than complex abstract graphics. Too much abstraction could cause some difficulty. For example, in the map application, animated arrows for depicting one- or two-way traffic was not readily understood. When the arrows were replaced with small icons of cars, subjects immediately understood the meaning. actions may require a visual representation, or they would be taken as static representations of location or object. We found that they were better able to identify activities as actions when the cartoon included standard visual cues for indicating motion - water running in a faucet, steam puffing out from a kettle, and so forth (see Figure 2). Without these action elements, subjects felt the drawings represented objects or locations (e.g., kitchen), rather than the associated action

Overall, participants were totally unable to make any sense of the text-based user interfaces for either application. None of the individual subjects, nor any of the subjects tested in groups were able to navigate the text-based Uls, even with prompting and encouragement. Most of the subjects were simply unable to read the text at all, and even those who could read isolated words were not able to read fluently enough to put what was written into the context of the scenario. For the map application, none of the participants were even able to locate the important landmarks for any of the tasks in the text-based map. Moreover, without the voice feedback, even users who had seen the text-free UI first, did not realize without significant prompting that one could click on text to cause an action (and with prompting, they still did not understand what they were clicking on). Throughout the study, they found that we needed to prompt and encourage subjects to try out things on screen. It is possible that a few encouraging voice instructions telling users how to operate the tool would be helpful.

A significant requirement that previous work does not mention is the need for abundant and consistently available help instructions. As they found themselves repeating the same background material and instructions to our subjects each time we visited them, we thought that this material could be encapsulated into the application itself, and this addition had a profound impact on the subjects' sense of autonomy. Even a single icon missing voice annotation, for example, causes confusion, as subjects expect to be able to point to any graphical component and find out what it represents. Similarly, help, if it is made available, must be available all the time, or it will cause a loss in confidence among subjects who tend to blame themselves for the interface's shortcomings.

Text is not the enemy:

How illiterates’ use their mobile phones By Hendrik Knoche, Jeffrey Huang

Conducting ethnographic studies of illiterate mobile users in several Asian countries, Chipchase reported that while illiterates could successfully turn on their phones and accept incoming calls, dialing numbers to make outgoing calls proved difficult for some. Nor could the illiterate users send text messages or use the address book on their phone. participants employed a number of strategies to retrieve telephone numbers of contacts that were stored on paper, e.g. business cards, telephone registers and pieces of paper that they could tell apart from spatial arrangements locations in note- or telephone registers, special colors of ink or paper, patterns, doodles, or shapes of paper, business cards (some enhanced with the owners’ pictures). The speed at which they traversed the phone menus was the same as for literate people. They mastered important functionality through rote learning: “After I have clicked on this icon I need to go down twice and then – click! - I’m done.” Continuous help was necessary for smart phone owners to download apps, games, music, ringtones, install customizations.

In phonebook, many entries were either approximated by the first letter of the name supplemented with invented characters, e.g. Mx#--, or just the first letter of the name. However, in many cases, and especially with duplicates resulting from the first letter strategy, they had to open each contact individually and try to recognize the contact. They did this by checking the country code, city code, other salient features in the telephone number, e.g. repetitions “222” or, in many cases, its last three digits. The more duplicates, the more onerous this task became due to the way the address books treat entries that contain first or last names. A large proportion of the participants made use of the alarm functions on their phones and one of them had extended this to keeping alarms in his iPhone as reminders for meetings, events and appointments.

The only information stored in the alarm was the time. All the techniques illiterates used in the physical world to manage information, e.g., spatial arrangements, shapes, colors, handwriting and doodles, were absent from contact lists and calendars in their phones. Pictures in address books were one of the few exceptions but these are not always available for contacts, are not searchable and cannot filter large lists.

A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India

By Neil Patel, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, Tapan S. Parikh

In this paper they present a study comparing speech and dialed input voice user interfaces for farmers in Gujarat, India. they ran a controlled, between-subjects experiment with 45 participants. They found that the task completion rates were significantly higher with dialed input, particularly for subjects under age 30 and those with less than an eighth grade education.

Dual-tone multi-frequency (DTMF) is a mechanism for navigating voice user interfaces using the phone’s numeric keypad. With a voice user interface (VUI) design that accepts a small number of distinct single word utterances at each node in the application (isolated word speech input).

They designed Avaaj Otalo (“voice-based community forum”), a Gujarati language application allowing farmers to access agricultural information over the phone. Functionality was laid out in hierarchical menus, and all tasks were linear. We limited all navigational nodes in the application to two or three options. To avoid command ambiguity, only directive-style prompts were used, telling the user specifically what commands they could give.

Avaaj Otalo was built and deployed using IBM Research India’s WWTW [5] platform. For the speech recognition, Gujarati commands were converted to lexicons using the American English phoneme set. In our experiment, the system performed with a recognition accuracy of 94%. Although this is lower than Plauche’sTamil system (98% accuracy), the difference reflects the cost of a larger command vocabulary for limited resource languages.

Participants were first introduced to the system and its features, and were assured that it was the system that was being tested, not them. Each participant completed three tasks with Avaaj Otalo corresponding to its three features (listening to announcements, listening to archived radio program recordings, and posting questions), ordered by increasing difficulty. They designed Avaaj Otalo to be responsive to input errors. If the system could not recognize user input, or if the user was silent, a follow-up prompt would ask the user to try again. If input was again not recognized, the system reverted to a series of yes-or-no prompts, offering each option serially. They classified a task as failed if the user either navigated to a part of the application that was not called for by the task, or failed to get passed the yes-or-no prompts after several attempts with no sign of recovery.

The overall task completion rate with DTMF was significantly higher than with speech. Participants using the DTMF interface also demonstrated a significantly greater performance improvement between the first and third task. Observations indicated two main reasons why speech input was less successful. First, users expressed discomfort speaking single word commands, which was perceived as unnatural. “Talking to the computer” was an unfamiliar idea; DTMF users may have had an easier time forming a mental model of the system. The second reason was difficulty in recovering from errors made by either the system (recognition error) or the user (bad or no input). With speech input, the task completion rate was 42% when one or more recognition errors occurred, compared to 67% when no errors occurred.

Due to the difficulty and expense of providing training, an interface that is easy to learn and understand is a key design consideration for information services serving remote populations. No users expressed difficulty in understanding how to operate the system through dialed input, including several fully illiterate participants. However, one difficulty with the DTMF interface was in transitioning between dialed input and speaking, which was required in the final task for recording the user’s question and personal information. A difficulty across both modalities was navigating command-driven menus and knowing when to provide input. Every spoken prompt was followed by a beep to indicate that input was requested. The prompts did not explicitly mention the beep, and many users either gave input too early or not at all.

The main limitation of the study is its external validity. The study was conducted in optimal conditions for both accurate speech recognition (a calm, quiet environment) and easy dialing (placing the dial pad in front of users).

An Approach to Help Functionally Illiterate People with Graphical Reading Aids By Marcel Goetze and Thomas Strothotte

This paper raises issues related to web access for users who have very poor language skills (functionally illiterate people). They present the design and a prototypical implementation of a browser which provides interactive reading aids to such persons.

In case the individual degree of language competence is below the essential and from the society expected competence, the person is functionally illiterate. The term functional illiteracy gives a relation between existing and required language competence. But it also means that people have received at least basic knowledge about a language.

Our work was inspired by the PIC-System (PIC-Pictogram Ideogram Communication) developed by Subhas C. Maharay. The system uses pictograms and ideograms for interpersonal communication. It was developed to enable communication between people who are unable to speak and for people who are able to speak but hard to understand. The PIC-Symbol material consists of 400 terms each described by a pictogram.

Another remarkable piece of work was done by Alvin H. Sacks and Richard Steele in 1984. They developed the ”Lingraphica” system. It was designed based on a database of ”word-concepts” connected with an icon to enable communication for people with aphasia. Patients can point on these icons and drag them together on storyboards. Lingraphica automatically translates these sentence-like constructions into text and spoken words.

Assumption 1: Functionally illiterate people are able to find a web page of interest. Assumption 2: Functionally illiterate users have at least some basic language skills. Assumption 3: Functionally illiterate users have at least some basic experience using a pen and paper. Assumption 4: Functionally illiterate users have never used a computer before.

Principles for deployment.

Principle 1: The main target group are functionally illiterate people. Principle 2: Text has to be explained not replaced. Principle 3: Giving graphical reading aids. Principle 4: Choosing the right type of picture. Principle 5: The User-Interface. Principle 6: Interaction techniques.

In this paper we presented a way of developing a web browser tailored to the needs of functionally illiterate people which is usable both for information retrieval and learning. Our approach focused both on using techniques of connecting images and text and interface design. The most important technique of displaying graphical reading aids for the participants was the dynamically shown pictograms by moving the pen over, or pointing on the word. The other techniques have been rated positive but has not been used as extensive as the dynamically shown picture.

Dialogue-based human-computer interfaces and active language understanding

By Will Fitzgerald, R. James Firby

To support human-computer dialogue effectively, architectures must support active language understanding; that is, they must support the close integration of dialogue planning and execution with general task planning and execution.Dialogue is an activity which involves the interlocutors’ goals and plans to achieve those goals.

The need for dialogue-based human-computer interfaces. it is true that visualization tools and graphical human-computer interfaces have come to dominate the ways computer users interact with computers, especially in the ways people interact with a “personal computer,” that is, one person interacting with a desktop or laptop computer with a high-resolution display. Effective speech-based interfaces will also require the ability for humans and computers to engage in dialogue: that is, people and computers will need to interact via spoken language (and other modalities) to make requests of one another, to question one another, to inform one another, and so forth. Equally importantly, people and computers will need to interact to clarify, deny, cast doubt on, etc., the requests, questions, informings, and so forth, being made. In other words, much of the richness available in human-human interaction will be important for human-computer interaction. Because human dialogue is typically fluent and unreflexive, once computers engage to any extent in dialogue, people start by assuming computers have the same fluency in dialogue they themselves have. Computers will not have such fluency any time soon (if ever), and thus the resources for handling conversational breakdowns available to humans (for example, asking for clarification) will need to be available to machines as well

Architectures for dialogue and action. Dialogue shares many of the same requirements with other kinds of task execution. For example: Agents can have multiple goals active simultaneously, Agents can be attempting to achieve these goals simultaneously, The world in which agents act is often unpredictable and often changing, Tasks and goals can be described hierarchically, Resources available for achieving goals vary, Goals can have differing priorities, which may vary dynamically, Agents will sometimes fail to achieve their goals.

Task execution is best done actively. When the environment in which tasks are carried out is relatively simple and static–either due to the nature of the environment or because it has been engineered to be simple and static–traditional approaches to planning and plan execution may suffice. In other cases, the environment in which tasks are carried out is often complex and dynamic, however. By “complex,” we mean that the future state of the environment is often unknowable, even given full knowledge of the current state. By “dynamic,” we mean that the environment is likely to change without our knowledge, which is just another way of saying that we cannot have full knowledge of the state of an environment. Several features of a successful task execution system: Plans can be sketchy, Plans can have multiple ways to succeed, Plans must carry their success and failure conditions, Plans can provide for breakdown.

Dialogue-based human-computer interfaces require active architectures. Dialogue requires active task execution entails the requirement for what we call active language understanding. Active language understanding is a view of pragmatic interpretation which emphasizes the close relationship that language and dialogue have to other kinds of activity, including the planful nature of language. If dialogue is a kind of task, and task execution requires a task execution architecture that is active, it follows that the successful execution of dialogue also requires a task execution architecture. In addition, an agent engages in dialogue to help achieve other goals, not (typically) just to engage in dialogue. Hence, dialogue tasks must take place in the context of other tasks (both linguistic and non-linguistic), sharing state, as well as “cognitive” and physical resources.

eScreening: Developing an Electronic Screening Tool for Rural Primary Care By K. Doruk Akan,

Sarah P. Farrell, Lisa M. Zerull, Irma H. Mahone, and Stephanie Guerlain, Senior Member, IEEE

The eScreening tool provides a graphical user interface with audio outputs for users who may be functionally or computer illiterate. The interface is a Macromedia Flash movie shown on a web site. A database automatically and anonymously records the screening data.

Each question in the screening requires separate frames for the unanswered situation. When the screen for a question first appears, the system uses a human voice to read the question aloud. When the user selects an answer, the box for that answer fills in green and the corresponding text becomes green. Other features on a typical screen include the back button and audio tools. The back button allows users to change answers they believe they chose mistakenly. When pressed, the back button displays the previous question with the answer the user had chosen. An important function of the eScreening system is its ability to store screening answers in a database. At the end of the screening, the system updates the database by sending the answers as data through an ASP script. Administrators can then analyze the summary results of the screenings to identify any trends in depression and alcohol abuse among primary care patients.

Audio problems included difficulties with the volume and a malfunctioning sound toggle button. Also of concern to users was the inconsistency in volume of the voice-over in the movie clips at the start of each question. Also, some of the voice-overs were difficult to hear because of static. Other issues included uncertainty with the movie clips and the Next button. Some users were uncertain whether they must listen to the entire clip before selecting an answer. Also, users did not notice or understand the function of the Next button. They addressed the latter issue by flashing the Next arrow once it appeared to capture the user’s attention. Design issues that arose were minor problems. In the tutorial, users were confused when the Next button did not appear when it was discussed.

This paper describes our efforts to design and implement the eScreening tool. The outputs are text and audio to accommodate possibly illiterate patients. Graphical touch-upswill help make the interface look more professional and appealing. Also, enhancement of the audio components will improve the quality of the patient eScreening experience.

From Illiteracy to Computer Literacy:

Teaching and Learning Using Information Technology By Mohamad Adnan Al-Alaoui

This paper addresses the use of information and communications technology, ICT or IT for brevity, to combat illiteracy and to move illiterates directly from illiteracy to computer literacy. The approach provides interactive learning, self-paced and autonomous learning, entertainment learning, ease of information updating, ease of entry and exit, and ease of application to E-Learning.

They propose that the software be divided intothree parts: one teaching innumeracy, the other teaching literacy, and the third teaching computer literacy. Each part will in turn be divided into different levels. Each level will be accessed after the user has passed the test at the end of the previous level. Each level will include a presentation/lesson, interactive exercises and finally a quiz/test. Each level will teach a certain notion or skill.

A. Literacy: This part of the program will contain 5 levels: writing and reading letters, grouping letters into words, spelling, grouping words into sentences, and grammar.

B. Innumeracy: This part of the program will contain 6 levels: writing and reading numbers, counting in groups of 2, 3, 5, 10…,addition, subtraction, multiplication, and division.

C. Computer Literacy This part of the program will contain 6 levels: familiarity with the input devices (Keyboard, mouse, diskette, CD, microphone, web Camera, familiarity with the output devices (Soft copy devices and hardcopy devices), familiarity with word processing and data entry, learning to obtain, copy, and move files, navigate the internet using Netscape, and using email.

It is recommended that teaching Literacy, Innumeracy, and Computer Literacy should be interleaved. Thus Level 1 Literacy can be followed by Level 1 Innumeracy which in turn should be followed by Level 1 Computer Literacy.

ICONS AS HELPERS IN THE INTERACTION OF ILLITERATE USERS WITH COMPUTERS

By Ismael Mattos A. Ávila, Ricardo Ribeiro Gudwin

This paper addresses both the use of icons in supporting digital content navigation and their role in scaffolding limited reading skills when accessing digital contents.

in general voice descriptions help in the understanding and that although the representations with audio were more intelligible than those without audio, bimodal (audiovisual) information can also be confusing to such an illiterate audience. On the other hand, visually richer information does not forcefully mean better intelligibility, and cartoons can prove to be more effective than photo-realistic representations. using pictures can improve the intelligibility of the presented concepts and facilitate the inferences that illiterate users make, but the quest for perfect icons is meaningless, since the iconicity depends on the previous experiences of every individual, and what is “perfect” for someone may be inadequate or insufficient for others.

The first test evaluated two types of image (photos and drawings) in terms of their intelligibility and the credibility they create among the viewers, as well as some aspects related to their interaction with texts, trying to correlate such factors with socio demographic variables. In the test they used images of daily objects and themes. Given that the first e-gov services developed in our project will cover the areas of health care and social security, we selected images related to such areas, such as medical specialties (obstetrician, pediatrician, dentist, oculist and generalist doctor), documents (Id. Card, Social Security card and birth register) and common objects or themes (“dog”, “TV set” and “couple”).

They indicate a clear preference of the users towards photos, but this preference does not mean, in the same extent, higher intelligibility. The higher credibility of photographic images could result from the fact that photos retain a physical link with the real world, in the sense that they show real doctors, real documents and so on, then giving more credibility to the representation.

The iconicity research described here was intended to shed some light on how icons can facilitate or make possible the access of low-literacy individuals to e-gov services. Test showed that, even though the use of photographs lends more credibility to the iconic representations, the intelligibility of drawings is not significantly lower than that of photos. The second test showed that the use of icons could reinforce and scaffold the reading skills of semiliterate users. They saw that the combined effect of icons and texts was, in some cases, greater than the sum of the isolated effects of each type of sign.