RESEARCH REPORT SERIES
(Survey Methodology #2009-02)

2039: A Day in the Life of a Usability Engineer

Theresa A. O'Connell [1]
Elizabeth D. Murphy
Renate J. Roske–Shelton[2]

[1] National Institute of Standards and Technology

[2] Food and Drug Administration, Web Communications – Usability

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: February 24, 2009

2039: A Day in the Life of a Usability Engineer

Theresa A. O'Connell[1]

National Institute of Standards and Technology

theresa.oconnell@nist.gov

Elizabeth D. Murphy[2]

U.S. Census Bureau

elizabeth.d.murphy@census.gov

Renate J. Roske–Shelton

Food and Drug Administration, Web Communications – Usability

RoskeHF@gmail.com

---

[1] Certain commercial equipment, instruments, materials, services or companies are identified in this paper. This in no way implies endorsement or recommendation by NIST.

[2] This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Abstract

As we approach the end of the fourth decade of the twenty–first century, intelligent agents will handle most of the time–consuming drudge work of usability engineering, freeing usability engineers (UEs) to focus on the mental work of analysis and usability recommendations. Just as gamers today meet in synthetic worlds, UEs will enter customized synthetic environments to investigate human cognition and behavioral interactions with new software capabilities. Immersed in synthetic worlds, UEs will employ a range of technologies at every step of usability evaluation. We'll be working smarter and more efficiently to measure and improve human experiences with technology. The future holds promise for UEs and, by extension, for users of the technologies we are beginning to imagine.

Story

> *Work becomes more abstract as it depends upon understanding and manipulating information. This marks the beginning of new forms of mastery and provides an opportunity to imbue jobs with more comprehensive meaning.*
>
> *Zuboff (1988: 6)*

Morning. Monday, October 4, 2039. Time to check the weather report delivered to my Blueberry by Wex, the weather agent. It looks like a beautiful day – great weather for travelling. Oops, Traff, my traffic agent, tells me the third level of the outer loop is closed for construction. The FedShuttle will have to detour. Its driver will be guided along its route by positioning system telemetry updates. but there will still be traffic jams. This means a longer drive to the airport along the second level but more time to get some work done in the u–eval–space. I don't usually go to conferences in person, but the Government Usability Engineers are having one of their rare face–to–face, in–person meetings.

It's one thing to exchange technical information in electronic meetings, but we all crave some actual human interaction. Some would argue that the most important work of a conference gets done outside of the technical sessions, over lunch or dinner, or just during the coffee breaks between sessions. That's certainly the time for networking, which gets lost in a virtual conference. I'm looking forward to seeing old friends and meeting new people.

As I board the FedShuttle, I pass other people at work already in their synthetic environments (SE). I find a cubicle at the back, settle into the fully adjustable comfort chair with safety bar, and start my session.

Pausing slightly to position my right eye straight in front of the wall mounted iris recognizers until I hear the confirmation tone, I press a button and the simulated usability–evaluation space opens ready to go, since we finished all the prep work yesterday. I'm suddenly surrounded by a 360 degree synthetic world (syworld) display with full directional audio.

The u–eval–space recreates our lab at our Center's office, which we don't use much anymore since most of our work is done remotely and on the go. But the u–eval–space goes way beyond the capabilities of our physical lab.

Something I really like about the syworld is that we no longer have to worry about test participants getting lost in the maze of our physical campus and parking garages. The u–eval–space finds the participants wherever they are, so they don't need to find us. In the old days, we lost some participants who got so disoriented that they gave up on trying to find the visitor's entrance. Sometimes they would show up too late to participate in collaborative testing. The u–eval–space also does away with the need to supply participants with step–by–step directions of how to get to us. And we don't need to go down and escort them through the secure bricks–and–mortar building. The synthetic testing space lets us fully concentrate on the participants, their thought processes, attentional behaviors, emotions and motivational factors, and on developing recommendations to resolve the tough usability issues.

A problem with the u–eval–space, or any syworld, is that participants sometimes experience a different kind of disorientation. They can become overwhelmed by the vast amounts of information in the streaming visualizations that surround them. Some people experience physical effects like nausea or dizziness; others report euphoria and manic moods. Combined with information overload, these phantom symptoms can make it very difficult for information analysts to do their jobs in a syworld. Since we're working within a fully realized version of the People Capabilities Maturity Model (PCMM) (Curtis 2008), however, information analysts and others who work in SEs are given training and practice to familiarize themselves with the psychophysical side effects of the work environment.

Today we're testing an upgrade to an immersive visual analytic tool with our last two participants. Like all our other users for this study, they are information analysts who are American citizens; but Tony and Greta live and work outside the U.S. Ira, the intelligent recruiting agent, has picked them from our database and scheduled them both for overlapping sessions today; so one of my colleagues, Sam, will conduct his testing in an adjoining room within the u–eval–space.

Before we get started with the testing sessions, Sam and I will meet to review our plan for the day's activities. Sam joins me in our regular cyber meeting space. Right now he is physically in his apartment in Georgetown. Via our avatars, we are meeting to review the test plan and to confirm the day's schedule. Our literature reviewing agent, Rex, has found a new measure of disorientation mentioned in an article by Dr. Kim Edwards. Since the measure has not yet been published, we have invited Dr. Edwards to join our discussion. It appears that disorientation is highly correlated with cognitive overload.

During an analytic session in a syworld, the visualizations become larger, more complex and information dense as they receive huge dynamic data sets from disparate sources. Although additional interactive functionality becomes available as needed, many users find it difficult to sustain the cumulative mental workload. Having to keep the products of mental manipulations in mind from screen to screen is particularly difficult, just as it was in the old days when system operators in spacecraft control centers had to make mental comparisons between actual streams of telemetry data and the nominal ranges for multiple parameters. Battery, solar panels, antenna status – they seemed like a lot in the old days. Now the computer has taken over these comparisons, but has given us a new information tsunami to handle with our same, limited, human neurological and psychological capabilities. We're interested in learning whether information analysts can cope with the intellectual workload of using visualizations of these massive data sets in syworlds.

To assess neurological and psychological effects, we outfit our participants with a smart, hooded jumpsuit that can take physiological measures, including brain scans, without distracting the participant or causing any discomfort. They wear glasses that collect eye tracking data, even as the participants move around in the syworld. This is full 360 degree, unobtrusive eye tracking, which has advanced to the point that we understand the cognitive sources of the various measures. Results from eye tracking help us measure user success (accuracy and efficiency). Using other measures, we assess the participant's satisfaction with the immersive experience of working in the syworld. All participants receive test kits with these items included and return pickups pre–scheduled.

The upgrade that we're testing is designed to help the analyst with the mental integration and interpretation of key findings that emerge from the automated analysis of multiple data streams. It puts the user in control of the data streaming, empowering the novice information analyst to slow down, or even pause the data streams. The expert analyst has the option of taking data streams in fast time, experiencing data elements in a compressed flow unbroken by dead time and examining the data at a fast rate. Using this tool, information analysts can systematically subtract selected streams from the analysis or add others to the mix. They can apply a variety of data–transformation filters. Known as WHIF, the tool provides an externalized "what if" capability, using visualizations to relieve some of the burden on analysts' working memories.

Today's testing will complete data collection on the WHIF upgrade.

Sam's participant lives and works on the International Space Station, rotating out after a six–month stint focused on analysis of information about the space–faring nations, including new members as they join the club. Other observers from the development team will join us, via their avatars, from time to time. Of course, the sessions will be fully recorded, and will be available to any authorized staff member.

Whenever there is a form to be filled out by a participant, such as a consent form or a satisfaction survey, the form appears holographically, and the participant can vocalize answers to populate the form. These data are instantly digitized and routed to usability analysts working from different perspectives.

My test participant is here, that is, his avatar has popped up in the syworld. Tony works in Rajististan. He has passed all the physiometric checks we have in place, so we're sure it's really Tony. Our avatars greet each other, and I turn things over to Trog, the intro agent, which already knows everything we know about Tony. Trog speaks in a well–modulated, female voice, which goes well with its tailored female avatar:

"Thanks for being here today, Tony. You'll be helping us evaluate WHIF, an upgrade to the syworld that you currently use in your work ... Please make sure you are wearing all the test kit items …."

Even though he's a native speaker of AmerEnglish, Tony prefers working in the language of his host country. So, we'll use the Rajististani version of WHIF, and he'll think aloud in Rajististani. Through my headset, I'll hear the questions and think–aloud commentary in English. All vocalizations will be captured not just by audio recorders, but also by a speech–recognition capability for the source language. Then the language system will translate all of Tony's spoken comments into English and synthesize that translation into a male voice similar to Tony’s, for me to hear. At the same time, it will send Tony’s comments in Rajististani to a natural language analyzer which will extract quotes and then send those quotes to machine translation from Rajististani into English for later incorporation with other data from the session.

Continuing the instructions, Trog explains to Tony that thinking aloud is a technique we use for getting at the user’s thought process as he works:

“Please talk about the decisions you make before you take any action that would result in changes to the data displayed; that is, please “think aloud” about your assessment of the situation, what you expect will happen if you take a particular course of action or an alternative action, and your reasons for the choices you make. If you stop talking for more than 30 seconds, I’ll remind you to keep talking. I can’t answer questions, but encourage you to ask questions aloud if that is part of your decision–making process. Next, we’ll do a brief practice on thinking aloud …”

Having the agent deliver these messages allows us to give the introduction and instructions in exactly the same way and in the same words to each participant, even though the agent–appropriate voices vary across avatars. The verbal protocol provided by the participant helps us to interpret the eye–tracking data and other measures of cognitive activity collected by the smart jumpsuit.

Checking highway conditions, displayed periodically by Traff, I see we're about halfway to the airport; so there's plenty of time to complete this session. It's a big help to have our human effort augmented by the intelligent software agents. They're all part of the team.

From previous testing and observation, it's not clear that the upgrade is usable. We've found indications of usability problems, such as disorientation, nausea, and dizziness, as well as cognitive overload. Resolving these problems is going to require some redesign, which will be time consuming. That's unfortunate because the WHIF upgrade was supposed to be launched next week. I wish the developers had requested testing earlier.

This new version of the analytic space used by information analysts has been touted as the answer to their need for ways to support their worldwide collaboration in managing widely diverse data sets, in searching visual informational spaces for keys to understanding, and in developing more insightful hypotheses. It may be that expectations were too high and that further capabilities will be needed to support interworld collaboration with scientists at the lunar colony.

The code for the prototype upgrade has long since been run through Usachex, the automated usability checker, a tool with capabilities to evaluate the look and feel of syworlds, as well as more day–to–day, mundane user interfaces. Although the upgrade passed this level of evaluation, we started finding issues when human participants, all information analysts, began to interact with the higher–powered syworld. For example, Usachex would not reliably predict nausea and disorientation. This is why we insist on keeping human test participants in the process. Personas and user models have their place in system design, and automated evaluators free us from some of the drudgery of old–style usability testing, but there's no substitute for the real thing in testing highly complex, interactive syworlds. The real user will invariably do something unexpected or experience unexpected side effects, and that's the true test of a flexible syworld.

One task in the evaluation of the WHIF upgrade asks the participant to monitor several sources of information about troop movements in an occupied city. One source of information provides aerial views of the city at different moments in time. Most participants have had trouble with this task because the troops have been cloaked by a stealth frequency that makes them invisible to the aerial cameras. Tony reasons that – if such a frequency is in use – it can't be effective from every possible angle of view; so he rotates the visualization until he can detect armored vehicles and troops moving through the streets. In the past, this strategy, when used in a syworld, resulted in disorientation. Tony reports that he does not experience disorientation or cognitive overload while performing this task. Electroencephalogram, electrocardiogram, and galvanic skin response feedback from the smart jumpsuit back this up.

Meanwhile, Sam's participant, Greta, is working on her simulated tasks, which involve simultaneous monitoring of unmanned satellite launches from Earth, checking on the progress of automated resupply missions to the current International Space Station (ISS4), and communicating with a simulated manned Mars mission. In the midst of a communication from Mars, Greta receives an emergency alert from an Automated Transfer Vehicle about to dock with ISS4. Forwarding the alert to a station mate, she stays on the line with Mars, only to hear that several members of the mission are suffering from exposure to cosmic radiation.

Later, when rating this task set for mental workload, Greta gives it the highest possible rating for cognitive demand. We'll have to check physiometric feedback from the smart suit to see if it backs this up. If there had been cognitive overload, we'd expect to see a spike in feedback from the central executive in working memory. In the debriefing, Greta comments that monitoring the visualized data streams made her feel slightly queasy. Our intelligent statistical agent, Isa, will correlate eye–tracking data with physiometric feedback so we can assess whether frequent eye movements or elongated saccades are associated with Greta's queasiness and sudden drop in blood pressure. This quantitative physiometric feedback sets our work above the best we could do thirty years ago. With hard data, we can validate or call into question subjective user comments. But, we still need the human in the loop; even if the data say that a phenomenon such as nausea wasn’t severe, if the user interprets and reports that it was, our user–centered recommendations call for a high–priority remedy. One thing remains the same in 2039: user impressions and experiences drive software acceptance today just as much as they ever did. We expect that they always will.

This completes the test sessions.

When we arrive at the airport, I close down the syworld and exit the FedShuttle. As I make myself comfortable in the waiting area, simulation sensors predict that, had Greta continued working with this WHIF version for just five more minutes, there is a high likelihood that she would have experienced something akin to an epileptic seizure. The autonomous agents simulate this condition with a surrogate persona (a program that has the relevant user characteristics of Greta). They devise a design solution that changes the flashing rates for a specific type of data display update. The screen will now refresh at a frequency that will not trigger an epileptic seizure. Then they rerun the tests, using the redesigned display, with four additional personas that simulate tendencies to epileptic seizures. They’ve eliminated the cause of the seizure.

But there is a new open question: Has the modification impacted usability? Because we are certain that we have removed the risk of causing a seizure with the software, I approve further testing. Ira recruits four more human users. While I’m in flight, these users run through the scenario Greta used. Sam monitors the sessions, ready to intervene if necessary. The agents and Ira have demonstrated their worth. Usability scores remain in the same ranges as those from all other users before the minor modification.

While I'm en route to the conference hotel, the data are aggregated automatically, so that when I fire up the syworld again, at the hotel, all the data have been rolled up, and standard statistical measures have been calculated (e.g., means and standard deviations, along with lostness functions, mental workload graphs, and a disorientation history). In the old days, we rarely tested with samples large enough to yield statistically significant results. Now, in the u–eval–space with our agent colleagues and new technologies, we can work with more participants, thereby generating much more data and testing hypotheses almost immediately.

Once the data analysis is completed, our widely distributed network of usability engineers (UE) collaborates on a holographic report. Working in my hotel room, but back in the syworld, I'm surrounded by visualizations of data from all the testing sessions we have conducted. Sam and our other human collaborators are remote. The report outline has been automatically generated based on our research design. Each section has authors assigned. Some sections are assigned to autonomous intelligent agents, and others to me, Sam, and our human collaborators.

The autonomous intelligent agents populate charts and conduct straightforward data reporting in real time. As we humans watch parts of the report being filled in, we do the heavy analytic work and come to conclusions. The visualization leads us through the evaluation process, step by step, starting with receiving data that the autonomous agents deem important, evaluating the data, coordinating the data into groups of facts, hypothesizing, and collaborating on refuting or accepting the research hypotheses. We're standing in the synthetic space, talking to each other in real time, building the report that surrounds us.

We can turn to the right to check out the user profiles; turn to the left to see the error reports; look toward the floor to see the critical incident report; and look toward the ceiling to see the summary of usability issues and recommendations. Our colleague, Jim, has a speech disability and cannot speak. So, instead of inputting to the report with speech as some of our colleagues do, he's using his invisible keyboard. Such tools have opened the door to UE careers for many new colleagues with disabilities. Jim's avatar looks like he's typing on thin air as a new paragraph appears in the findings section of the report.

Our colleague, Mario, is thinking aloud: "I wonder if we should swap places for Sections A and B?" The sections immediately trade places. He shakes his head, and they return to their original locations.

Major issues are discovered in WHIF disorientation and cognitive overload. We recommend that the upgrade not be released until these problems are resolved. We stress the need for additional semi–autonomous agents to take on the burden of monitoring data streams, a task not well suited for human analysts. As the lead author, Sam reviews the draft report for consistency. Once our final report is delivered to the sponsors and all data archived in the UE library, I log out and head for the welcoming reception, ready to greet old friends and meet new ones.

Where We Are Today

Reality is merely an illusion, albeit a very persistent one.

Attributed to Albert Einstein (1879–1955)

Powerful new tools will empower us to do better usability engineering. These tools will result from trends that are part of the user experience today – the integration of many technologies into a single device and the integration of many technologies into end–to–end systems. A better understanding of human–syworld interaction (HSI) will inform tool design. Usability challenges are already surfacing: human interaction with technologies in complex interactive environments that change with each interaction; the new and wider range of user experiences possible in syworlds; and continual technological advances that outmode usability findings with each new version. Today we are beginning to understand the usability issues behind HSI. For example, gaming research is informing the usability of syworld analytical tools such as WHIF (O'Connell et al. 2008).

In this section, we look at four types of technologies encountered in our story and discuss human experiences with them today, keeping an eye open to the implications for 2039.

Synthetic Environments

> *The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. We refer to this experience as immersion…*
>
> *Murray (1997: 98–99)*

In our story, both UEs and users experience some degree of immersion in syworlds. The classic usability lab with its one–way glass has given way to the syworld of the u–eval–space. The UE's toolbox has expanded with contents derived from research in progress today. The test participants experience new technologies, e.g., hologram–survey forms and WHIF, an advanced visual analytic tool.

In reality, this immersion does not yet exist to the seamless extent portrayed in the story. In our story, the UE simply fires up the syworld using her own laptop, from any place she may happen to be. When today's users enter the closest thing we have to such an environment, a CAVE (Computer–Assisted Virtual Environment), they must don special glasses to experience the multi–dimensional graphics. The CAVE requires a large dedicated physical space with projectors, electromagnetic sensors, and speakers, all to create the impression of immersion in the multi–dimensional syworld.

Today, holograms reconstruct images so that they appear three–dimensional. The image can only be updated by showing a different view of an existing image. The hologram report in the story uses holography in a different way. Here users interact with the hologram construction by providing data input. How this input becomes part of an integrated updated hologram is purely speculative. Perhaps the user's speech is converted to digital text and then to a digital representation, which is fed to the hologram. Pattern recognition, analysis, and synchronous publication (rendering) are the necessary technological process elements. Holography is advancing to the point where a holograph will accept updates: the required increased memory capacity is within reach and erasing and rewriting capabilities are already under research (Steere 2008).

The virtual keyboard used by Jim, the UE with a speaking disability, is currently under research by Rashid and Smith (2008). Although they intended to facilitate typing into small mobile devices and to save screen real estate on touch screens, the leap to inputting to a virtual report is a short one. Mario moves report content by shaking his head. Using so–called natural gestural interaction to communicate with a computer is not new. To take an example from gaming, players gesture to communicate with the Nintendo Wii™. However, they are restricted to using a handheld wireless motion sensor, which has disappeared in 2039.

Without lifting a finger, the UEs watch their avatars move through the synthetic world. As the UE team builds its report, Mario appears to return report content to its original location at will. Today it is possible to use haptic interfaces to move objects from one part of an interface to another, but Mario and the UEs have done this via thought sensing. Such abilities are currently emerging in games. The first steps have already been taken toward understand the human experience with brain–computer interfaces in virtual worlds (Bayliss & Auernheimer 2001). Today, in video games, brainwave information is collected and used to set the pace of the game in response to a player’s emotional state or attention level (e.g., Yang 2008, NeuroSky 2008). Companies such as Emotiv Systems (2008) use information about facial expressions and emotional states, as well as users’ thoughts, to manipulate onscreen objects. Of course, this doesn’t happen to the degree described in the story, but the path has been forged to empower users to immerse themselves in the creative intellectual aspects of building a report without having to share cognitive resources with such mundane activities as manipulating input devices.

Intelligent Agents

> *Work must be smart, appropriately targeted, and adapted to the particular circumstances of the process and the customer. Imagination, flexibility, and commitment to results are needed ... You are accountable for results, not for effort ...*
>
> *Hammer (1997: 28)*

We already have the technology to support the behavior of the personal, autonomous agents in the story (e.g., Traff, Wex). Less sophisticated technologies exist today in specific household applications for vacuuming, floor sweeping or gutter cleaning. Each of the autonomous agents in the story is programmed to do a specific job, as specified by the user: gathering the current traffic report or weather forecast for a specified area daily at the same times. The story extends these concepts into areas in which autonomous agents could be useful, such as the area of recruiting participants for usability tests. In this case, we want the agent to be semi–autonomous, because there is always a need to coordinate schedules and exceptions with the human usability analyst. Franklin and Graesser (1997) have developed a systematic inventory of agent classifications based on such goal directed functionality.

The role of Ira, the intelligent recruiting agent, is to take over the tedious tasks of contacting potential participants and scheduling them for time slots specified by usability analysts in published calendars. Ira uses variants of email and mobile messaging that we anticipate will persist into the 2030s to communicate associated social niceties. Ira's message gives potential participants a choice of available dates and times. Individuals sign up or decline. It is part of Ira's job to remove slots as they are filled. Ira also annotates the database of potential participants with the outcome of every attempt to contact a particular person and with information about the person's actual participation (e.g., date, time, name of study). The technology exists today, but a price barrier still exists before wide–spread commercial adoption can occur.

The technology is already in place to conduct an automated evaluation of user–interface code, looking for violations of usability principles. Various tools have been developed for this purpose, for example, AIDE (Sears 1995) and KALDI (Al–Qaimari & McRostie 1999). Ivory and Hearst (2000) present a detailed review of the theoretical background, methods, models, and techniques developed for automated user–interface evaluation.

A proof–of–concept appeared in the early–to–mid–1990s in research conducted for the NASA–Goddard Space Flight Center (e.g., Jiang et al. 1993). The prototype tool was called CHIMES, for Computer–Human Interaction Models. CHIMES was based on the concept of checking user–interface–design code against coded human factors guidelines on such topics as color use, font variations use, and screen density. A conceptually similar prototype tool, called TUNE (Nordqvist 1996), evaluated compliance with human factors guidelines and style guides. Automated tools are also used to assess software accessibility for people with disabilities (SSB Technologies Inc. 2006, World Wide Web Consortium 2007, 2008). In our story, Usachex follows in this tradition.

A difficulty for early automated usability tools was how to deal with graphical user interfaces, that is, how to quantify the use of graphical elements without penalizing the design for using elements that actually decrease burdens on the user's cognitive capabilities. A promising set of style–consistency checking tools, known as SHERLOCK (Mahajan & Shneiderman 1997), tackled some of these issues. A task–analytic approach gets around the issue of quantifying graphical elements by focusing on the user's performance of tasks (e.g., Lecerof & Paterno 1998). In the story, we imagine that obstacles to an automated evaluation have been overcome and that automated user–interface evaluators, such as Usachex, can deal with subtle, yet quantifiable, effects of design strategies on the user experience in SEs. Progress toward that goal is evident in work reported by Okada and Asahi (2000) on their tool, GUITESTER2, which checks the properties of graphical widgets for consistency. Research on tangible user interfaces (e.g., Merrill, Kalanithi, & Maes 2007) has the potential to bring cognitive overload under user control in SEs.

Natural Language Technologies

> *"There is, of course, some engineering involved. … Yet, something is missing….As adults, we focused on constricted problems… But as children, we encountered the world in all its broad diversity, and we learned our relation to the world, and that of every other entity and concept. We learned context."*
>
> *Kurzweil (1999: 89)*

The end–to–end systems that handle the linguistic components of interactions between the UE and the users are possible today. Speech recognition (SR), machine translation (MT), synthesis, and natural language understanding (NLU) are commercially available. However, they do not have the accuracy required to carry off the scenario in the story. Context is the reason why. Natural language technologies have made great strides on the levels of phonemes, morphemes, words and syntax, but have not conquered the full context of human communication.

Pragmatics is the study of language in the context of human social interactions. The environment in which discourse occurs provides contextual clues that shed light on meaning. Harnessing these clues is currently a focus of computational linguistics, but much work remains to be done. Integrating statistical models of language into natural language technologies is a current approach to resolve context, but it has become apparent that this integration is only part of the solution.

Systems of integrated computational linguistic components are commonplace today. In the story, the transition takes two paths, both starting when Tony speaks, although, of course, if the UE replied, there would also be a path from spoken English to spoken Rajististani. (We’re assuming that the autonomous agent, Trog, is multilingual, interacting with Tony in Rajististani and with Greta in English.) The first path leads to synthesis for the UE and the second to NLU where Tony’s comments are understood and salient elements are extracted. For example, in the second path, extraction can produce a list of places Tony mentioned. Thus, participants’ statements are passed to the agents for use as anecdotal evidence to explain quantitative data in the report.

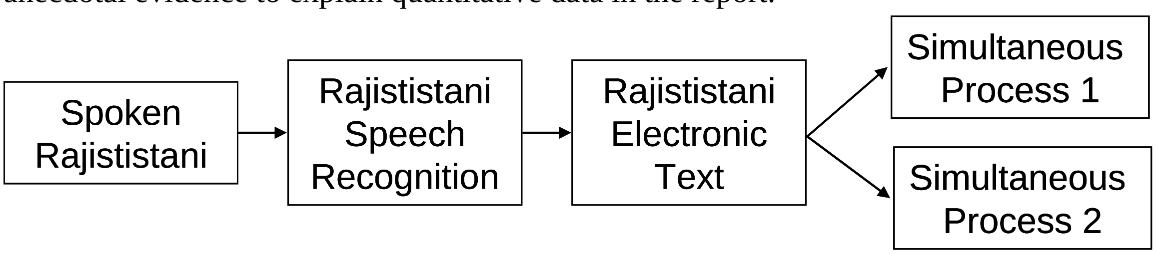

Figure 1. Spoken Rajististani undergoes speech recognition to become electronic text to pass on to two simultaneous downstream processes.

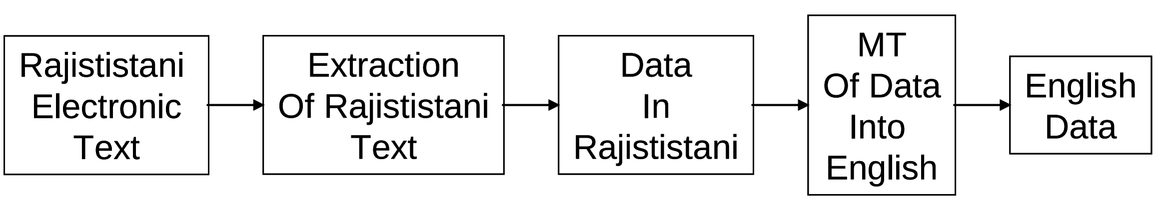

Figure 2. In Simultaneous Process 1, Rajististani electronic text undergoes extraction followed by machine translation (MT).

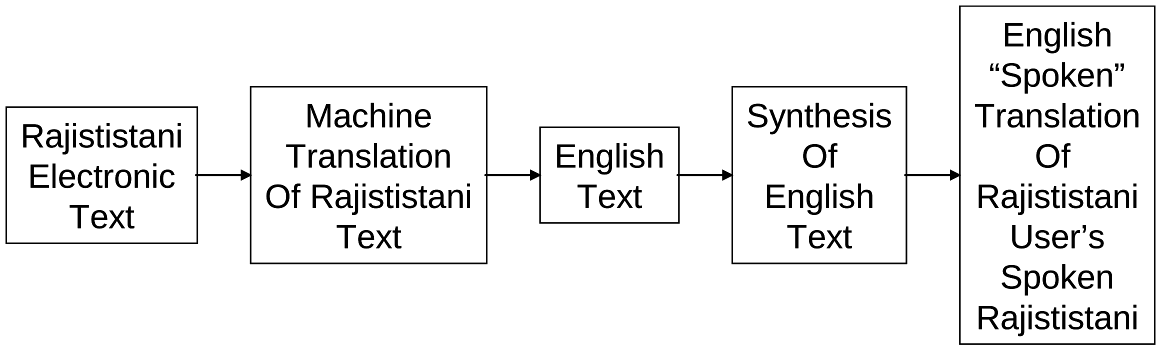

Figure 3. In Simultaneous Process 2, the Rajististani electronic text is automatically translated into English and synthesized to sound like spoken English.

However, Tony is not speaking his native language, AmerEnglish, but speaking Rajististani with a foreign accent, introducing allophones beyond the Rajististani allophonic range. Today, this input would impair the SR at the beginning of the process and pass tainted data to all subsequent steps. Given this confounding factor and the current state of NLU accuracy, the result would not be accurate enough to produce data to inform usability judgments.

Speech recognition with immediate input to the form, even with today’s state of the technology, will be more reliable. If the system automatically fills in stored unusual words such as the user’s name, street, and city before the user starts to give responses, the words in the user’s responses will most likely fall within the range of a defined vocabulary (e.g., numbers, yes/no) that, even without training each user, should be recognized.

Today, MT performs best in applications designed for specific linguistic domains, e.g., weather or travel. Non–domain–specific MT continues its more than 60–year quest for perfection. Because of the complexity of human language in context, MT remains an application that can successfully transfer the gist from one language to the other, but often at the expense of stylistics and fluency. These require human post–editing.

The state of today's technology is that SR will send input to MT that includes recognition errors. Since our scenario does not include human post–editing of the machine translation, the MT would propagate SR errors, adding its own. The output quotes would not be accurate enough for use in a report.

In *Star Trek*, *The Next Generation*, the android, Mr. Data, cannot use contractions. The agents in this story use contractions to make their speech more natural and engaging. Working at the pragmatic level of linguistics and taking contextual clues into account, when appropriate, the agents use contractions. A minor victory for the androids, and not a stretch from today's computational linguistics capabilities.

Physiometrics

> *I believe in evidence. I believe in observation, measurement, and reasoning, confirmed by independent observers*.
>
> *Asimov (1997: 43)*

By 2039, biometrics has evolved into physiometrics, measuring a wider range of physiological attributes. This work has already started. A physiometric project, called the VIEW of the Future, measured neurophysiological and psychophysiological factors of users interacting with a virtual environment (VE) (Karaseitanidis et al. 2006). The eSense project (Presser et al. 2006) addressed the miniaturization and robustness required for physiometric sensors such as those in the story.

The smart jumpsuit in the story draws on technologies in common use today, and presupposes a miniaturization and everyday integration of “wearable computing” that we are already on the path to achieving. Our jumpsuit is fitted with a tiny electroencephalogram (EEG). An EEG collects neurological information about brain activities through electrodes placed on a person’s head.

An EEG indicated the possibility of Greta’s impending epileptic seizure. Today, an EEG can also supply information on a user’s state of mind to help them control objects in a syworld or the real world. For example, to play the game Mindball (Interactive ProductLine 2008), players wear a headband with an embedded EEG and electrodes that connect to a physical gaming table. Physiometrics on electrical activity in the players’ brains become input to the game that empowers players to move a physical object, a small ball, in the real world. In real time, visualizations of analysis of players’ brain activity show players’ relaxation levels. Sending the EEG feedback to an intelligent agent is surely possible.

The jumpsuit has an embedded electrocardiogram (ECG) that, like today's ECGs, provides heart-rate information to inform stress analyses. This information helped the UE team in the story validate that Tony had not experienced the stress of disorientation. The jumpsuit contains galvanic skin response biosensors which, like today's lie detectors, use factors such as sweat to indicate emotion. We assume that the smart jumpsuit has automatically self–calibrated to the individual user. Excluding the extreme miniaturization factor, these are all today's capabilities.

What Needs to Change

> *It is change, continuing change, inevitable change, that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be ... our everyman must take on a science fictional way of thinking.*
>
> *Asimov (1978: 5)*

Usability engineering is all about improving human experiences with technology. Therefore, we are interested in what needs to change to bring the story's technologies to realization, but we are also interested in what has to change to build the bridge to human acceptance of these technologies and to foster human ability to use the technologies. The first change that has to happen is that cognitive psychology must rethink its theories of “psychological reality” to incorporate the integral role of computers in human experience (Norman 2008: 5). Cyberpsychology will pave the way for a more holistic approach to usability engineering.

New technologies will continue to face the same obstacle that they face today: user acceptance. Although tomorrow’s shelfware may be virtual, nonetheless, shelfware it will be if users do not accept it. Usability will remain a primary driver behind acceptance, but as the story shows, usability will take on new dimensions because of the syworld context of use. The call has already gone out for us to rethink our approaches (e.g., Cooper 1999). In this section, we discuss what must change in both usability engineering and in technology to make our science fiction reality.

The drivers behind the way we evaluate HSI will include a factor currently under investigation: in 2039, analysts will differ from their predecessors in more than their techno–savvy work styles. There is evidence that the digital native gamers of our current generation can assimilate larger amounts of information and handle more complex displays than their predecessor could; and they can interact with each other in syworlds to a degree unknown to their predecessors (Jonas–Dwyer & Pospisil 2004). Therefore, they may be able to achieve work products beyond the capabilities of their predecessors. Tony and Greta will feel quite at home in the syworld they test. The 2039 generation of UEs will be comfortable collaborating on a report while immersed in a visualization of mountains of usability data.

New user profiles and changed technologies will generate new interaction behaviors. We must be able to measure human interaction with evolving technologies in the context of syworld use. Perhaps the longest leap in our story is the commonplace and ubiquitous use of synthetic worlds. Unless these technologies and the associated metrics mature in a way that fosters usability, UEs and information analysts will not be able to realize the benefits that the characters in our story enjoy. To this end, work is underway to identify usability principles to inform the metrics that UEs will use in 2039 (e.g., O'Connell & Choong 2008, Wilson & D'Cruz 2006). Some will be traditional, others innovative.

The story assumes that we have already evaluated the technologies that the UEs use. For example, we have tested Ira using metrics such as the ratio of the number of participants recruited to the number of participants who actually perform the evaluation; the degree to which participants conform to the desired user profile; and participants' subjective ratings on the experience of being recruited by an Ira. While these do go beyond today's stress on time and errors, they are not completely innovative. To be innovative, our new metrics must address the innovative aspects of the new technologies. We are already identifying questions that will lead to these metrics (Choong & O'Connell 2008).

For example, how do you measure cognitive workload when all you see directly is the avatar? How do you measure perceptual overload when the user is interacting with a hologram or working in a fully immersive synthetic world? How do you measure the user's total psychological experience in a synthetic environment? Such questions need answering to take usability engineering to the story's level of immersion and innovation.

Sensitive metrics and refined methods will measure human interaction with technologies with more precision than we can in the first decade of the 21st century. We will measure the usability of such tools by the quality of the work products – is the UE report more complete and insightful than those produced back in the first decade of the millennium? For example, how many aspects of the user experience did we measure? How quickly did we deliver the report? And, perhaps most interestingly, how are we going to measure insight in a UE report?

Wilson (2006) has predicted that evaluating human interaction with VEs will be "extraordinarily difficult." He identifies a steeply rising number of variables and application possibilities that will, in turn, mean larger interaction spaces and a heavier analysis workload for the evaluator. We expect that the syworlds themselves will alleviate some of this increased workload and empower UEs to go beyond the analytical accomplishments of their predecessors.

Perhaps the u–eval–space itself is the most interesting evaluation challenge. Does the UE become disoriented or nauseated in this environment? Do we need to assess a privacy factor? The UE employs vocal, gestural and perhaps traditional input. The u–eval–space provides audio and visual feedback. *In situ* evaluation may become more important in the future as we tackle these new concerns. If we can generate a syworld anyplace, we will be doing work in public places, but still wanting to protect the anonymity of the user and the independent integrity of the UE's work.

The technologies in the story will have to change. Today, to achieve the interaction that the UEs enjoy with their syworld, users must wear a headpiece to interact "hands free" with a gaming syworld. Miniaturization will be a primary goal to make devices small enough to be embedded in clothing or accessories, out of sight, out of mind and not interfering with immersion in the synthetic world.

Intelligent agents are already commonplace, and the uses described in this paper are logical extensions of existing technology. In our story, the agent works as a UE assistant, a cyber–colleague; it goes beyond traditional information delivery; it is doing intelligent literature searches, connecting us to others in the field by bringing literature to our attention – it closes the communication loop. We already know that trust is a vital aspect in measuring acceptance of intelligent agent technology (Maes 1994, O'Connell 1990), and we know how to measure users' sense of an agent's trustworthiness. However, before we accept and use such a tool, we need to measure factors such as the relevancy of the literature the agent brings to the reporting and the willingness of our colleagues to interact with agent technology.

Our testing scenario relies heavily on natural language technologies that now exist. However, the accuracy and fluency in our story require progress in accommodating non–linguistic factors such as context.

In our story, physiometric information is collected through a jumpsuit that users wear during the usability evaluation. ICT (2008) have already identified some of the things that must change to make the smart jumpsuit a reality. Factors that impact the human experience include miniaturization of hardware to the point where it is unobtrusive; wireless communication systems that will relieve users of their tethers to physiometric collection devices; and secure data systems. The UEs in our story would stress that these data systems must be secure enough to protect the user’s privacy. In addition, the UEs would add that the interfaces between the human and the jumpsuit will have to be unobtrusive and transparent to the user experience (Jain & Pankanti 2008).

There are two sure conclusions to our examination of how things are now and what needs to change to take us to the future we envision for usability engineering. First, we will have to continue to research evaluation metrics. Then we must test them against the evolving technologies that will empower better usability engineering. To achieve the changes necessary to make our story reality, we will start with what we know; from there, we will develop new evaluation metrics and methods.

Many alternate futures are possible, but, in 2039, thanks to new developments in synthetic worlds, usability engineering should be going strong. Up-and-coming syworlds are going to be better places for both UEs and users. We'll work smarter. We'll be more efficient and effective. But we still won't be satisfied; we'll be wondering about working in the syworlds of 2069.

References

1. Al–Qaimari, G., and McRostie, D. (1999). KALDI: A Computer–aided usability engineering tool for supporting testing and analysis of human–computer interaction. In J. Vanderdonckt and A. Puerto (Eds.), *Proceedings of the 3rd International Conference on Computer–Aided Design of User Interfaces* (pp. 337–355), Louvain–la–Neuve, Belgium. Norwell, MA, U.S.A.: Kluwer.
2. Asimov, I. (1978). My own view. In R. Holdstock (Ed.), *The Encyclopedia of Science Fiction* (p. 5). London: Octopus.
3. Asimov, I. (1997). *The roving mind*. Amherst, NY: Prometheus.
4. Bayliss, J. D. and Auernheimer, B. (2001). Using a brain–computer interface in virtual and real worlds. In M. J. Smith, G. Salvendy, D. Harris, and R. J. Koubek (Eds.), *Usability Evaluation and Interface Design:*

Cognitive Engineering, Intelligent Agents and Virtual Reality (Vol. 1, pp. 312–316). Mahwah, NJ, U.S.A.: Erlbaum.

5. Choong, Y. and O'Connell, T. (2008). Planning user–centered evaluations for interactive information visualizations. *Proceedings of Usability Professionals Association International Conference,* June 16–20, Baltimore, Maryland, U.S.A.
6. Cooper, A. (1999). *The inmates are running the asylum*. Indianapolis, IN, U.S.A.: Sams.
7. Curtis, B. *People capability maturity model* (Version 2.0). Carnegie Mellon University Software Engineering Institute. Pittsburgh, PA, U.S.A.: Carnegie Mellon.
8. Emotiv Systems. (2008). Home page. Retrieved September 17, 2008 from http://emotiv.com/INDS_3/inds_3.html .
9. Franklin, S. and Graesser, A. (1997). Is it an agent, or just a program?: A Taxonomy for autonomous agents. In J. P. Muller, M. J. Wooldridge, and N. R. Jennings (Eds.), *Intelligent Agents III: Agent Theories, Architechtures, and Languages.* (pp. 21–35). Berlin: Springer Verlag.
10. Hammer, M. (1997). The soul of the new organization. In F. Hesselbein, M. Goldsmith, and R. Beckhard (Eds.). *The Organization of the Future* (pp. 25–31). San Francisco, CA, U. S. A.: Jossey–Bass.

11. ICT Results (01/07/08). Ambient Intelligence: Snowboarding to the new frontier. Retrieved September 11, 2008 from http://cordis.europa.eu/ictresults/index.cfm/section/news/tpl/article/id/89427

12. Interactive ProductLine (2008). EG Biofeedback. Retrieved September 17, 2008 from http://www.mindball.se/egg_biofeedback.html .

13. Ivory, M. Y., and Hearst, M. A. (2000). The State of the art in automated usability evaluation of user interfaces. Retrieved August 9, 2008 from http://www.eecs.berkeley.edu/Pubs/TechRpts/2000/CSD–00–1105.pdf

14. Jain, A. K., and Pankanti, S. (2008). Beyond fingerprinting. *Scientific American, September*, 78–81.

15. Jiang, J., Murphy, E. D., Carter, L. E., Bailin, S. C., and Truszkowski, W. (1993). Knowledge–based evaluation of GUI color usage. *Proceedings of MOTIF '93 and COSE: International User Conference* (pp. 78–87). Bethesda, MD, U.S.A.: Open Systems.

16. Jonas–Dwyer, D. and Pospisil, R. (2004). The millennial effect: Implications for academic development. *Proceedings of the 2004 Annual International Conference of the Higher Education Research and Development Society of Australia (HERDSA, pp. 194–207)*, Sarawak, Malaysia. Crawley, WA, Australia: University of Western Australia.

17. Karaseitanidis, I., Amditis, A., Patel, H. Sharples, S., Bekiaris, E., Bullinger, A., and Tromp, J. (2006). Evaluation of virtual reality products and

applications from individual, organizational and societal perspectives—The "VIEW" case study. *International Journal of Human–Computer Studies*, *64*, 251–266.

18. Kurzweil, R. 1999. *The Age of spiritual machines*. New York : Penguin.
19. Lecerof, A., and Paterno, F. (1998). Automatic support for usability evaluation. *IEEE Transactions on Software Engineering*, *24*, 863–888.
20. Mahajan, R., and Shneiderman, B. (1997). Visual and textual consistency checking tools for graphical user interfaces. *IEEE Transactions on Software Engineering*, *23*, 722–735.
21. Maes, P. (1994) Agents that reduce work and information overload. *Communications of the ACM*, *37*(7), 31–40.
22. Merrill, D., Kalanithi, J., and Maes, P. (2007). Siftables: Towards sensor network user interfaces. *Proceedings of the First International Conference on Tangible and Embedded Interaction* (TEI'07). Baton Rouge, Louisiana, U.S.A.
23. Murray, J. H. (1997). *Hamlet on the holodeck*. Cambridge, MA: MIT Press.
24. NeuroSky. Product Summary. Retrieved September 17, 2008 from http://www.neurosky.com/menu/main/technology/product_summary/.
25. Nordqvist, T. (1996). TUNE: A tool for user interface evaluation. *Proceedings of the Sixth Australian Conference on Computer–Human Interaction*, Hamilton, New Zealand, 129–134.

26. Norman, K.L. (2008). *Cyberpsychology: An introduction to human–computer interaction.* New York: Cambridge University Press.
27. O'Connell, T.A. (1990). A simplistic approach to internationalization: Design considerations for an autonomous intelligent agent. *Proceedings of the 6th ERCIM Workshop, "User Interfaces for All"* (pp. 15–27). Florence, Italy.
28. O'Connell, T., and Choong, Y–Y. (2008). Metrics for measuring human interaction with interactive visualizations for information analysis. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1493–1496). Florence, Italy.
29. O'Connell, T., Choong, Y–Y., Grantham, J., Moriarty, M. and Wong, W. (2008). From a Video Game in a Virtual World to Collaborative Visual Analytic Tools. *Journal of Virtual World Research*, North America, 1 1 07 2008. Retrieved November 2, 2008 from http://www.jvwresearch.org/.
30. Okada, H., and Asahi, T. (2000). Automatic consistency evaluation tool for graphical user interfaces. *NEC Research and Development*, *41*, 377–382.
31. Presser, M., Gluhak, A., Babb, D., Herault, L. and Tafazolli, R. (2006). e–SENSE – Capturing ambient intelligence for mobile communications through wireless sensor networks. *Proceedings of the 15th Information Science Technologies (IST) Summit* – Myconos, Greece, Workshop.

32. Rashid, D. and Smith, M. (2008) Relative keyboard input system. In J. Bradshaw, H. Lieberman, and S. Staab (Eds.), *Proceedings of the 13th International Conference on Intelligent User Interfaces* (IUI 2008, Maspalomas, Gran Canaria, Spain, pp. 397–400). NewYork: Association for Computing Machinery (ACM).

33. Sears, A. (1995). AIDE: A Step toward metric–based interface development tools. *Proceedings of the 8th ACM Symposium on User Interface Software and Technology* (pp. 101–110). New York: ACM Press.

34. SSB Technologies, Inc. (2006). InSight/InFocus 508 Suite. Retrieved November 3, 2008 from http://www.ssbtechnologies.com/products/infocus/indix.html

35. Steere, M. (2008). Scientist: Holographic television to become reality. Retrieved November 3, 2008 from http://www.cnn.com/2008/TECH/science/10/06/holographic.television/index.html.

36. World Wide Web Consortium (W3C). (2008). CSS Validation Service. Retrieved November 3, 2008 from http://jigsaw.w3.org/css–validator/.

37. World Wide Web Consortium (W3C). (2007). W3C Markup Validator (v0.8.3). Retrieved November 3, 2008 from http://validator.w3.org/.

38. Wilson, J. (2006). Interaction with virtual environments. *International Journal of Human–Computer Studies, 64*, 157.

39. Wilson, J. and D'Cruz, M. (2006). Virtual and interactive environments for work of the future. *International Journal of Human–Computer Studies, 64,*158–169.

40. Yang, S. (2008). Interview. Retrieved September 17, 2008 from http://www.streamocracy.com/watch/f5a9133504de568f5f6d/Neurosky–Stanley–Yang–CEO.

41. Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.