Proceedings of the Second IASTED International Conference
HUMAN-COMPUTER INTERACTION
March 14–16, 2007, Chamonix, France
ISBN Hardcopy: 978-0-88986-654-6 / CD: 978-0-88986-655-3

FINDING INFORMATION AND FINDING LOCATIONS IN A MULTIMODAL INTERFACE: A CASE STUDY OF AN INTELLIGENT KIOSK

Loel Kim, PhD
Thomas L. McCauley, PhD
Melanie Polkosky, PhD
Sidney D'Mello
Sarah Craig
Bistra Nikiforova
The University of Memphis
Memphis, TN
USA
loelkim@memphis.edu, tmccauley@memphis.edu, mpolkosky@cmamemphis.com, sdmello@memphis.edu, scraig@memphis.edu, bnikifrv@memphis.edu

ABSTRACT
Increasingly, technology developers are turning to interactive, intelligent kiosks to provide routine communicative functions such as greeting and informing people as they enter public, corporate, retail, or healthcare spaces. A number of studies have found intelligent kiosks to be usable with study participants reporting them to be appealing, useful, and even entertaining. However, the field still lacks insight into the ways in which people use multimodal interfaces to seek information and accomplish tasks. The Memphis Intelligent Kiosk Initiative project, or MIKI, was designed for multimodal use and although in usability testing it exemplified good interface design in a number of areas, the complexity of multiple modalities—including animated graphics, speech technology and an avatar greeter—complicated usability testing, leaving developers seeking improved instruments. In particular, factors such as gender and technical background of the user seemed to change the way that various kiosk tasks were perceived, deficiencies were observed in speech interaction as well as the location information in a 3D animated map.

KEY WORDS
Intelligent kiosk, multimodality, usability evaluation instrument.

1. Introduction

Increasingly, technology developers have turned to interactive, intelligent kiosks to provide routine communicative functions such as greeting and informing people as they enter public, corporate, retail, or healthcare spaces. A number of studies have found intelligent kiosks—often including avatars—to be usable, with study participants reporting them to be appealing, useful, and even entertaining [1], [2], [3], [4]. People are becoming increasingly familiar with avatars populating their informational spaces, virtual and real: Video games and online communities such as Yahoo!®, as well as museums, schools, and other public or institutionalized spaces offer avatars as guides, narrators, and virtual companions [5], [6], [7]. However, as the means for handling information tasks with communication-rich, multimodal interfaces are becoming more feasible, our understanding of the ways in which people select and use the modalities is yet impoverished.
The Memphis Intelligent Kiosk Initiative (MIKI) project offers a good case study illustrating the development and testing of a multimodal interface.

1.1 Research Background

Even ordinary tasks that an information kiosk could be expected to handle can pose design and usability issues that complicate standard user-testing efforts. Deeper discussions of information design for multimedia and complex problem-solving have been discussed elsewhere in technical communication research [8], [9], but HCI studies indicate that a successful measure for one aspect of a multimodal system can be dependent on multiple characteristics. For example, social perceptions shaping interaction quality was identified by Stocky and Cassell as contingent on the fit between personality of the avatar and the user, and the consistency of verbal and non-verbal cues across information [10],. and suggest increased complexity in capturing valid usability measures
Research is still at the beginning stages of developing guidelines for successfully incorporating multiple communication modalities, such as written and spoken language content, graphics, an avatar, and speech, into one, seamlessly functioning interface.
Understanding the nature of interactivity is key to designing a useful kiosk. The August spoken dialog system is a kiosk that helps users find their way around Stockholm, Sweden, using an on-screen street map. Much like a sideshow barker at a carnival, August uses speech to elicit a conversation from the user [11], [12], which engages the user with the information at a heightened level than text and map could do. However, this adds

another dimension of design considerations: The prevailing belief for speech technology is that users prefer human speech over synthetic speech, or text-to-speech (TTS), but it is impractical to record a complete range of possible responses needed in dynamic voice interfaces with human voice talent. Thus, a combination of human and TTS are typically used. However, Gong and Lai noted in their study that *consistency*, or a seamless match between features of speech quality and perceived personality is key to users' willingness to interact, and ability to comprehend and process information smoothly [13]. Further, Lee and Nass' study shows that increased social presence is perceived when the personality of the speaker and the verbal content match [14].

In terms of making design decisions, what personality type or combinations of characteristics will effectively reach most people in a public setting? One approach to ensuring usability is through user training: The MINNELLI system is designed to interact with bank customers through the use of short animated cartoons that present information about bank services [15]. However, training probably adds an unreasonable burden for a casual user of a general kiosk.

Another successful kiosk with a broader scope is the MACK system [16]. Stocky & Cassell developed MACK as an embodied conversational information kiosk that took into consideration the context of the surrounding space, allowing users to optimize knowledge drawn from the environment as they interacted with each other, a type of intelligence they call "spatial intelligence" (p. 224).

These are but a few of the concerns facing the development team as they set out to design and build an intelligent kiosk, but it was clear that as the "reach" of multimodal interfaces increasingly extends beyond the immediate screen of a monitor, effective design must include consideration of the physical presence and fit of the object, in this case, a kiosk, and the environment in which it resides, and is used.

1.2 Background - MIKI

MIKI is an interactive, natural language information kiosk, located in the lobby at the main entry to the FedEx Institute of Technology (FIT), at The University of Memphis. Through a touch screen, MIKI delivers information about people, offices and research centers located there, plus a listing of events.

In many ways serving as a showcase for the university, the FIT is a multi-use building, housing classrooms, research centers, and supercomputing facilities, but also offering sleek, state-of-the-art conference facilities. Thus, in addition to students, faculty, and researchers, people from public and corporate organizations frequent the building. Mary Bartlett, conference director, noted 15,000 guest visits from July 2005 to June 2006 (personal communication, September 12, 2006), so traffic is relatively heavy, making this a good location for an information kiosk.

When entering the FedEx Institute of Technology through the main entrance, MIKI is immediately visible to the visitor, directly ahead and to the left. MIKI is mounted on a large granite wall, opening on the left to the larger part of the lobby and, flanked to the right by a large set of stairs to the mezzanine. The kiosk competes visually with a large, circular information desk visible to the left.

At rest, the screen cycles through general building information and text invitations to approach and use the kiosk. A small, FireWire webcam is installed above the display and serves two main purposes. The first is to identify individuals who approach the kiosk through face recognition; the system then determines the size of that face to ascertain the distance from the monitor. The second purpose of the camera is to record interactions with the kiosk. When the kiosk registers that a person is approaching, the avatar greets and invites the passer-by to ask a question (see Figure 1).

Figure 1. The main screen of MIKI, shows the male or female (shown) avatar in the lower, left-hand corner of the 52-inch plasma screen.

In addition to the camera, an Acoustic Magic® phased-array microphone is installed below the display. The microphone can isolate a person's voice from a distance and provide high quality audio input. Although still not as good as a head mounted microphone, it provides unobtrusive filtered audio input to the speech recognition system.

MIKI handles a variety of questions:

- History and general information about the FedEx Institute of Technology (FIT)
- Location of people and research centers housed at the FIT
- Events taking place at the FIT
- Directions to rooms within the building

In designing the kiosk, a number of questions emerged that drove our design choices:

When faced with a rich, multi-modal interface, how do people select and use modalities for basic information seeking tasks?

How do people select and use modalities for location-seeking tasks? Can richer graphical information (3D) help people navigate the building more successfully than limited graphical information (2D)?

What makes an appealing avatar? Do different people respond differently to being greeted and given information by avatars of different gender, ethnicity, and perceived social standing?

Along with the richness of offering an avatar, come many design decisions about both the obviously visible as well as the subtle features of the humanlike companion: physical appearance including gender; skin color; hair texture, length, color, and style; eye shape and color; body type, and even how much of the body to include. In addition, subtle features of human behaviour that informs our "reading" of the verbal messages we convey when in person, such as eye gaze, body movements, voice quality, accent, and pacing.

When given an avatar "greeter," complex information delivery and use is a technical communication issue addressed as an increasingly important issue as information is displayed, shaped, and delivered via multiple modalities. A usability evaluation tested 38 users' abilities to accomplish the most common tasks MIKI was designed to support. Testing methodology, results, and discussion follow.

2. MIKI Usability Testing

2.1 Methodology

The usability test was a within-subjects, 2 x 2 x 2 repeated measures design (see Table 1).

Table 1. Design Factors

Factor	Level Names	# Participants
Gender	Female	23
	Male	15
Avatar persona	Khadejah (K)	21
	Vince (V)	17
Discipline	Humanities	28
	Engineering	10
Task Order	Place→Event→Person	11
	Place→Person→Event	0
	Event→Place→Person	5
	Event→Person→Place	4
	Person→Place→Event	10
	Person→Event→Place	8

Tasks & Measures. Each participant was verbally instructed to complete three tasks with the Intelligent Kiosk:

1. Find a person
2. Find a place
3. Find an event

These tasks were representative of the most commonly requested information by visitors to the FedEx Institute of Technology.

Seven measures were used in usability testing, three scaled instruments: (1) After Task Questionnaire (ATQ), (2) Usability Scale for Speech Interfaces (USSI), (3) Post-Scenario System usability Questionnaire (PSSUQ); three observational measures: (4) Task Completion Measures, (5) Observed Usability Problems, (6) Observed Interaction Preferences; and (7) Qualitative Interviews, using a cued recall technique.

Participants. Thirty-eight participants (approximately 60% female and 40% male) were recruited from two summer communication course sections at the University of Memphis (see Table 2). Ten participants were drawn from a course comprised of engineering students, a technology-intensive discipline, and twenty-eight were drawn from a section of humanities majors, a relatively technology-non-intensive discipline.

Table 2. Participant Makeup by Gender

Discipline	Male	Female	Total
Engineering	8	2	10
Humanities	7	21	28

Participants received extra course credit for their participation in the study.

2.1.1 Usability Test Measures

Measurement consisted of a variety of observational measures and rating scales, as well as participant responses to interview items. The measures presented to each participant were:

1. After Task Questionnaire (ATQ) – The ATQ is a validated 3-item, 7-point scale that measures the user's immediate satisfaction of the task just completed [17] Users filled out one for each of the three tasks. To complete the scale, participants rated their relative agreement with each item:

I am satisfied with the ease of completing this task

I am satisfied with the amount of time it took to complete this task

I am satisfied with the support information (online help, messages, documentation) when completing this task

2. The Usability Scale for Speech Interfaces – This scale is a 25-item, 7-point measure that assesses the usability of speech technologies [18]. It uses four factor scores—U**ser Goal Orientation, Speech Characteristics, Customer Service Behavior, Verbosity—**with 6 – 8 items for each factor:

Sample items

The IK made me feel like I was in control.

I could find what I needed without any difficulty.

The avatar's voice was pleasant

This avatar used everyday words.

The avatar spoke at a pace that was easy to follow

The messages were repetitive.

The avatar was too talkative.

Participants filled out the scale after all three tasks were completed, rating their relative agreement with each item.

3. Post-Scenario System Usability Questionnaire (PSSUQ) – The PSSUQ is a validated 16-item, 7-point scale that measures usability [19]. It provides three factor scores—**System Usefulness**, **Interface Quality, and Information Quality**—as well as an overall usability score. Participants filled out the scale after all three tasks were completed, rating their relative agreement with each item.

Sample items

I am satisfied with how easy it is to use the IK.
It was simple to use this system.

4. Task Completion – For each task, the evaluators recorded time on task, and whether or not the participant successfully completed the task.

5. Observed Usability Problems – As each participant completed each task, two evaluators observed participant behavior to describe and record any usability problems encountered during the task. After the evaluation, each usability problem was ranked according to severity:

1 = no usability problems observed
2 = mild confusion <1min
independent task completion
3 = confusion >1min
independent task completion
4 = confusion with task stoppage
recovery using provided supports (e.g., help)
5 = task failure or abandonment

6. Interaction Preference – During each task, evaluators recorded whether or not participants used either the graphic user interface (GUI), the speech user interface (SUI), or both.

7. Qualitative Interview Items Two evaluators interviewed participants, prompting them by showing them problem screens noted in 5. Observed Usability Problems. Cued recall interviews helped flesh out details of participants' likes/dislikes, their interaction preferences, places they thought necessary information was missing, avatar-related, and other design changes. When appropriate, they were asked to describe the perceived source of any confusion or usability problems they encountered during task completion.

2.1.2 Usability Test Results

An ANOVA with order cast as a between subjects variable indicated a non-significant effect ($p > 0.05$); therefore, order effects were not present and were not considered in subsequent analyses.

1. After Task Questionnaire (ATQ) – Participants rated their satisfaction levels for the Find a Person and Find an Event tasks relatively positively with the Find a Place task rated below the midpoint of the scale.

Table 3. After Task Questionnaire (7-pt scale)

Task	mean	sd
Find Person	6.65	0.50
Find Place	3.73	1.16
Find Event	6.90	0.32

A three-way mixed model repeated measure ANOVA was performed with participant ratings on the ATQ as the within-subjects factor and the participant's gender, discipline (engineering or humanities), and the avatar persona (Khadejah or Vince) as between-subject factors. The main effect of task was statistically significant, $F(2, 62) = 43.077$, $MSe = 1.523$, $p < .01$. Bonferroni post hoc tests revealed that participant ratings after locating a person and an event were on par and significantly higher ($p < .01$) than the ratings associated with finding a place. The main effects for the between subject factors including participant gender, participant discipline, and avatar persona were not statistically significant ($p > 0.4$). Additionally, higher order interactions between the ATQ and the other two between-subject factors were not statistically significant.

2. The Usability Scale for Speech Interfaces – Table 4 provides the mean and standard deviation for each of the four factors of the Usability Scale for Speech Interfaces.

Table 4. Usability Scale for Speech Interface

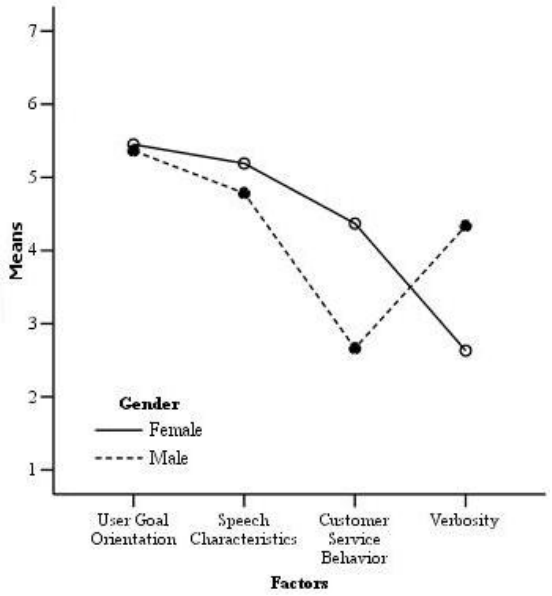
Speech Categories	mean	sd
User Goal Orientation	5.28	1.46
Speech Characteristics	5.04	1.74
Customer Service Behavior	6.02	1.25
Verbosity	3.52	1.91

A three-way mixed model repeated measure ANOVA was performed with participant ratings on the USSI as the within-subjects factor and the participant's gender, discipline (engineering or humanities), and the avatar persona (K or V) as between-subject factors. The main effect of factor was statistically significant, $F(3, 93) = 13.696$, $MSe = 1.752$, $p < .01$ (partial $\eta^2 = .306$), suggesting that participant ratings significantly differed across the four factors of the USSI. Main effects for the between subject factors including participant gender, participant discipline, and avatar persona were not statistically significant ($p > 0.05$).

Higher order interactions between the USSI and participant gender and discipline were not statistically significant ($p > .05$). However, a statistically significant interaction between participant ratings on the USSI and gender was discovered, $F(3, 93) = 7.692$, $p < 0.01$ (partial $\eta 2 = .199$). Participants of both genders provided similar ratings with respect to the User Goal Orientation factor; however, ratings by males for the Speech Characteristics and Customer Service Behavior factors were quantitatively lower than their female counterparts (see

Figure 2). Additionally, females provided lower Verbosity ratings than males.

Figure 2. Means for participant ratings on the USSI segregated by participant gender.



3. Post-Scenario System Usability Questionnaire (PSSUQ) – A three-way mixed method repeated measure ANOVA was performed, with participant ratings on the PSSUQ as the within-subjects factor, and the participant's gender, discipline (engineering or humanities), and the avatar persona (K or V) as between-subject factors. Participant ratings across the 3 factors of the scales were similar (see Table 5).

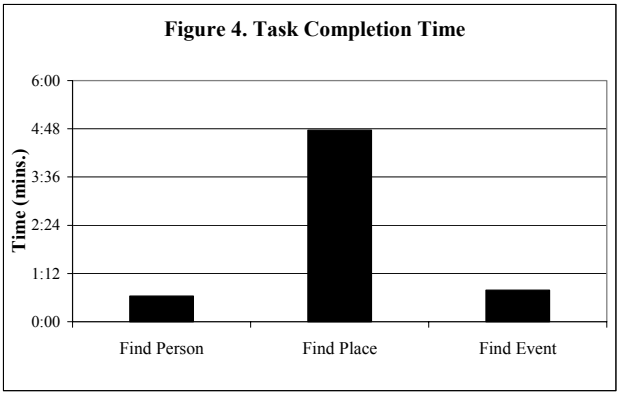
Table 5. Post-Scenario System Usability

Interface Categories	mean	sd
System Usefulness	5.54	1.48
Interface Quality	5.56	1.59
Information Quality	5.04	1.79
Overall	5.36	1.64

Main effects for the three between subject factors including participant gender, participant discipline, and avatar persona were not statistically significant ($p > 0.05$). Additionally, higher order interactions between participant ratings on the PSSUQ and the other two between subject factors were not statistically significant ($p > 0.05$).

4. Task Completion – Overall, task completion rates were high (see Figure 4): All participants successfully completed the Find a Person task, and only one person failed to find an event. The lowest completion rate was shown by the Find a Place task, in which 5 of 38 participants failed. Additionally, completion time for the Find a Place task was also elevated in comparison to the other two tasks.

Figure 4. Task Completion Time



5. Observed Usability Problems – By far, people exhibited the most problems trying to accomplish the Find a Place task. We observed a number of problems contributed to this result:

First, tab labels were unclear to most people. Further unclear language for lists, such as Specialty Rooms misled 33 out of 38 participants.

Once the language failed them in the first two to three steps, twenty-nine of 38 participants were reduced to cycling through links to the four lists for the building's four floors. However, the lists did not contain enough information, giving room numbers, but not names of centers, further frustrating participants.

Figure 5. MIKI Directions screen showing 3D floor plan of the building. Floor plan continually rotates as long as screen is visible.

Navigational expectations were unmet for many: Fifteen of the 38 participants looked for a direct link from the Directions screen to the center's home page. The text and navigational problems were then compounded when they failed to find adequate help (17/38), and, finally, when 22 participants turned to the SUI, the voice recognition failed 14 of them.

Finally, observational notes indicated that participants did not find the 3D animated floor plan helpful. One participant's comment summarizes well the observed participant experience: "Too many things going on with directions—animation, voice, text—confusing."

6. Interaction Preference – A count of participants who used the GUI, SUI, or both interfaces showed a strong preference for the GUI interface, as shown in Table 6.

Notably, for the Find a Place task, the majority of participants used both interfaces, possibly due to the number of usability problems encountered during this task.

Table 6. Interaction Preference

Task	GUI	SUI	Both
Find Person	33		5
Find Place	1		22
Find Event	33	4	1

When asked which interface they prefer, participants also expressed a strong preference for the touch screen/GUI (84% of participants), as compared with the voice interface (5%) and both interfaces (11%).

3. Conclusion

From the usability tests, MIKI exemplifies much of good interface design, primarily the quality of the graphics, screen layout, and organization of most of the information. However, a number of lessons were learned in this project that can be applied to further development of MIKI and of other kiosk interfaces that are intended to address information tasks of varying complexity for a public audience.

The data indicated that participants needed more time task and encountered more problems with the Find a Place task. This finding appeared to be at least partly due to the lack of visual orientation of the 3D building floor plan. Adding clearly marked start and end points would help anchor the image of the building to the visitor's sense of the physical context—both the immediate physical surroundings of the lobby and kiosk area, as well as the building beyond.

In addition, although the spinning 3D floor plan may have added visual interest to the screen, the user could not control the animation—either the speed of spinning, or the capability to stop, reverse, or zoom in. Thus, the user could not view at an individual pace, or stop the spinning to study details. We believe that any potential to enhance the user's sense of the building by showing it from all angles was at best, diminished. We do not know how much the added visual busy-ness may have impaired the user's ability to understand the information or, ultimately, to accomplish the task. Would a 2D image been easier to grasp than 3D? What kind of cognitive load did this bear on the user as he or she was juggling other processing tasks? Accordingly, next phase development will include adding points of reference (e.g., "You are here" and "Your destination is here"), further exploration of 2D vs. 3D, and value added by the ability to spin the image.

Finally, we do not know how much better participants would have performed this task had they had some type of aide memoire of the information once they found the correct directions. We did not specifically measure how long they could retain the directions in memory, but our informal observations of the participants' confusion after leaving the kiosk suggest this may have been the case. Possibly a printout or a downloadable set of directions would be necessary to fully help users with this task.

Making the speech feature more robust is clearly necessary, and a more focused analysis of the advantage of speech for certain information tasks would be a first step. A more fine-grained application of speech technology in which we target the points at which people turn to speech of the other modalities offered, would be most helpful would not only alleviate the burden on a whole-system speech integration, but would take better advantage of the strengths of offering information through speech, as opposed to pictures or text.

As far as developing best practices for using avatars in multimodal interfaces, much more work remains on the interplay of persona features. Initial studies into the cultural impact of our choices of avatar gender, ethnicity, social standing, and culturally shaped behaviors such as eye gaze, not only mediate the interaction and perceived quality of the communication [20], but also convey paralinguistic information shaping the message and perpetuating cultural attitudes [21].

With additional focused research and development, a robust, interactive information kiosk can be successfully deployed in a number of different domains. Some examples include retail outlets such as malls and "big-box" type stores. Additionally, healthcare institutions and corporate office buildings can also make use of this type of kiosk. The key limitation, at this point, rests in the general applicability of speech recognition for a broad audience and a diverse conversational domain. Our research also points to the difficulty of using speech as an interface medium within a public space. Further research needs to be conducted to better quantify this effect.

Even so, this technology shows great promise. Therefore, the next time you're in your local mall don't be surprised if a large video display tries to strike up the conversation.

Acknowledgement

The MIKI project was funded by the 2005 FedEx Institute of Technology Innovation Fund.

References

[1] A.D. Christian & B.L. Avery. Speak out and annoy someone: Experiences with intelligent kiosks, *Proceedings of CHI '98*, 2002, 313-320.

[2] C. Guinn & R. Hubal. An evaluation of virtual human technology in informational kiosks, *ICMI '04*, October 13-15, 2004, State College, Pennsylvania, USA, 2004.

[3] E. Mäkinen, S. Patomäki & R. Raisamo. (). Experiences on a multimodal information kiosk with an interactive agent, *ACM International Conference Proceeding Series, 31*, *NordiCHI October 19-23*, Ärhus, Denmark, 2002, 275-278.

[4] Cavalluzi, B. De Carolis, S. Pizzutilo & G. Cozzologno. Interacting with embodied agents in public environments, *Proceedings of AVI*, May 2004: 240-243.

[5] Yahoo!® Canada Avatars. Accessed online September 2006: http://ca.avatars.yahoo.com/

[6] M. Roussou, et al. Experiences from the use of a robotic avatar in a museum setting, *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, Nov. 28-30, Glyfada, Greece*; 2001, 153-160.

[7] C. Guinn & R Hubal. An evaluation of virtual human technology in informational kiosks. *Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, USA,* 2004, 297-302.

[8] J.T. Hackos, M. Hammar & A. Elser. Customer partnering: Data gathering for complex online documentation, Com Tech Services, 1997. Accessed online: http://www.comtech-serv.com/pdfs/Customer%20Partnering%20IEEE.pdf ComTech.

[9] M. Albers. Design considerations for complex problem-solving. *STC Proceedings*, 2002. Accessed online: http://www.stc.org/confproceed/2002/PDFs/STC49-00011.pdf

[10] T. Stocky & J. Cassell. Shared reality: Spatial intelligence in intuitive user interfaces, *IUI '02*, January 13-16, 2002, San Francisco, California, USA.

[11] J. Gustafson, N. Lindberg, & M. Lundeberg. *The August spoken dialogue system*. in *Eurospeech '99*. 1999.

[12] J. Gustafson, M. Lundeberg, & J. Liljencrants. *Experiences from the development of August - a multimodal spoken dialogue system*. in *IDS*. 1999

[13] L. Gong & J. Lai. Shall we mix synthetic speech and human speech? Impact on user's performance, perception, and attitude. *Proceedings of SIGCHI '01, March 31- April 4, Seattle, WA*, *3*(1), 2001, 158-165.

[14] K.M. Lee & C. Nass. Desinging social presence of social actors in human computer interaction. *CHI 2003, April 5-10, Ft. Lauderdale, Florida, USA 5*(1), 289-296.

[15] P. Steiger & B.A. Suter. *MINELLI - Experiences with an Interactive Information Kiosk for Casual Users*. in *UBILAB '94*. 1994. Zurich.

[16] J. Cassell, et al. *MACK: Media lab Autonomous Conversational Kiosk*. in *Imagina '02*. 2002. Monte Carlo

[17] J.R. Lewis. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, *International Journal of Human-Computer Interaction*, *7*, 1995, 57-78.

[18] M. Polkosky. *Toward a psychology of speech technology: Affective responses to speech-based e-service,* Doctoral dissertation: University of South Florida, 2005.

[19] J.R. Lewis. Psychometric evaluation of the PSSUQ using data from five years of usability studies, *International Journal of Human-Computer Interaction*, *14*, 2002, 463-488.

[20] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, & M.A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment, *CHI Proc. 5*(1), Ft. Lauderdale, FL, USA, 529-536.

[21] S. Zdenek. "Just roll your mouse over me": Designing virtual women for customer service on the web, *Technical Communication Quarterly*, forthcoming 2007.