INTL. JOURNAL OF HUMAN–COMPUTER INTERACTION, *25*(2), 134–154, 2009

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447310802629561

Taylor & Francis
Taylor & Francis Group

The Effects of Different Presentation Modalities on a Person Identification Task: A Study of Wireless Multimedia PDA Use in a Security Context

Dan J. Kim[1], Charles Steinfield[2], and Pamela Whitten[2]
[1]University of Houston – Clear Lake, Houston, Texas
[2]Michigan State University, East Lansing, Michigan

The convergence of affordable multimedia capabilities with local and wide area wireless data networking on handheld devices is a particularly important trend. Using off-the-shelf technology, we conducted a simple experiment that required participants to engage in a person-identification task while posing as security agents in a field setting. Participants attempted to identify a target person in a public setting under different conditions with or without the aid of video information distributed to them via wireless-enabled multimedia PDAs. We developed a research framework and expectations based on previous research in the areas of the dual processing theory of working memory, information richness theory, and the role of video in grounding. Our findings suggest that dual modes of presentation (i.e., phonological and visuospatial information) simultaneously incorporated by a specific presentation significantly influence user perception and enhance task performance.

1. INTRODUCTION

Advances in mobile computing and wireless communications are appearing at a dizzying pace, creating the possibilities for new types of communication situations, new business applications, and new consumer services. The convergence of affordable multimedia capabilities with local and wide area wireless data networking on handheld devices is a particularly important trend. The statistics are telling: according to the International Telecommunications Union (ITU, 2008) at the end of 2006, about 76% of the more than 2.6 billion mobile telephones in the world

The authors thank Michigan State University, which supported this work through an Intramural Research Grant as well as through the MSU CyberSecurity Initiative. The authors thank Zoo H. Chae, Peter Murray, Christiana Sargent, and Amar Shool for their assistance in the field experiment. We also thank the management of the mall that hosted this study for their cooperation.

Correspondence should be addressed to Dan J. Kim, University of Houston Clear Lake, 2700 Bay Area Boulevard, Delta 169, Houston, TX 77058. E-mail: kimdan@uhcl.edu

were digital. Increasingly, these wireless mobile devices are multimedia, including a built-in camera that makes it easy for people to take and share images or video clips over the Internet for all to see. Already major events like the Virginia Tech massacre have been brought to the eyes of international viewers with grainy pictures and footage captured by phones.

Although multimedia messaging (MMS) is still in its infancy, the sheer numbers of users now possessing camera-equipped cell phones and PDAs suggests that it may follow in the footsteps of short message service (SMS), especially among younger people. According to a recent press release (CNN, 2007), research by Gartner Group indicates that in 2010 over 1 billion camera-equipped cell phones will be sold worldwide, nearly double the 589 million predicted to sell in year 2007. As the world's cellular operators roll out high-speed data transmission services, the sending of still images and video clips has become one of the initial service targets.

With the low cost and widespread availability of wireless-enabled multimedia devices will come new applications in both the consumer and business market.[1] This widespread proliferation represents an important research challenge for communication researchers. Although there is a long history of research, especially on video conferencing (Egido, 1990), the systems studied were never truly ubiquitous and mass market technologies. Rather, much of the early research on the importance of video in mediated interactions used experimental prototypes or expensive systems only available in large companies rather than off-the-shelf consumer technologies. This early research often relied on concepts from social presence (Short, Williams, & Christie, 1976) and information and media richness (Daft & Lengel, 1984) and questioned whether video channels were essential to most tasks in organizational settings. More recent work focusing on dyadic and group collaboration, however, illustrates the importance of video information in collaborative tasks (Fussell, Kraut, & Siegel, 2000). Video information in this latter work complements audio information to help establish common ground.

Another reason to revisit questions about the importance of multimedia-supported interactions results from the evolution of communication technologies from usage in stationary contexts with larger devices to mobile environments with handheld equipment (Zhang & Adipat, 2005). Differences in portability, ease of use, screen size, and user interfaces call for new empirical work to better understand the utility of video information in the new types of contexts in which usage can now occur. Multimedia communication aids will find application in a much wider range of everyday tasks, in both business and personal settings. To this end, in this article we explore the contribution of still and moving video on users' performance and perceptions in a simulated security-related task. Although our goal is to broadly explore the potential usefulness of video images in everyday

[1]For a look at the many types of applications and services under consideration and in early trials, visit such Web sites and blogs as Multimedia Messaging.com (http://www.multimediamessaging.com), Picturephoning.com (http://textually.org/picturephoning), and MobileMMS.com (http://www.mobilemms.com).

tasks and activities, there are clearly important public safety applications that represent near-term targets for such devices.[2]

Using off-the-shelf technology, we conducted a simple experiment that required participants to engage in a person-identification task while posing as security agents in a field setting. Participants attempted to identify a target person in a public setting under different conditions with or without the aid of video information distributed to them via wireless-enabled multimedia PDAs. We develop expectations based on previous research in the areas of dual-modal information presentation, information richness, and the role of video in grounding. Our findings confirm the effectiveness of users' information comprehension using the dual-modal information presentation and suggest new directions for research on this new generation of video-enabled devices.

The remainder of the article is organized as follows. Section 2 reviews prior research, primarily in the areas of dual processing theory of working memory, visual and auditory information process, and information richness, that can be used to justify expectations for an effect of dual-modality (i.e., audio and visual information) on task performance and user perceptions. This review is summarized with a conceptual research model and supporting hypotheses. In section 3, we describe the research setting, data collection methods, and analysis strategy. Results are summarized in section 4, and the final section provides a discussion of the findings, limitations, and implications for theory and practice.

2. REVIEW OF LITERATURE

To provide a theoretical framework of the study, we have reviewed a few streams of research: working memory, dual processing theory of working memory, and representation of information in terms of information richness and social presence.

Human information processing is a cognitive process that involves the encoding, organization, storage, retrieval, and comparing of information (Dillon & Pellegrino, 1991). During information presentation, people must construct a mental representation of situations conveyed by the words and images contained in the presentation. The working memory model proposed by Baddeley and Hitch (1974) argues that humans have separate visual and verbal information processing systems. The model depicts three components: central executive, visuospatial sketchpad, and phonological loop (see Figure 1). According to the model, human working memory has two subsystems for storage: visuospatial sketchpad and phonological loop. The visuospatial sketchpad is responsible for visual coding and handling spatial imagery information in analog forms, whereas the phonological loop refers to acoustic or phonological coding, which plays an important role in reading, vocabulary acquisition, and language comprehension. The phonological loop and

[2]For example, law enforcement agents are using camera phones to send photos of offenders or missing children for quick identification on the scene and there have been numerous reports of citizens using camera phones to capture crimes in progress (CBSNews, 2003).

FIGURE 1 The working memory model proposed by Baddeley and Hitch (1974).

visuospatial sketchpad are able to simultaneously hold verbal and imagery information without interference. The central executive is the control system that supervises and coordinates the information retrieved from the two storage subsystems for further integration.

Baddeley and Hitch's (1974) model has been used and confirmed by many studies. For example, Mousavi, Low, and Sweller (1995) found that a combined verbal and graphics presentation reduced the cognitive load in solving geometry problems, resulting in superior learning. They also suggested that distributing relevant information in visual and auditory modalities might effectively increase working memory.

Drawing from Baddeley and Hitch's (1974) model, the dual-coding theory of working memory proposed by Paivo (1986) suggests that human information processing can take place in a verbal working memory and/or in a visual working memory workspace. These two working memory workspaces facilitate the encoding of information as functionally independent but interconnected systems. The visual working memory system is considered to be more efficient than the verbal working memory system, because the verbal system encodes information by sequential processing of clauses that are combined to convey a set of concepts and the verbal information must be transformed into imagery that illustrates spatial arrangement, physical characteristics, and temporal/causal order of events (Paivo, 1986).

Archer, Head, Wollersheim, and Yuan (1996) compared the user's preferences and the effectiveness of information delivery in visual, auditory, and visual-auditory modes. They suggest that information should be organized according to its perceived importance to the user, who should also have flexible information access at different levels of abstraction. Research on interaction between sound, written words, and the image of objects shows that when different sources of information are integrated, a learner's cognitive overload remains light and does not limit learning (Dubois & Vial, 2000). Andres and Petersen (2002) suggest that multimedia presentations invoke the use of both the verbal and visual working memory channels, resulting in a reduction of the cognitive load imposed by increased information complexity. The previous studies on visual and auditory dual-model interfaces are well summarized in the study conduced by Fang, Xu, Brzezinski, and Chan (2006). They investigate the feasibility of simultaneously presenting distinct textual information through dual channels (i.e., visual and auditory) by examining two multimodal interfaces. The findings suggest that "users can attend to auditory information while visually browsing textual information and that information overlap may reduce distraction" (p. 3). They conclude the study with a possibility that multimodalities (visual plus auditory modes) present more information than in a single modality.

In organizational research, there is a tradition in which organizational communication media are viewed in terms of their capacity to carry information (Thomas & Griffin, 1983; Thompson, 1967). Organization and information systems researchers studying the importance of media characteristics on communication tasks in the workplace emphasize the role of information richness in task performance (Daft & Lengel, 1984, 1986; Daft, Lengel, & Trevino, 1987; Dennis & Kinney, 1998; Suh, 1999). Information richness theory (IRT) or media richness theory is based on the information processing model of organizational communication and is an extension of contingency theory (Suh, 1999), which states that organizational information processing must match the complexity of environmental inputs in order for the organization to endure and prosper. Rich media are viewed as critical to support tasks characterized by a high degree of equivocality, whereas lean media are more efficient for routine and unambiguous situations.

Researchers have long studied the affects of mediated communication, and especially video interaction, on task performance (Burke, Aytes, & Laku, 2001; Daft & Lengel, 1984; Dennis & Kinney, 1998; Fulk, Steinfield, Schmitz, & Power, 1987; Kahai & Cooper, 2003; Markus, 1987; Robert & Dennis, 2005; Suh, 1999). A common thread in earlier work is that richer media such as video do convey more *social presence*—loosely defined as the extent to which the medium permits users to perceive each other to be psychologically present during interactions—but may not be necessary for all types of tasks (Short et al., 1976). Rather, the early findings emphasized the importance of video for person perception, and in interpersonally involving tasks where there may be negotiation or conflict but not in cooperative problem solving situations (Short et al., 1976). Although results were largely from laboratory studies, early fieldwork appeared to confirm that the addition of a video channel did not necessarily enhance the outcomes from group interactions (Johansen, 1984). Egido (1990) summarizes the early, and somewhat disappointing, findings for video conferencing, as well as its limited acceptance in the market but argues that as the technology becomes more available and people start using it for tasks where the video information really matters, it will become more valuable and widely accepted.

Daft and Lengel (1984) proposed that media vary along a continuum of information richness, based on four criteria: the speed of feedback, the range of modalities supported, the personalness of the source, and the relative ability to carry natural language. From the perspective of IRT, face-to-face communication is the richest medium followed by telephone, E-mail, letters, notes, memos, special reports, and fliers and bulletins (Daft & Lengel, 1986). Later work by Trevino, Lengel, and Daft (1987; 1990) provided some support for IRT, in that the more effective managers were found to be those who tended to select the most appropriate medium for a given task. However, Fulk and colleagues (Fulk, Schmitz, & Steinfield, 1990; Fulk et al., 1987; Schmitz & Fulk, 1991) criticized both social presence and information richness theory for failing to take the social and organizational context into account. Recently, Robert and Dennis (2005) argued the paradox of richness; the use of rich media high in social presence induces increased motivation but decreases the ability to process information, whereas the use of lean media low in social presence induces decreased motivation but increases the ability to process information.

These researchers emphasize the role of social influences on actual media choice patterns exhibited by organizational participants, which may or may not conform to the expectations derived from social presence or information richness theory. Arguing from a social construction perspective, they propose that a combination of influences from the social and organizational context, as well as individuals' previous experiences, shape perceptions about how lean or rich any medium might be and in what situations it is appropriate to use a medium.

Over the past decade, information systems researchers have tested basic propositions from IRT with quite mixed results. Table 1 summarizes the research issues, independent, dependent variables, and key findings of a group of selected experiments testing IRT-derived propositions. Most examined information richness theory in the context of two different types of task (equivocal/non-equivocal, high-equivocal/low-equivocal, cognitive/intellective, or negotiation/intellective) and multiple levels (two to four) of dyadic communication cues (e.g., face-to-face, audio, and computer-mediated text modes). As dependent variables, almost all studies measured task performance (e.g., decision time, quality, and consensus) and users' task perceptions (e.g., process and outcome satisfaction).

This review suggests that caution is advised when applying IRT in practice. Not only do studies in the field find that media choices cannot be fully explained by a strict media-task fit model, but laboratory studies are mixed in terms of the hypothesized relationship between the media-task fit and performance. Certainly social influences can help explain the mixed results, particularly in field settings. Although these types of effects ought to be more limited in laboratory research, group composition factors still seem to overwhelm any effect of the medium (Hollingshead, McGrath, & O'Connor, 1993).

One explanation for the pattern of performance results can be derived from Egido's (1990) opinion that at least up to the point of her review, there was a lack of real utility for video. It may be that researchers would find greater effects with more of a focus on situations where there is an obvious need for video information. Fussell et al.'s (2000) work on the role of video in helping two people establish common ground attempted to explore such a task. They examined a common type of situation—an expert helping a novice repair something, which in their study was a bicycle. The expert had a clear need to see the bicycle in order to better direct the actions of the novice. Their research focused not only on task performance but explored the verbal interactions that made use of the shared context supplied by the video information. Task performance effects were difficult to find, as the video-aided group did not perform as well as a side-by-side group and no better than an audio-only group. However, Fussell et al. (2000) show how the conversation of the video-group relied on shared visual context to focus attention, monitor understanding, and improve conversational efficiency. Limitations in the video system—e.g., the expert had no way to point at something in the helper's room, he had to verbally refer to it—were used to explain the lack of performance findings. More recent work by these researchers has demonstrated the importance of video information in collaborative tasks (Kraut, Fussell, & Siegel, 2003; Kraut, Gergle, & Fussell, 2002).

We thus face considerable uncertainty in developing expectations for the effects of use of multimedia PDAs on tasks requiring video information. Earlier

Table 1: Selected IRT Studies

Research issues	*Independent variables*	*Dependent variables*	*Key findings*	*Authors*
The effect of medium and task on dyadic communication	Three levels of dyadic communication (face-to-face, audio, and computer-mediated text modes) High-equivocal task (a budget allocation problem) Low-equivocal task (GRE problem)	Decision time, consensus change, and communication satisfaction	Decision time varied by the medium	Kinney and Watson (1992)
Study of group task performance and communication technology	Two levels of cues (face-to-face and computer-mediated groups) Negotiation and intellective tasks	Group performance	No statistical differences were indicated between two the groups. Group membership and experience appeared to have more effects on group performance	Hollingshead et al. (1993)
The effect of media richness on decision-making	Three levels of multiplicity of cues (face-to-face, audio-video, and computer-mediated text), two of immediacy of feedback (immediate vs. delayed) High-equivocal task (college admissions) Low-equivocal task (SAT problems)	Decision quality, consensus change or communication satisfaction	The varying cues and feedback had no effect on dependent variables	Kinney and Dennis (1994)
The study on the task and media fit	Four levels of cues (face-to-face, video, audio, and computer-mediated text) Cognitive conflict (a budget-allocation problem) and intellective (a directory and map-searching problem)	Satisfaction	A predicted pattern of higher richness led to higher satisfaction for the cognitive conflict task, but no pattern showed for the intellective task	Valacich, Mennecke, Wachter, and Wheeler (1994)
The effects of media richness on decision making in two-person teams using new media	Four media conditions (video-immediate feedback, video-delayed feedback, computer mediated communication-immediate feedback, and computer mediated communication-immediate-delayed feedback) Multiplicity of cues (high, low), immediacy of feedback (immediate, delayed), task equivocality (high, low), treatment order (higher equivocal task first, lower equivocal task first)	Individual level perceptions (communication satisfaction and richness, equivocality and task complexity), dyadic level (time, consensus, and decision quality)	Use of media providing fewer cues led to slower decisions and more so for the less equivocal task. Matching media richness to task equivocality did not improve performance	Dennis and Kinney (1998)

Impact of communication medium on task performance and satisfaction	Media characteristics (face-to-face, video, audio, text) Task characteristics (intellective, negotiation)	Task performance (decision time and quality), task satisfaction (process and outcome)	There were no task-medium interaction effects on either decision quality or decision time	Suh (1999)
Media effects on the development of cohesion and process satisfaction in computer-supported workgroups	Three levels (and more) of dyadic communication cues (face-to-face, computer mediated, video, etc.) Group (proximity, size, and skills), task (structure and complexity), and context (duration, rewards, history, and culture)	Group performance (cohesion development and process satisfaction)	Initial levels of cohesion and process satisfaction differ depending on the medium's inherent richness and cohesion. Process satisfaction increase over time in all types of electronic support	Burke et al. (2001)
Impact of cue multiplicity and feedback immediacy on decision quality	Cue multiplicity and feedback immediacy (unsupported face-to-face meeting, supported face-to-face meeting, electronic conferencing, and electronic mail)	Social perceptions, message clarity, evaluations of others and decision quality	Richer media can have significantly positive impacts on decision quality when participants' task relevant knowledge is high. Effects of participant deception can be mitigated by employing richer media	Kahai and Cooper (2003)
Paradox of richness: a cognitive model of media choice	Four levels of cues (face-to-face, video, audio, computer-mediated text)	Task performance (decision time, quality), task satisfaction (process), group performance	The use of rich media high in social presence induces increased motivation but decreases the ability to process information and vice versa	Robert and Dennis (2005)

work heavily relied on costly prototypes in fixed (that is, not mobile) contexts, not low-cost, off-the-shelf and portable systems. Previous research emphasized complex group decision and problem-solving tasks in the context of intense interactions, not the conduct of day-to-day tasks where there is a well-defined purpose for visual information. However, early work did find video to be particularly important in person perception tasks, and from an information richness perspective, there is a basis for expecting that provision of such information should improve task performance. Later work has established the utility of video for collaboration across distances among dyads or small groups. Hence, the general research hypothesis explored in this study is that given a task that requires visual information, performance will be higher when it is provided. Moreover, in keeping with prior research on dual processing theory of working memory and information richness theory, the main research question of the study is:

> Can the dual modes of presentation (i.e., phonological and visuospatial information) simultaneously incorporated by a specific presentation influence user perception and enhance task performance?

3. RESEARCH MODEL AND HYPOTHESES

As a context for exploring video applications with multimedia PDAs, we settled on a security-related person identification task. We arrived at the specific task after interviewing local police, as well as management at a local mall. We specifically asked for examples of commonly occurring situations where the provision of some form of visual information would improve the ability of security personnel to do their jobs. Interviewees all described instances where an agent in the field could be remotely supplied with current visual information in order to improve his ability to correctly identify and apprehend a suspect. We learned further that one of the most significant problems facing retailers is shoplifting, which annually accounts for approximately a third of the more than $30 billion in inventory losses in the U.S. retail sector (Hollinger & Davis, 2003). A particular problem for mall security was to respond to reports that someone previously caught shoplifting and barred from the mall has returned and been observed by store personnel. Agents may not know in advance what the individual looks like and yet must apprehend and escort these prior shoplifters from the premises. Clearly, there is significant penalty in terms of customer relations if they apprehend the wrong person. Currently these tasks are usually handled by having someone in the security office verbally describe the suspect to the remote agent using a walkie-talkie. Using a wireless PDA to distribute a photograph to a remote agent at the mall should therefore improve identification accuracy. Such a photograph might have been previously taken and stored in the security office files. In addition, a "richer" form of information that could be available to agents is a video clip of the suspected shoplifter captured that same day by an on-premise surveillance camera. Conceivably, when the initial report is made by a store worker, the security office could examine the store's camera recordings, find a short clip of the suspect, and transmit it to the remote agent. Such a video should be even more helpful than a

previously taken photograph since it would show what the suspect was currently wearing, his current hairstyle, style of movement, and even some background visual information showing the last known area he visited at the mall. We discussed these scenarios with the mall management, who agreed that these represented commonly occurring situations and potentially feasible technical solutions.

We thus have a context for a simple media richness experiment that satisfies three basic requirements: (a) an application for off-the-shelf wireless equipped multimedia PDAs, (b) a task that clearly benefits from the addition of video information, and (c) the potential to vary the degree of presentation modality in the media used by participants. There are essentially three different presentation modality conditions:

- Low presentation modality: the baseline walkie-talkie condition, in which someone from the security office calls the remote agent, charges him with the task, and provides a verbal description of the target (single mode, low information).
- Medium presentation modality: a richer, still image condition, in which the security office person calls the remote agent via walkie-talkie, charges him with the task, and then provides a photograph of the target that the remote agent can view on his PDA (dual mode, medium information).
- High presentation modality: the most rich, video clip condition, in which the security office person calls the remote agent via walkie-talkie, charges him with the task, and then provides a video clip of the target that the remote agent can view on his PDA (dual mode, high information).

In all three conditions, the remote agent can seek clarification via the walkie-talkie, providing us with the opportunity to observe whether there are differences in the nature of requests for information based on the modality in which information is presented. Figure 2 illustrates this continuum of presentation modality conditions, which serves as the independent variable in the experiment.

Two groups of dependent variables derived from prior research are measured in this experiment: user perceptions of the performance and actual task performance.

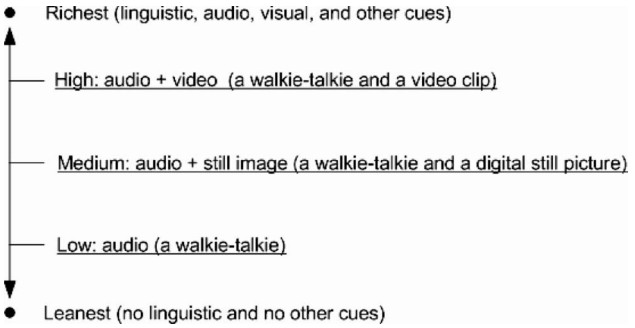

FIGURE 2 Presentation modality conditions.

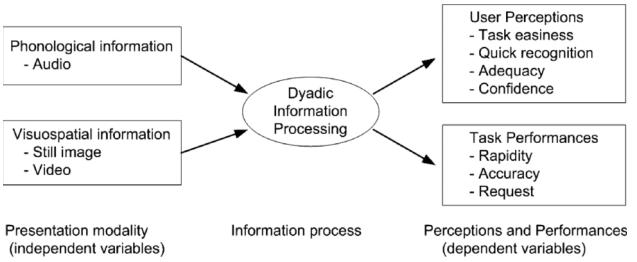

FIGURE 3 The research model.

Three measures of performance include the time it takes an agent to complete the identification task, accuracy of the identification, and the extent to which the agent requests additional information or help. We defined four user perceptions related to the task including: perceived difficulty of locating the target, perceived ease of recognizing the target once located, perceived adequacy of the information supplied, and perceived confidence that the person identified was the correct person. Figure 3 shows the research framework for the study including independent and dependent variables.

We restate our general propositions below in hypothesis form.

Hypothesis 1: An increase in presentation modality (i.e., audio → audio + still image → audio + video) will result in a greater positive impact on perception about the task.

H1a: As presentation modality increases, participants will perceive greater ease of the finding the target.

H1b: As presentation modality increases, participants will perceive greater ability to recognize the target.

H1c: As presentation modality increases, participants will perceive greater adequacy of the information supplied.

H1d: As presentation modality increases, participants will perceive greater confidence that the correct target was found.

Hypothesis 2: An increase in presentation modality (i.e., audio → audio + still image → audio + video) will result in better actual task performances.

H2a: As presentation modality increases, participants will identify the target fast.

H2b: As presentation modality increases, participants will find the target correctly.

H2c: As presentation modality increases, participants will request less additional information to help locate the target.

4. RESEARCH METHODS

We examined the above hypotheses using a field experiment design. The field setting was a local shopping mall that agreed to allow us to set up a Wi-Fi base

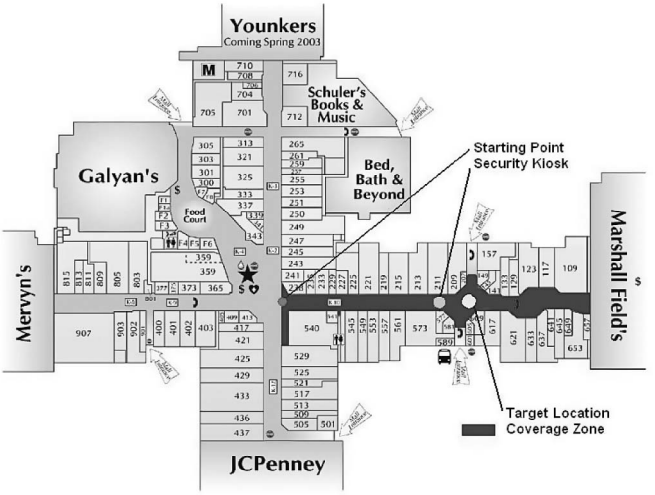

FIGURE 4 The field experiment setting.

station[3] in a busy corridor. The mall owners provided us with an empty kiosk that masked the access point equipment and served as the base for a study confederate who played the role of the security office manager (see Figure 4). Participants each had an appointed time when they arrived at the mall over a 2-week period. We selected time periods where there was a higher probability of traffic based on information from the mall owner, including 6:00 p.m. to 8:30 p.m. on weekday evenings and 1:00 p.m. to 6 p.m. on weekends. Participants were told they were to play the role of a security guard at the mall. The general scenario provided to each participant was that a store clerk had just called and reported that a former shoplifter was observed. The task was to find and correctly identify the shoplifter.

Undergraduate students in a large public Midwestern university were recruited for this experiment. On arrival at the mall, participants were randomly assigned into one of three groups: low, medium, and high presentation modality. Those in the low presentation modality group were supplied with a walkie-talkie, and those in the other two groups were supplied with a multimedia PDA[4] in addition to the walkie-talkie. They received training on the use of the equipment they would use and then were escorted by a study confederate to the start of the research area (see Figure 4). A study confederate served as the shoplifter to be

[3]Wi-Fi (or wireless fidelity) is a trademarked name from the Wireless Broadband Alliance that refers to equipment compatible with the IEEE 802.11b (11 Mbps transmission rate) wireless local area networking standard. We used a base station and external booster antenna to provide signal coverage throughout a corridor in the mall.

[4]For the field experiment, we used SONY PEG-NZ90 CLIE PDA devices with wireless LAN (802.11b) card. This latest model has lot of built-in multimedia communication features including a Bluetooth module for cable-free communication, a 2-megapixel camera supporting UXGA resolution for high-quality still photos, video recording (MPEG-4) and playback, an MP3 audio player, etc. The device also has 64k TFT color display with 480 × 320 pixels resolution and 3.9-inch touch screen.

"found" by the participants. She sat in a seating area in the center of the corridor in order to minimize any effects of her location on participants' performance. We confirmed that participants were not previously acquainted with the target and did not know of her location ahead of time.[5]

At the start of each trial, the participant received a call on his or her walkie-talkie, announcing that the shoplifter had been spotted in the corridor. In condition 1 (single modal, low information richness), a brief verbal description of the participant was supplied, including gender, approximate age, and hair color. In condition 2 (dual modal, medium information richness), the participants received the announcement and were instructed to download a photo still image onto the PDA from the security Web site. The photo was a head shot of the target taken before the research project commenced. In condition 3 (dual modal, rich information richness), participants received the same announcement and were instructed to open a video clip stored on the PDA.[6] The video clip was an overhead shot of the target looking at some merchandise in a store taken at the outset of the research project. A unique characteristic of the target was not used for the experiment, because a person identification is a relatively easy task for people when they acknowledge a distinctive feature (e.g., a bald white man wearing black sunglasses) of the target. At each trial, the target wore the same clothing as she did during the shooting of the clip. Participants then were told to walk the length of the corridor in each direction and to call the security desk as soon as they believed they had spotted the shoplifter.

The three performance dependent variables were operationalized as follows. The study confederate staffing the security office kiosk recorded the elapsed time between the start of the trial and the participant's report that he or she had found the target. The identification accuracy measured whether the participants successfully found the target or not (some walked the course without finding anyone, and others identified the wrong target). Also recorded was the number of times the participant called back to the security office confederate to ask for help or additional information.

At the conclusion of the trial, each participant was briefly interviewed about the experience and also filled out a short survey asking how easy the task was, how quickly he or she was able to recognize the target once observed, how adequate the information was, and how confident he or she was about the identification made. These were assessed with single-item 5-point Likert agree-disagree scales. After excluding three participants due to prior knowledge of the target or equipment malfunctions, a total of 95 valid participants was used, split among the three conditions. Fifty-one were male (53.7%) and 44 were female (46.3%). The average age of the participants was 21.46, with a minimum of 19 and a maximum of 27.

[5]One participant was previously acquainted with the target and dropped from the analysis.

[6]In order to minimize delays from downloading and opening the video clip, we prestored it on the PDA. We recognize that download delays would have made the task more realistic and provided a fairer test but found that the operating system was not quite user-friendly enough to avoid artificial delays for this task.

5. RESULTS

A simple analysis of variance, complemented with planned contrast tests using SPSS v. 10, was used to examine the effects of the presentation modality condition on performance and perception outcomes. Table 2 provides summary statistics and ANOVA results on each variable. Table 3 reports the correlations found among the dependent measures.

The results show there is evidence that:

1. The mean differences are statistically significant among the three levels of presentation modality for all four participant perceptions (task ease, recognition, adequacy, and confidence).
2. The mean differences are statistically significant among the three levels of presentation modality for two of the three performance measures. The elapsed time (rapidity) did not differ among the groups, but the proportion of successful identifications and frequency of requests for additional information did.

In order to find out which presentation modality group means were significantly different from each other, we conducted additional planned (a priori) contrast tests with an assumption of equal variances (Tabachnick & Fidell, 1996). Table 4 presents the results of these tests.

The pattern of contrast results is fairly consistent. In all dependent measures except perceived confidence, the walkie-talkie-only group was significantly lower in performance and perception than the still image– and video clip–aided groups. Those with the video clip were significantly more confident than the walkie-talkie group, however. Surprisingly, in every measure but the frequency of additional information requests, there was no significant difference between the still image– and video clip–aided groups. We found that the video clip did not generally improve either performance or perception over and above the gains offered by having a still photo of the target. Table 5 summarizes the extent to which all hypotheses were supported or not.

6. DISCUSSION AND CONCLUSIONS

The results offer some support for the application of dual processing theory of working memory and information richness theory to the use of multimedia mobile devices in the public safety context described here. Although there were no statistically significant differences in the amount of time it took participants in each of the three conditions to locate the target shoplifter, we did find other clear performance differences. Only about half of the participants in the low presentation modality condition were able to successfully find the target. However, more than 80% of the participants in the other two higher presentation modality groups found the target. This success rate difference also shows up when we examine the number of incorrect identifications. Nearly a fourth (22% or 5 of 22) of the low presentation modality group participants who did report finding the target

Table 2: Summary of Means, Standard Deviations, F-Ratios, and R-Squares of ANOVA Analysis

	Presentation modality[a]			*F-ratios*	
	Low	*Medium*	*High*	*(p value)*	*R-squares*[b]
Perceptions (1, *strongly disagree*; 5, *strongly agree*)					
Task ease: It was easy to locate the person I was asked to find	2.75 (1.22)[+]	3.69 (1.18)	3.87 (1.12)	**8.353 (.000)*****	.133
Quick recognition: I recognized the person I was asked to find as soon as I saw her or him	2.59 (1.32)	3.28 (1.28)	3.71 (1.30)	**5.956 (.004)****	.113
Adequacy: I was supplied with enough information to identify the missing-person	2.06 (1.11)	4.03 (.93)	4.19 (1.17)	**38.946 (.000)*****	.370
Confidence: I am confident that I identified the missing person correctly	2.88 (1.31)	3.53 (1.48)	4.10 (1.08)	**7.267 (.001)****	.136
Task performance (actual measurement)					
Rapidity: the time required for participants to locate the person (in seconds)	276.13 (113.95)	269.00 (115.04)	234.32 (120.15)	1.157 (.319)	.021
Accuracy: do they find the right person (%)	.50 (.51)	.84 (.37)	.84 (.37)	**6.928 (.002)****	.097
Request for information: how many times participant asks for additional information	1.56 (1.50)	.63 (.87)	.13 (.34)	**15.909 (.000)*****	.249
Sample characteristics				Total	
Sample size	32	32	31	95	
Gender (male/female)	18/14	17/15	16/15	51/44	
Average age	21.53	21.69	21.26	21.46	

Note. The bold values show significant difference with $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; [+]Mean (standard deviation).

[a]Low presentation modality (audio), medium presentation modality (audio + a still image), and high presentation modality (audio + a video clip).

[b]R-squares measure the goodness of fit of a linear model.

Table 3: Correlations of Variables

	Task ease	*Quick recognition*	*Adequacy*	*Confidence*	*Rapidity*[+]	*Accuracy*
Quick recognition	.782**					
Adequacy	.625**	.566**				
Confidence	.663**	.656**	.604**			
Rapidity[a]	−.525**	−.476**	−.291**	−.491**		
Accuracy	.494**	.417**	.444**	.497**	−.444**	
Request	−.258*	−.192	−.352**	−.178	.133	−.076

Note. [a]Reverse coded.
*$p < 0.05$, **$p < 0.01$.

Table 4: Summary *t*-Statistics of Contrast Tests

Contrast	*Mean difference*	*Standard error*	*t-Stat*	*Significant (2-tailed)*
Perception: Task ease				
Medium-Low	.938	.299	3.132	**.003***
High-Low	1.121	.295	3.802	**.000*****
High-Medium	.183	.289	.634	.528
Perception: Quick Recognition				
Medium-Low	.688	.324	2.121	**.038***
High-Low	1.116	.329	3.390	**.001****
High-Medium	.428	.324	1.322	.191
Perception: Adequacy				
Medium-Low	1.969	.256	7.700	**.000*****
High-Low	2.131	.286	7.444	**.000*****
High-Medium	.162	.266	.611	.544
Perception: Confidence				
Medium-Low	.656	.350	1.875	.065
High-Low	1.254	.304	4.120	**.000*****
High-Medium	.598	.328	1.821	.073
Task performance: Accuracy				
Medium-Low	.344	.111	3.097	**.003***
High-Low	.339	.113	3.006	**.004***
High-Medium	.005	.094	.054	.957
Task performance: Request for information				
Medium-Low	−.938	.307	−3.056	**.003***
High-Low	−1.433	.276	−5.187	**.000*****
High-Medium	−.496	.168	−2.959	**.004****

Note. The bold values show significant difference with *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

actually found the wrong person. None of the participants in the medium or high presentation modality condition group selected an incorrect target. Given the public relations costs of improperly accusing someone of being a shoplifter, this gain in accuracy is potentially important. Interestingly, although none of the medium presentation modality group participants identified the wrong person, they were not significantly more confident in their choice then the low presentation

Table 5: Summary of Hypothesis Tests

Hypothesis	*F-ratio or t-statistics*	*Significance*	*Conclusion*[a]
User perceptions			
H1a: Task ease	8.353	**.000**	Accept
Medium > Low	3.132	**.003**	Accept
High > Low	3.802	**.000**	Accept
High > Medium	.634	.528	Reject
H1b: Quick Recognition	5.956	**.004**	Accept
Medium > Low	2.121	**.038**	Accept
High > Low	3.390	**.001**	Accept
High > Medium	1.322	.128	Reject
H1c: Adequacy	38.946	**.000**	Accept
Medium > Low	7.700	**.000**	Accept
High > Low	7.444	**.000**	Accept
High > Medium	.611	.544	Reject
H1b: Confidence	7.267	**.001**	Accept
Medium > Low	1.875	.065	Reject
High > Low	4.120	**.000**	Accept
High > Medium	1.821	.073	Reject
Task performance			
H2a: Rapidity	1.157	.319	Reject
H2b: Accuracy	6.928	**.002**	Accept
Medium > Low	3.097	**.003**	Accept
High > Low	3.006	**.004**	Accept
High > Medium	.054	.857	Reject
H2c: Requests for Information	15.909	**.000**	Accept
Medium > Low	–3.056	**.003**	Accept
High > Low	–5.187	**.000**	Accept
High > Medium	–2.959	**.004**	Accept

Note. [a]Accept or reject null hypothesis with 0.05 significant level. The bold values are p-values showing significant difference with 0.05 significant level.

modality group. Only those with a video clip (i.e., high presentation modality group), which had additional confirming data such as the clothing the participant was wearing that day, were significantly more confident in their choice than the low presentation modality (i.e., audio) group.

We should note that even though the elapsed time measures were not significantly different from each other, the trends are in the expected directions. It may be that we simply did not have enough power in our test to find significant differences. Our study design, however, may have unintentionally limited the value of this measure. By limiting participants to one pass up and down the corridor in each direction, we artificially constrained the maximum time each participant could take. The average time participants took to locate the target was less than 5 minutes (259.46 seconds).

Given the nature of the task, we certainly are not surprised that those aided by video images of the target are more likely to correctly identify the target, found the task to be easier, and were more confident of the outcome. More interesting is the finding that there are no significant differences between the medium

presentation modality (i.e., audio + a still image) and the high presentation modality (i.e., audio + a video clip) groups in performance or perception, other than the number of times someone asked for additional help. This set of findings is somewhat consistent with a fit-oriented conceptualization of richness (McGrath, 1993; Mennecke, Valacich, & Wheeler, 2000). More accurately, it appears that there is a form of threshold effect—up to a point, additional video information aided in the completion of this task. That is, the relationship between richness and performance is nonlinear—once the minimum degree of richness is supplied, there is no incremental improvement in task performance or perception. The test results for linearity (i.e., weak R-squares) support the nonlinear relationship. This certainly suggests some investment implications, in that there may be higher costs associated with the provision of the richer, moving video information.

Our ability to interpret and generalize from this study is limited by a number of factors. First, the use of students who were unfamiliar with the technology except for a brief training session, and who had no real experience in security-related work, certainly limits generalizability. The relatively small sample size resulted in low statistical power. Our relatively gross differences in presentation modality levels between the conditions may also have masked more subtle effects of richness on task performance. Moreover, although this study mainly focuses on the effect of three different modalities (i.e., audio only, audio + still image, and audio + moving image) with different information richness, there is a possibility of confounding the effect of information richness with information content of different modalities of the experiment. Therefore, we must admit that it is not clear whether the difference between modalities actually comes from the richness of information itself or from the dual presentation modality. In future work it would be interesting to specifically manipulate the compounding effect and test the equivocality of the different modalities with different information richness, including two additional single-modality (i.e., still image only and moving image only) conditions. It is important to point out that these data say nothing about media choices or any social or organization influences on adoption and use. Finally, with regard to user perception and task performance, it is possible that different modalities favor different types of tasks, which were identified by Lee and Lai (2005). Thus, there are many areas with different tasks that we need to explore within the domain of multimedia presentation modality. An altogether different study might examine the factors that influence how agents elect to adopt and use such devices as employed here. It is important not to confuse these very different issues.

Despite these limitations, this study has demonstrated some initial potential applications for wireless, handheld multimedia devices and illustrated a clear payoff in performance in a relatively common public safety task. Apparently there are many potential applications of wireless multimedia mobile technologies in a security, military, and emergency service context. For instance, one interesting call-for-proposal (SBIR, 2007) is about the development of a wiki-based system (i.e., wiki-berry) that is capable of using numerous collection sources (voice, image, document, and video collection, etc.) to operate on a rugged, portable, and Internet-capable PDA-sized computer. The system should allow collecting

cultural intelligence on peoples of interest and transmitting the data to a large central wiki database for collaboration. There is already a fair amount of preliminary research and development projects underway to look at the usage of mobile handheld computing devices. Green, Stanton, Walker and Salmon (2005), for example, investigated the applicability of wireless communication systems with PDAs for a virtual reality command and control center integrating reconnaissance unmanned systems. This study has also confirmed the general proposition of the dual processing theory of working memory (i.e., the dual modes of presentation simultaneously incorporated by a specific presentation positively influence user perception and enhance task performance at the same time). In addition, it may be more useful to conceive of richness impacts as following a threshold model, with limited incremental improvements once the minimum required information is presented.

REFERENCES

Andres, H. P., & Petersen, C. (2002). Presentation media, information complexity, and learning outcomes. *Journal of Educational Technology Systems*, *30*(3), 225–246.

Archer, N., Head, M., Wollersheim, J., & Yuan, Y. (1996). Investigation of voice and text output modes with abstraction in a computer interface. *Interacting with Computers*, *8*(4), 323–345.

Baddeley, A. D., & Hitch, G. J. (Eds.). (1974). *Working memory* (Vol. 8). New York: Academic Press.

Burke, K., Aytes, K., & Laku, C. (2001). Media effects on the development of cohesion and process satisfaction in computer-supported workgroups: An analysis of results from two longitudinal studies. *Information Technology & People*, *14*(2), 122–142.

CBSNews. (July 9, 2003). Camera phone etiquette abuses. http://www.cbsnews.com/stories/2003/07/09/tech/main562434.shtml (last accessed: January 16, 2008)

CNN. (2007). Camera phones take over [electronic version]. Switched. Retrieved January 1, 2008 from http://www.switched.com/2007/05/21/camera-phones-take-over

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior 6* (pp. 191–233). Homewood, IL: JAI Press.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly*, *11*(3), 355–366.

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, *9*(3), 256–274.

Dillon, R. F., & Pellegrino, J. W. (1991). *Instruction: Theoretical and applied perspectives*. New York: Praeger.

Dubois, M., & Vial, I. (2000). Multimedia design: The effects of relating multi-modal information. *Journal of Computer Assisted Learning*, *16*, 157–165.

Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), *Intellectual teamwork* (pp. 351–371). Hillsdale, NJ: Lawrence Erlbaum.

Fang, X., Xu, S., Brzezinski, J., & Chan, S. S. (2006). A study of the feasibility and effectiveness of dual-modal information presentations. *International Journal of Human–Computer Interaction*, *20*(1), 3–17.

Fulk, J., Schmitz, J., & Steinfield, C. W. (1990). A social influence model of technology use. In F. J. & C. W. Steinfield (Eds.), *Organizations and communication technology* (pp. 117–140). Newbury Park, CA: Sage.

Fulk, J., Steinfield, C. W., Schmitz, J., & Power, J. G. (1987). Power, a social information processing model of media use in organizations. *Communication Research*, *14*(5), 529–551.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). *Coordination of communication: Effects of shared visual context on collaborative work*. Paper presented at the Conference on Computer Supported Cooperative Work 2000, December 2000, Philadelphia, PA, USA.

Green, D., Stanton, N., Walker, G., & Salmon, P. (2005). Using wireless technology to develop a virtual reality command and control centre. *Virtual Reality*, *8*(3), 147–155.

Hollinger, R. C., & Davis, J. L. (2003). *2002 National retail security survey*. Gainesville, FL: Center for Studies in Criminology and Law.

Hollingshead, A. B., McGrath, J. E., & O'Connor, K. M. (1993). Group task performance and communication technology: A longitudinal study of computer-mediated versus face-to-face work groups. *Small Group Research, 24*(3), 307–333.

ITU. (2008). Yearbook of statistics - Telecommunication Services (chronological time series 1997–2006). 34th Ed. Geneva, Switzerland.

Johansen, R. (1984). *Teleconferencing and beyond: Communications in the office of the future*. New York: McGraw-Hill.

Kahai, S. S., & Cooper, R. B. (2003). Exploring the core concepts of media richness theory: The impact of cue multiplicity and feedback immediacy on decision quality. *Journal of Management Information Systems*, *20*(1), 263–299.

Kinney, S. T., & Dennis, A. R. (1994). *Reevaluating media richness: Cues, feedback and task*. Paper presented at the Twenty-Seventh Annual Hawaii International Conference on System Sciences, January, 1994, Wailea, HI, USA.

Kinney, S. T., & Watson, R. T. (1992). *The effect of medium and task on dyadic communication*. Paper presented at the International Conference on Information Systems, December, 1992, Dallas, Texas, USA.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction*, *18*(1&2), 13–49.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). *The use of visual information in shared visual spaces: informing the development of virtual co-presence*. Paper presented at the 2002 ACM Conference on Computer Supported Cooperative Work, November, 2002, New Orleans, LA.

Lee, K. M., & Lai, J. (2005). Speech versus touch: A comparative study of the use of speech and DTMF keypad for navigation. *International Journal of Human–Computer Interaction*, *19*(3), 343–360.

Markus, M. L. (1987). Toward a critical mass theory of interactive media universal access, interdependence, and diffusion. *Communication Research*, *14*(5), 491–511.

McGrath, J. E. (1993). Time, task, and technology in work groups: The JEMCO workshop study. *Small Group Research*, *24*(3), 283–421.

Mennecke, B. E., Valacich, J. S., & Wheeler, B. C. (2000). The effects of media and task on user performance: A test of the task-media fit hypothesis. *Group Decision and Negotiation*, *9*, 507–529.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Education Psychology*, *87*(2), 319–334.

Paivo, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Robert, L. P., & Dennis, A. R. (2005). Paradox of richness: A cognitive model of media choice. *IEEE Transactions on Professional Communication*, *48*(1), 10–21.

SBIR. (2007). Cultural intelligence wiki-berry. Retrieved January 16, 2008, from http://www.dodsbir.net/SITIS/archives_display_topic.asp?Bookmark=29953

Schmitz, J., & Fulk, J. (1991). Organizational colleagues, information richness, and electronic mail: A test of the social influence model of technology use. *Communication Research*, *18*, 487–523.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London. New York: John Wiley and Sons.

Suh, K. S. (1999). Impact of communication medium on task performance and satisfaction: An examination of media-richness theory. *Information & Management*, *35*(5), 295–312.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins Publishers.

Thomas, J., & Griffin, R. (1983). The social information processing model of task design: A review of the literature. *Academy of Management Review*, *8*, 672–682.

Thompson, J. (1967). *Organizations in action*. New York: McGraw-Hill.

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness, and media choice in organizations. *Communication Research*, *14*(5), 553–574.

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1990). Understanding managers' media choices: A symbolic interactionist perspective. In J. Fulk & C. Steinfield (Eds.), *Organizations and communication technology* (pp. 71–94). Newbury Park, CA: Sage.

Valacich, J. S., Mennecke, B. E., Wachter, R. M., & Wheeler, B. C. (1994). *Extensions to media richness theory: A test of the task-media fit hypothesis*. Paper presented at the Twenty-Seventh Annual Hawaii International Conference on System Sciences, January, 1994, Wailea, HI, USA.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human–Computer Interaction*, *18*(3), 293–308.