Concept evaluation and usability testing of a TV based video communications system

D.J. Wheatley *, S. Basapur

Motorola Applied Research Center, 1295 E. Algonquin Rd., Schaumburg, IL 60196, USA

ARTICLE INFO

ABSTRACT

Attempts to successfully commercialize video telephony have thus far failed, however technical developments in broadband networks, video encoding, imaging and processing are now making TV based video telephony both technically and commercially viable. This paper describes two empirical studies carried out to evaluate such a concept. A first study assessed the user value of TV based video telephony by means of a comparative evaluation against a PC/webcam solution and face to face communication using subject dyads and structured audio/visual tasks. Significant differences were found between all three conditions; while pre- and post-test Likert scales indicated that ratings for the TV condition increased post-experience and were not significantly different from the face to face condition ratings. Two prototype systems were then developed which enabled TV to TV video telephony calls and a second study was carried out to evaluate in greater depth, the usability and acceptability of the feature sets and their respective ease of access. The studies indicated that TV-based video communications does have intrinsic user value and also has the potential to approach the richness of face to face communications, but that certain control and privacy functions need to be implemented in the UI before this can be fully realized. Such functions included; control over the callers with whom video would be used, control over who could access the videotelephony system, control over the recording of calls, the ability to turn off the self-view and, for total privacy, physical occlusion of the camera when not in use.

© 2011 International Federation for Information Processing Published by Elsevier B.V. All rights reserved.

1. Introduction

Video telephony via the television has now become technically and economically viable due to performance improvements in broadband networks, audio/video encoding standards and the increasing power and reducing costs of digital imaging and computing platforms. However, previous attempts to introduce video telephony have failed to achieve success and furthermore, the commercial viability of such a concept is (and was) based on the implicit assumption that it would be of intrinsic user value. It would also need to deliver a better user experience than current PC/webcam solutions. Two user studies are described here. The first set out to evaluate the relative merits of TV based video telephony experience in comparison to a PC/webcam based video communications experience and a face to face communications experience. The aim was to investigate the acceptability of a TV based video telephony and the criteria for user acceptance. In order to investigate participant preconceptions, a quantitative comparison was also carried out of their expectations prior to the experience and the value judgments expressed after the experience. Open-ended questions were also used to elicit information on

expected usage scenarios for TV based video telephony. A prototype video telephony application was developed based on the findings of this initial concept study and two variants of the system were evaluated in a second study. This set out to test the conclusions of the first study in terms of the features and functions which should be provided to the user and also to evaluate the usability and intuitiveness of the prototype user interface design.

2. Background

Since the first video conference call was made in 1930 between AT&T headquarters and Bell Laboratory in New York City [3], video communication has come a long way in terms of audio/video technology, infrastructure/bandwidth and the penetration of technology into people's daily life. AT&T's PicturePhone debuted in 1964 [10] and in 2001 the first 3G cell phone based video call was made [20]. Recently, the first video calls were made over LTE Technology (Long Term Evolution, also known as 4G) using cell phones [23,27]. Since its inception, video communication has always attracted considerable critique as a useful communications medium and has struggled to find its niche in the market [2].

Carmen Egido reviewed some of the failures of video conferencing as a technology to support group work [6]. She stated that despite sound intuitive feeling, video conference is not the closest

---

* Corresponding author. Tel.: +1 847 668 0131; fax: +1 847 576 2792.
 E-mail address: David.j.Wheatley@motorola.com (D.J. Wheatley).

thing to "*being there*". Video conferencing, according to Egido, is not exactly the half way point between a telephone conversation and a face to face meeting. She also noted that the success of the video conferencing depends more on the nature of the application for which it is introduced than on the system details and features. The factors, she thought (in 1988) that might influence acceptance would be the decreasing cost of bandwidth, automation and technology development, fuel shortage and travel disruptions and related costs. In the present economic climate, fuel costs and economic issues have made these very factors more relevant than ever. Also, after the failure of the AT&T PicturePhone concept, research in this area focused mainly on the office setting and work related video communication between two or more parties that were remotely located [10,21]. Unfortunately, little research was done on the use of video communications in home and domestic settings and for everyday social needs, even though it has been long acknowledged that understanding people's need for technology is important, especially communication needs within the home environment [18].

In addition to technological advances, other research in communication and basic conversation mechanisms has also matured over time [11,12,15,26]. For instance, a key factor influencing the effectiveness of video telephony is that of eye-contact, and previous research has indicated that users are very sensitive to this. Studies [1,22] show that people can identify accurately (84%) when a person is looking directly at them (the nose bridge being the point of reference), when they are face to face at a distance of 2 m. Also people can identify with more than 60% accuracy when two people facing each other are directly looking at each other's nose bridges even from a 1.5 m distant, right angled observation point. Vertegaal et al. [21,24] have also shown that participants synchronize turn taking using eye contact as one of the cues. However in a comparison of two videoconference systems with face to face interaction, O'Conaill and Whittaker [17] found from informal feedback that users experienced difficulty in achieving mutual eye gaze and, because subjects stared fixedly at the screen, perceived this to be confrontational, and Bruce [5] also concluded that video communication enables people to use non-verbal cues such as eye gaze, gestures and body language to help in understanding each other.

When eye contact was studied using a desktop video conferencing system, Grayson and Monk [13], found that "*some kind*" of mutual gaze may be possible with the desktop systems. Although this cannot be assumed to be "eye contact" as such, it is still conceivable that this type of mutual gaze awareness could provide similar communicative functions as can true mutual gaze or eye contact. Therefore to optimize people's ability to achieve at least some kind of mutual gaze or gaze awareness, the camera should be located close to the image of the other person. In light of this background, the displays in both of the present studies were set up to maximize the degree of eye contact which was attainable in each of the three conditions.

Other research [13,14] points out that we should not pre-suppose that "*if (a) video telephony solution efficiently and sufficiently imitates face-to-face communication then it will be a success*". This research iterates that video telephony is not a substitute for face-to-face communication and it should not be positioned as such, but as a solution which provides a rich means of communication in the absence of a face-to-face communication opportunity.

Early research concluded that face to face communication and video telephony are both inherently superior to Plain Old Telephone service (POTs telephony) because they enable more channels of communication [8], but that, compared with face to face, both the telephone and CMC (Computer Mediated Communication) limit feedback to the speaker and provide fewer sensory cues. The authors stressed the difficulties in establishing "common ground" between the communicating parties when not co-located

and hypothesized that audiovisual availability, synchrony and channel symmetry (perfectly enabled in face to face communication) heavily influence the communication. A study similar in approach to the current study was also carried out by O'Malley et al. [19]. This study hypothesized that firstly, there are advantages for communication in seeing the face, and secondly, that video-mediated conversations are very much like face to face. While there is empirical evidence of the first opinion, the case for similarities between video-mediated and face to face conversation was not established. In order to compare these two means of communication, the authors carried out a series of experiments using the 'map task'. The task was for the person with a route marked on his or her map (the information-giver) to convey information sufficiently clearly to enable another person (the information-follower) to reproduce the route on his or her map as accurately as possible. In the case of the video-mediated conversation technology, they used 'video tunnels' which emulated face to face interaction as closely as possible without the participants actually being co-present. The "video tunnel" was a physical arrangement of camera and monitor with a half-silvered mirror in front of the monitor and a full mirror in front of the camera. This apparatus was enclosed in a box so that only the monitor could be seen, this arrangement of hardware therefore enabled direct eye contact between participants in the study.

They conducted three experiments; the first compared video-mediated vs. audio-only interaction. They found no effects of medium on performance of the task, however, in terms of 'efficiency of communication', video-mediated conversations were less efficient or effective than either audio-only or co-present face to face communication.

The second experiment measured the effects of size of video image in order to determine whether or not participants were using visual cues mainly from the face (gaze, expression, lip-movement) or more global cues such as posture and gesture (shrugging of shoulders). This study was inconclusive.

The third experiment measured the effects of both the audio and video signal delay. They compared performance on the map task using 'video tunnels' with 'videophones'. Videophones result in a delay of the audio and video signals due to the low bandwidth of analogue telephone lines. The manipulation of delay had a significant effect on task performance. Delay in the audio signal produced around 36% poorer performance than no delay and seeing the face did not help participants overcome the effects of delay.

Previous research therefore seems to suggest that eye contact is crucial in video-mediated communications since it enables non-verbal communication and a face to face like communication experience. Also, it appears that the existence of a self view while in video communication with others improves meta-cognitive behaviors [25,28]. Similarly, we hypothesize that a wider angle of view (as might be achieved by a video camera on top of the TV) and the lean back experience afforded by the TV will contribute to an experience closer to that of face to face communication than can be achieved using a PC and webcam. Also, previous research has shown that audio/video latencies are critical to the video telephony experience [16,17]. With the advancements in audio–video technology and enhancements in user experience design, we may now be able to address some of these issues and develop a solution which may start to approach the richness of face to face communications.

3. Concept study methodology

In the first concept study, equipment was set up to enable bidirectional audio/video communications between pairs of participants. In the PC and TV conditions, the equipment was set up in separate rooms so that no face to face communication was possible