Immersive Group Telepresence and the Perception of Eye Contact

Steve McNelley, Ph.D.
Global Telepresence Laboratory
Digital Video Enterprises, Inc.
19200 Von Karman Ave. Suite 400, Irvine CA 92612
www.DVEtelepresence.com

ABSTRACT

Parameters of an immersive telepresence audio/visual system was first defined by a review of the literature and also a preliminary qualitative analysis with live subjects. The definition that emerged was drawn upon as criteria for the audio/visual design and equipment installation used for the present experiment. The definition is simulating a conferencing image of a person sitting on the other side of the meeting table in the conversational space around the table, and also includes a life-size image and eye contact.

Specifically, eye contact and the resulting parallax angle of the life-size image placed in the conversational space as if around a meeting table was the subject of the experiment. Two images of live interactive conferencing video streams were sent to participant subjects. Image signal "A" was a true eye contact image and image signal "B" was a parallax camera angle with a camera mounted at bezel height on a 50" plasma. The subjects were asked to identify the difference in the images and were asked which they would prefer to conference with. Ninety-three percent (93%) of the subjects correctly identified the true eye contact image. Ninety-three percent (93%) of the subjects stated that they would prefer to communicate with the real eye contact image.

The study confirms the necessity for an eye contact solution in an immersive telepresence experience. The study confirms the difficulty of traditional and common audio/visual equipment to provide an immersive videoconferencing experience with eye contact. Equipment suggestions are made to overcome this eye gaze problem for immersive telepresence conferencing.

KEYWORDS

Immersive, Telepresence, Proxemics, Eye Contact

INTRODUCTION

In recent years, there has been an increased awareness that something is missing from the experience provided by traditional group VTC (video teleconferencing) systems. These traditional systems typically include TVs on carts with set-top cameras. In years past, the TVs utilized have been CRTs from about 27"-36" diagonally measured. More recently, plasma and LCD technology are replacing the CRT. This common group room videoconferencing equipment arrangement is being challenged for its validity to provide an immersive telepresence experience.

Immersive telepresence is generally referred to as a conferencing experience that more closely simulates the experience as if the imaged conferee on the display is sitting around the meeting table. The traditional VTC cart system usually resides against the wall and is therefore outside the normal conversational space of the table. It is as if the person on the screen was placed at a chair outside the conversational dynamic that occurs at a meeting table.

Eye contact is an essential aspect of human communication according to Argyle and Cook [3]. Several studies have been conducted on the importance of eye contact and only a handful on the perception of eye contact while videoconferencing. Two directions of gaze have been analyzed: up and down (vertical) and left and right (horizontal) which is effected by where one looks in relation to the camera capturing their image. This study limits its focus specifically to the issue of vertical eye contact and challenges the common placement of the camera above the monitor as is common in nearly all VTC group conferencing systems from major conferencing manufacturers. Horizontal eye contact was not the subject of this study which would be relevant if multiple life-size people were being seen on the screen or many screens. These multiple conferee arrangements have other sets of eye gaze issues to contend with, such as the appearance that people are looking too far left or too far right when directing attention to a specific person.

While previous eye gaze studies have set out to apply a general rule of noticeable eye gaze within common VTC design, this study specifically focuses on the immersive telepresence arrangement of simulating a life-size person sitting around a meeting table. The previous studies do not approach the subject within the totality of immersive telepresence as defined by the experimental parameters of this study. This study should also be read in the light of the researcher's doctoral dissertation which explored a similar issue in regards to desktop videoconferencing [15].

By ignoring one aspect of immersive telepresence, such as a life-size image, then eye contact can be achieved but immersive telepresence is then sacrificed. For example, certainly a little monitor on the other side of the room has an insignificant parallax angle and with the small picture will leave one with the impression of eye contact. Context is needed of the other variables in equipment design in order to accurately assess the importance of eye contact with life-size people appearing to sit in (not outside) the conversational space of a meeting table.

PRIOR STUDIES IN EYE CONTACT PERCEPTION

The field of eye contact studies has a broad and in-depth history in the field of psychology where numerous attributes have been attributed to people who either provide appropriate eye contact or who are gaze avoidant. A natural outgrowth of the research was the development of experimental methods to assess a person's gaze direction. Noted studies were done by the following researchers: Gibson and Pick [8], Cline [5], Kruger and Huckstedt [13], and Ellgring [7]. Much of this work has been drawn upon by this researcher and other researchers to asses the nature of, the recognition of, and the significance of various eye gaze angles while videoconferencing.

Perceiving Eye Contact While Videoconferencing

Eye gaze has proven to be a very complex and rich area of study, especially as it relates to perceptions while videoconferencing. Numerous studies support the contention that gaze can be detected within one to a few degrees [2] [5] [8] [11] [18] [7] [13] [19].

Acker and Levitt [1] attempted the difficult task of assessing satisfaction while videoconferencing with eye contact. While they failed to show the increased satisfaction with their experimental design, they did comment that "improved eye contact apparently allowed participants to evaluate more confidently their counterparts, and to participate more comfortably in exchanging information."

McNelley[15] studied specifically desktop videoconferencing and also found it difficult to establish an increase in satisfaction, but did clearly demonstrate that 88% of the subjects did prefer an eye contact image over a non-eye contact image in a specific desktop arrangement. Given that the overwhelming majority preferred an eye contact image and that 68% noticed the eye contact issue without knowing that the study was about eye contact confirmed that the eye contact problem in that particular desktop arrangement was significant.

Several experimental methods in assessing eye gaze while videoconferencing have been explored by the researcher. Utilizing photographs in a side-by-side comparison with one photo illustrating an extreme angle of gaze and the other with a lesser degree of gaze (or even true eye contact) was considered, but rejected due to the failure of the photographs to simulate the dynamic motion of gaze which assists in cueing an observer to the gaze direction in real-time. Live or recorded images of lookers who are non-conversational have also been rejected since it is unnatural for people to engage eye gaze perception with a "stiff" non-interactive subject.

In the cases above, if varying degrees of gaze are presented, except for a true eye contact image, fielding qualitative responses presents too many confounding variables where participants are asked to select which has better eye contact when in reality none may. People will naturally gravitate toward affirming the image with "better" eye contact as "true" eye contact, especially if they have no other choice.

In an unpublished study by this researcher, participants were asked to respond to a videoconference of whether they sensed "eye contact" from two differing image signals. One signal had perfect eye contact and the other varying degrees of parallax. A few of the subjects chose the non eye contact image as being the eye contact image. The same participants where presented the same visual stimuli several hours later and asked if they sensed "that I {the interviewer) was looking into your eye." By simply changing the language from "eye contact" to "looking into your eye" all the respondents chose correctly the true eye contact image.

The researcher concluded that asking someone to discern "eye contact" may not be sufficiently accurate. Apparently some people express they have eye contact even when someone is actually looking at their nose, mouth, chin, forehead or neck. Perhaps the psychological transparency of perfect eye contact makes some people uncomfortable, and perhaps even more so in a controlled experiment environment, so they prefer a "generalized eye contact" as gazing toward someone's face. This may be partly explained by cultural, personality, and relational styles. As a result, the researcher specifies to participants the definition of eye contact as "looking into the eyes" when involved in quantitative and qualitative analyses to asses perception of eye gaze.

In another unpublished study by the researcher, various ambient light conditions and spot light positions effecting spectral reflection positions on the eyes were shown to affect a person's ability to accurately judge eye gaze direction. Further research is needed to determine the level of significance lighting may have on gaze direction experimental research design.

Chen [4] concludes that the perception of eye contact in the down direction can be at a threshold of about 5 degrees as an equipment design parameter. He extrapolates from that a rule of thumb for all videoconferencing systems from PDAs to large screen rooms. Chen points out the subjectivity of the perception of eye contact and observers tend to perceive eye contact when in actuality eye contact was not presented within the lesser parallax angle. For Chen this presents opportunities to design videoconferencing systems within the threshold of perceivable eye contact.

To illustrate his point, Chen configures off-the-shelf components to resolve the eye contact problem by keeping the angle of parallax below 5 degrees. In a desktop mock up he demonstrates such an apparatus in his figure 7. One cannot help but notice that the picture shows an image of the person that only partially fills the screen and the person's head is "cut off" at the top so the eyes can be closer to the camera above. Still further, the observer is asked to sit back 36" which is further then people usually view a desktop computer monitor. It is questionable how effective such a solution would be with a small picture relative to screen size, needing to push back to view the screen, and take the unusual measure of chopping off the top of the imaged person's head. Such measures for commercially viable products for desktops and even group systems are doubtful.

Several questions remain regarding the Chen study. The first question is the relative size of the imaged conferee to the screen size which would determine how much effective resolution was applicable to the image of the conferee. Also, were life-size images presented which match our common way of relating to people in person? Another question is the quality of video presentation. A front projection screen is used in the "gaze measuring room." Since the connection is live interactive, a camera also captured an image of the observer viewing the looker. Ambient light and even low ambient light would have reduced the contrast and clarity of the front projected image that observers used. A higher quality display presentation is preferred. An LCD, a rear projection or a plasma would have matched more closely a real world conferencing display presentation.

It is commonly recognized among audio/visual designers that front projection is a very poor medium for videoconferencing. This may in part explain the dichotomous results of the accuracy of gaze direction between the classic studies which assert a lower threshold, including the Bell Labs study [19]. This may also explain why, in the face to face control in the Chen study, observers could much more accurately detect eye direction. Further study is needed to determine whether front projection was a confounding variable.

Another question is that the Chen study does not simulate the camera angle shooting down on the observer. The looker's camera is at eye level and the looker is asked to look at targets above and below eye level. Videoconferencing cameras when placed on a common VTC cart are not at eye level. The camera is above eye level on top of the cart's TV display. While the parallax angle may be the same, the higher origination point of the camera will create a downward perspective of the conferee. This downward perspective was not simulated in the Chen study and may be a confounding variable. Further research is needed.

IMMERSIVE GROUP TELEPRESENCE PARAMETERS

The following discussion highlights three primary factors of immersive telepresence which includes a life-size image, correct proxemics, and eye contact. Certainly immersive telepresence includes also television quality video, high quality audio, and effective conferee lighting, all of which are assumed essential to this experiment. The combination and interaction of these three factors is what determines the qualification as an immersive telepresence.

Life-Size Image

The goal of immersive telepresence is to simulate the experience that an imaged conferee is actually in the meeting room within the conversational space of the meeting table – exactly what would be if that conferee was there in person. Obviously a life-size image is required – ratio 1:1. Rose and Clarke [17] concluded, among other things, that a life-size image is important so that the "image of the remote participant,…[has] more easily recognizable body language." Certainly, conversing with tiny images of people or giant images of people presents a whole "other reality" to conferees. The conferees are forced to project themselves beyond the strange presentation of the conferee on the screen in order to interact naturally.

A life-size image as well, if presented correctly, will show at least upper body and arm gestures. In a doctoral study by McNelley [15] the lack of body language was the second most important factor next to the lack of eye contact that was defined as being "awkward".

Proxemics

The word "proxemics" coined by Hall [9] refers to "the study of man's unconscious structuring of microspace." Hargie [10] proposes that proxemic includes territoriality, body orientation and proximity of interpersonal space. Hall [9] maintains that proxemics is also highly derived by one's culture. Other studies have found that distance to another can be associated to a variety of emotional responses such as friendliness, aggressiveness, dominance, etc. (Paterson and Sechrest [16]).

Immersive telepresence, as defined by this researcher, seeks to simulate most accurately whatever distance is common in an in-person interaction and to closely maintain that distance for a videoconference (i.e., sitting around a table). By arbitrarily placing images of people farther out or closer in, in relation to the distance of their common conversational experiences, introduces a whole host of possible psychological cues and representations that are not intended by those engaged in the conversation.

Eye Contact

The literature on the importance of eye contact and the nature of eye contact is voluminous. According to Kendon [12], gaze is vital in the flow of natural communication monitoring of feedback, regulating turn taking, and punctuating emotion. The lack of eye contact shows timidity, embarrassment, shyness, uncertainty, and social awkwardness (Edelmann and Hampson [6]). In men, gaze avoidance has been correlated to being emotionally inhibited, over-controlled, and having psychosomatic and physical symptoms; and gaze avoidant women have been associated with higher degrees of psychopathy, hysteria, and traditional femininity (Larsen and Shackelford [14]). The research is extensive and reasonable to conclude that a person who is videoconferencing and is presented as gaze avoidant, due to a camera parallax angle, will be perceived negatively.

METHODOLOGY

The perception of eye contact while engaged in an immersive telepresence designed room was the central focus of this study. Specifically, the audio and visual equipment was designed around the parameters of a mid-sized corporate meeting room and a meeting table that seats up to 6 people comfortably. This environment was selected, because it represents a significant majority of existing VTC installations.

Immersive Group Telepresence Room Arrangement

Life-Size: A 50-inch plasma display panel was selected to display a life-size upper torso shot of one conferee. The 50-inch plasma is ideal since the screen width and height measurements allow a person's upper body language to be seen, including enough width for hand gestures with the wide 16:9 aspect ratio screen. The upper torso shot was a classic broadcaster framing. The entire head, shoulders, and upper body were seen on the display and centered in the image.

Proxemic: A meeting table 5 feet wide was selected. On one side the table was a plasma panel mounted 2 feet from the table edge. The bottom of the picture to the floor was 30". On the other side of the table, directly across the plasma, was the seated position of the participant subject. The participant subject's eyes are approximately 1 foot from the table edge. The eyes of the person imaged on the plasma display to the eyes of the observer on the other side of the table was approximately 8 feet. This simulates a person sitting on the other side of a large meeting table and inside the normal conversational space.

A pilot study was conducted prior to the present study to confirm the placement of the plasma display with qualitative interview analysis with subjects from a pool of ten people. While a conference was being conducted with one life-size image of the interviewer on the plasma screen the plasma display was moved to various locations 9, 10 and 11 feet from the participant's eyes. The participants responded that when the image was closer they had a greater sense of "immersion." This confirmed the 8 feet distance selected. The plasma could have been brought in even closer to the table, but for the sake of analyzing the perception of eye contact the researcher elected to challenge the perception of the parallax angle with the lesser degree.

Eye Contact Image Capturing Room

The Eye Contact Image Capturing Room was described as stated above for the room arrangement.

Interviewer Cameras: In the Eye Contact Image Capturing Room two cameras were mounted to the plasma. One common VTC PTZ camera was mounted as low as possible on top of the plasma and another directly over the eyes of participant seen on the screen. The VTC PTZ camera selected for top of the plasma had the lowest profile commercially available to minimize as much as possible the parallax angle. Since a broadcaster framing shot was selected, the camera lens was placed over the eyes 35% down vertically from the top of the screen. These two live video feeds were sent to the Eye Contact Observing Room.

Audio System: A high quality microphone and speaker was used for audio interaction between the interviewer and the participant.

FIGURE 1
Eye Contact Image Capturing Room
Low profile PTZ camera on the top bezel and a second PTZ camera mounted over the eyes of the image of the participant subjects.

Eye Contact Image Observing Room

The Eye Contact Image Observing Room was described as stated above for the room arrangement.

Participant Camera: In the Eye Contact Image Observing Room a common VTC PTZ camera was mounted as low as possible on top of the plasma.

Observer Video Switch: One video switch was provided on the table so the participants could switch between the two camera feeds to view an eye contact image and non-eye contact image.

Audio System: A high quality microphone and speaker was used for audio interaction between the interviewer and the participant.

Experiment 1: Gaze Perception in Immersive Group Telepresence

The researcher was the interviewer that presented two simultaneous images to the participant subjects. One image originated from a NTSC camera mounted directly in front of the interviewer's plasma at the exact eye line which produced image signal "A." A second camera mounted as low as possible on top of the center of the plasma and captured an image of the interviewer which is identified as image signal "B".

The participants from a selected pool of 43 subjects were engaged by the interviewer with a structured line of questions in dialog format seeking demographic information from the participants. The participants were business professionals working in a large corporate building complex. This dialog gave the participant the needed conversational interaction within the conference experience without the possible confounding variables of random topics of conversation. Before the structured interview, the participants were asked to regularly select from their video switch the two different angles and at the end of the interview were asked for their impression of the difference between image signal "A" and image signal "B". The two images were color corrected so they looked virtually identical except for the camera angle. The participants were asked to select which image they perceived the interviewer was looking at them in the eye. Also, qualitative responses were sought to elaborate on their reasoning for their selection.

Experiment 2: Gaze Preference in Immersive Group Telepresence

The same subjects of experiment 1 were asked, after completion of experiment 1, to identify their preference between eye contact image signal "A" and image signal "B", if they were going to use this interactive video communication system on a regular basis.

RESULTS

Experiment 1: Ninety-three percent (93%) of the participant subjects selected the eye contact image of image signal "A". The qualitative responses included supportive comments for image signal "A" such as "feels more intense," "it really feels like he is looking at me," "definitely eye contact," "feels more immersive and engaging." Responses for the non-eye contact image signal "B" were typically "he is looking down," "he is looking below my head," "it feels that he is not talking to me," "it seems unnatural."

Experiment 2: Ninety-three percent (93%) of the subjects preferred the eye contact image signal "A" in the context of asking them which image signal they would prefer to see if they were to use this interactive communication system on a regular basis.

Discussion: The study conducted utilized the researcher as the interviewer who is male, fair skinned, and has blue eyes. Further research with interviewers with differing appearances would assist in expanding the research.

The image signals between rooms were analog. The confounding variable of image compression artifacts was not a design parameter of this study. However, Chen [4] found insignificant differences between the analog and high quality compressed sources. For this study, the sources were NTSC resolution with 4:3 aspect ratio cameras displayed on 16:9 aspect ratio plasmas. The plasmas were set in zoom image mode so the entire 16:9 screen image was filled. As a result a certain amount of resolution was lost to fit the 4:3 camera images on the 16:9 screen.

With the emergence of HDTV conferencing and native 16:9 cameras and displays, the effective resolution of future videoconferencing systems will have even greater resolution then what this study provided. The improved definition will certainly even enhance the perception of eye gaze. It may even rival the face to face perception of eye gaze as presented by Chen [4] which demonstrated a significant increase in the awareness eye gaze angle in a face to face as compared to a videoconference. Further study is needed.

NECESSITY FOR AN IMMERSIVE TELPRESENCE EYE CONTACT SOLUTION

In the preliminary study, participants commented that the plasma display, when placed closer around the table, "felt" more "immersive." The table selected was actually wider then common meeting tables which range from 3 feet wide to 4 feet wide. The researcher selected a 5 foot wide table to challenge the lesser parallax angle for this study. Also, the plasma was located 2 feet beyond the table yet within the conversational space of the meeting table. The plasma could have been brought in closer to the table, but again the researcher wanted to challenge the lesser parallax angle.

Based on this proxemic placement of the plasma screen relative to the participant's eyes, the parallax was sufficient for the participants to overwhelmingly recognize and prefer the eye contact image. This leaves a dilemma for traditional audio/visual group conferencing designers. How does one meet both the proxemic requirement to stay within the conversational space of the table, maintain a life-size image and at the same time enable eye contact? The solution provided by Chen [4] of chopping off the top of the head of the conferee to get the eyes closer to the top-mounted camera is not acceptable in immersive telepresence system design because of its awkwardness.

FIGURE 2
Telepresence 50 Immersive Group
Eye line camera mounted behind the eyes of the imaged conferee, 50 inch plasma for life size images framed with a broadcaster's shot to see body language.

Currently, none of the major conferencing codec manufactures provide an eye contact solution for immersive telepresence. A specialty product is needed to enable immersive telepresence. One product from Digital Video Enterprises that provides true eye contact can effectively achieve immersive group telepresence (Figure 2).

This study was specifically with one life-size image. Certainly, if multiple conferees were at the distance site live switching between participants would retain immersive telepresence as defined by this study.

ACKNOWLEDGEMENTS

I want to thank the many busy business professionals that took time out of their schedules to serve as participants for this study. Also a special thanks to Mike Costa for his willingness to drop everything and help with several details of this project. Lastly, James Ewing Ph.D., my dissertation chair from another life ago, continues in my memory to leave fresh thoughts to stir my drive for this subject.

REFERENCES

[1] Acker, S.R., and Levitt, S.R.
Designing Videoconferencing Facilities For Improved Eye Contact. *Journal of Broadcasting and Electric Media, 31,* pages 181-191, 1987.

[2] S. Antis, J. Mayhew, T. Morley
The Perception of where a Face or Television 'Portrait' is Looking. *American Journal of Psychology*, pages 474-489, 1969.

[3] M. Argyle and M. Cook.
Gaze and Mutual Gaze.
Cambridge University Press, 1976.

[4] M. Chen
Leveraging the Asymmetric Eye Contact for Videoconferencing.
CHI 2002, April 20-21, pages 1-8, 2002

[5] M. Cline.
The Perception of Where a Person is Looking.
American Journal of Psychology, pages 41-50, 1967.

[6] R. Edelmann and S. Hampson
Changes in Non-Vebal Behaviour During Embarrassment.
British Journal of Social and Clinical Psychology, 18, 385-390, 1979

[7] J. Ellgring.
Die Beurteilung des Blickes auf Punkte innerhalb des Gesichtes. *Zeitschrift fur experimentelle und angewandte pshchologie*, pages 600-607, 1970.

[8] J. Gibson and A. Pick.
Perception of Another Person's Looking Behavior.
American Journal of Psychology, pages 386-394, 1963.

[9] E. Hall
The Hidden Dimension of Culture.
New York, Double Day, 1966.

[10] O. Hargie, C. Sanders, and D. Dickson.
Social Skills in Interpersonal Communications (3rd ed.) New York, Rutledge, 1994.

[11] J. Jaspers, et al.
Het observen van ogencontact.
Nederlands Tijdschrift voor de Psychologie, 28, pages 67-81.

[12] A. Kendon.

Some Functions of Gaze Direction in Social Interaction. *Acta Psychologica*, 26, pages 22-63, 1967.

[13] K. Kruger and B. Huckstedt.
Die Beurteilung von Blickrichtungen.
Zeitschrift fur experimentelle und angewandte psychologie, pages 452-472, 1969.

[14] R. Larson and T. Schackelford.
Gaze Avoidance: Personality and Social Judgments of People Who Avoid Direct Face to Face Contact. Personality and Individual Differences, 21, pages 907-917, 1996.

[15] S. McNelley
The Significance of Eye Contact While Desktop Videoconferencing.
Ph.D. Dissertation, United States International University, pages 52-81, 2000.

[16] M. Patterson and L. Sechrest.
Interpersonal Distance and Impression Formation. *Journal of Personality* 38, pages 161-166, 1970.

[17] D. Rose and P. Clarke.
A Review of Eye to Eye Videoconferencing Techniques. *BT Technology Journal*, 13 (4), pages 127-131, 1995

[18] B. Stapley.
Visual Enhancement of Telephone Conversations.
Ph.D. Thesis, University of London, 1972.

[19] R. Stokes.
Human Factors and Appearance Design Considerations of the Mod II PicturePhone Station Set. *IEEE Transactions on Communication Technology*, pages 318-323, 1969.