The influence of gaze direction on social presence of the remote participant

L.L. van der Leeden
l.l.vanderleeden@student.utwente.nl
Department of EMCS, University of Twente

ABSTRACT
This study analyses gaze direction in a videoconferencing experiment set up by the University of Twente. Examined is whether there are differences in a 2D versus a 2.5D setup with 3 local participants and 1 remote participant. In the 2D setup, gaze direction is distorted, whereas in the 2.5D condition, gaze direction closely approaches the actual gaze direction. To examine the reliability of the data, part is retagged and inter-annotator disagreement is discussed and examined.

Keywords
video conferencing, social presence, gaze, 2.5D

1. INTRODUCTION
Video conferencing has made communication over a distance much easier, people can look at each other to better understand what the other person means [11]. Video conferencing however, is still not the same as face to face interaction. Delay causes confusion in interpreting nonverbal cues and the person on the other side seems to be looking just beside you instead of straight at you. The whole experience seems artificial, less personal than a face-to-face conversation. Two important factors influence this experience, the degree of social presence and nonverbal cues, of which gaze is of great influence in structuring conversations.

1.1 Social presence
Social presence is sometimes used as part of the concept ‘presence’, the other part being defined as ‘telepresence’, the sense of ‘being there’ in the virtual environment that is created in mediated spaces such as video conferences. Social presence on the other hand focusses more on the feeling of ‘being there together with another’[2].

The measurement of social presence is a subjective measurement that has been attempted in various ways. Short et al. were the first to define the term as “degree of salience of the other person”. Their research was set up to determine the effectiveness of various media channels, thus focussing on the opinion of the user on the medium itself. Their approach was to determine the rating of telecommunication systems on a series of seven-point bipolar pairs such as “impersonal - personal”, “cold - warm”. A high degree of social presence would be rated as warm, personal, sensitive and sociable [5]. One can understand the higher the richness of the medium, the higher the social presence rating has proven to be.

Biocca, Harms and Gregg have more recently defined mediated social presence as “the moment-by moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other...” they believe that social presence is “.. the outcome of the other’s cognitive, emotional and behavioral dispositions”[1]. Based on extensive literature study they have determined several dimensions of social presence (see figure 1).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2010, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.
jective measurement, one has to decide which of the fac-
tors it can support. It may be clear that by using gaze
direction as a objective measurement tool, we will focus
on the behavioral aspects of social presence rather than
the cognitive and emotional aspects. To determine what
gaze influences, we will first look at its functions.

1.2 Gaze
Research has shown that gaze holds important functions in
conversation. In dyadic (two-person) conversations, gaze
direction holds at least four functions according to Kendon
[7]; (1) to provide visual feedback; (2) to regulate the flow
of conversation; (3) to communicate emotions and rela-
tionships and (4) to improve concentration by restriction
of visual input. The regulation of turn taking in multi-
party situations is more complicated, since there are more
possible speakers to take the next turn in the conversa-
tion. Gaze direction, however, is still an important non-
verbal communication tool to determine who’s next. The
research of Vertegaal et al. in mediated triadic conver-
sations [14] shows that subjects indicate that they could
more easily perceive whom was talking to whom when the
percentage of gaze was higher.

Grayson has shown that time spent gazing at the indi-
vidual one is talking to or listening to, is significantly
more than the time spent looking at other individuals [3],
which makes gaze an excellent predictor of conversational
attention in multiparty conversations [13]. Furthermore,
research into both primates and humans has established
that while the direction of social attention is determined
by other cues than gaze direction alone, gaze cues override
both head gestures and body postures, making gaze direc-
tion one of the most important cues for social attention
[9].

While gaze has been an important topic of research over the
years, its use as an objective measure for aspects of
social presence has not been widely investigated.

1.3 Video conferencing
There are various factors that negatively influence the ex-
prience of traditional video conferencing; the absence of
eye-contact, the lack of a shared social and physical con-
text, the limited possibility for informal communication,
a lack of media richness and support for verbal and non-
verbal communication [5].

The camera placement in regular 2D conferencing systems
such as Skype distorts mutual gaze; without the use of spe-
cial mirrors, it is not possible for two participants to look
directly into each others eyes, since the camera is always
placed elsewhere. Turn taking and regulating the flow in
mediated conversation is often less clear and more difficult
to determine for the user than in face to face conversa-
tions. Gaze direction is an important factor in resolving
such issues.

Advances in technology have made it possible to develop
more sophisticated programs such as 3D conferencing sys-
tems that accurately facilitate eye contact [6], however,
the systems developed are expensive and not in wide use.

Grayson and Monk’s research suggests that such elaborate
workarounds may not be necessary; participants can learn
to interpret gaze direction very accurately given optimal
configuration of low cost desktop equipment [3].

2. EXPERIMENT
The university of Twente has set up an experiment do-
ing exactly that. Participants are located in two rooms,
three co-located participants ($LP_{1,2,3}$) in one room, and a
remote participant (RP) in a separate room. In the first
condition, the RP can see the LP’s in a regular 2D video
conferencing setting such as Skype. The images of the
three LP’s are placed next to each other on a large LCD.
The camera placement in the room of the LP’s is such,
that the RP cannot deduce whom is looking at whom.

In the second setting, the RP is shown a more immersive
2.5D representation on the LCD (see figure2); images are
positioned as if the three LP’s are situated around the
table and the RP is at the head of the table. Although
the image used is actually 2D, it is stretched and somewhat
distorted to suggest a 3D image, this concept requires less
computational power than generating a 3D image on a 2D
screen and is known as pseudo-3D or 2.5D [10].

The cameras in the room of the LP’s are placed to more
accurately represent the real direction in which the partic-
ipants are gazing, so that when one of the LP’s is gazing
at another, this is reflected in the video images of the RP
(see figure 2).

11 groups held two 15 minute discussions, one in each con-
dition, after which they filled in a questionnaire similar in
setup to Biocca’s.

Finally, they discussed their experience. During the ses-
sions, there was one observer for each participant behind
a one-way window, who annotated when and how long
his subject was looking at which participant. These anno-
taters were positioned in front of the video image of the RP
in the room of the LP’s. As a result, one of the annotators
had to annotate gaze direction from behind and slightly
to the side of $LP_2$, probably making these annotations the
least accurate of the experiment.

![Figure 2. Setup of the 2.5D situation](image-url)
Questions to be answered:

• how accurate does the data reflect the gaze direction and duration of the participants?

• how do the 2D and 2.5D systems compare when using an objective measure of social presence?

  – how does the (lack of) gaze direction of the LP’s affect the RP?

  – how does the (lack of) gaze direction of the RP affect the LP’s?

3. DATA ANALYSIS

To determine the reliability of the data we will first explain the type of data we have and methods of analysing it. The annotations that are available from the experiment were made during the experiment. As explained, for each participant there was one observer that noted when a participant started looking at someone or something and when he/she stopped. Both participants and observers changed during settings. Since every meeting was annotated once, there is no possibility to replicate the exact circumstances. However, video (and audio) images are available from every meeting and every participant, frontal view, as is an adapted version of the annotation tool that was used.

Reliability of research results depend largely on the reliability of the data, in this case depends on the accuracy of the observer. To determine the quality of the annotations, one has to look at inter-annotator reliability.

3.1 Inter-annotator reliability

The more errors annotators make, the more ‘noise’ will occur in the data, negatively influencing the accuracy and reliability of research results. A limited amount of data can be annotated several times by different annotators to quantify the number of errors, which will show up as inter-annotator disagreement. The following information on inter-annotator (dis)agreement is freely summarized from the work of [12].

This presents a problem for our particular kind of data, since the annotations we have are not simple units to be labeled, but segments of which the annotator has to both determine start and ending besides labeling the segment. There are several ways to solve this problem. Quek et al. determined the percentage of agreement using frames of video images, effectively defining each frame as a unit that was correctly labeled or not. What this method does not show is whether disagreements stem from incorrect identification of time segments or because the segments are labeled differently. One can imagine that mislabeling, or not labeling at all as in 3.1 mistake B or C by annotators 1.1 and 2.1 is a rather large mistake, whereas example B of annotators 1.2 and 2.2 could be considered to be hardly a mistake at all. The last type of disagreement can be resolved by first identifying which segments have been found by both annotators, and by introducing a threshold by which the start and end times may differ.

What this method will not solve are errors such as error C the second example in 3.1. Applied to our data, it would seem that annotator 1.2 determined that his subject kept looking at someone, while annotator 2.2 was of the opinion that the participant glanced away for a few short periods of time. While the effects are clear, the cause of the problem is not; the annotator could not have understood correctly how accurate he/she should annotate, or the explanation was insufficient, causing both annotators to interpret the instructions differently. It is also possible that the error stems from a lack of training, an honest mistake, or perhaps it was simply hard to see or determine which way the subject was gazing.

While it can be useful to determine the causes of annotator disagreement, the fact that it is not possible to replicate the circumstances under which the data was annotated, limits both the accuracy with which we can determine inter-annotator disagreement and therefor the use of the data.

Data with low inter-annotator agreement has been used in previous research with usable results by specifying some constraints, as we will also do. While it is as of yet uncertain whether our data is extremely reliable or not, we will impose some constraints because of the difference in setup of the experiments.

Considering the types of errors discussed above, we cannot verify the exact instructions regarding the annotations of frequency of gaze. Since the setup of the experiment for the second annotations differs from the original setup, we can also not determine whether shifts in boundaries of start and end of segments are caused by the change in setup or not. We will therefor limit our reliability analysis to the duration of gaze from one participant to the other over the timeframe that we have chosen to annotate.

3.2 Statistics

For each participant in a meeting, we have re-annotated
about two-thirds of two meetings. Since the view of LP₂ was the worst, and the direction of gaze of the remote participant was harder to estimate because the images of the other participants were closer together than for the other participants, we expect that the results for these two participants to be worse than for LP₁ and LP₃.

To determine the inter-annotator agreement we will use a chance corrected reliability metric. Such a metric assumes that when two annotators assign labels randomly, one would expect a certain amount of chance agreement. Chance corrected reliability metrics normalize the levels of agreement by expected chance. While such metrics assume annotations defined as labels assigned to units, Cohen’s κ uses percentages of observed and expected agreement. While some discussion exists about the use of this particular metric, it is still widely used and considered to be an overly cautious measure of inter-annotator agreement.

Since we are also dealing with percentages, Cohen’s κ can be applied, and is defined as: $\kappa = \frac{P(A) - P(E)}{1 - P(E)}$, in which $P(A)$ is defined as the observed agreement between annotators, and $P(E)$ as the expected agreement between annotators. $\kappa = 0$ when the amount of agreement equals the amount expected by chance. Landis and Koch [8] gave the following interpretation of the value of $\kappa$ between 0.61 and 0.80 as substantial agreement. While there is no clear reasoning for this measure, since 0.5 is exactly midway between chance and perfect agreement, this seems a fair assumption, and we will accept .61 as the threshold for acceptable quality of data.

### 3.3 Reliability results

As expected, the results of LP₂ and the remote participant are slightly lower than of LP₁ and LP₃, but all results are above the threshold (3.3), and we can thus continue using the data within the defined boundaries.

![Figure 5. Reliability results](image)

### 4. MEETING ANALYSIS

Unfortunately further analysis of the data did not reveal significant differences in gaze between the 2.5D setup and the 2D setup. Since the time span of the research allowed only a cursory analysis of the data, further research into differences between participants and the relation between data and questionnaire should still be examined. Idealiser, the data should be annotated again, and reliability should be determined with the new set of annotations to make full use of the possibilities of the data.

### 5. DISCUSSION

While the experiment generated useful data for further research, a few improvements could be made: the view of LP₂ was not optimal, and the setup of the experiment prevented the possibility of double annotations to make a thorough reliability analysis. However, reliability analysis has so far shown fairly promising results, indicating that (with a few constraints) the data can be used for further research. Unfortunately the results of the analysis of the data between the 2.5D and the 3D situation were less promising, further in depth analysis should be done to confirm these results.

### 6. FUTURE RESEARCH

Research has been done on interpreting gaze direction in dyadic video conferencing, but not on the differences between group video conferencing systems with one remote participant using very poor versus rather accurate representation of gaze direction. This research can show whether gaze direction with the RP differs from that with the participants in the same room, the influence of the accuracy of the gaze of the RP, and the influence of gaze (or lack of it) that the RP experiences in both settings. Future research can use these results to further examine the influence of such a system on social presence of both parties and the use of gaze direction as an objective measurement of some factors of social presence as defined in the Networked Minds measure.

### 7. REFERENCES


APPENDIX

A. DATA OVERVIEW

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Std Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>t</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1.61</td>
<td>10</td>
<td>1.46</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.74</td>
<td>10</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>3.35</td>
<td>20</td>
<td>19.43</td>
<td>0.24</td>
</tr>
</tbody>
</table>

| 4  | 0  | 1   | 3.35  | 10  | 19.43         | 0.24     |
|    |    |     | 1.68  | 10  | 2.95          | 0.08     |
|    |    |     | 3.78  | 10  | 1.98          | 0.04     |
|    |    |     | 1.53  | 10  | 0.01          | 0.00     |
|    |    | Total| 2.50  | 40  | 12.94         | 0.51     |

| Total| 5  | 0   | 4.35  | 40  | 13.99         | 0.22     |
|      |    |     | 1.62  | 30  | 0.09          | 0.00     |
|      |    |     | 2.15  | 30  | 1.40          | 0.22     |
|      |    |     | 1.74  | 30  | 0.00          | 0.00     |
|      |    |     | 1.63  | 30  | 0.00          | 0.00     |
|      |    | Total| 2.42  | 110 | 15.94         | 0.25     |

| 3  | 0  | 2   | 4.92  | 20  | 1.49          | 2.21     |
|    |    |     | 1.23  | 20  | 0.59          | 0.04     |
|    |    |     | 2.23  | 20  | 0.09          | 0.00     |
|    |    |     | 1.84  | 20  | 0.03          | 0.00     |
|    |    | Total| 3.49  | 80  | 15.14         | 0.23     |

| 2  | 0  | 4   | 4.40  | 20  | 1.86          | 0.33     |
|    |    |     | 1.79  | 20  | 0.57          | 0.03     |
|    |    |     | 1.92  | 20  | 1.11          | 0.12     |
|    |    |     | 1.70  | 20  | 1.10          | 0.13     |
|    |    | Total| 2.43  | 80  | 19.52         | 0.23     |

| 1  | 0  | 5   | 0.49  | 20  | 1.19          | 0.14     |
|    |    |     | 2.93  | 20  | 1.01          | 0.04     |
|    |    |     | 1.95  | 20  | 1.22          | 0.12     |
|    |    |     | 1.89  | 20  | 0.04          | 0.00     |
|    |    | Total| 2.43  | 80  | 19.52         | 0.23     |

| 4  | 0  | 1   | 3.17  | 20  | 1.17          | 0.14     |
|    |    |     | 1.65  | 20  | 0.55          | 0.03     |
|    |    |     | 3.52  | 20  | 1.00          | 0.12     |
|    |    |     | 1.86  | 20  | 0.00          | 0.00     |
|    |    | Total| 2.49  | 80  | 12.91         | 0.18     |

| Total| 5  | 0   | 4.75  | 80  | 16.12         | 0.24     |
|      |    |     | 1.55  | 60  | 0.99          | 0.02     |
|      |    |     | 2.85  | 60  | 1.44          | 0.20     |
|      |    |     | 1.90  | 60  | 0.00          | 0.00     |
|      |    |     | 1.74  | 60  | 0.00          | 0.00     |
|      |    | Total| 2.49  | 240 | 16.93         | 0.27     |