

# Non-Zero-Sum Gaze in Immersive Virtual Environments

*Andrew C. Beall, Jeremy N. Bailenson, Jack Loomis, Jim Blascovich & Christopher S. Rex*

Research Center for Virtual Environments and Behavior  
Department of Psychology, University of California  
Santa Barbara, CA 93106  
beall@psych.ucsb.edu

## Abstract

We discuss the theoretical notion of augmenting social interaction during computer-mediated communication. When people communicate using immersive virtual reality technology (IVET), the behaviors of each interactant (i.e., speech, head movements, posture, etc.) are tracked in real time and then rendered into a collaborative virtual environment (CVE). However, it is possible to change those behaviors online, and to render these changed behaviors for strategic purposes. In the current paper we discuss one such augmentation: non-zero-sum gaze (NZSG). An interactant utilizing NZSG can make direct eye contact with more than one other interactant at a time. In other words, regardless of that interactant's physical behavior, IVET enables him to maintain simultaneous eye contact with any number of other interactants, who each in turn may perceive that he or she is the sole recipient of this gaze. We discuss a study in which an experimenter attempted to persuade two participants in an CVE, and manipulated whether gaze was natural (i.e., rendered without transformation), augmented (i.e., each participant received direct gaze for 100% of the time) or reduced (i.e., neither participant received any gaze). We measured participants' head movements, subjective perceptions of the experimenter's gaze, the attitude change for the persuasion topic, and recall of information. Results indicated that participants were unaware of augmented and reduced gaze behaviors despite the fact that the participants' own gaze behavior changed in reaction to those conditions. We discuss these results in terms of understanding mediated communication and nonverbal behavior.

## 1 Overview and Rationale

Real-time augmentation of one's social behavior during interaction is an appealing, albeit Orwellian, prospect that is made possible by recent advances in immersive virtual environment technology (IVET). Using this technology, which tracks and renders a person's nonverbal behavior, one can use intelligent social algorithms to enhance the manner in which an individual's nonverbal behaviors are conveyed to others.

IVET also allows us to examine complex patterns of visual nonverbal behaviors within realistic social contexts with nearly perfect experimental control and high precision. In the current study, we augment gaze in Collaborative Virtual Environments (CVEs). Mutual gaze occurs when individuals look at one another's eyes during discourse. In face-to-face conversation, gaze is zero-sum. If interactant A maintains eye contact with interactant B for 60 percent of the time, it is not possible for A to maintain eye contact with interactant C for more than 40 percent of the time. However, CVEs are not bound by this constraint; in a virtual interaction with *avatars* (virtual hu-

man representations), A can be made to appear to maintain mutual gaze to both B and C for a majority of the conversation. In the following, we describe a paradigm that allows interactants to achieve non-zero sum gaze (NZSG). Figure 1 demonstrates the concept of NZSG.

Gaze in general is one of the most thoroughly studied nonverbal gestures in psychology (Gibson and Pick, 1963; Anstis, Mayhew, and Morley, 1969; Rutter, 1984; Kleinke, 1986). According to Kendon (1977), speakers use gaze to regulate the conversation. Gaze can provide cues for intimacy, agreement, and interest (Arygle, 1988). Consequently, a CVE that augments interactants' capacity to transmit gaze can provide an excellent tool to study social interaction.

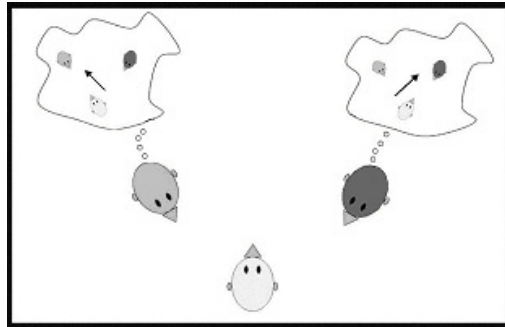
Gaze can be expressed by both head and eye movements. In previous work, we argued that both cues are important sources of information (Bailenson, Beall, & Blascovich, 2002) in social interaction. Head and eye direction are highly correlated and therefore, with caution, head pose can be used to estimate focus of attention. Head pose also conveys unique symbolic information, such as indications of agreement or disagreement.

In real face-to-face interaction, gaze has been shown to significantly enhance performance of information recall. This positive effect has been shown for both children (Ottenson & Ottenson, 1979) and adults (Fry and Smith, 1975; Sherwood, 1987) using a simple fact-recall task. The authors of these studies generally attribute the enhanced performance to there being an increased sense of intimacy between interactants, which in turn better captures attention.

Realizing accurate gaze in CVEs is challenging. Video Teleconferencing often fails to convey effective gaze information because the camera's lens and the monitor's image of interactant's eyes are not optically aligned. To overcome this, various ingenious techniques either optically align camera and monitor (Buxton & Moran, 1990; Ishii, Kobayashi, & Grudin, 1993) or alter the display to "correct" the gaze (Vertegaal, 1999; Gemmell, Zitnick, Kang, & Toyama, 2000).

In IVEs, however, assessment of performance as a result of gaze is still very much work in progress. Recently, Gale and Monk (2002) devised a two person CVE after the ClearBoard demonstration of Ishii et al (1993). They found that gaze behavior traded off with other communication channels, reducing the number of turns and speech required to complete the collaborative task. In other experiments, subjective ratings made by the interactants indicated significant enhancements to the social communication when gaze information was conveyed (Müller, Troitzsch, and Kempf, 2002; Bailenson, Beall, and Blascovich, 2002).

Thus far, computer science and behavioral research has focused on the difference between having gaze cues available and not. What is novel about the work here is a possibility that emerges during n-way interactions with more than two persons, namely NZSG. Consider the fact-recall task. Normally, the constraint of zero-sum gaze imposes a hard limit on a speaker's ability to capture the attention of individual listeners via gaze. As such, we speculate that average fact recall after a group presentation would be worse than the same average recall had the speaker presented the material to each person dyadically. We believe CVEs offer an intermediate possibility, namely that even in simultaneous n-way interaction, each interactant can be led to believe that she is being gazed upon more than in reality. Specifically, we hypothesize that this form of augmented gaze can serve to enhance performance as compared to either natural gaze (zero sum) or gaze absent conditions. While this hypothesis provided the motivation for the current study, we recognize that augmenting interaction with such a simple social algorithms may in fact fail as a result of not cor-



**Figure 1: A conceptualization of Non-Zero-Sum Gaze. The balloons above each person represents his or her belief state concerning the experimenter's gaze.**

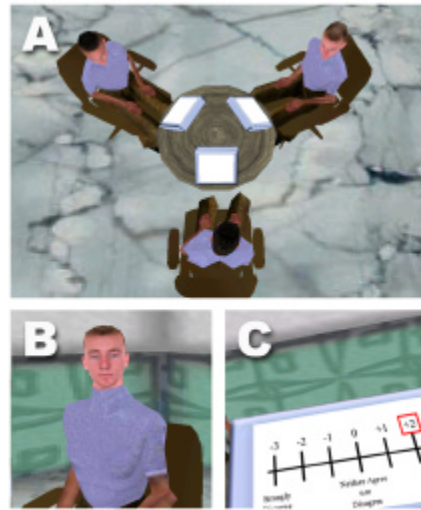
rectly capturing the repertoire of complex and linked head and eye motions that individuals employ to convey intent and meaning. If this study does in fact find that our social algorithm fails, it will show that gaze cannot be blindly amplified, but likely requires a more sophisticated algorithm to be realized.

## 2 Experimental Design and Procedure

In this study, 27 groups of three people (two participants and one experimenter) interacted in the same CVE which resembled a conference room. The 54 participants were told that the purpose of the experiment was to test a CVE in which an experimenter was going to lead a discussion. Group gender was always matched across all three interactants. We employed two 2 male experimenters and 2 female experimenters. Figure 2 shows images of the conference room.

All three interactants were placed in physically different rooms with the door closed and remained seated throughout the study. Each participant's perspective view of the virtual environment was rendered stereoscopically and updated at 60 Hz. Head orientation and position were tracked by a hybrid inertial/optical tracking system with low latencies (less than 5 and 20 ms, respectively). A full duplex intercom system provided natural audio communications among all participants. Mouth movements were tracked via a microphone that sensed sound amplitude, which in turn controlled simple mouth animations of each person's avatar. Figure 3 shows a participant in his own room wearing the head-mounted display (HMD). We chose for both scientific and technical reasons (i.e., the challenge of accurately tracking eye movements in IVEs) to use avatars in which head and eye directions are always locked together.

We manipulated interactants' perception of gaze in three conditions. The first was *natural interaction* (head movements of all interactants were veridically rendered). The second was *augmented gaze* (each participant saw the experimenter's avatar making direct gaze for 100% of the time). The third condition was *reduced gaze* (neither participant received any gaze from the experimenter's avatar). There were 9 groups (i.e., 18 experimental participants) in each condition. Participants were never told of the gaze manipulation and the experimenters themselves were kept blind to condition to ensure that experimenters behaved similarly. We encouraged our experimenters to be as persuasive as possible and to use as much eye contact as possible. To implement the augmented and reduced conditions, our software scaled the experimenter's actual head motions by a factor of 20 and re-centered the effective straight-



**Figure 2: Scenes from the CVE. Panel A - bird's eye view. Panel B - avatar close-up. Panel C - Likert response screen on each computer monitor.**



**Figure 3: A participant uses an HMD, intercom, and gamepad.**

ahead position to point either at the participant's head or the experimenter's screen, respectively.

Participants went into their own physical room without meeting the experimenter. We demonstrated how to use the equipment and respond to the questionnaires. Once the three were immersed and online, the experimenter read two passages to the participants. We measured as dependent variables: 1) head movements, 2) subjective estimation of experimenter's gaze direction, 3) information recall, and 4) persuasion for the passages.

### 3 Experimental Results and Conclusions

One of the most striking findings of this study is that participants did not detect either the augmentation or the reduction of gaze. Despite the fact that, from a given participant's point of view, the other participant received absolutely no gaze from the experimenter in the augmented condition, participants did not notice. After the study, we asked each of the participants to estimate the percentage of time that the experimenter looked at each participant. Figure 4 demonstrates those differences by condition and participant. Estimation in every condition and participant was statistically different from zero. Consequently, participants did not notice the lack of gaze given to their counterparts in the augmented condition.

Next, we analyzed the head movements of our participants. If they accepted the augmented gaze as real gaze, then we would predict that participants would return the gaze (i.e., look towards the experimenter) most often in the augmented condition. Figure 5 demonstrates the percentage of time that participants looked (oriented the head) toward the experimenter or the other interactant.

To test the significance of this difference in looking we ran a two factor ANOVA: gaze condition (natural, augmented, and reduced) and head orientation (towards experimenter or other participant). The predicted interaction was significant ( $F(2,49)=3.70$ ,  $p<.05$ ), demonstrating that the difference in looking percentage between experimenter and other interactant was greatest in the augmented condition.

Our other dependent measures such as persuasion and recall did not show a discernible pattern across gaze conditions. We have no compelling explanation for

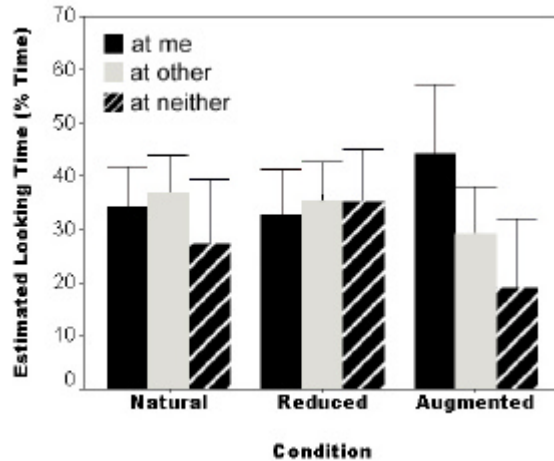


Figure 4: Participants' estimation of where the experimenter was looking by condition

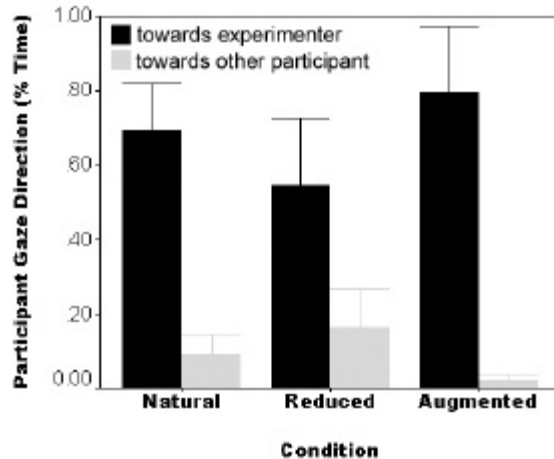


Figure 5: Participant's gaze direction during passage presentation

this. It is possible that this study lacked the power to find differences that have been found in previous real face-to-face and IVE interactions. At worst, social augmentation algorithms such as ours may be too simple and the lack of real connectedness between experimenter and participant may have undermined its potential effectiveness.

However, we feel our data suggests such algorithms may succeed. Participants were not aware in the augmented condition that their partner was being entirely ignored. Equally important, the augmented was as successful if not better at capturing their attention compared to the natural condition. In sum, this study points out the possibility for augmenting social interaction within computer mediated environments and shows that technology available today can usefully investigate this phenomenon.

## References

- Anstis, S., Mayhew, J., & Morley, T (1969). The Perception of where a Face or Television 'Portrait' is Looking. American Journal of Psychology, p. 474-489.
- Argyle, M. (1988). Bodily communication (2nd ed.). London, England UK: Methuen.
- Bailenson, J.N., Beall, A.C., & Blascovich, J. (2002). Mutual gaze and task performance in shared virtual environments. Journal of Visualization and Computer Animation, 13, 1-8.
- Buxton, B. & Moran, T. (1990). EuroPARC's integrated interactive intermedia facility (IIIF): early experiences, Proceedings of the IFIP WG 8.4, p.11-34, Heraklion, Crete, Greece.
- Cline, M. (1967). Perception of Where a Person is Looking. American J. of Psychology, p. 41-50.
- Fry, R. & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. Journal of Social Psychology, 96, 145-146.
- Gale, C. & Monk, A.F. (2002) 'A look is worth a thousand words: full gaze awareness in video-mediated conversation'. Discourse Processes 2002, Volume 33, Number 1.
- Gemmell, J., Zitnick, L., Kang, T., and Toyama, K., (2000). Software-enabled Gaze-aware Video-conferencing, IEEE Multimedia, 7(4), 26-35.
- Gibson, J., & Pick, A.(1963). Perception of Another Person's Looking Behavior. American Journal of Psychology, 76, 386-394.
- Ishii, H., Kobayashi, M. & Grudin, J.(1993). Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments, ACM Transactions on Information Systems (TOIS), 11(4), 349-375.
- Kendon, A. (1977). Studies in the Behavior of Social Interaction. IU: Bloomington.
- Kleinke, L. (1986). Gaze and eye contact: A research review. Psychological Bulletin, 100, 78-100.
- Otteson, J. P., and Otteson, C. R. (1979). Effect of teacher's gaze on children's story recall. Perceptual and Motor Skills, 50, 35-42.
- Müller, K. Troitzsch, H. & Kempf, F. (2002). The role of nonverbal behavior in 3D-multiuser-environments. Proceedings of the 4th International Conference on New Educational Environments 2.1, 35-38.
- Rutter, D. R. (1984). Looking and Seeing: The Role of Visual Communication in Social Interaction. John Wiley & Sons: Suffolk.
- Sherwood, J. V. (1987). Facilitative effects of gaze upon learning. Perceptual and Motor Skills, 64, 1275-1278.
- Vertegaal, R. (1999). The GAZE groupware system: mediating joint attention in multiparty communication and collaboration, Proceeding of the CHI 99 conference on Human factors in computing systems: the CHI is the limit, 294-301.