Multimodal Transformed Social Interaction

Matthew Turk*, Jeremy Bailenson†, Andrew Beall‡, Jim Blascovich‡, Rosanna Guadagno‡

*Dept. of Computer Science
University of California
Santa Barbara, CA 93106
mturk@cs.ucsb.edu

†Dept. of Communication
Stanford University
Stanford, CA 94305
bailenson@stanford.edu

‡Dept. of Psychology
University of California
Santa Barbara, CA 93106
{beall, guadagno, blascovi}
@psych.ucsb.edu

ABSTRACT

Understanding human-human interaction is fundamental to the long-term pursuit of powerful and natural multimodal interfaces. Nonverbal communication, including body posture, gesture, facial expression, and eye gaze, is an important aspect of human-human interaction. We introduce a paradigm for studying multimodal and nonverbal communication in collaborative virtual environments (CVEs) called Transformed Social Interaction (TSI), in which a user's visual representation is rendered in a way that strategically filters selected communication behaviors in order to change the nature of a social interaction. To achieve this, TSI must employ technology to detect, recognize, and manipulate behaviors of interest, such as facial expressions, gestures, and eye gaze. In [13] we presented a TSI experiment called *non-zero-sum gaze* (NZSG) to determine the effect of manipulated eye gaze on persuasion in a small group setting. Eye gaze was manipulated so that each participant in a three-person CVE received eye gaze from a presenter that was normal, less than normal, or greater than normal. We review this experiment and discuss the implications of TSI for multimodal interfaces.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Artificial, augmented, and virtual realities.

General Terms

Experimentation, Human Factors

Keywords

Multimodal processing, computer-mediated communication, transformed social interaction

1. INTRODUCTION

Technology has long facilitated human communication and social interaction, from cave wall markings and smoke signals to written letters, telegraph, telephone, email, videoconferencing, and instant messaging. Historically, these communication technologies have been primarily unimodal, providing the participants with a single channel with which to convey sometimes complex and abstract information. From the list above, only videoconferencing is fully multimodal, as participants in a conversation can use words, intonation and prosody, non-speech sounds, facial expressions, eye gaze, and body posture and gesture to convey many (sometimes conflicting) messages at once. In face to face interaction, visual nonverbal communication is quite meaningful, with much information conveyed by subtle actions such as mutual gaze, a brief and slight smile or raised eyebrow, nods, shrugs, etc.

While there is a very rich history of research in spoken communication and in visual communication systems such as sign languages, much less is known about multimodal and nonverbal communication. However, there has been significant interest in these areas in recent years (e.g., see [1] and [2]), in part due to the emergence of new communication technologies that enable multimodal interaction. In particular, an understanding of the relationship between gesture and speech [3] has progressed significantly. In order to develop effective and powerful multimodal computer interfaces, it is vital that the research community continues to pursue an understanding of multimodal human communication.

Collaborative Virtual Environments (CVEs) facilitate social interaction in immersive virtual environments. Virtual environment technology (such as head-mounted displays, trackers, 3D rendering capabilities) has advanced to the point where it is possible to simulate physical and social environments, completely controlled by researchers in order to investigate human behavior and to create new communication technologies. Because it is much easier to modify parameters of a virtual world than in the physical world, CVEs have become an important tool in behavioral and perceptual psychology.

In this paper, we describe our paradigm of *transformed social interaction* (TSI) in collaborative virtual environments to study multimodal and nonverbal communication and present an initial experiment on the effects of *non-zero-sum gaze* on persuasion. We discuss the implications of this work for multimodal interfaces and environments and close with comments on future research directions for multimodal TSI.

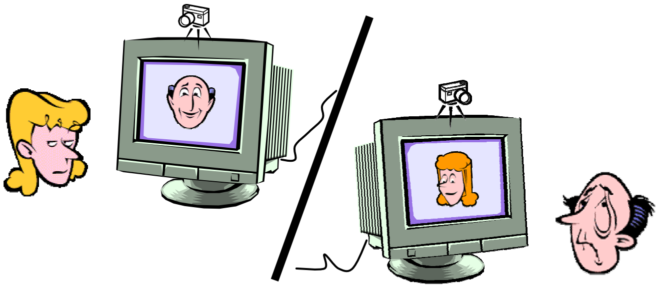

Figure 1. A depiction of a simple Transformed Social Interaction (TSI) scenario. Two people converse over a multimedia network, and certain aspects of a user's appearance or behavior are modified (in this case, facial expression and hairstyle). This TSI scenario is not in a CVE.

2. TRANSFORMED SOCIAL INTERACTION IN COLLABORATIVE VIRTUAL ENVIRONMENTS

We have been studying multimodal and nonverbal communication in the context of Collaborative Virtual Environments (CVEs) through the paradigm of Transformed Social Interaction (TSI). CVEs are computer mediated communication systems that allow geographically separated individuals to interact verbally and nonverbally in a shared virtual space in real time. Unlike videoconferencing, in which audio and video streams are transmitted directly, CVEs represent each human interactant visually by a dynamic digital representation – a 3D computer generated avatar. Tracking technology (optical sensors, mechanical devices, and/or cameras) monitors the movements and behaviors of the individual interactants, and the CVE renders those behaviors via the avatars, which are visible to others in the environment.

CVEs promise to enable more effective interactions than videoconferencing by giving interactants a richer environment and opportunities for a wider range of interaction styles and experiences ([5], [6]). CVEs are also quite effective experimental laboratories for research in social interaction [7], allowing a reasonable measure of ecological validity (a realistic environment) while maintaining a high level of experimental control.

Because interactants in a CVE view avatars of one another rather than direct video feeds, the opportunity arises for new interaction strategies based on manipulating the visual appearance and behavior of the avatars. (Although we currently render the audio directly, there are such opportunities in that channel as well.) For example, a CVE system could be designed to render a "spruced up" avatar of an interactant who just woke up and hasn't shaved, showered, or dressed, or to change visual appearance, identity, gender, etc. Similarly, the CVE could selectively filter, amplify, delete, or transform the nonverbal behaviors of an interactant, either with or without his or her knowledge.

Transformed Social Interaction (TSI) [7] describes the strategic filtering of communicative behaviors in order to change the nature of social interaction. Such changes can have positive, negative, or neutral effects for a given interactant. TSI can purposefully enhance or degrade interpersonal communication, and can do so selectively, since the generation of nonverbal behavior is decoupled from its rendering. In a CVE with three people, for example, the appearance and nonverbal behavior of interactant A may be rendered in one way to interactant B and in a quite different way to interactant C, with interactants B and C both believing that they see the "true" rendering. This kind of selective transformation may also be done without the knowledge of interactant A.

A TSI could conceivably detect, recognize, and modify many visual nonverbal communication cues in the context of social interaction. Smiles could be inverted to frowns, gestures could be edited for cultural sensitivity, yawns could be suppressed, and eye gaze could be redirected. (See Figure 1.) With sufficiently accurate sensing and rendering technologies, the opportunities to effect a variety of visual and behavioral transformations are huge. Indeed, the opportunities to *misuse* such transformations are also huge, raising significant ethical questions about the use of TSI. We believe that this decoupling of rendered behavior from actual behavior is inevitable as the technology progresses (due to many compelling positive uses, as well as less positive ones), so it is important to understand the possibilities and the limitations of these technologies (and also to be able to detect their use) so that it will be possible to ethically use and manage TSI in the future.

CVEs are fundamentally multimodal environments, conveying several kinds of information via audio and visual channels: spoken speech, non-speech sound, general visual appearance, visual identity, and a range of body postures, movements, and

Figure 2. A Collaborative Virtual Environment (CVE). Images of users are used to build 3D head models; interactants in the CVE wear head-mounted displays (HMDs) and experience a computer generated scene with an avatar representing each interactant.

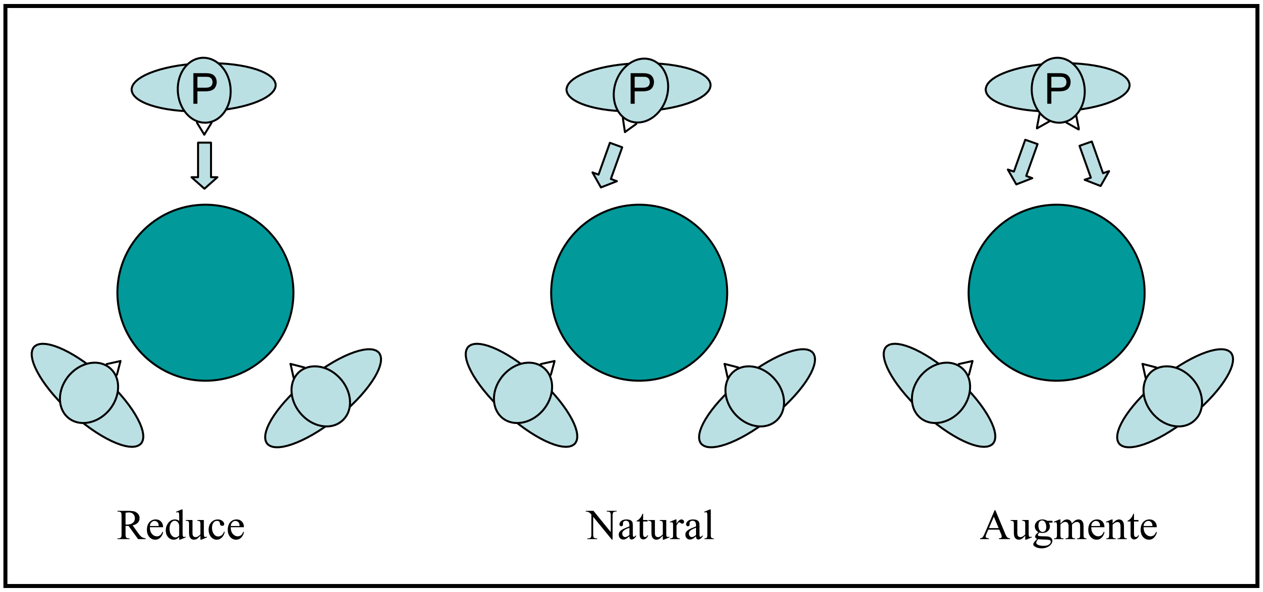

Figure 3. The three gaze conditions in the experiments, showing the presenter's (P) gaze directions. Reduced: presenter achieves no mutual gaze with the participants. Natural: presenter achieves a natural amount of mutual gaze with each of the participants. Augmented: presenter achieves 100% mutual gaze with both participants simultaneously. The "reduced" and "augmented" conditions exhibit non-zero-sum-gaze (NZSG).

gestures (including facial expression and eye gaze). TSI enables the controlled study of multimodal, and especially nonverbal, communication to better understand the relative importance of various modalities and the possibilities for detection, recognition, and selective modification of semantically meaningful actions. These studies should inform the study and development of multimodal computer interfaces. In fact, a TSI *is* a multimodal interface in the context of a collaborative virtual environment. Besides being an experimental platform for social interaction, TSI can lead to richer, more powerful interfaces for immersive environments.

In the following sections, we describe a TSI experiment that manipulates the amount of eye gaze devoted to multiple recipients in order to investigate its effect in a CVE.

3. NON-ZERO-SUM GAZE

Mutual gaze between people is a powerful and meaningful aspect of face-to-face communication. Normally, gaze can be directed to only one individual at a time in a social interaction, and the sum of the portions of time that one's gaze is directed to various individuals cannot exceed 100%. This is the *zero-sum* situation, where the amount of gaze that was "given" by a person equals the sum of the amounts of gaze "received" by others. A collaborative virtual environment gives us the ability to implement a TSI called *non-zero-sum gaze* (NZSG), where the amount of gaze given does not equal the amount received. At one extreme, one's avatar can be made to give each interactant 100% gaze (as is the case with the news anchor looking into the camera when rendered on a 2D TV screen). At the other extreme, the avatar can be made to give each interactant 0% gaze – i.e., to never look directly at them.

Of course, anywhere between those extremes is also possible. In an environment with N other participants, one's expected average of $1/N$ of the gaze time may be slightly increased, or significantly increased, for each interactant – perhaps making each participant feel that he or she is receiving special attention from the others.

We ran an initial study [9] on NZSG, featuring three interactants in a CVE, as depicted in Figures 2 and 3. One of the interactants, the presenter, led a discussion concerning designated topics with two listeners while the presenter's gaze was manipulated experimentally. The presenter's avatar exhibited one of three conditions: natural gaze (each interactant received an approximately equal amount), reduced gaze (mostly directed at neither interactant), or augmented gaze (mostly directed at both interactants simultaneously). The experimental conditions were unknown to both the presenter and the listeners.

In the initial study [9], we demonstrated that: (1) participants were unaware that we augmented the presenter's gaze under conditions of NZSG (reduced or augmented gaze), and (2) participants directed their attention (i.e., looking direction) more towards the presenter when under NZSG (altered gaze) conditions than natural gaze conditions. In terms of perceived responsiveness of the presenter, the interaction suffered only slightly due to the TSI's interference with natural nonverbal cues. In the augmented gaze condition, the increased gaze of participants toward the presenter implies that the presenter was *implicitly* perceived by our subjects as hyper-responsive, since he or she appeared to dedicate large amounts of nonverbal resources to each of them. In essence, the gaze abilities of the presenter were doubled at the cost of true responsiveness, but that cost in terms of a functioning interaction (implicit responsiveness) was minimal.

4. EXPERIMENT: PERSUASION IN NZSG

In a recent study on non-zero-sum gaze [13], we investigated whether or not a presenter in a CVE implementing NZSG conditions (augmented and reduced) would be more or less persuasive than in natural gaze conditions. In that study, the passages read by the presenter were chosen to allow for examination of participants' agreement with the subject matter. Via pretesting, we developed passages that allowed for persuasion – i.e., either an increase or decrease in agreement from the normative agreement of our participant populations. We began with four major hypotheses:

Hypothesis 1. We predicted that NZSG manipulations would decrease participants' explicit ratings of the presenters' social presence and awareness. Because presenters in either condition of NZSG should be viewed as unresponsive [10] and in violation the normal expectancies of an interaction, participants should rate augmented gaze presenters as having lower awareness and social presence than other presenters.

Hypothesis 2. We predicted that augmented NZSG would increase the level of agreement by listeners to the presenter's message. Previous research has demonstrated that eye-gaze increases persuasion in natural conversation [11]. While NZSG may cause a decrease in social presence and awareness, we have not found these costs to be overly disruptive in previous research on NZSG [9]. Despite the cost to interactivity, receiving an abundance of apparently direct eye gaze from a presenter should increase that presenter's ability to influence an audience.

Hypothesis 3. Since previous research has demonstrated that instructor-gaze increases learning [12], we predicted that NZSG should increase memory (recall) for the material presented.

Hypothesis 4. Finally, we predicted that participants would not explicitly detect the NZSG manipulations (despite being influenced by them). It may be the case that TSI in general, including NZSG, will not be an effective strategy if interactants in a given discussion do not believe that the verbal and nonverbal behaviors they are seeing have in fact actually occurred. One can imagine a future in which the use of TSI becomes so prevalent during computer-mediated communication that the interaction loses all practical function and meaning. Consequently, studying the human ability to detect transformed behavior, and the effects of this awareness, are uniquely interesting theoretical and applied questions.

4.1 Design, Apparatus, and Procedure

We selected-for or manipulated two between-subjects variables: *participant gender* (male or female) and *presenter gaze condition* (natural, augmented, or reduced). Although gender did not figure in our main hypotheses, previous experience led us to expect gender effects. In a given session, the presenter and both participants, as well as their three avatars, were either all male or all female. The presenters were encouraged to use head movements in order to engage the other participants. In the *natural* (normal gaze) condition, the presenter's avatar veridically displayed his or her head movements. In the *augmented* condition, the presenter's avatar directed his or her gaze at each of the two participants at all times. In the *reduced* condition, the presenter's avatar looked down at his or her computer monitor 100 percent of the time and did not direct gaze at either participant. These manipulations (the augmented and reduced gaze conditions) were implemented by transforming the veridical head movements of the presenter by scaling the magnitude of those movements down by a factor of 20 (e.g., if a presenter moved their head 20 degree to the left, the participants would only see a 1 degree movement), and by re-centering the effective straight-ahead position of the presenter's head such that the presenter's eyes were looking directly at the eyes of each participant (augmented) or directly down at the computer screen (reduced). We scaled down the actual movements in order to provide slight head movements in the transformed conditions, so that the presenter did not appear frozen. The presenters and the listeners were always blind to the transformations; they were not aware of the manipulations.

In all three conditions (natural, reduced, and augmented), the listeners' head movements were rendered veridically. All three avatars blinked randomly, according to a computer algorithm we developed based on human blinking behavior; they also exhibited lip movements driven in real time by the amplitude of the sound of each interactant's speech as measured by a small microphone place near the mouth. There were no other changes in facial expression, so other than lip movements, head movements, and eye-blinks, no other behaviors were exhibited by the participants' avatars. We employed three male presenters and two female presenters, each of whom was cycled across the three gaze conditions approximately an equal number of instances.

The virtual room contained a round table around which the three avatars sat for a typical meeting environment (see Figures 2 and 3). Participants could see the other avatars in the room as well as part of their own torsos, if they looked straight down. Each participant wore a head-mounted display (HMD) and an intercom device. The HMD contained a separate display monitor for each eye, and the virtual scene was rendered by the graphics system separately for each eye (in order to provide stereoscopic depth) at approximately 60 Hz. Using an inertial tracking system for orientation with low latencies, it was possible for participants to experience realistically dynamic visual input – the delay between a user's movement and the system's detection of that movement was less than 40 ms.

We recruited participants from introductory psychology classes for pay or for experimental credit. Twelve participants, ranging in age from 18 to 25, were used in each of the six between-subjects conditions resulting from crossing participant gender and gaze condition for a total of 72 participants.

Participants did not meet each other or the presenter before the session. A second experimenter escorted each participant into his or her separate room and read a paragraph describing the environment and the task, which was ostensibly to test out how smoothly an interaction can flow in a virtual conference room.

The two participants then received instructions regarding use of the technology: the HMD, the integrated intercom, and the game-pad used to record responses during the experiment session. Once they began using the system, there was an introduction phase, during which the presenter's head movements were rendered veridically. Finally, once participants adjusted to the environment, the appropriate manipulation condition (NZSG or normal gaze) commenced. At this point, the presenter read the persuasive passage, which was broken up into four sections. Each section took approximately twenty seconds to read; after the presenter completed each section, he or she facilitated a discussion concerning that section of the passage for about 90 seconds.

After the passage was read, the presenter verbally asked the participants to answer three Likert-scale agreement questions and three multiple choice memory questions concerning the passage. Participants responded by using their game-pad device to move a cursor. After completing these agreement and recall questions, the participants removed their HMDs and completed pen-and-paper questionnaires. The questionnaires measured *social presence*, the degree to which participants regarded the presenter's avatar as a sentient person who was aware of the interaction. Next, each participant estimated the percentage of time in which the experimenter looked at him or her, at the other participant, and at neither one. Finally, participants wrote three paragraphs (of at least five sentences each): one about the presenter, one about the other participant, and one about the virtual conference experience.

4.2 Measures

Presenter Awareness/Social Presence. To test hypothesis 1, we calculated a score to measure how much the participants felt that the presenter was aware of the participants and how real the presenter's avatar behaved during the interaction.

Persuasion. To test Hypothesis 2, we used an estimated measure of persuasion, intended to measure the degree to which participants agreed with the statements summarizing the passage read by the presenter.

Memory. To test Hypothesis 3, we calculated a score to measure the participants' recall for information from the passage by averaging the responses (1 for correct, 0 for incorrect) for three multiple-choice memory responses, where higher scores indicate greater memory.

Presenter Gaze estimation. To test Hypothesis 4, participants estimated the amount of time that the presenter looked at them, the other participant, and at no one. Each was estimated on a scale of 0 to 100 percent.

Explicit NZSG Detection. In the open-ended paragraph that the participants wrote about the presenter, two independent coders blind to the experimental condition analyzed the text, looking for instances in which the participants explicitly detected the NZSG. The coders looked for any wording concerning the redirection of head movements, fake or artificial gaze, nonverbal tampering, etc. This measure was also used to test Hypothesis 4.

4.3 Results

Presenter Awareness/Social Presence. To determine perceived awareness of the participant by the presenter, we examined the average scores from the questions the participants were asked. We ran an ANOVA with gender and gaze condition as the independent variables and presenter awareness as the dependent variable. Female subjects reported higher presenter awareness in the natural condition than did male participants.

Persuasion. We ran an ANOVA with gender and gaze condition as between-subjects independent variables and with agreement score as the dependent variable. There was a main effect of gaze condition, indicating that the augmented condition was significantly higher than either the reduced condition or the natural condition. However, there was also an interaction between gender and gaze condition, such that female participants were more persuaded by augmented gaze than by the other gaze conditions while males did not show a reliable difference between conditions. In addition, there was a main effect of gender, such that females were less persuaded than males overall. In summary, females were less persuaded than males as a group; however, females in the augmented gaze condition were more persuaded than females in the other gaze conditions.

Memory. We ran an ANOVA with gender and gaze condition as the independent variables and recall score as the dependent variable. There was a main effect of gender, with males scoring higher than females. The memory scores of males were significantly above chance, whereas female scores did not differ from chance.

Presenter Gaze Estimation. The estimation by participants of the percentages of the presenter's gaze provided data regarding participants' awareness of the presenter's gaze behavior that we used as a manipulation check. These estimations also allowed us to test whether the participants were aware of our manipulations (e.g., if the responses totaled above 100% for augmented gaze).

We conducted an ANOVA with participant gender and gaze condition as between-subjects variables, gaze direction as a within-subjects variable and percent time as the dependent variable. There was a main effect of gaze direction; participants estimated they received more of the presenter's gaze than did the other participant. There were no significant main effects for gender or gaze condition. In addition, there was an interaction between gaze condition and gaze direction; participants in the augmented gaze condition perceived more gaze directed toward themselves than toward the other participant, compared to the other two gaze conditions.

We summed each participant's responses to the three gaze estimation questions: "looking at me," "looking at the other participant," and "looking at neither." *Total gaze*, the sum of this subjective estimation measure, should be approximately 100% if participants assumed the natural constraint of zero-sum gaze. The mean of total gaze across participants was not statistically different from 100 in a paired t-test. We cannot draw conclusions from null results, but we were encouraged by how close the total gaze figure was to 100 percent.

In sum, the participants generally perceived that they received more of the presenter's gaze than did the other interactant in their group, regardless of the actual gaze condition. This effect was most apparent in the augmented condition, where the presenter actually did look at them constantly. In addition, participants did not detect the non-zero-sum gaze in the augmented or reduced condition, as determined by their ratings of gaze estimation. In this analysis, as well as in all subsequent analyses, there was no interaction effect between the particular presenter of the day and the other independent variables.

Explicit NZSG Detection. There were no instances in which a participant explicitly detected that the human gaze had been modified by either coder, so the condition of NZSG was undetected.

See [13] for a more detailed description of the results with all the appropriate experimental measures and statistics.

5. DISCUSSION

Previous authors have pointed out the unique ability of a computer mediated virtual environment to be used as a tool of persuasion (e.g., [14]). In the second hypothesis, we predicted that participants would demonstrate more persuasion (i.e., higher agreement with the presented passage) in the augmented condition than in the other two conditions. This hypothesis was confirmed for female participants much more than for male participants. The fact that males did not demonstrate a higher response to the nonverbal gaze behavior is consistent with previous findings concerning gender and the reception of gaze cues, both in physical and in immersive virtual environments. Furthermore, studies specifically intended to understand computer-mediated persuasion have also demonstrated similar gender effects [15]. One possible explanation for this trend is that, in the current study, male participants may have focused primarily on the verbal content of the message. This suggestion is supported by both their higher recall for the actual content of the message than females, and their smaller differences in persuasion and Presenter Awareness/Social Presence ratings by gaze behavior condition than females. Another possibility is that females may have focused primarily on the head movements of the presenters. This hypothesis is supported by their lower recall for the verbal content of the message than males and their higher degree of persuasion and lower ratings of Presenter Awareness/Social Presence ratings in the augmented condition.

The somewhat fascinating result of this study is that, despite reporting the clearly problematic aspects of the presenter's responsiveness in the 100% augmented gaze condition on an explicit level, female participants were still persuaded by the visual input of augmented eye gaze on a low, implicit level. In other words, even though female participants were well aware on some level that the presenter's behavior was not socially appropriate, they were still persuaded most after receiving the higher amounts of eye gaze.

The results of this study have several implications. On a positive note, while we did not find any significant facilitation of memory recall in the current study (i.e., Hypothesis 3 was not confirmed), the potential for NZSG to assist instructors in distance learning CVEs is still an intriguing possibility. Instructors should be able to provide students with more personalized nonverbal attention using NZSG in addition to other transformed social behaviors. On a less positive note, NZSG may turn out to be an attractive strategy for advertisers, salespeople, politicians and others who seek to gain influence on others by manipulating the social environment. In the current study, participants were not able to detect the transformed gaze. However, one hopes that as CVEs become more widespread, people will become more aware of the possibilities of misusing such rendered representations. In addition, it may be relatively simple to develop algorithms to detect NZSG and other forms of transformed social interaction [7].

The current NZSG study has a number of limitations. First, we transformed gaze in perhaps the coarsest way possible: all or none. In future studies we plan to develop more subtle transformations of gaze direction, varying percentages of direct gaze as a function of conversational flow. Making the gaze more interactive (e.g., briefly turning the presenter's head away from one participant when the other one begins to speak) may dramatically increase the overall effectiveness of the TSI. Second, our manipulation of eye gaze was completely linked to head orientation in the current study: as participants moved their heads, their eyes were always facing straight ahead. A more effective manipulation would be to transform both head direction and eye direction; in future studies we plan to employ eye-tracking technology in order to achieve this goal. In addition, the length and specificity of our persuasive passages (relatively short passages of only a single thematic topic) as well as the way in which we presented the argument (verbal passages followed by discussion) do not generalize to all types of persuasive situations. Future research should definitely examine TSI in all sorts of persuasive contexts. Moreover, virtual reality simulations are currently limiting because the equipment is expensive and cumbersome. In the past few years, however, improvement in the technology has accelerated drastically [16], and there is no reason to suspect that this trend will not continue.

Finally, in future studies, our plans are to run conditions of "explicit" NZSG manipulation, informing participants before the study of the specific transformed behavior. The goals in these experiments will be twofold: to test whether or not participants can detect NZSG when they are aware of its potential, and to determine if the low level, implicit effects of gaze response are strong enough to override the explicit knowledge that the gaze behavior is manipulated.

6. IMPLICATIONS AND FUTURE WORK

Gaze is but one of many aspects of human-to-human interaction that can be manipulated in a transformed social interaction paradigm, and the persuasion effect of non-zero-sum gaze is but one of many aspects of gaze to study. A thorough understanding of visual nonverbal communication and how it impacts human interaction is ultimately an important scientific foundation for the study of multimodal interfaces, which in the broad view seeks to develop interaction technologies that leverage fundamental aspects of human communication styles and abilities. Nonverbal information in multimodal interfaces can be in both directions, both automatically perceiving the nonverbal behaviors of interactants and generating interfaces that exhibit nonverbal behaviors of their own.

In addition to the basic understanding sought by TSI experiments in collaborative virtual environments, TSIs are multimodal interfaces themselves, combining audio with graphical representations of users (output) that are generated by sending, tracking, and recognition technologies (input). Eventually, semantic processing of the audio signal will also be an important part of TSIs.

In addition to continuing to pursue NZSG experiments and to develop visual tracking and recognition technologies to support advanced TSIs, we are beginning to pursue TSI in non-immersive environments such as traditional videoconferencing and CSCW environments. Transforming appearance and behavior in face-to-face video interactions has the potential for significant impact.

7. ACKNOWLEDGMENTS

This research was sponsored by NSF ITR Award 0205740 and performed at the Research Center for Virtual Environments and Behavior (ReCVEB) and the Four Eyes Lab at UCSB.

8. REFERENCES

[1] Argyle, M. *Bodily Communication* (2nd edition). London, England, UK: Methuen, 1988.

[2] Depaulo, B. M., and Friedman, H. S. Nonverbal communication. In D. T. Gilbert, S. T. Fiske, and G. Lindzey (eds.), *The Handbook of Social Psychology* (4th edition), pp. 3-40, Vol. 2, Boston: McGraw-Hill, 1998.

[3] Quek, F., McNeill, D., Bryll, B., Duncan, S., Ma, X., Kirbas, C., McCullough, K-E, and Ansari, R. Multimodal Human Discourse: Gesture and Speech, *ACM Transactions on Computer-Human Interaction*, Vol. 9, No. 3, September 2002, pp. 171-193.

[4] Oviatt, S. L. Multimodal interface research: A science without borders. *Proceedings of the International Conference of Spoken Language Processing (ICSLP'2000),* ed. by B. Yuan, T. Huang & X. Tang, Beijing: Chinese Friendship Publishers, 2000, vol. 3, 1-6.

[5] Zhang, X. and Furnas, G. Social interactions in multiscale CVEs. *Proceedings of the ACM Conference on Collaborative Virtual Environments*, 2002.

[6] Normand, V. et al. The COVEN project: exploring applicative, technical and usage dimensions of collaborative virtual environment. *Presence: Teleoperators and Virtual Environments*, 8(2), 1999.

[7] Bailenson, J., Beall, A., Blascovich, J., Loomis, J. and Turk, M. Transformed social interaction: decoupling representation from behavior and form in collaborative virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, MIT press, Vol. 13, No. 4, August 2004.

[8] Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry, 13*, 146-149, 2001.

[9] Beall, A.C., Bailenson, J.N., Loomis, J., Blascovich, J., and Rex, C. Non-Zero-Sum Mutual Gaze in Collaborative Virtual Environments. *Proceedings of HCI International,* Crete, 2003.

[10] Cappella, J. N. and Pelachaud, C. Rules for responsive robots: Using human interactions to build virtual interactions. In Vangelisti, A.L., Reis, H.T., and Fitzpatrick, M. (Eds.), *Stability and Change in Relationships*, pp. 325-354, Cambridge, MA: Cambridge University Press, 2002.

[11] Morton, G. Effect of eye contact and distance on the verbal reinforcement of attitude. *Journal of Social Psychology, 111*, pp. 73-78, 1980.

[12] Sherwood, J. V. Facilitative effects of gaze upon learning. *Perceptual and Motor Skills, 64*, pp. 1275-1278, 1987.

[13] Bailenson, J., Beall, A., Blascovich, J., Loomis, J., and Turk, M. Non-zero-sum gaze and persuasion in immersive virtual environments. Manuscript under review.

[14] Gigorovici, G. Persuasive effects of presence in immersive virtual environments. In G. Riva, F. Davide (Eds), *Being There: concepts effects and measurement of user presence in synthetic environments*, Amsterdam, The Netherlands: IOS Press, online, 2003.

[15] Guadagno, R. E. and Cialdini, R. B. Online persuasion: An examination of gender differences in computer-mediated interpersonal influence. *Group Dynamics: Theory Research and Practice. Special Issue on Internet Research, 6*, pp. 38-51, 2002.

[16] Lanier, J. Virtually there. *Scientific American*, April, 2001.